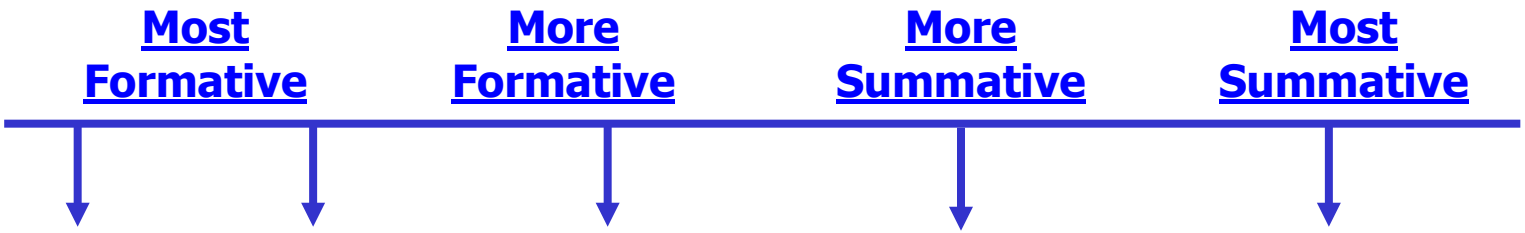
The Assessment Continuum Resource Tool

Introduction: The purpose of the Assessment Continuum Resource is to provide District administrators, principals, and teachers with access to detailed descriptions of specific assessments aligned to the key element found in the Assessment Continuum.



Minute to Minute Classroom Assessments

Whiteboard responses to key questions

Teacher/Student and Student/Student Questions in Class

Research

Research Resources

Samples

Teacher team developed short cycle assessments

Teacher developed quizzes and curriculum based assessments

Exit Tickets

Research

Research Resources

Samples

Grade-level team developed common assessments

Screening, diagnostic, ongoing monitoring

Performance-based assessments

Research

Research Resources

Samples

District-developed Mid-Year Benchmark Assessments

Teacher developed End of Unit Assessments

NWEA Assessments

Research

Research Resources

Samples

California Standards Test (CST)

California High School Exit Exam (CAHSEE)

SAT College Entrance Exam

California English Language Development Test (CELDT)

Research

Research Resources

Samples

Most Formative Assessments

http://assessmentcontinuum.wetpaint.com/page/Minute+to+Minute+Assessments

Description
Most formative assessments are closest to instruction. They take place within the actual instructional cycle and provide students and teachers with immediate feedback about the degree to which all students are achieving the instructional targets. Results from these assessments can be used to address student learning needs quickly. They are also motivational for students because they can be constructed in ways that provide joy to students. For example, a just in time assessment called Crumple It Up found in the resource called *Science Formative Assessments* by Paige Keeley engage student in responding to a question developed by the teacher about an important learning target. Students write their responses on half sheets of paper, crumple them up and then toss them to fellow students. Students can then read their fellow students responses aloud and provide their own interpretations of the responses.

The development of high quality most formative assessments would involve the following key steps:

- Identify the Standard that you want to teach
- Deconstruct the Standard into smaller learning targets
    - Content
    - Thinking skills
    - Science inquiry or process skills
    - Potential student products or performances
- Identify potential student misconceptions for the learning targets
- Design student formative assessments that clearly identify the learning targets and provide opportunities for students to reflect on their strengths, evidence of strengths, areas for improvement, support.
- Design or acquire very short cycle formative assessments that provide opportunities for students to demonstrate their understanding of the learning targets or their misconceptions. These assessments should be used with all students.
- Interpret the results from the short cycle assessments and modify instruction or provide individualized support to students.
- Share the assessments and student work with your teacher team and make modifications to the system as necessary. Share instructional strategies to better support students who have misconceptions.

Research Support:
There is abundant research evidence that supports the effectiveness of frequent formative assessments in improving student academic achievement. Robert Marzano, in his book called Classroom Assessment Grading that Work synthesizes the work of many educational researchers who have conducted comprehensive reviews of the effectiveness of formative assessments. Key findings from this research and reported by Marzano include:

- Feedback from classroom assessments should give students a clear picture of their progress on learning goals and how they might improve.
- Feedback on classroom assessments should encourage students to improve.
- Classroom assessment should be formative in nature.
- Formative classroom assessment should be frequent.

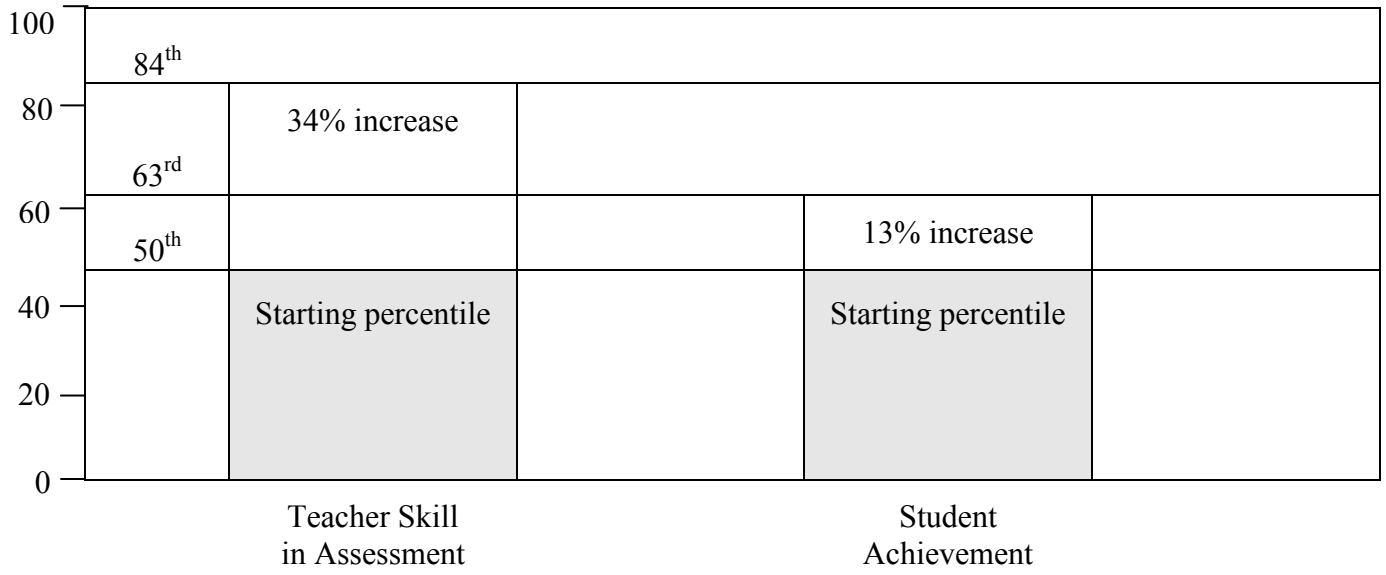
The figures below demonstrate the research results that support the effectiveness of formative assessments.

Click Here to Return to Main Page

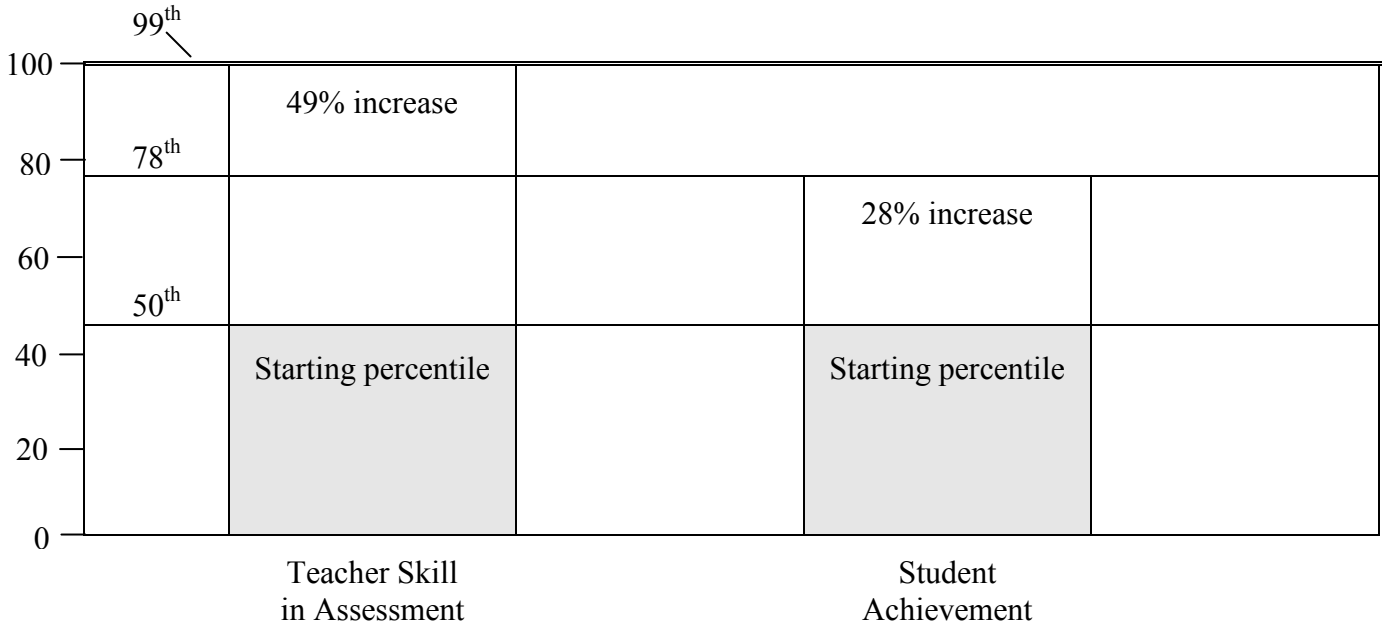
Figure 1.1: This figure taken from Classroom Assessment that works shows significant improvement in student improvement when teachers' ability to use formative assessment improves.

FIGURE 1.1
Effect of Teacher's Increased Skill in Classroom Assessment on Student Achievement

Predicted increase in student achievement when teacher's skill in classroom assessment increases from 50th to 84th percentile.



Predicted increase in student achievement when teacher's skill in classroom assessment increases from 50th to 99th percentile.



Click Here to Return to Main Page

Figure 1.2: This figure taken from Classroom Assessment that works demonstrates the effect sizes of various types of formative assessment in improving student achievement. A 0.2 Effect size is considered small, a 0.4 Effect Size is considered moderate, Effect sizes above 0.6 are considered strong. When teachers evaluate student work using clearly defined rules that are described in a rubric, there is the strongest effect size of 0.91 or a potential percentile gain of 32 points.

FIGURE 1.2
Findings on the Effects of Different Types of Feedback

Source	Characteristics of Feedback from Classroom Assessment	Number of Studies	Effect Size	Percentile Gain or Loss in Student Achievement
Bangert-Drowns, Kulik, Kulik, & Morgan (1991)	Right/wrong	6	-.08	-3
	Provide correct answer	39	.22	8.5
	Criteria understood by students vs. not understood	30	.41	16
	Explain	9	.53	20
	Repeat until correct	4	.53	20
Fuchs & Fuchs (1986)	Displaying results graphically	89	.70	26
	Evaluation (interpretation) by rule	49	.91	32

*Indicates the number of studies that were examined by the researchers to compute an effect size. See Technical Note 1.2 for discussion of an effect size.

Click Here to Return to Main Page

Figure 1.3: This figure taken from Classroom Assessment that works demonstrates the effect sizes for the increased use of formative assessments. There is a dramatic increase in the number of formative assessments used from 1 to 10 with a slower increase in effect size after 10 formative assessment administrations.

CLASSROOM ASSESSMENT & GRADING THAT WORK

FIGURE 1.3
Gain Associated with Number of Assessments over 15 Weeks

Number of Assessments	Effect Size	Percentile-Point Gain
0	0	0
1	.34	13.5
5	.53	20.0
10	.60	22.5
15	.66	24.5
20	.71	26.0
25	.78	28.5
30	.80	29.0

Note: Effect sizes from data reported by Barugers-Drowns, Kulik, and Kulik (1991).

Click Here to Return to Main Page

Research Resources for Most Formative Assessments:

- Inside the B lack Box: Raising Standards through Classroom Assessments. Paul Black and Dylan Wiliam. Phi Delta Kappan, October 1998. (http://www.edsource.org/edu_ass_res_black.cfm)
- Classroom Assessments *for* Student Learning by Rick Stiggins, Judith Arter, Jan Chappuis, and Steve Chappuis
- Classroom Assessment & Grading that Work by Robert J. Marzano. Association for Supervision and Curriculum Development. Alexandria, VA. 2006.
- PD360 Formative Assessment Professional Development: http://www.pd360.com/pd360.cfm?
- Test Better, Teach Better: the Instructional Role of Assessment by W. James Popham. Association for Curriculum and Supervision. Alexandria, VA. 2003.
- *Science Formative Assessment:75 Practical Strategies for Linking Assessment, Instruction, and Learning* by Page Keeley. Corwin Press. London. 2008.

Click Here to Return to Main Page

More Formative Assessments

http://assessmentcontinuum.wetpaint.com/page/Screening%2FDiagnostic+Assessments

Description:
When teachers assess for learning using more formative assessments, they use a classroom assessment process to generate a continuous flow of information about student achievement to advance and not merely check in on student learning. This kind of formative assessment provides teachers with important information about the degree to which individual and groups of students achieve specific standards-aligned instructional targets that they are teaching. More formative assessments will help to inform key elements of the instructional process including how to differentiate instruction; plan and conduct academic conferences and data-informed grade level meetings. When combined with technology systems that provide timely and organized results, along with focused instructional interventions, more formative assessments provide teachers with a powerful integrated system for improving student achievement during the instructional process.

These assessments are different than most formative assessments in that they are not instantaneous and require more time to collect, organize visualize and interpret the data in support of student learning. Since they are used to inform and improve student learning during the instructional process, they are not graded. Examples of more formative assessments include the following:

- **Common Assessments**: http://assessmentcontinuum.wetpaint.com/page/Common+Assessments
A common assessment is a uniform tool administered consistently by all teachers in a grade level or course within a given time frame. These assessments use various levels of depth of knowledge and various types of questions to measure students' understanding of team developed learning targets and/or standards. Scoring is done collaboratively using a uniform and mutually developed scoring guide which is understood by students and staff. Results are shared among the teacher team to identify student strengths and areas for improvement. Additionally the results can be used to identify effective teaching practices. The key qualities of common assessments: · Grade-level and/or department teacher teams identify common learning targets and/or skills that they will teach over a relatively short period of time.

  Teacher teams collaborate on building a common assessment that aligns with the identified instructional learning targets. · Teachers teach the unit and then administer the common assessment. · Teachers review the results and then to identify student strengths and areas for improvement. They also identify successful teaching practices. · Teacher teams work collaboratively to develop new instructional strategies to support those students who did not achieve the instructional targets.

  Key processes involved in the development of common assessments include:
  - Identify the Standard that you want to teach.
  - Deconstruct the Standard into smaller learning targets
    - Content
    - Thinking skills
    - Science inquiry or process skills
    - Prerequisite skills
  - **Identify potential student misconceptions for the learning targets. Check on your own possible misconceptions as well.**
  - Design student formative assessments that clearly identify the learning targets and provide opportunities for students to reflect on their strengths, evidence of strengths, areas for improvement, and support.

Click Here to Return to Main Page

- Design or acquire very short cycle formative assessments that provide opportunities for students to demonstrate their understanding of the learning targets or their misconceptions. These assessments should be used with all students.
- Interpret the results from the short cycle assessments and modify instruction or provide individualized support to students.
- Share the assessments and student work with your teacher team and make modifications to the system as necessary. Share instructional strategies to better support students who have misconceptions.
- Do the formative process again by observing and monitoring the implementation of the enhanced formative assessment.
- Celebrate your success and student success!

Screening assessments:

Screening assessments support the identification of students who may be at risk within an academic area. The DIBELS test is an example of a screening test used within the District to identify the students who might be at risk in the key reading elements of phonemic awareness, phonics, vocabulary, fluency, and comprehension. Results from screening assessment can be to identify students who are on track and students who are at risk. Students who are at risk should participate in diagnostic assessments that can pinpoint specific student learning needs. There are three key goals to early reading screening assessments:

  - Goal 1: Identify children who are failing to make progress in acquiring crucial early literacy skills.
  - Goal 2: Monitor reading development to determine whether children are keeping up with grade level expectations.
  - Goal 3: Monitor the progress of individual students in prevention and intervention programs so that educational programming can be modified if needed.

- **Diagnostic assessments**: Diagnostic assessments are comprehensive assessments that are administered to identify specific key skills where students need support. Diagnostic assessments can be administered to students who have been identified as at-risk through a screening assessment. If a screening assessment, determined that a group of 1st grade students had difficulty in phonological awareness, a diagnostic assessment like the Comprehensive Test of Phonological Processing (CTOPP) might be used to identify specific phonological skills that individual students might be having difficulty with. Combined with focused instruction geared to address the specific phonological needs of students, will result in improved reading skills.

- **Ongoing monitoring assessments**: Ongoing-monitoring assessments are used by teachers in conjunction with the main instructional program to determine the extent to which all students demonstrate the standards as well as the key ideas and skills embedded within the standards. There are three important types of ongoing monitoring assessments. These include interim assessment; short-cycle, and curriculum-embedded assessments. These assessments are described below:

Click Here to Return to Main Page

Research Support for Common Assessment;
Much of the research supporting the effectiveness of common assessments has been done by Rick DuFour, Becky DuFour, and Robert Eaker. A synthesis of their findings using a key ideas format can be found below:

1. **Team-developed common assessments are more efficient.**
   If five teachers teaching the same course or grade level are responsible for ensuring all students acquire the same knowledge and skills, it make sense those teachers would work together to determine the best methods to assess student learning. A team of teachers could divide responsibilities for creating a unit and developing assessments. Teachers working in isolation replicate and duplicate effort. They work hard, but they do not work smart.

2. **Team-developed common assessments are more equitable.**
   The use of common assessments increases the likelihood that students will have access to the same curriculum, acquire the same essential knowledge and skills, take assessments of the same rigor, and have their work judged according to the same criteria. We have witnessed repeated examples of teachers who were emphatic about the need for consistency, equity, and fairness in terms of how they were dealt with as adults, being completely unconcerned about the inconsistency, inequity, and lack of fairness that characterized the assessment of student learning in their school. If every teacher has license to assess whatever and however he or she determines, according to criteria unique to and often known only by that teacher, schools will never be institutions that truly model a commitment to equity.

3. **Team-developed common formative assessments are more effective in monitoring and improving student learning.**
   We have cited several researchers who have concluded that team-developed common formative assessments are one of the most powerful strategies available to educators for improving student achievement. We know of no research concluding the formal assessments created by individual teachers working in isolation advance student learning.

4. **Team-developed common formative assessments can inform and improve the practice of both individual teachers and teams of teachers.**
   Teachers do not suffer from a lack of data. Virtually every time a teacher gives an assessment of any kind, the teacher is able to generate data – mean, mode, median, standard deviation, percentage failing, percentage passing, and so on. As Robert Waterman (1987) advised, however, data alone do not inform practice. Data cannot help educators identify the strengths and weaknesses of their strategies. Data inform only when they are presented in context, which almost always requires a basis of comparison.
   Most educators can teach an entire career and not know if they teach a particular concept more or less effectively than the teacher next door because the assessments they generate for their isolated classrooms never provide them with a basis of comparison. Most educators can assess their students year after year, get consistently low results in a particular area, and not be certain if those results reflect his or her teaching strategies, a weakness in the curriculum, a failure on the part of teachers in earlier grades to ensure students develop a prerequisite skill, or any other cause. In short, most educators operate within the confines of data, which means they operate in the dark. But in a PLC, collaborative teams create a series of common assessments, and therefore every teacher receives ongoing feedback regarding the proficiency of his or her students, in achieving a standard the team has agreed is essential, on an assessment the team has agreed represents a valid way to assesses what members intend for all students to learn, in comparison to other students attempting to achieve the same standard. That basis of comparison transforms data into information.

   Furthermore, as Richard Elmore (2006) wrote, “teachers have to feel that there is some compelling reason for them to practice differently, with the best direct evidence being that students learn better” (p. 38). When

teachers are presented with clear evidence their students are not becoming proficient in skills they agreed were essential, as measured on an assessment they helped to create, and that similar students taught by their colleagues have demonstrated proficiency on the same assessment, they are open to exploring new practices. When the performance of their students consistently prevents their team from achieving its goals, they are typically willing to address the problem. In fact, we consider team-developed common formative assessments one of the most powerful motivators for stimulating teachers to consider changes in their practice.

5. **Team-developed common formative assessments can build the capacity of the team to achieve at higher levels.**
   As Wiliam and Thompson (2007) found, the conversations surrounding the creation of common formative assessments are a powerful tool for professional development. When schools ensure every teacher has been engaged in a process to clarify what students are to learn and how their learning will be assessed, they promote the clarity essential to effective teaching. When teachers have access to each other's ideas, methods, and materials they can expand their repertoire of skills. When a team discovers the current curriculum and their existing instructional strategies are ineffective in helping students acquire essential skills, its members are able to pursue the most powerful professional development because it is specific, job-embedded and relevant to the context of their content, their strategies, their team, and their students.

6. **Team-developed common formative assessments are essential to systematic interventions when students do not learn.**
   We argue that if educators were truly committed to high levels of learning for all students, they would not leave the question, "what happens when some students do learn" to chance. They would instead, work together to create systems of intervention to ensure any student who struggles receives additional time and support for learning in a timely and directive way. Team-developed common formative assessments are a critical element of that system of intervention.

Not every assessment should be a common assessment. There is still a place for individual teachers to create their own formal assessments. Team-developed common assessments will never eliminate the need for individual teachers to monitor student learning each day through a wide variety of strategies that check for understanding. But if schools are ever to take full advantage of the power of assessment to impact student learning in a positive way, they must include common formative assessments in their arsenal. Professional learning communities will make team-developed common formative assessments a cornerstone of their work.

Click Here to Return to Main Page

Research Support for Screening, Diagnostic and Ongoing Monitoring Assessments:
Phonemic Awareness
"The best predictor of reading difficulty in kindergarten or first grade is the inability to segment words and syllables into constituent sound units (phonemic awareness)" (Lyon, 1995).

The ability to hear and manipulate phonemes plays a causal role in the acquisition of beginning reading skills (Smith, Simmons, & Kame'enui, 1998).

The effects of training phonological awareness and learning to read are mutually supportive. "Reading and phonemic awareness are mutually reinforcing: Phonemic awareness is necessary for reading, and reading, in turn, improves phonemic awareness still further." (Shaywitz)

Phonological awareness is teachable and promoted by attention to instructional variables (Smith, Simmons, & Kame'enui, 1998).

Alphabetic Principles
Letter-sound knowledge is prerequisite to effective word identification. A primary difference between good and poor readers is the ability to use letter-sound correspondence to identify words (Juel, 1991).

Because our language is alphabetic, decoding is an essential and primary means of recognizing words. There are simply too many words in the English language to rely on memorization as a primary word identification strategy (Bay Area Reading Task Force, 1996).

The table below illustrates the important correlation between the ability to decode words and reading comprehension.

TABLE 1
Correlations Between Decoding and Comprehension
in the Connecticut Longitudinal Study

	DECODING								
Compre-hension	**GR. 1**	**GR. 2**	**GR. 3**	**GR. 4**	**GR. 5**	**GR. 6**	**GR. 7**	**GR. 8**	**GR. 9**
Grade 1	.89								
Grade 2	.75	.83							
Grade 3	.70	.74	.77						
Grade 4	.64	.71	.74	.73					
Grade 5	.58	.63	.68	.67	.70				
Grade 6	.59	.65	.67	.68	.66	.69			
Grade 7	.53	.61	.65	.65	.67	.68	.69		
Grade 8	.49	.58	.62	.62	.64	.65	.65	.63	
Grade 9	.52	.58	.60	.62	.60	.63	.63	.61	.63

Note: All correlations are significant at $p<.001$ and sample sizes range from 390 to 403.

(Foorman, et. al., 1997)

Click Here to Return to Main Page

Fluency and Accuracy with Text
Successful readers...

- Rely primarily on the letters in the word rather than context or pictures to identify familiar and unfamiliar words.
- Process virtually every letter.
- Use letter-sound correspondences to identify words.
- Have a reliable strategy for decoding words.
- Read words for a sufficient number of times for words to become automatic.

(Hasbrouck, 1998)

Proficient readers are so automatic with each component skill (phonological awareness, decoding, and vocabulary) that they focus their attention on constructing meaning from the print (Kuhn & Stahl, 2000).

Vocabulary
Children enter school with "meaningful differences" in vocabulary knowledge.

- **What doesn't matter:** race/ethnicity, gender, birth order.
- **What does matter**: relative economic advantage.

1. Emergence of the Problem

In a typical hour, the average child hears:

Family Status	Actual Differences in Quantity of Words Heard	Actual Differences in Quality of Words Heard
Welfare	616 words	5 affirmations, 11 prohibitions
Working Class	1,251 words	12 affirmations, 7 prohibitions
Professional	2,153 words	32 affirmations, 5 prohibitions

2. Cumulative Vocabulary Experiences

Family Status	Words heard per hour	Words heard in a 100-hour week	Words heard in a 5,200 hour year	Words heard in 4 years
Welfare	616	62,000	3 million	13 million
Working Class	1,251	125,000	6 million	26 million
Professional	2,153	215,000	11 million	45 million

Click Here to Return to Main Page

3. **Meaningful Differences**

By the time the children were 3 years old, parents in less economically favored circumstances had said fewer different words in their cumulative monthly vocabularies than the children in the most economically advantaged families in the same period of time.

Cumulative Vocabulary	
Children from welfare families:	500 words
Children from working class families:	700 words
Children from professional families:	1,100 words

(Hart and Risley, 1995)

Comprehension

Causes of Reading Comprehension Failure

- Inadequate instruction
- Insufficient exposure and practice
- Deficient word recognition skills
- Deficient memory capacity and functioning
- Significant language deficiencies
- Inadequate comprehension monitoring and self-evaluation
- Unfamiliarity with text features and task demands
- Undeveloped attentional strategies
- Inadequate cognitive development and reading experiences

Kame'enui & Simmons, 1990

Types of Comprehension Instruction that Have Evidence of Improving Comprehension

- Comprehension monitoring
- Cooperative learning
- Multiple strategies
- Mental imagery / mnemonics
- Graphic organizers
- Summarization
- Semantic organizers including:
    - story maps
    - question answering
    - question generation

Kame'enui & Simmons, 1990

Click Here to Return to Main Page

More Formative Assessments Research Resources:

Common Assessments
Research Support for Common Assessments:
Rick DuFour, Robert Eaker, and Rebecca DuFour. 2005. *On Common Ground – The Power of Professional Learning Communities*. Solution Tree. Bloomington, IN.

Screening, Diagnostic, and Ongoing Monitoring Assessments
Bay Area Reading Task Force (1997). A reading-writing-language source book for the primary grades. San Francisco, CA: University School Support for Educational Reform.

Foorman, B. R., Francis, D. J., Shaywitz, S. E., Shaywitz, B. A., & Fletcher, J. M. (1997). The case for early reading intervention. Hillsdale, NJ: Erlbaum.

Hart, B., & Risley, R. T. (1995). Meaningful differences in the everyday experience of young American children. Baltimore: Paul H. Brookes.

Hasbrouck (1998). Reading fluency: Principles for instruction and progress monitoring. Professional Development Guide. Austin, TX: Texas Center for Reading and Language Arts, University of Texas at Austin.

Juel, C. (1991). Beginning reading. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), Handbook of reading research (pp. 759-788). New York: Longman.

Kame'enui, E. J. & Simmons, D. C. (1990). Designing instructional strategies: The prevention of academic learning problems. Columbus, OH: Merrill Publishing Company.

PD360 Formative Assessment Professional Development: http://www.pd360.com/pd360.cfm?

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J.M., & Makuch, R. (1992). Distribution and temporal stability of dyslexia in an epidemiological sample of 414 children followed longitudinally. New England Journal of Medicine, 326, 145-150.

Smith S. B., Simmons, D. C., & Kame'enui, E. J. (1998). Phonological awareness: Instructional and curricular basics and implications. In D. C. Simmons & E. J. Kame'enui (eds.), What reading research tells us about children with diverse learning needs: Bases and basics. Mahwah, NJ: Lawrence Erlbaum Associates.

Lyon, G. R. (1995). Toward a definition of dyslexia. Annals of Dyslexia, 45, 3-27.

Click Here to Return to Main Page

More Summative Assessments

http://assessmentcontinuum.wetpaint.com/page/Summative+Assessments

Description:
More summative assessments are used at the end of a longer instructional cycle of 6-8 weeks and are meant to measure student achievement of learning targets after the student has had many opportunities to achieve and demonstrate mastery of the targets.

Benchmark Assessments:
http://assessmentcontinuum.wetpaint.com/page/Benchmark+Assessments
Benchmark assessments are more summative and assess student overall achievement of key standards and/or skills at near the end of each grading period. The results of these assessments provide both school and district staff with strategic information about student achievement that can inform key indicators of school and district strategic focus including curriculum development, professional development, and mid-course corrections necessary to achieve strategic goals. Assessment items are aligned to the standards taught in ELA and/or Math. Assessment results are used within results-based inquiry protocols at the school and district levels to inform the success of key initiatives and programs.

Northwest Evaluation Assessment (NWEA):
http://www.nwea.org/
What are Measures of Academic Progress, or MAP tests?
MAP assessments are adaptive achievement tests in Mathematics, Reading, Language Usage, and Science that are taken on a computer.
What are computerized adaptive tests?
Computerized adaptive tests are taken on a computer. The difficulty of a test is adjusted to the student's performance so each student sees different test questions. The difficulty of each question is based on how well the student has answered the questions up to that point. As the student answers correctly, the questions become more difficult. If the student answers incorrectly, the questions become easier.
How often can a student be tested?
Districts have the option of testing their students up to four times a year. Students typically take tests at the beginning of the school year and at the end of the school year. Some districts may also choose to test students in winter and summer.
Do all students in the same grade take the same test?
No. The computer adjusts the difficulty of the questions so that each student takes a unique test.
What types of scores are reported on NWEA MAP assessments?
Percentile scores
Achievement scores
Growth scores
What is the RIT scale?
The scale NWEA uses to measure a student's progress is called the RIT scale, short for Rasch Unit. The RIT scale is an equal-interval scale much like centimeters on a meter stick. It is used to chart a student's academic growth from year to year.
If you would like more information on the RIT scale, refer to the following materials:

Hambleto, R.K. , Swaminathan, H., & Rogers, J. (1991). Fundamentals of Item Response Theory. Newbury Park: Sage.

Ingebo, G. (1997). Probability in the Measure of Achievement. Chicago: MESA Press

Wright, B.D. & Stone, M. (1979). Best Test Design. Chicago: MESA Press.

Visit http://edres.org/irt/ for introductory material about the item response theory and the Rasch model.
Click Here to Return to Main Page

Research Support: (Source: WestEd):
http://www.wested.org/cs/we/view/feat/159)
Schools across the nation are using benchmark assessments to boost student learning, but does the practice work? "There hasn't been a lot of research on the impact of benchmarks," says WestEd Senior Research Associate Susan Henderson. In response to that need, Henderson and a team of researchers analyzed math achievement data from middle schools across Massachusetts over a five-year period, culminating in 2006, the end of the schools' first year implementing benchmark assessments. Their report has just been released.

Regular use of benchmark assessments, particularly when aligned with state content standards, is widely seen as having potential to improve student performance. While annual state testing provides summative measures of achievement, the results are available only after students have moved to the next grade. In contrast, benchmarks are scored immediately, providing valuable information that can alert teachers and administrators to learning gaps before students move on. In a 2005 survey, approximately 70 percent of school superintendents reported their districts used benchmark assessments.*

The Massachusetts Department of Education backs that trend. The department is "committed to exploring strategies for supporting interim and benchmark assessment practices," says Carrie Conaway, the department's Director of Planning, Research, and Evaluation. "Benchmark assessments are an important part of our strategy to establish support systems that enable all students to reach proficiency."

Funded by the U.S. Department of Education through the Northeast and Islands Regional Educational Laboratory (NEIREL), Henderson and coauthors examined data from 66 middle schools, representing eight Massachusetts districts, to determine if they could measure impact of benchmark assessments on student achievement. Twenty-two of the schools had received grants to develop and use benchmark assessments. The recipients serve high-poverty populations and had been identified as needing significant improvement in math performance. The funding was to help the schools use a data management system to create their own grade-level assessments aligned with Massachusetts Curriculum Frameworks.

Before the state asked NEIREL to evaluate the initiative's impact, Massachusetts had already selected which schools would participate, so researchers could not conduct a randomized trial. "We had to use other tools in the research toolbox," notes Anthony Petrosino, WestEd Senior Research Associate and one of the study coauthors. "We ended up using a statistical matching procedure to equate the schools that didn't get the intervention to those that did."

Through covariate matching, the researchers identified 44 Massachusetts schools that did not receive grants to implement benchmark assessments but did share income, demographic, and social characteristics, and whose students' math scores were nearly identical to those of the program schools prior to benchmark implementation. Thomas Hanson, a WestEd expert in research methodology, helped design "an interrupted time-series analysis" for the study. It involved examining scores on the math portion of the Massachusetts State Comprehensive Assessment System (MCAS) for multiple years prior to the benchmark assessment intervention to determine a trend in achievement, then analyzing whether the intervention "interrupted" or impacted that trend.

Results were inconclusive at the end of the first year of benchmark implementation, with no statistical difference in math scores between schools that administered the benchmarks and the comparison schools. Average scores in both groups fell within the "needs improvement" designation.

But the data indicate the possibility that a second year of benchmarks may bring more tangible gains, Henderson says, a result the WestEd group is preparing to study further. One interesting factor the researchers note: As schools increasingly respond to a push for more frequent testing, even the comparison schools may have been using benchmark assessments during the study.

Click Here to Return to Main Page

The Massachusetts Department of Education has found the research compelling enough to warrant another round. "The NEIREL study provides a valuable perspective on the impact of benchmark assessments in the pilot schools," Conaway says. "We would've been surprised to see conclusive results after just one year of implementation, so we intend to continue to review results as the program progresses."

Henderson also sees this initial year as a starting point. "We'd love to look at how schools are using data to inform instructional practice," she says, including changes to instructional calendars, curriculum mapping, re-teaching, and other classroom strategies based on what benchmark assessments reveal. Henderson and colleagues also hope to study how teachers are "unpacking" state standards to identify the "essential" standards for testing. Then, she says, they can track changes in skill sets — measuring performance in specific operations, for example, as opposed to global math scores.

In Massachusetts and across the country, educators are eager for a deeper understanding of how benchmark assessment improves student learning. "These initial results, along with other evaluation and research findings," says Conaway, "will guide future policy decisions and targeted assistance activities."

For more information, contact Susan Henderson at 781.481.1118 or shender@WestEd.org or Stephen Hamilton at 781.481.1104 or shamilt@WestEd.org.

* Olson, L. (2005, November 30). Benchmark Assessments Offer Regular Checkups on Student Achievement. Education Week, pp. 13-14.

Click Here to Return to Main Page

More Summative Assessments Research Resources:

Abedi, J., & Gándara, P. (2006). Performance of English language learners as a subgroup in large-scale assessment: interaction of research and policy. Educational Measurement: Issues and Practice, 25(4), 36–46.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1).

Black, P., & Wiliam, D. (1998b). Inside the black box: raising standards through classroom assessment. Phi Delta Kappan, 80(2). Retrieved June 21, 2007.

Bloom, H. S. (2003). Using "short" interrupted time-series analysis to measure the impacts of whole-school reforms. Evaluation Review, 27(1), 3–49.

Boston, C. (2002). The concept of formative assessment. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Olson, L. (2005, November 30). Benchmark Assessments Offer Regular Checkups on Student Achievement. Education Week, pp. 13-14.

Susan Henderson. Anthony Petrosino. Sarah Guckenburg. Stephen Hamilton. 2005. Measuring How Benchmark Assessments Affect Student Achievement. Institute for Educational Sciences. U. S. Department of Education.

Click Here to Return to Main Page

Most Summative Assessments

http://assessmentcontinuum.wetpaint.com/page/Summative+Assessments

Description:
More summative assessments are used at the end of a longer instructional cycle of 6-8 weeks and are meant to measure student achievement of learning targets after the student has had many opportunities to achieve and demonstrate mastery of the targets monitored with many non-graded formative assessments. Most summative assessments are usually used at the end of a semester or at the end of the year. Results can be used to make key decisions for students. Results are also used to inform the success of the adults in helping students achieve key learning standards.

Assessments of learning are summative assessments that assess student overall achievement of key standards and/or skills on a regular basis throughout the school year. The results of these assessments provide both school and district staff with strategic information about student achievement that can inform key indicators of school and district strategic focus including curriculum development, professional development, and mid-course corrections necessary to achieve strategic goals. Since these assessments are designed to assess a wide range of standards and skills, they are less useful as formative opportunities to diagnose individual student academic successes and needs. Three types of assessments of learning include State Tests, District Benchmark assessments, and norm-referenced assessments. These assessments are described below.

- **State Assessments:** State Assessments are criterion-referenced assessments aligned to state standards. This means that the assessment items align to specific standards or standard strands for a given academic area like English Language Arts or Mathematics and for a specific grade level. The assessments are built to the specifications outlined in a blueprint for the test that shows the number and percentage of assessment items that align to specific standards and standard strands. The California Standards Test (CST) and the California High School Exit Examination (CAHSEE) are state assessments.

- **California English Language Development Test CELDT:**
  http://assessmentcontinuum.wetpaint.com/page/California+English+Language+Development+Test+%28CELDT%29
  State law (Education Code sections 313 and 60810) and federal law (Title III of the Elementary and Secondary Education Act [ESEA]) require that school districts administer a state test of English language proficiency (1) to newly enrolled students whose primary language is not English and (2) to students who are English learners as an annual assessment. For California public school students, this test is the California English Language Development Test (CELDT).

  The CELDT (instituted by Education Code sections 313 and 60810[d]) has three purposes:
  - To identify students who are limited English proficient
  - To determine the level of English language proficiency of students who are limited English proficient
  - To assess the progress of limited English proficient students in acquiring the skills of listening, reading, speaking, and writing in English.

Click Here to Return to Main Page

- **District Benchmark Assessments:** District benchmark assessments are built using the same blueprints that the state uses to construct the state assessment for a given academic area and grade level. If it can be verified that there is a close correlation between the district benchmark assessment and the state assessment then district benchmark assessments can play an important role in predicting student performance on the state assessment. Districts will usually acquire or build three forms of a benchmark assessment and then administer these assessments quarterly in combination with the state assessment. Since these assessments attempt to assess a wide range of key standards, there will be fewer assessment items per standard and thus this assessment will have less value as a formative opportunity to diagnose individual student achievement of specific standards.
- **Norm-referenced Assessments:** Norm-referenced are assessments that use a large representative sample of the general population such as national sample of 3rd grade students. This sample represents the norming group. Future performance of individuals and groups on this assessment are then compared to the performance of the norming group. The content for norm-referenced assessment represents a range of items in terms of academic content and difficulty. This assessment can provide information about how well individual and groups of students perform in a given academic area compared to a national group of similar students. Since it is not aligned to specific standards and there are a limited number of assessment items per skill, this kind of assessment has less usefulness as a formative tool. The California Achievement Test, 6th Edition Survey (CAT/6 Survey) is a norm-referenced assessment.

Most Summative Assessment
Research Support:

Robert Linn: Educational Accountability Systems. 2008
http://www.cse.ucla.edu/products/summary.asp?report=687

Test-based educational accountability systems have considerable appeal to politicians, policymakers, and the general public. Such systems have been widely used by states for more than a decade and with the enactment of the No Child Left Behind Act of 2001 all states must now implement an accountability system that uses results from assessments in mathematics and English/language arts and is administered each year in grades 3 through 8 plus one high school grade. A wide variety of test-based accountability systems of states and as required by NCLB are described and their strengths and weaknesses are evaluated. It is argued that the sanctions for schools that are part of the accountability system require causal inferences about school effectiveness. It is concluded, however, that basing causal inferences about school quality on the results that can be obtained from the existing school accountability systems is not scientifically defensible. It would be better to view accountability results as a source of descriptive information about schools and the basis of hypotheses that can be evaluated by gathering addition information about instructional staff and practice.

Robert L. Linn. Test-based Educational Accountability in the Era of No Child Left Behind 2005.
http://www.cse.ucla.edu/products/summary.asp?report=651

The ever-increasing reliance on student performance on tests as a way of holding schools and educators accountable is discussed. Comparisons are made between state accountability requirements and the accountability requirements of the No Child Left Behind (NCLB) Act of 2001. The resulting mixed messages being given by the two systems are discussed. Features of NCLB accountability and state accountability systems that contribute to the identification of a school as meeting goals according to NCLB but failing to do so according to the state accountability system, or vise versa, are discussed. These include the multiple hurdles of NCLB, the comparison of performance against a fixed target rather than changes in achievement, and the definition of performance goals. Some suggestions are provided for improving the NCLB accountability system.
Click Here to Return to Main Page

Robert L. Linn. Accountability: Responsibility and Reasonable Expectations. 2003
http://www.cse.ucla.edu/products/summary.asp?report=601

Some of the central features of current educational accountability systems are discussed using the requirements under the No Child Left Behind (NCLB) Act of 2001 as the primary example. It is argued that broadly shared responsibility is needed for accountability systems to contribute to improved education. It is also suggested that systems need to be designed in ways that are consistent with research and past experience. This requires the setting of ambitious performance standards and improvement targets, but ones that can reasonably be achieved given sufficient effort and supporting resources. These design features are contrasted with the NCLB requirements. Illustrations are provided of some of the state responses to the NCLB demands that attempt to avoid the over-identification of schools for improvement and sanctions.

Most Summative Assessments Research Resources

Ballou, D. (2004). Rejoinder. *Journal of Educational and Behavioral Statistics, 29*(1), 131-134.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37-65.

Bhola, D. S., Impara, J. C., & Buckendahl, W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice, 22*(3), 21-29.

California Department of Education. (2005, October). *2004-05 Academic performance index growth report: Information guide.* Sacramento, CA: Author. Available at http://www.cde.ca.gov/ta/ac/ap/documents/infoguide05g.pdf

Center on Education Policy. (2005). *States test limits of federal AYP flexibility*. Washington, DC: Author.

Education Week. (2005, January 5). Quality Counts 2005: No small change: Targeting money toward student performance. *Education Week, 24*(17).

Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational Researcher, 31*(8), 4-14.
Goals 2000: Educate America Act of 1994, Pub. L. No. 103-227 (1994).aertel, E. H., & Herman, J. L. (2005). A historical perspective on validity arguments for accountability testing. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data in accountability testing. Yearbook of the National Society for the Study of Education* (Vol. 104, Part I, pp. 1-34). Boston, MA: Blackwell Publishing.
Hamilton, L. (2003). Assessment as a policy tool. In R. L. Floden (Ed.), *Review of Research in Education, 27*, 25-68.

Hanushek, E. A., Kain, J. F., Harkman, J. M., & Rivkin, S. G. (2005). Does peer ability affect student achievement? *Journal of Applied Economics, 18*(5), 527-544.

Click Here to Return to Main Page

Hill, R. K., & DePascale, C. A. (2003). Reliability of No Child Left Behind accountability designs. *Educational Measurement: Issues and Practice, 22*(3), 12-20.

Improving America's Schools Act of 1994, Pub. L. No. 103-382 (1994).
Kane, M. (In press). *Validation*. In R. L. Brennan (Ed.), Educational measurement (4th ed.). Westport, CT: American Council on Education/Praeger.
25

Kim, J. S., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher, 34*(8), 3-13.

Klein, S. P., Hamilton, L. S., McCaffrey, D. F., & Stecher, B. M. (2000). *What do test scores in Texas tell us?* Santa Monica, CA: RAND.

Koretz, D. (2005). *Alignment, high stakes, and the inflation of test scores*. In J. L. Herman & E. H. Haertel (Eds.), Uses and misuses of data in accountability testing.

Yearbook of the National Society for the Study of Education (Vol. 104, Part I, pp. 99-118). Boston, MA: Blackwell Publishing.

Koretz, D. & Barron, S. I. (1998). *The validity of gains on the Kentucky Instructional Results Information System (KIRIS).* Santa Monica, CA: RAND.

Koretz, D., Linn, R. L., Dunbar, S. B., & Shepard, L. A. (1991). *The effects of high-stakes testing on achievement: Preliminary findings about the generalization of findings across tests.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4-14.

Linn, R. L. (2003, September 1). Performance standards: Utility for different uses of assessments. *Education Policy Analysis Archives, 11*(31). Retrieved September 1, 2003 from http://epaa.asu.edu/epaa/v11n31/

Linn, R. L. (2005a, June 28). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Educational Policy Analysis Archives, 13*(33). Retrieved June 30, 2005 from http://epaa.asu.edu/epaa/v13n33/.

Linn, R. L. (2005b). Issues in the design of accountability systems. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data in accountability testing. Yearbook of the National Society for the Study of Education* (Vol. 104, Part I, pp. 78-98). Boston, MA: Blackwell Publishing.

Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher, 31*(6), 3-16.

Click Here to Return to Main Page

Linn, R. L., Graue, M. E, & Sanders, N. M (1990). Comparing state and district test results to national norms: The validity of claims that "everyone is above average." *Educational Measurement: Issues and Practice, 9*(3), 5-14.
Linn, R. L., & Haug, C. (2002). The stability of school building scores and gains. *Educational Evaluation and Policy Analysis, 24*(1), 27-36.
26

Linn, R. L., & Herman, J. L. (1997). *A policymaker's guide to standards and assessment.* Denver, CO: Education Commission of the States. Available at

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational Statistics, 29*, 67-101.

McDonnell, L. M. (2004). *Politics, persuasion, and educational testing*. Cambridge, MA: Harvard University Press.

McDonnell, L. M. (2005). Assessment and accountability from the policymakers' perspective. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data in accountability testing. Yearbook of the National Society for the Study of Education* (Vol. 104, Part I, pp. 35-54). Boston, MA: Blackwell Publishing.

Meyers, R. H. (2000). *Value-added indicators: A powerful tool for evaluating science and mathematics programs and policies* (NISE Brief 3, No. 3). Madison, WI: National Center for Improving Science Education, University of Wisconsin-Madison.
National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: U.S. Government Printing Office.

No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2002).

Olson, L. (2002, January 9). Testing systems in most states not ESEA-ready. *Education Week, 31*(16), 1, 26-27.

Olson, L. (2005, September 2). Defying predictions, state trends prove mixed on schools making NCLB targets. *Education Week, 25*(2), 1, 26-27.

Olson, L. & Hoff, D. J. (2005). U.S. to pilot new gauge of growth: Ed. Dept. to permit shifts in how states track gains. *Education Week, 25*(13), 1, 16.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3-14.

Porter, A. C., Linn, R. L., & Trimble, S. (2005). The effects of state decisions about NCLB Adequate Yearly Progress targets. *Educational Measurement: Issues and Practice, 24*(4), 32-39.

Click Here to Return to Main Page

Raudenbush, S. W. (2004a). *Schooling, statistics, and poverty: Can we measure school improvement? The ninth annual William H Angoff Memorial Lecture*. Princeton, NJ: Educational Testing Service.

Raudenbush, S. W. (2004b). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics, 29*(1), 121-129.

Rouse, C. E. (2005). Accounting for schools: Economic issues in measuring school quality. In C. A. Dwyer (Ed.), *Measurement and research in the accountability era* (pp. 275-298). Mahwah, NJ: Lawrence Erlbaum Associates.

Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004). A potential outcomes view of value-added assessment. *Journal of Educational and Behavioral Statistics, 29*(1), 103-116.

Sanders, W., & Horn, S. (1998). Research findings from the Tennessee value added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12*(3), 247-256.

Spellings, M. (November 21, 2005). Letter to Chief State School Officers, announcing growth model pilot program, with enclosures. Available at: http://www.ed.gov/policy/elsec/guid/secletter/051121.html

U.S. Department of Education. (2004). Standards and assessments peer review guidance: Information and examples for meeting the requirements of the No Child Left Behind Act of 2001, April 28. Washington, DC: Author.
Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states* (Research Monograph No. 18). Madison WI: National Institute for Science Education.

The White House. (2001). Press conference with President George W. Bush and Education Secretary Rod Paige to Introduce the President's Education Program, Office of the Press Secretary, January 23.
http://www.whitehouse.gov/news/releases/2001/01/20010123-2.html

Zvoch, K. & Stevens, J. J. (2006, January 20). Successive student cohorts and longitudinal growth models: An investigation of elementary school mathematics performance. *Education Policy Analysis Archives, 14*(2). Retrieved January 20, 2006 from http://epaa.asu.edu/epaa/v14n2/

Click Here to Return to Main Page

Most Formative Assessment Samples:
http://assessmentcontinuum.wetpaint.com/page/Minute+to+Minute+Assessments

Agreement Circles

This is a sample formative assessment taken from the book called *Science Formative Assessment: 75 Practical Strategies Linking Assessment, Instruction, and Learning* by Page Keeley. This assessment not only provides the teacher with instant information about student understanding of learning targets but it is also motivational in that students enjoy the activity. It is originally intended for a science class but can easily be adapted to other subject areas

Description:
Agreement Circles provide a kinesthetic way to activate thinking and engage students in scientific argumentation. Students stand in a circle as the teacher reads a statement. The students who agree with the statement step to the center of the circle. They face their peers still standing in the circle and then match themselves up in small groups of students who agree and disagree. The small groups engage in discussion to defend their thinking. After discussion, the students are given an opportunity to reposition themselves with those who now agree standing in the center of the circle. The idea is to get every child either inside the circle or on the circumference. This process is repeated with several rounds of statements related to the same topic, each time with students starting by standing along the circumference of a circle.

How this Most Formative Assessment Supports Student Learning:
Agreement Circles activate students' thinking about science ideas related to a topic that they are studying. As the statements are mad, students access their existing knowledge. They must justify their thinking to their peers about why they agree or disagree with the statement. As they engage in a scientific argument with their "opposing partners" still standing on the circle, student may change their original ideas as new information convinces them that their original ideas may need adjustment and either step into or out of the circle.

How This Fact Informs Instruction:
This Most Formative Assessment can be used prior to instruction or during the concept development stage when formally introduced concepts may need reinforcement. The teacher can get a quick visual sense of students' understanding according to which part of the circle they are in. As the teacher circulates and listens to students' arguments, information about student thinking is revealed that can be used to design further learning experiences or revisit prior experiences aimed at developing conceptual understanding. Giving students an opportunity to change their position after discussion indicates the extent to which small group discussions may have changed some student initial thinking.

Click Here to Return to Main Page

Design and Administration:
Develop a set of three to five conceptually challenging statements related to the learning targets of instruction. Statements should be a combination of true and false. False statements can be developed based on examining the research on students' commonly held misconceptions or alternative views. For example, a set of eighth-grade statements used to elicit students' ideas about energy might be as follows:

- Energy is a material that is stored in an object.
- When energy changes from one form to another, heat is usually given off.
- Energy can never be created or destroyed
- Something has to move in order to have energy
- Energy is a type of fuel.

Begin by having students form a large circle. Read the first statement and then give students five to ten seconds to think. Ask students to move to the center of the circle if they agree with the statement and stay on the outside if they disagree. Match students up 1:2, 1:3, 1:4 1:5, or whatever the proportion of agree/disagree indicates and give them a few minutes to defend their ideas in small groups. Call time, read the question again, and have students reposition themselves according to whether their ideas have changed or stayed the same. Students who agree with the statement move to the inside of the circle. Note any changes and then have students go back to the circle for another round. When finished with all rounds, the next step depends o the stage of instruction. If the assessment was used to activate and elicit student thinking, the next step is to plan and provide lessons that will help the student explore ideas further and formulate understandings. If the assessment was used during the conceptual development stage, provide an opportunity for whole class discussion to resolve conceptual conflicts, formalize development of key ideas, and solidify understandings.

Click Here to Return to Main Page

Most Formative Assessment Samples:

http://assessmentcontinuum.wetpaint.com/page/Secondary+Formative+Assessment

Teacher Tool to Deconstruct the Standard into Learning Targets

In order for teachers to effectively use most formative assessments, they will need first to deconstruct the standard into its key elements including content. The tool below is a sample tool that teachers could use to deconstruct the standard. This tool was developed by Keith Nuthall and his team at the San Diego County Office of Education.

DECONSTRUCTING STANDARDS

<table>
<tr><td colspan="2">Content Area / Course:</td><td colspan="2">Grade Level:</td></tr>
<tr><td colspan="4">Strand / Cluster:</td></tr>
<tr><td colspan="4">Standard:

Student Friendly:</td></tr>
<tr><td>√ Product</td><td>☐ Skill</td><td>☐ Reasoning</td><td>☐ Knowledge</td></tr>
<tr><td colspan="4">Learning Targets
What are the knowledge, reasoning, skill and/or product expectations underpinning the standard?</td></tr>
<tr><td>Product Target</td><td>Skill Target</td><td>Reasoning Target</td><td>Knowledge Target</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>

- Highlight key phrases / terms that you feel will limit students' understanding of the target.

Click Here to Return to Main Page

Most Formative Assessment Samples:

http://assessmentcontinuum.wetpaint.com/page/Secondary+Formative+Assessment

Student Tool to Help Students Reflect on their Understanding of the Learning Targets

In order for students to effectively demonstrate their understanding of key learning targets they need to be able to explicitly see the learning targets and reflect on their understanding of the target, provide evidence that they have achieved the target, or identify areas where they need continued assistance and support. The tool below is a sample tool that students could use to deconstruct the standard. This tool was developed by Keith Nuthall and his team at the San Diego County Office of Education.

Student Goal/Weekly Reflection Sheet

Unit of Instruction:				
Learning Target(s) I am working on…	**Date**	**The evidence in my work that shows I clearly understand the target is…**	**The stretch that shows in my work (what I need to work on next) is…**	**The person that helped me reflect on this was…**

Click Here to Return to Main Page

Most Formative Assessment Samples:

Short Cycle Assessments:
http://assessmentcontinuum.wetpaint.com/page/Short+Cycle+Assessments

Short cycle assessments are examples of most formative assessments because the results of these assessments can be used immediately by teachers to diagnose student learning needs and intervene to support students on a continuum of thinking skills associated with the learning targets for the short cycle assessment. Short-cycle assessments are more like "diagnostic testlets". Each of our short-cycle assessments aligns with one or two standards and contains relatively few items. Short-cycle assessments are administered at the point in the instructional cycle where the items in the assessment specifically match the content of the classroom instruction. In fourth grade mathematics there will be 26 short-cycle assessments of 8 to 12 items each

Short-cycle assessments serve at least four purposes.

1. They provide teachers and students with high quality exemplars of the range and variety of assessment items implied in each of the standards.
2. They provide opportunities to help identify potential student misconceptions or naïve beliefs about content or skills implied within the standards.
3. If they are administered as pre-tests, they provide teachers with diagnostic information that can employed to make defensible decisions about on-going differential instruction and flexible grouping.
4. If they are administered as posttests, they generate summative information about student achievement that can be employed as a reliable and valid source of information for Standards-Based Report Cards.

Click Here to Return to Main Page

Califomia Standards
Sixth Grade Mathematics

By the end of grade six, students have mastered the four arithmetic operations with whole numbers, positive fractions, positive decimals, and positive and negative integers; they accurately compute and solve problems. They apply their knowledge to statistics and probability. Students understand the concepts of mean, median, and mode of data sets and how to calculate the range. They analyze data and sampling processes for possible bias and misleading conclusions; they use addition and multiplication of fractions routinely to calculate the probabilities for compound events. Students conceptually understand and work with ratios and proportions; they compute percentages (e.g., tax, tips, interest). Students know about pi and the formulas for the circumference and area of a circle. They use letters for numbers in formulas involving geometric shapes and in ratios to represent an unknown part of an expression. They solve one-step linear equations.

Measurement and Geometry

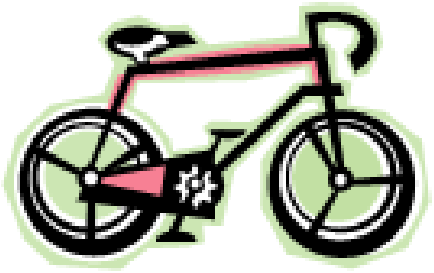
1.0 Students deepen their understanding of the measurement of plane and solid shapes and use this understanding to solve problems:

1.1 Understand the concept of a constant such as p; know the formulas for the circumference and area of a circle.

1.2 Know common estimates of p (3.14; 22/7) and use these values to estimate and calculate the circumference and the area of circles; compare with actual measurements.

1.3 Know and use the formulas for the volume of triangular prisms and cylinders (area of base x height); compare these formulas and explain the similarity between them and the formula for the volume of a rectangular solid.

Click Here to Return to Main Page



1. Kevin needs to know the circumference of his bicycle tire so he can attach streamers for a parade. He uses a piece of string to measure the diameter of the tire. The string is 20 inches long. Which equation would be the best choice to figure out the estimated circumference of the tire?

O a. $3 \times 20 = C$

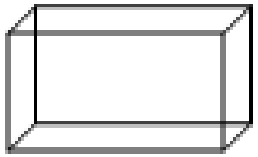
O b. $\left(\frac{1}{2} * 20\right)^2 \times 3 = C$

O c. $20 \times 20 = C$

O d. $(20 \times 20)^2 \div 3 = C$

Click Here to Return to Main Page

4.



Sonia wanted to package her 1 ½ inch cubed candies in a rectangular box. How could she estimate the volume of the box without filling it completely with inch cubes?

O a. Fill the bottom layer with inch cubes and multiply the number of cubes by 1 ½.

O b. Fill the bottom layer with inch cubes. Determine how many layers high will fit in the box and multiply the number of cubes in the bottom layer by the number of layers.

O c. Place the inch cubes in a row in the bottom of the box and try and guess how many total rows it would take to fill the box.

O d. Randomly place the inch cubes in the box and count how many there were total.

Click Here to Return to Main Page

Sixth Grade Short-Cycle Assessment-Measurement Standard-3

8.

56 inches high
About 34 inches wide
27 inches deep

Can A

Can B

54 inches high
31 inches in diameter

Hint: Volume of a rectangular prism= b x h x d
Volume of a cylinder= $\pi \cdot r^2 \cdot h$

Estimate the volume of each garbage can to determine which one would hold the most garbage. Elaborate on your reasoning.

Show your work and explain your reasoning below:

[Click Here to Return to Main Page]

8. Rubric for extended response.

4-points Estimates the volume of both garbage cans using the formulas and either rounding or front-end estimation. Based on their estimates uses correct mathematical terminology to choose the garbage can that has the larger volume.

Sample response **may** be:

Can A: $50 \times 30 = 1{,}500\ and\ 1{,}500 \times 20 = 30{,}000\ in.cubed$

Or, $60 \times 30 = 1{,}800\ and\ 1{,}800 \times 30 = 54{,}000\ in cubed$

Can B: $3 \bullet (15)^2 \bullet 50 = 33{,}750\ in cubed$

Or, $3 \bullet (16)^2 \bullet 50 = 38{,}400\ in cubed$

Explanation should demonstrate an understanding that the garbage can with the largest volume would hold the most garbage. This would be determined by the type of estimating that they did. For most estimates, Can A would have the largest volume, therefore hold the most garbage.

3-points Estimates the volume of both garbage cans with minor errors, includes a mathematically correct explanation based on the estimates they calculated.

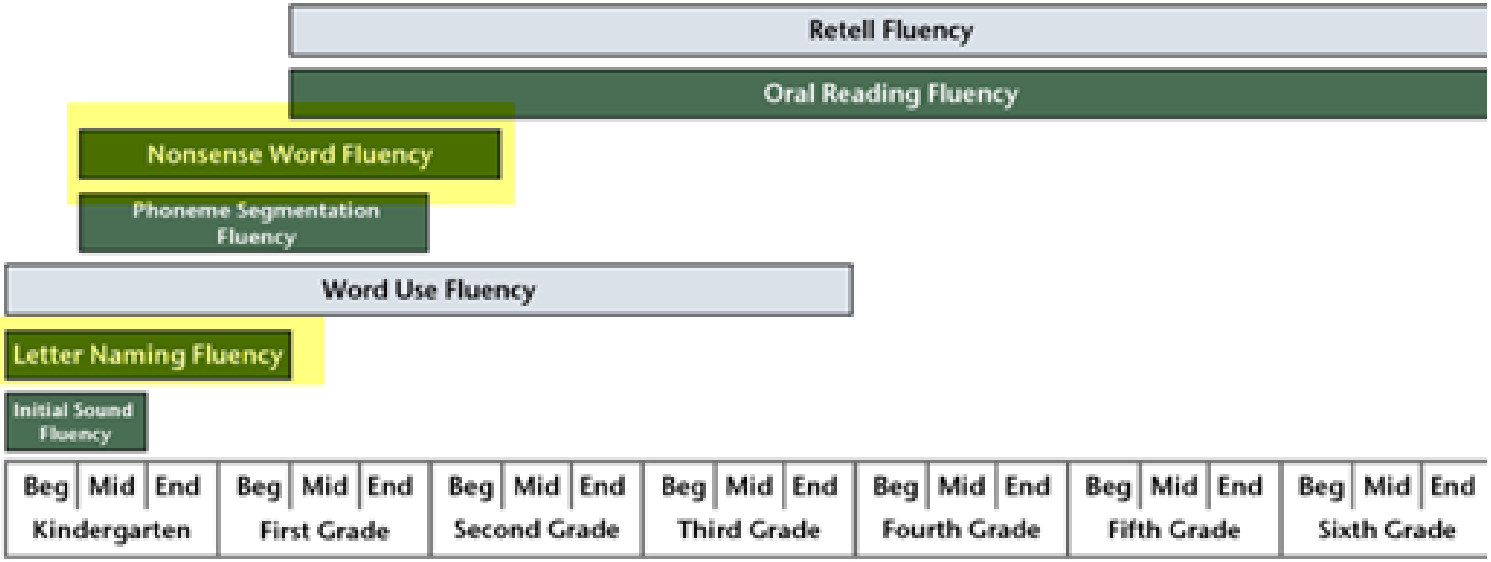
2-points Solves for an exact answer and then estimates **OR** attempts to solve the problem with multiple errors. Gives a correct explanation without showing any computation. Gets computations correct, but does not answer the question.

1-point Makes a minimal attempt to solve problem showing very little understanding of the task.

0-points No attempt, or attempt demonstrates no understanding of the concept.

More Formative Assessment Samples

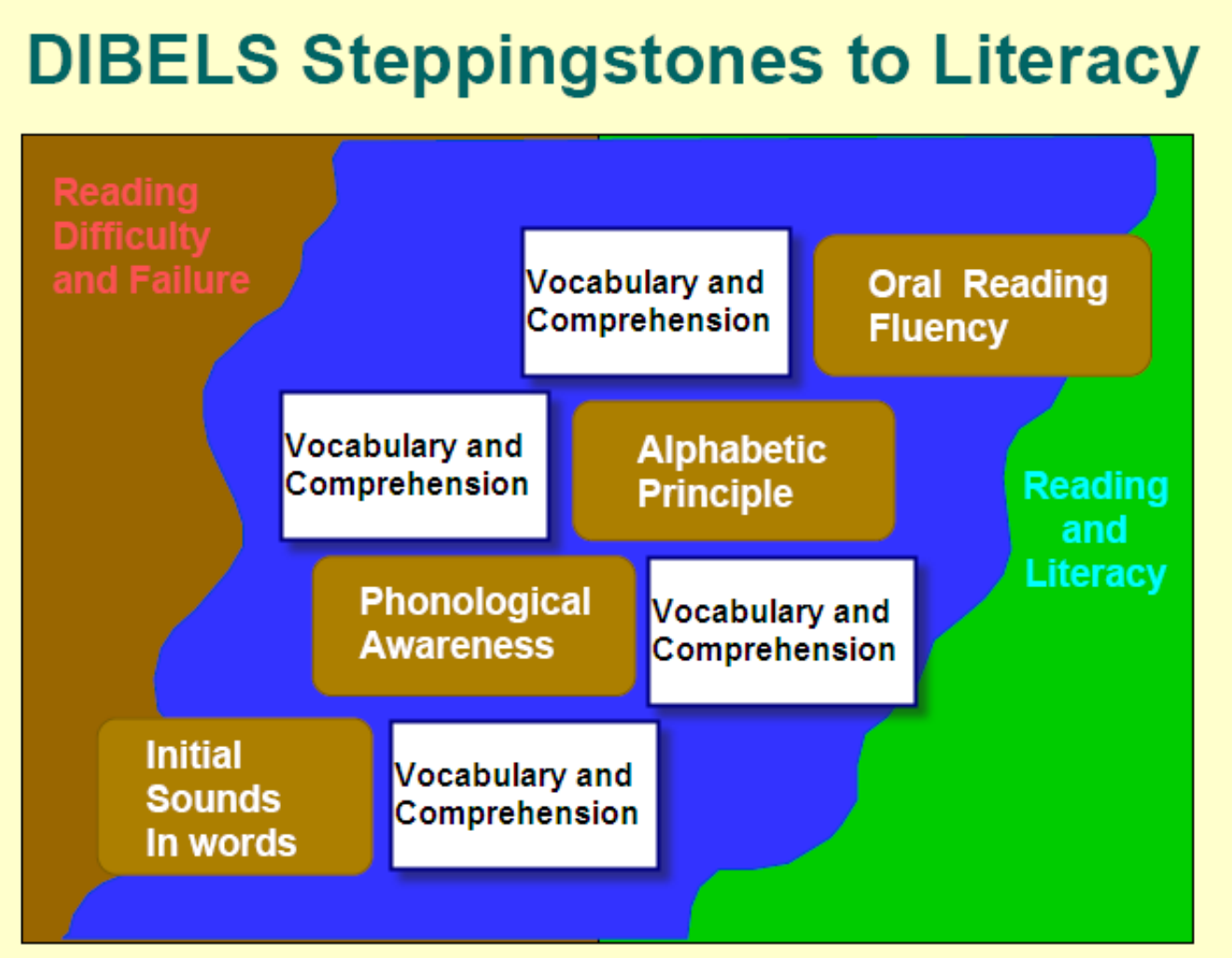
Screening Assessment Overview and Samples:
http://assessmentcontinuum.wetpaint.com/page/Screening%2FDiagnostic+Assessments



Source: Dynamic Indicators of Basic Early Literacy Skills Web site: https://dibels.uoregon.edu/

Click Here to Return to Main Page

More Formative Assessment Samples
Screening Assessment Overview:
http://assessmentcontinuum.wetpaint.com/page/Screening%2FDiagnostic+Assessments



Click Here to Return to Main Page

More Formative Assessment Samples

Screening Assessment Sample – Nonsense Word Fluency (Phonemic Awareness):
http://assessmentcontinuum.wetpaint.com/page/Screening%2FDiagnostic+Assessments

DIBELS® Training Institute

Role: Examiner

Essential Workshop: NWF

Nonsense Word Fluency
Short Form Directions

Look at this word (point to the first word on the practice probe). ***It's a make-believe word. Watch me read the word: /s/ /i/ /m/ "sim"*** (point to each letter then run your finger fast beneath the whole word). ***I can say the sounds of the letters, /s/ /i/ /m/*** (point to each letter), ***or I can read the whole word "sim"*** (run your finger fast beneath the whole word).

Your turn to read a make-believe word. Read this word the best you can (point to the word "lut"). ***Make sure you say any sounds you know.***

CORRECT RESPONSE: If the child responds "lut" or with some or all of the sounds, say	INCORRECT OR NO RESPONSE: If the child does not respond within 3 seconds or responds incorrectly, say
That's right. The sounds are /l/ /u/ /t/ or "lut"	***Remember, you can say the sounds or you can say the whole word. Watch me: the sounds are /l/ /u/ /t/*** (point to each letter) ***or "lut"*** (run your finger fast through the whole word). ***Lets try again. Read this word the best you can*** (point to the word "lut").

Place the student copy of the probe in front of the child.

Here are some more make-believe words (point to the student probe). ***Start here*** (point to the first word) ***and go across the page*** (point across the page). ***When I say "begin", read the words the best you can. Point to each letter and tell me the sound or read the whole word. Read the words the best you can. Put your finger on the first word. Ready, begin.***

 Page 4

Benchmark 1
Nonsense Word Fluency

k i k	w o j	s i g	f a j	y i s	___/15
k a j	f e k	a v	z i n	z e z	___/14
l a n	n u l	z e m	o g	n o m	___/14
y u f	p o s	v o k	v i v	f e g	___/15
b u b	d i j	s i j	v u s	t o s	___/15
w u v	n i j	p i k	n o k	m o t	___/15
n i f	v e c	a l	b o j	n e n	___/14
s u v	y i g	d i t	t u m	j o j	___/15
y a j	z o f	u m	v i m	v e l	___/14
t i g	m a k	s o g	w o t	s a v	___/15

Total: _____

Error Pattern:

 Page 5

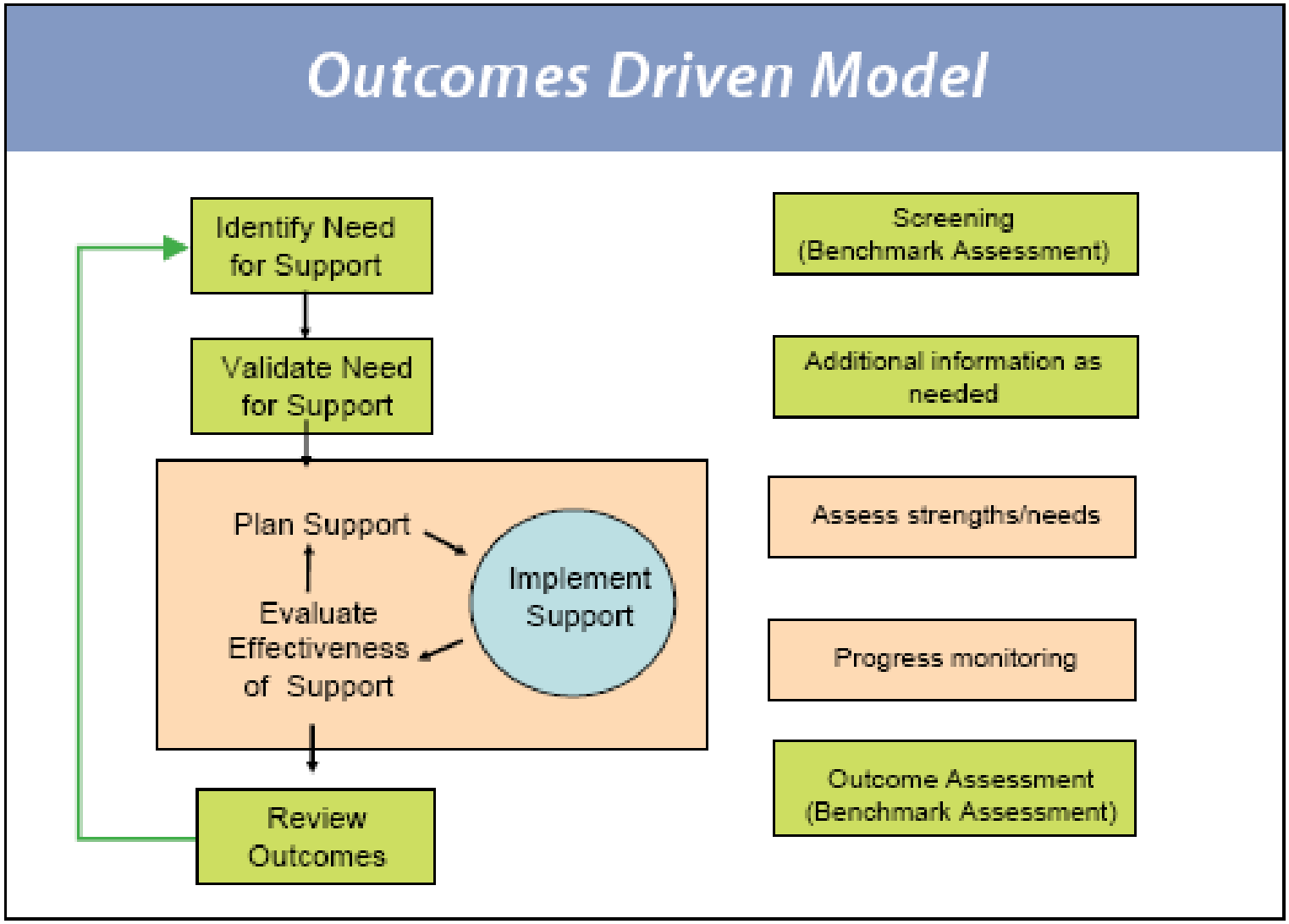
17

Click Here to Return to Main Page

More Formative Assessment Samples

Sample Protocol for Using the Results of Screening and Ongoing Monitoring Assessments

http://assessmentcontinuum.wetpaint.com/page/Screening%2FDiagnostic+Assessments



Click Here to Return to Main Page

More Summative Assessment Samples

This is a sample 7th grade math pacing guide for Hayward Unified School district. Pacing guides are the starting points for building Benchmark Assessments. Standards should be deconstructed into learning targets.
http://assessmentcontinuum.wetpaint.com/page/Benchmark+Assessments/file

HUSD **Suggested Timeline for Math 7** **2007-08 Draft Aug. 2007**
Textbook: *Course 2, McDougal, Littell*

Time	Topics: Content & Assessment	CA Standard
42 instructional days	**1st Quarter**	
13 days	**Chapter 1 – Operations with Numbers** Preliminary Class Activities/Set Up (2 days) 1.2 Expressions and Variables (2 days) 1.3 Powers and Exponents (2 days) 1.4 Order of Operations (2 days) 1.7 Commutative, Associative Properties (1 day) 1.8 Distributive Property (2 days) ***Review/Assessment (2 days)***	 NS 1.2, AF 1.3, 1.4 AF 2, 2.1, MG/MR 2.1 AF 1.2, 1.3 AF 1.3, (MR 1.2, 2.6) AF 1.3, MG 2.1 (MR 1.3, 2.4, 2.8)
9 days	**Chapter 2 – Operations in Algebra** 2.2 Combining Like Terms (2 days) 2.4 Translating Sentences into Equations (2 days) 2.5 Solving Equations Using Addition or Subtraction (1 day) 2.6 Solving Equations Using Multiplication or Division (1 day) 2.8 Solving Inequalities (1 days) ***Review/Assessment (2 days)***	 AF 1.3, 1.4 AF 1.1 AF 1.1, 1.3, 4.0 AF 1.1, 1.3, 4.0 MR 2.5 AF 4.0

Click Here to Return to Main Page

More Summative Assessment Samples

This is a sample test blueprint for Oakland Unified School District
http://assessmentcontinuum.wetpaint.com/page/Benchmark+Assessments/file

3rd Grade Mathematics Assessment Blueprint

Number Sense	PA1	SUM	PA2	PA3
		Exams		
1.0 Students understand the place value of whole numbers:				
1.1 Count, read, and write whole numbers to 10,000.				
1.2 Compare and order whole numbers to 10,000.				
***1.3** Identify the place value for each digit in numbers to 10,000.		5[1]	3	
1.4 Round off numbers to 10,000 to the nearest ten, hundred, and thousand.				
***1.5** Use expanded notation to represent numbers (e.g., 3,206 = 3,000 + 200 + 6).		4[2]	3	
2.0 Students calculate and solve problems involving addition, subtraction, multiplication, and division:				
***2.1** Find the sum or difference of two whole numbers between 0 and 10,000.		6[3]	5	
***2.2** Memorize to automaticity the multiplication table for numbers between 1 and 10.	4			
***2.3** Use the inverse relationship of multiplication and division to compute and check results.	4	3		3
***2.4** Solve simple problems involving multiplication of multidigit numbers by one-digit numbers (3,671 × 3 = __).				5
2.5 Solve division problems in which a multidigit number is evenly divided by a one-digit number (135 ÷ 5 = __).				
2.6 Understand the special properties of 0 and 1 in multiplication and division.	3			
2.7 Determine the unit cost when given the total cost and number of units.				
2.8 Solve problems that require two or more of the skills mentioned above.				

Click Here to Return to Main Page

More Summative Assessment Samples

These are sample problems from a 3rd grade OUSD Mid-Year Benchmark Assessment
http://assessmentcontinuum.wetpaint.com/page/Benchmark+Assessments/file

AF-2.1 • 2411

20. There are 15 pages in a notebook. Each page has 4 stickers on it. How many stickers are in the notebook?

A. 19

B. 60

C. 150

D. 154

AF-2.1 • 2440

21. Markers cost $12.00 per set. Mr. Green bought five sets of markers. How much did Mr. Green spend?

A. $7.00

B. $12.05

C. $17.00

D. $60.00

AF-1.1

22. One grasshopper has six legs. Which number sentence shows how to find the total number of legs on three grasshoppers?

A. 3 + 6 = ☐

B. 3 × 6 = ☐

C. 6 ÷ 3 = ☐

D. 6 – 3 = ☐

MG-2.2 • 2298

23. A triangle with three equal sides is called _____ triangle.

A. an equilateral

B. an isosceles

C. a right

D. an obtuse

Go on to the next page ➜

Click Here to Return to Main Page

More Summative Assessment Samples

This is a sample tool that can be used to judge benchmark assessment items
http://assessmentcontinuum.wetpaint.com/page/Benchmark+Assessments/file

Jurying Process
Sample Jurying Tool

Item Rating Sheet

Name:_____ **Subject:** _____ **Grade:** _____

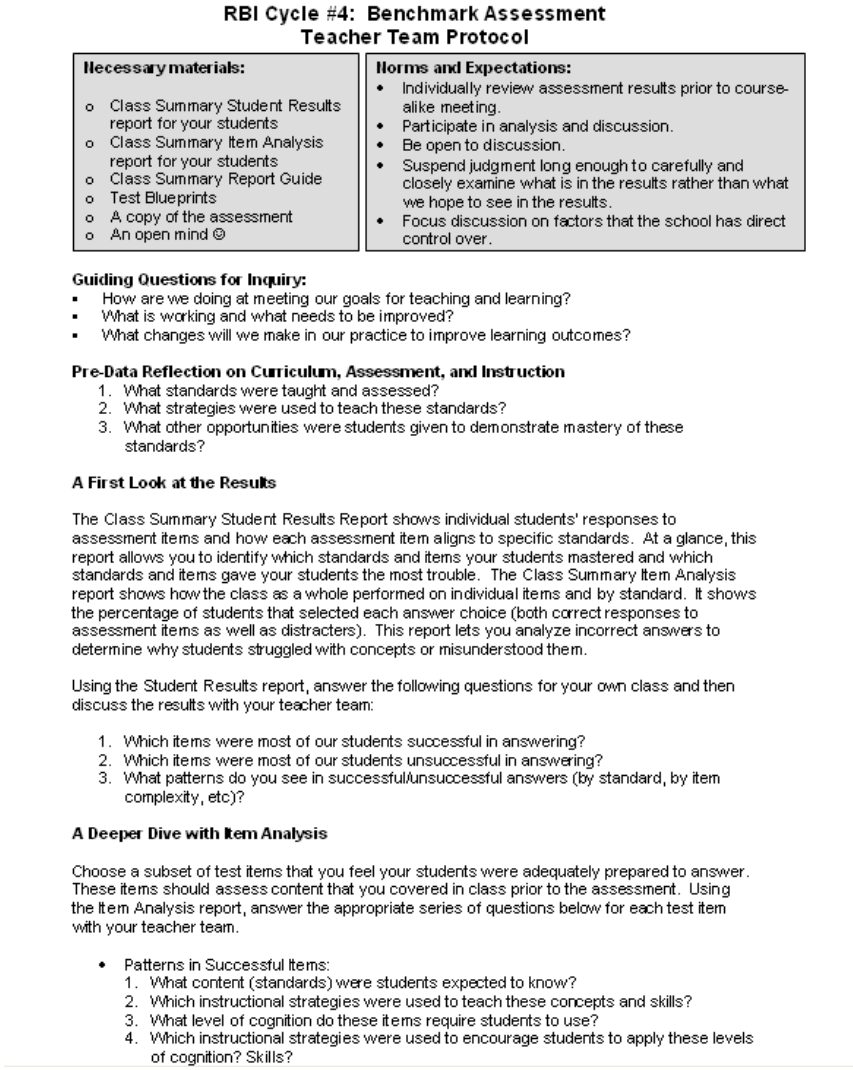
Item Number	The item is well-written and presents a clear question.					The item content is aligned to the designated learning standard.		The item content is central to the learning standard.					The indicated thinking skill is appropriate.		The item content is free of bias.		This item should be included on a test form.			Comments or Suggested Edits (Please make major edits on Item Sheets).
	SD	D	U	A	SA	Yes	No	SD	D	U	A	SA	Yes	No	Yes	No	Yes	No	Edit	Comments
1 ID# _____	1	2	3	4	5	Y	N	1	2	3	4	5	Y	N	Y	N	Y	N	With Edits	
2 ID# _____	1	2	3	4	5	Y	N	1	2	3	4	5	Y	N	Y	N	Y	N	With Edits	
3 ID# _____	1	2	3	4	5	Y	N	1	2	3	4	5	Y	N	Y	N	Y	N	With Edits	
4 ID# _____	1	2	3	4	5	Y	N	1	2	3	4	5	Y	N	Y	N	Y	N	With Edits	
5 ID# _____	1	2	3	4	5	Y	N	1	2	3	4	5	Y	N	Y	N	Y	N	With Edits	
6 ID# _____	1	2	3	4	5	Y	N	1	2	3	4	5	Y	N	Y	N	Y	N	With Edits	
7 ID# _____	1	2	3	4	5	Y	N	1	2	3	4	5	Y	N	Y	N	Y	N	With Edits	
8 ID# _____	1	2	3	4	5	Y	N	1	2	3	4	5	Y	N	Y	N	Y	N	With Edits	
9 ID# _____	1	2	3	4	5	Y	N	1	2	3	4	5	Y	N	Y	N	Y	N	With Edits	

Click Here to Return to Main Page

More Summative Assessment Samples

Teacher Team Protocol for Looking at Benchmark Assessment Results

http://assessmentcontinuum.wetpaint.com/page/Benchmark+Assessments/file



RBI Cycle #4: Benchmark Assessment
Teacher Team Protocol

Necessary materials:

- Class Summary Student Results report for your students
- Class Summary Item Analysis report for your students
- Class Summary Report Guide
- Test Blueprints
- A copy of the assessment
- An open mind ☺

Norms and Expectations:

- Individually review assessment results prior to course-alike meeting.
- Participate in analysis and discussion.
- Be open to discussion.
- Suspend judgment long enough to carefully and closely examine what is in the results rather than what we hope to see in the results.
- Focus discussion on factors that the school has direct control over.

Guiding Questions for Inquiry:

- How are we doing at meeting our goals for teaching and learning?
- What is working and what needs to be improved?
- What changes will we make in our practice to improve learning outcomes?

Pre-Data Reflection on Curriculum, Assessment, and Instruction

1. What standards were taught and assessed?
2. What strategies were used to teach these standards?
3. What other opportunities were students given to demonstrate mastery of these standards?

A First Look at the Results

The Class Summary Student Results Report shows individual students' responses to assessment items and how each assessment item aligns to specific standards. At a glance, this report allows you to identify which standards and items your students mastered and which standards and items gave your students the most trouble. The Class Summary Item Analysis report shows how the class as a whole performed on individual items and by standard. It shows the percentage of students that selected each answer choice (both correct responses to assessment items as well as distracters). This report lets you analyze incorrect answers to determine why students struggled with concepts or misunderstood them.

Using the Student Results report, answer the following questions for your own class and then discuss the results with your teacher team:

1. Which items were most of our students successful in answering?
2. Which items were most of our students unsuccessful in answering?
3. What patterns do you see in successful/unsuccessful answers (by standard, by item complexity, etc)?

A Deeper Dive with Item Analysis

Choose a subset of test items that you feel your students were adequately prepared to answer. These items should assess content that you covered in class prior to the assessment. Using the Item Analysis report, answer the appropriate series of questions below for each test item with your teacher team.

- Patterns in Successful Items:
  1. What content (standards) were students expected to know?
  2. Which instructional strategies were used to teach these concepts and skills?
  3. What level of cognition do these items require students to use?
  4. Which instructional strategies were used to encourage students to apply these levels of cognition? Skills?

Click Here to Return to Main Page

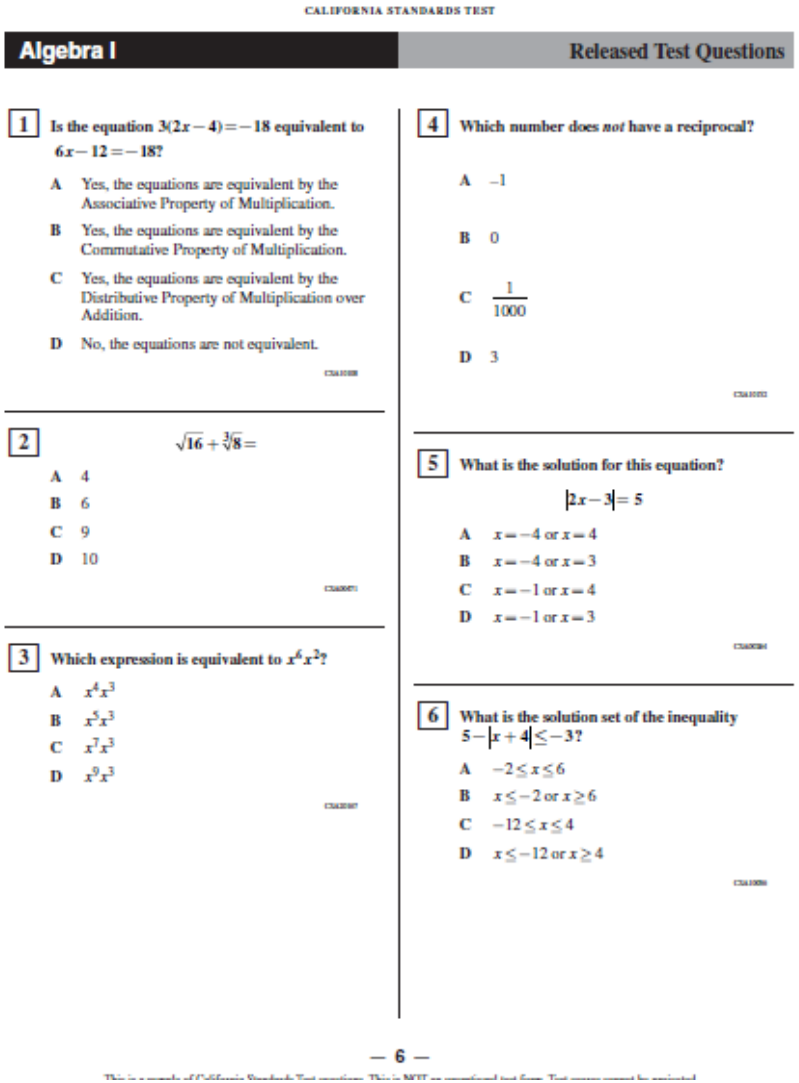
Most Summative Assessment Samples:
http://assessmentcontinuum.wetpaint.com/page/CST+and+CAHSEE+Assessments

CALIFORNIA STANDARDS TESTS
GRADE 2 ENGLISH-LANGUAGE ARTS
(Blueprints adopted by the State Board of Education 10/02)

CALIFORNIA CONTENT STANDARDS: READING	# of Items	%
1.0 WORD ANALYSIS, FLUENCY, AND SYSTEMATIC VOCABULARY DEVELOPMENT: Students understand the basic features of reading. They select letter patterns and know how to translate them into spoken language by using phonics, syllabication, and word parts. They apply this knowledge to achieve fluent oral and silent reading.	22	34%
1.1 Decoding and Word Recognition: recognize and use knowledge of spelling patterns (e.g., diphthongs, special vowel spellings) when reading	3	
1.2 Decoding and Word Recognition: apply knowledge of basic syllabication rules when reading (e.g., v/cv = su/per, vc/cv = sup/per)	3	
1.3 Decoding and Word Recognition: decode two-syllable nonsense words and regular multi-syllable words	3	
1.4 Decoding and Word Recognition: recognize common abbreviations (e.g., Jan., Sun., Mr., St.)	1	
1.5 Decoding and Word Recognition: identify and correctly use regular plurals (e.g., -s, -es, -ies) and irregular plurals (e.g., fly/flies, wife/wives)	2	
1.6 Decoding and Word Recognition: read aloud fluently and accurately, and with appropriate intonation and expression	NA*	
1.7 Vocabulary and Concept Development: understand and explain common antonyms and synonyms	3	
1.8 Vocabulary and Concept Development: use knowledge of individual words in unknown compound words to predict their meaning	2	
1.9 Vocabulary and Concept Development: know the meaning of simple prefixes and suffixes (e.g., over-, un-, -ing, -ly)	2	
1.10 Vocabulary and Concept Development: identify simple multiple-meaning words	3	

Click Here to Return to Main Page

Most Summative Assessment Samples Released CST Questions:
http://assessmentcontinuum.wetpaint.com/page/CST+and+CAHSEE+Assessments



CALIFORNIA STANDARDS TEST

Algebra I — Released Test Questions

1 Is the equation $3(2x-4)=-18$ equivalent to $6x-12=-18$?

A Yes, the equations are equivalent by the Associative Property of Multiplication.

B Yes, the equations are equivalent by the Commutative Property of Multiplication.

C Yes, the equations are equivalent by the Distributive Property of Multiplication over Addition.

D No, the equations are not equivalent.

2 $\sqrt{16}+\sqrt[3]{8}=$

A 4

B 6

C 9

D 10

3 Which expression is equivalent to x^6x^2?

A x^4x^3

B x^5x^3

C x^7x^3

D x^9x^3

4 Which number does *not* have a reciprocal?

A -1

B 0

C $\frac{1}{1000}$

D 3

5 What is the solution for this equation?

$|2x-3|=5$

A $x=-4$ or $x=4$

B $x=-4$ or $x=3$

C $x=-1$ or $x=4$

D $x=-1$ or $x=3$

6 What is the solution set of the inequality $5-|x+4|\leq -3$?

A $-2\leq x\leq 6$

B $x\leq -2$ or $x\geq 6$

C $-12\leq x\leq 4$

D $x\leq -12$ or $x\geq 4$

— 6 —

Click Here to Return to Main Page