DOINGWHATWORKS

SAMPLE MATERIAL

Universal Screening—Establishing District Benchmarks

Dave Heistad, Minneapolis Public Schools, Minnesota

Topic: Response to Intervention in Primary Grade Reading

Practice: Universal Screening

This 2009 PowerPoint presentation by Dr. Dave Heistad, Executive Director of Research, Evaluation and Assessment for the Minneapolis Public Schools, walks through understanding and using universal screening measures and establishing district benchmarks. Real-life examples and resources from schools are included. The presentation addresses six questions:

1. What is comprehensive screening?
2. What should screening instruments predict?
3. Why do we need to establish local benchmarks?
4. How are district benchmarks established?
5. What type of data/reports are generated by benchmarks?
6. How are screening data and benchmarks used within the RtI model?

This project has been funded at least in part with Federal funds from the U.S. Department of Education under contract number ED-PEP-11-C-0068. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

National Center on Response to Intervention
www.rti4success.org

RTI and Universal Screening: Establishing District Benchmarks

- March 25, 2009
- Dave Heistad, Ph.D.
- dheistad@mpls.k12.mn.us

IDEAs that Work U.S. Office of Special Education Programs

American Institutes for Research

Outline of the Webinar

- This presentation will focus on six key questions:
  1. What is comprehensive screening?
  2. What should screening instruments predict?
  3. Why do we need to establish local benchmarks?
  4. How are district benchmarks established?
  5. What type of data/reports are generated by benchmarks?
  6. How are screening data and benchmarks used within the RTI model?

What is Universal Screening?

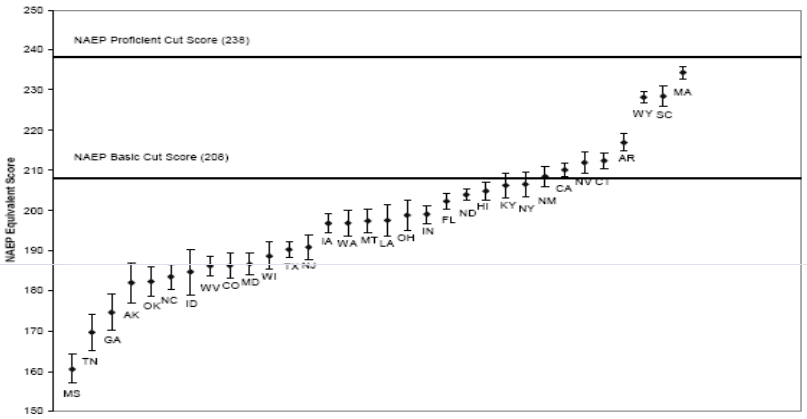
- Screening involves brief assessments that are valid, reliable and evidence based. They are conducted with all students or targeted groups of students to identify students who are at risk of academic failure and, therefore, likely to need additional or alternative forms of instruction to supplement the general education approach (National Center on Response to Intervention)

First Question: What criterion (outcome measure should be used?)

- Screeners should be used to predict success or need for addition support on some important outcome.
- Many school districts have established the goal that all students be able to read well by the end of third grade
- In the 1980s and early 1990s most districts used a National Norm-referenced multiple choice exam to measure reading achievement in third grade. Minneapolis used the Stanford Achievement test and later the California Achievement Test.
- Starting in the late 1990s and throughout this decade the focus has been on State Tests designed to measure State Standards in reading.

Not all state standards are created equal

Figure 2. NAEP score equivalents of states' proficiency standards for reading, grade 4: 2005



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Reading Assessment, and National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

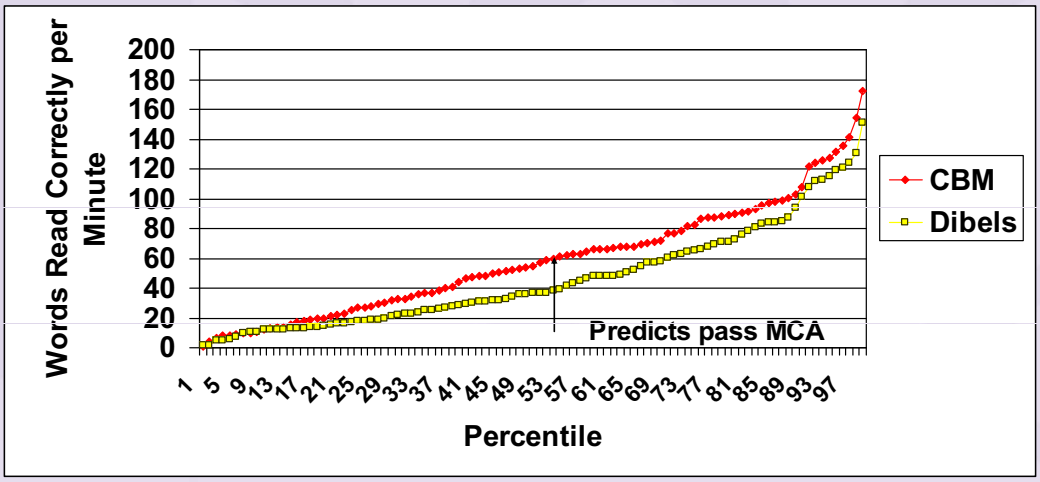
[12] Correlations calculated are standard Pearson correlations.

Not all screening measures are created equal

e.g., Grade 1 MPS-CBM vs. Dibels taken from Reading First study in MN

	Words Correct Per Minute (wcpms)	Dibels Oral Reading
Valid	193	193
Missing	0	0
Mean	58.9	46.5
Median	55	37

Grade 1 DIBELS much harder than Minneapolis CBM with different benchmarks for predicting success



Thus we need to establish local benchmarks

- Each screening instrument needs to be benchmarked against each state test
- Vendor information on cut-scores needs to be verified or modified
- Strength of association with criterion variables needs to be verified
- And information from the screener needs to be customized to the setting in which the data are used to drive instruction

How are local Benchmarks established?

- In Minneapolis Public Schools (MPS) we started with the criterion of success on the State test in reading, the Minnesota Comprehensive Assessment (MCA), by third grade
- The first screener we benchmarked was the Northwest Evaluation Association (NWEA) Adaptive Levels Test (NALT); now we are benchmarking the Measures of Academic Progress (MAP)
  - The MAP is a computer adaptive assessment
  - Items are linked to the State test with a customized item bank
  - Scores are reported on a continuous scale (i.e., the "RIT" scale) from Grade 2 to Grade 10
  - MPS has used the RIT scale to measure progress in reading and math
  - MAP tests are given in the fall, winter and spring

Benchmarking step 1: Establish the reliability of the screening score for each major source of measurement error.

- If the test has more than one item, establish the inter-item reliability and standard error of measurement
  - Coefficient Alpha
  - Generalizability Coefficient
  - IRT based
- Reliability is a correlation coefficient from 0.0 to 1.0.
- The acceptable standard for reliability is .8 or above; the high standard we strive for in Minneapolis is .9 or above
- The inter-item reliability for the MAP reported by the publisher by grade ranges from .94 to .95 with a median of .94.

Benchmarking step 1: Establish the reliability of the screening score for each major source of measurement error.

- Using screening instruments with high reliability insures that the students identified for intervention are consistent from one version of the assessment to another, from one time to another, and from one rater or scorer to another.
- Reliability is reported as a correlation coefficient which should be .8 or higher.

Reliability of the screening score(s)

- All screeners should report test-retest reliability
- The MAP is designed to be administered no more than 4 times per year
- The retest stability from fall to spring ranges from .84 to .89 with a median of .88.
- The MAP is computer administered and scores so inter-rater reliability is not calculated.
- When we get to CBM measures and other human administered instruments, inter-rater reliability is crucial.

Benchmarking step 2: Establish the validity of the screening score

- The key areas of validity for evaluating a screening measure are
  - Construct validity: The screener truly measures reading
  - Concurrent validity: The screener correlates highly with other accepted measures of reading given at the same time
  - Predictive validity: The screener predicts future performance on an accepted measure of reading
  - For the MAP/NALT concurrent validity with State reading tests across the country varied from .69 to .86 with a median of 45 coefficients = .81
  - The standard for predictive validity set by the National Center on Response to Intervention (RTI) = .70

Benchmarking step 2: Establish the validity of the screening score

- The key area of validity for evaluating a screening measure is predictive validity
  - Predictive validity: The screener predicts future performance on an accepted measure of reading
  - The standard for predictive validity set by the National Center on Response to Intervention (RTI) = .70

Benchmarking step 3: Run a benchmarking study to determine classification accuracy and to set cut scores

- MPS did a study of the grade 3 fall RIT score predicting the spring grade 3 MCA state test score in 2007. The first cut score established was "partially proficiency".
- The correlation between the RIT score and MCA was .86
- The overall classification accuracy at the partially proficient cut score was 87%
- The RIT score that predicted proficiency with 87% accuracy was a score of 173; a score of 182 predicted proficiency with 85% accuracy

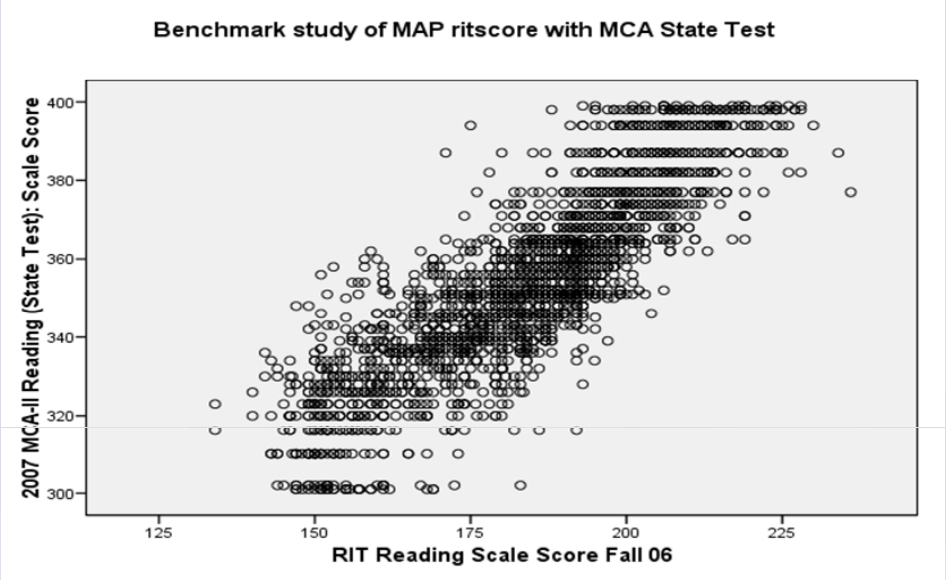
Questions

If you have a question please submit it using the Q&A tab at the top of your screen.

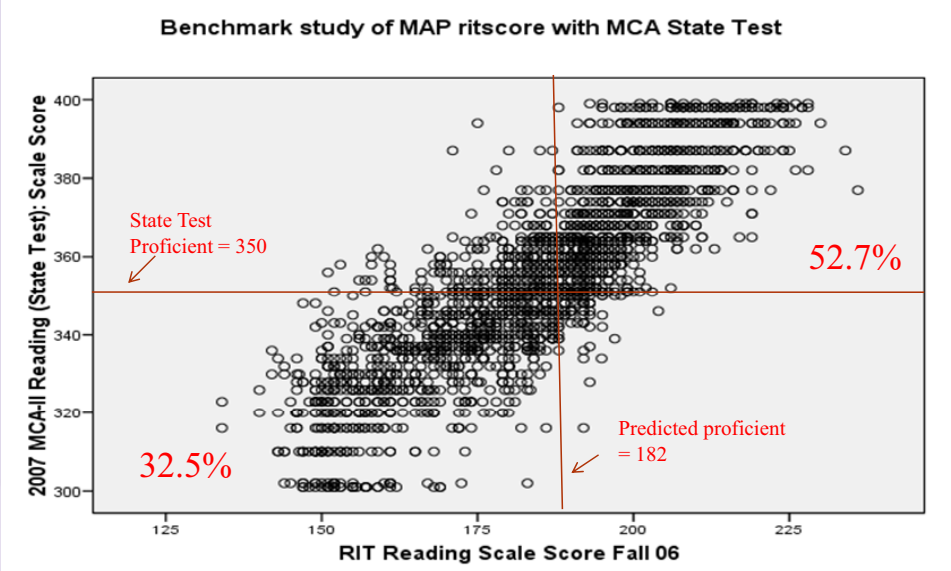
Benchmarking step 3: Run a benchmarking study to determine classification accuracy and to set cut scores

- MPS did a study of the grade 3 fall RIT score predicting the spring grade 3 MCA state test score in 2007. The first cut score established was "partially proficiency".
- The NWEA assessment was given to all 3rd grade students in the fall of the year and the MCA was given in the spring to all students.
- Only students with both test scores are included in the analysis
- The first result we look at is the correlation between the fall screener (NWEA) and the Spring criterion test (MCA)
- We want to see that high scores on the screener correspond with high scores on the criterion test (see next slide)

Correlation = .86



Overall Classification Accuracy = 52.7% + 32.5% = 85.2%



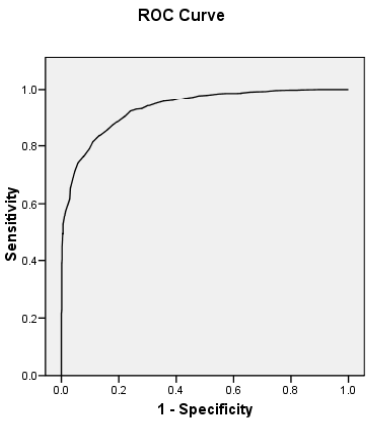
Statistics Packages will conduct a ROC (receiver operation characteristic) analysis which evaluates sensitivity and specificity at the same time

Area Under the Curve

Test Result Variable(s): RTI Reading Score Fall 06

Area
0.934

The standard for ROC area under the curve = .90



How to find the cut score – Three methods that usually yield similar results

- ROC analysis
- Discriminant Function Analysis (especially for composite scores) or Logistic Regression
- Equal Percentile Linking (most frequently used in MPS)
  - For example 100 students w/ screener and state test scores all lined up… 340 = partially proficient; 350 = proficient

- MCA …320 322 324 326 328 330 332 334 336 338 340 342 344 346 348 350 352 354 356…
- MAP….163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 182 184 186 188…

Partially proficient

Proficient

Gold Standard: Cross-validate the findings with a different sample (e.g.,the next year)

- In 2009 we redid the analysis and got a correlation between RIT score and MCA = .849
- Cut score at 182 predicted with 84.3% accuracy
- Area under the curve = .93
- Also, run the analysis at “Proficient” and consider dividing up the scores into three categories
  - Not “on course” for partially proficient (red)
  - On course for partially proficient but not proficient (yellow)
  - On course for proficient (green)

How are screening data and benchmarks used within the RTI model? Fall 2009 data:

Name	HR	Grade	Scale	% ile	Vocabulary	Literal Comprehension	Interpretive and Evaluative Comprehension	Chance	Growth
	06	06	184.00	4	181-190	171-180	191-200	DM	=1yr
	06	06	219.00	70	211-220	211-220	211-220	ME	=1yr
	06	06	214.00	55	221-230	211-220	201-210	ME	>1yr
	06	06	182.00	3	181-190	171-180	181-190	DM	<1yr
	06	06	206.00	31	201-210	201-210	201-210	PM	<1yr
	06	06	206.00	31	191-200	201-210	211-220	PM	<1yr
	06	06	199.00	17	191-200	191-200	191-200	PM	<1yr
	06	06	195.00	12	181-190	191-200	201-210	DM	>1yr
	06	06	207.00	34	191-200	201-210	211-220	PM	<1yr
	06	06	190.00	7	191-200	181-190	191-200	DM	<1yr
	06	06	162.00	1	171-180	171-180	171-180	DM	<1yr
	06	06	198.00	15	191-200	191-200	191-200	DM	<1yr
	06	06	175.00	1	171-180	171-180	171-180	DM	>1yr
	06	06	193.00	10	191-200	181-190	191-200	DM	<1yr
	06	06	210.00	42	201-210	201-210	201-210	ME	N/A
	06	06	199.00	17	201-210	191-200	201-210	PM	N/A
	06	06	207.00	34	211-220	201-210	191-200	PM	=1yr
	06	06	200.00	19	191-200	191-200	201-210	PM	>1yr
	06	06	214.00	55	211-220	211-220	211-220	ME	>1yr

National Center on Response to Intervention
www.rti4success.org

Name	Home Room	Read NALT Scale	Read NALT %tile	Math NALT Scale	Math NALT %tile	% Attend	Abs. Exc*	Abs. UnExc*	Tardy*	# Susp	CBM SPR
	202					75.6	5	15	0		
	201										
	202										
	201										
	201										
	202										
	201	148	2	183	31	98.8	0	2	0		49
	201	149	2	162	2	95.4	0	25	0		
	201	151	3	174	12	97.1	5	0	0		31
	202	152	3	162	3	93.0	9	3	0		23
	202	151	3	177	16	95.9	3	4	0		25
	202	151	3	179	20	97.7	1	3	0		29
	201	154	4	164	4	98.8	1	1	0		25
	201	157	6	184	34	99.4	1	0	0		30
	202	160	8	176	15	95.9	7	0	0		22
	202	160	8	174	12	90.6	11	5	0		35
	201	164	11	176	15	98.8	2	0	0		47
	201	165	12	179	21	98.0	3	1	0		
	202	167	13	179	20	92.7	5	8	0		49
	201	170	16	175	13	97.7	1	3	0		36
	201	172	18	190	50	94.2	8	2	0		50
	201	176	23	182	29	88.3	3	17	0		
	201	176	24	177	16	97.7	4	0	0		60
	201	177	26	178	19	92.4	13	0	0		63
	202	177	26	187	42	98.8	2	0	0		69

National Center on Response to Intervention
www.rti4success.org

Fall 2006 Reading NALT Data Reported By Strand

Name	HR	Grade	Scale	%ile	Vocabulary	Literal Comprehension	Interpretive and Evaluative Comprehension	Chance
	53	04	200.00	46	Avg	Avg	High	ME
	53	04	193.00	28	Avg	Low	Low	PM
	53	04	209.00	74	High	High	High	ME
	53	04	169.00	4	Low	Low	Low	DM
	53	04	193.00	28	Low	Avg	High	PM
	53	04	222.00	97	High	High	High	ME
	53	04	158.00	1	Low	Low	Low	DM
	53	04	203.00	55	Avg	Avg	High	ME
	53	04	191.00	25	Low	Low	Avg	PM
	53	04	196.00	35	Avg	Avg	Avg	PM
	53	04	196.00	35	Avg	Avg	Avg	PM
	53	04	190.00	23	Low	Avg	Avg	PM
	53	04	196.00	35	Avg	Avg	High	PM
	53	04	194.00	30	Avg	Avg	Avg	PM
	53	04	201.00	49	Avg	Avg	Avg	ME
	53	04	162.00	2	Low	Low	Low	DM
	53	04	194.00	30	Avg	Avg	Low	PM
	53	04	197.00	37	Avg	Avg	Avg	PM
	53	04	178.00	9	Low	Low	Low	DM
	53	04	185.00	16	Low	Low	Low	DM
	53	04	195.00	33	Avg	Avg	Avg	PM
	53	04	151.00	1	Low	Low	Low	DM
	53	04	193.00	28	Avg	Low	Avg	PM

Click Here to see skills of student scoring "Low" on Comprehension

Subject: Reading
Subcomponent: Literal Reading Comprehension
RIT Score Range: 151-160

Secured Skills	Emerging Skills RIT Range 151 and 160	Future Skills RIT Range 161 and 170
	Sequencing	*Reading Directions*
	Format: Read approximately 30 word passages	Format: Read short simple sentences
	Follow straightforward sequence of events	Follow 3-6 steps, no more than 60 words
	Use clue words supplied: first, next, last, before, after, later…	Find literal detail in simple directions
	Identify what happened after another event	Determine the purpose of simple directions
	Identify what happened last in a sequence of events	*Sequencing*
	Reading for Detail	Format: Read up to 70 word passages where clue words are rarely supplied
	Format: Read short passages – up to 40 words	Identify what happened after another event

Class By RIT Report

NWEA
Northwest Evaluation Association
Partnering to help all kids learn

Class Breakdown By Goal for Mathematics
St. Helens Elementary School - Fall 2007
Cridebring, Deran N. and TF070098 Cridebring Homeroom 1(A)

The following table shows how the class is broken down by RIT and goal.

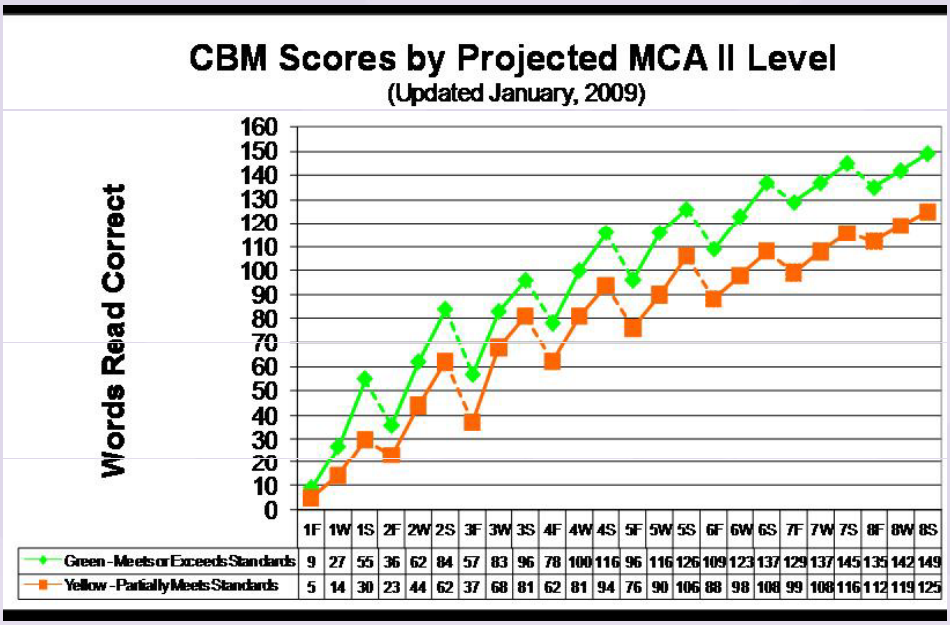
Test Name: Math Survey w/ Goals 2-5 CO V2

	< 171	171-180	181-190	191-200	201-210	211-220	221-230
Algebraic Methods		S. I. Chernyshev (180)	N. O. Nicolella (190) C. L. Pab?N (191) E. N. Hogie (192) S. R. White Wolf (198) M. D. Rhoden (202) T. A. Basley (203)	J. N. Urrego (200) Z. A. Wysocki (203) T. E. Heath (206)	F. A. Strommenger (197) A. M. Kyzar (201) A. N. Dresen (204) K. N. Pafilas (207) V. E. Brown-Gmahl (208)	G. A. Martinson (202) T. A. Castiglione (212) N. N. Foglio (219) A. C. Walterscheid (223)	L. N. Troub (211) T. O. Salisbury (221) C. R. Fidai (228)
Computation	S. I. Chernyshev (180) N. O. Nicolella (190)	S. R. White Wolf (198)	C. L. Pab?N (191) A. M. Kyzar (201) T. O. Salisbury (221)	E. N. Hogie (192) F. A. Strommenger (197) J. N. Urrego (200) G. A. Martinson (202) M. D. Rhoden (202) A. N. Dresen (204)	T. A. Basley (203) Z. A. Wysocki (203) T. E. Heath (206) K. N. Pafilas (207) V. E. Brown-Gmahl (208) L. N. Troub (211) T. A. Castiglione (212) N. N. Foglio (219)		C. R. Fidai (228)
Data Analysis & Probability			S. I. Chernyshev (180) N. O. Nicolella (190) A. M. Kyzar (201) A. C. Walterscheid (223)	C. L. Pab?N (191) E. N. Hogie (192) G. A. Martinson (202)	F. A. Strommenger (197) S. R. White Wolf (198) M. D. Rhoden (202) T. A. Basley (203) Z. A. Wysocki (203) A. N. Dresen (204) T. E. Heath (206) K. N. Pafilas (207) V. E. Brown-Gmahl (208) L. N. Troub (211) T. A. Castiglione (212)	J. N. Urrego (200) N. N. Foglio (219)	

Questions

If you have a question please submit it using the Q&A tab at the top of your screen.

CBM Benchmarks

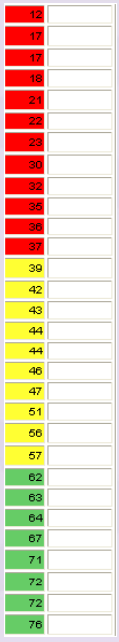


National Center on Response to Intervention
www.rti4success.org

Student Names

Words Read Correctly

Screening in Fall, Winter, and Spring On Words Read Correctly on Grade Level



National Center on Response to Intervention
www.rti4success.org

National Reading Panel Categories School Aggregate Report

NATIONAL READING REPORT
STUDENTS ABOVE PROFICIENT
Based on Spring 2006 Data

School Name	Grade	Alphabetic Principle		Phonemic Awareness		Fluency		Vocabulary		Literal Comprehension		Interpretive Comprehension	
		Prof	Total*	Prof	Total*	Prof	Total*	Prof	Total*	Prof	Total*	Prof	Total*
SHINGLE CREEK	00	0%	1	0%	1								
SHINGLE CREEK	01	81%	36	64%	36								
SHINGLE CREEK	02					68%	38						
SHINGLE CREEK	03					56%	27	48%	25	48%	25	52%	25
SHINGLE CREEK	04					33%	30	29%	24	21%	24	25%	24
SHINGLE CREEK	05					59%	29	24%	29	24%	29	41%	29
SHINGLE CREEK	55												

Prof= Percentage of Students at this Grade Level who Took the Test AND are Proficient

Total* = Total Students at this Grade Level who Took the Test

Alphabetic Principle and Phonemic Awareness from Spring 2006 Kindergarten Assessment

Fluency from Spring 2006 CBM

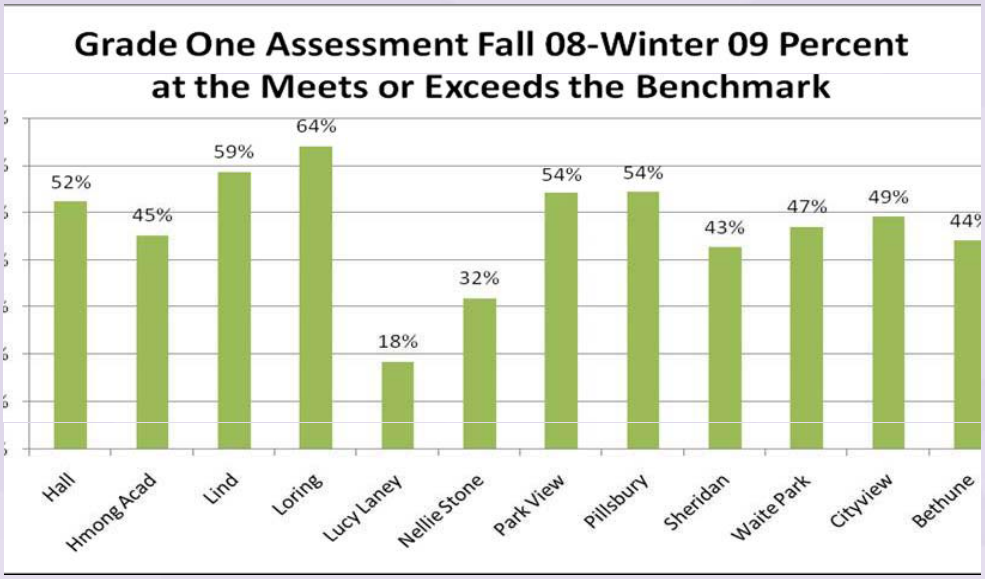
Vocabulary from Spring 2006 NALT

Literal Comprehension from Spring 2006 NALT

Interpretive Comprehension from Spring 2006 NALT

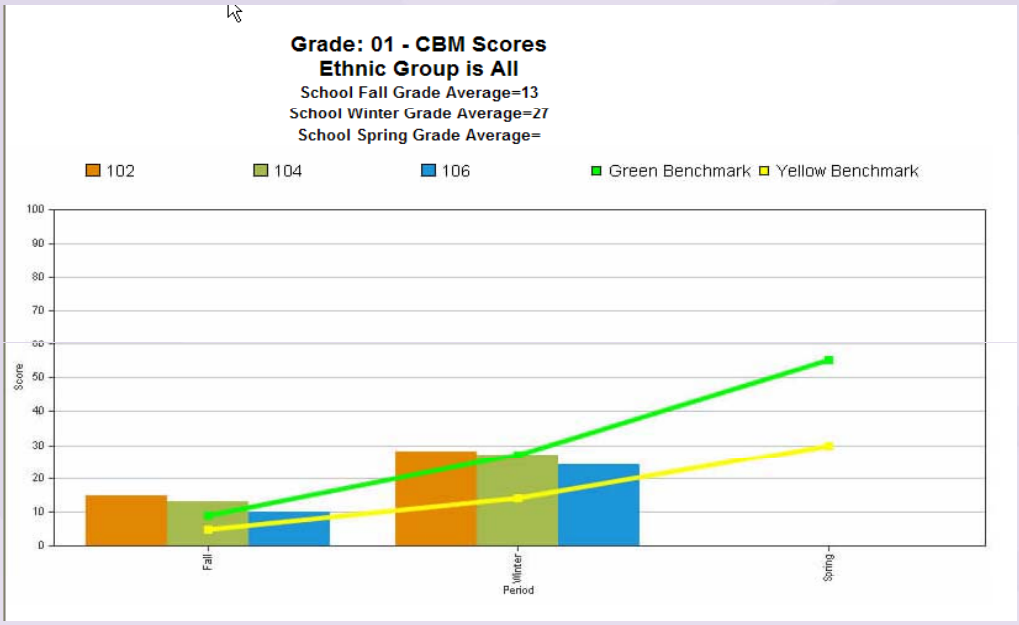
National Center on Response to Intervention
www.rti4success.org

Oral Reading Percent Making Benchmark



National Center on Response to Intervention
www.rti4success.org

Fall and Winter Grade 1 CBM Screening



Literacy Items on the Beginning of Kindergarten Assessment (BKA)

- Includes:
  - Picture vocabulary
  - Oral comprehension
  - Letter names
  - Letter sounds
  - Rhyming
  - Alliteration (initial sounds)
  - Concepts of Print
  - Total Composite Score

BKA Predicts Reading Well by Grade 3 (3 and ½ years later!)

- Correlation between BKA composite and NALT Grade 3 Reading= .67
- Correlation between BKA composite and MCA Grade 3 Reading= .61
- A BKA composite score of 85 or higher predicts with 75% accuracy that students will score at level 3 (1420) on the MCA Reading in 3rd grade

Early Literacy Screening Report

Winter 2009 Oral Reading/Math

		READING										MATH									
SIN	NAME	Words Read Correctly		Reading Expression		Compre-hension		Letter Sounds		Phoneme Seg-mentation		Number ID		Quantity Discri-mination		Quantity Array		Addition / Sub-traction		Addition Compu-tation	
		W	F	W	F	W	F	W	F	W	F	W	F	W	F	W	F	W	F	W	F
		30	13	1	1	56%	44%		46		36	50	33	33	21	20	10	71%	0%	20	7
		103	60	4	4	89%	100%		25		46	51	26	24	12	14	5	43%	0%	10	0
		6	4	1	1	11%	22%	30	30	19	11	21	12	21	13	15	6	14%	14%	12	4
		14		1		22%						33		34		19		57%		18	
		11	4	1	1	22%	11%		16		15	24	17	23	9	15	11	0%	0%	2	1
		20	9	1	1	11%	11%		43		26	56	51	39	21	17	12	0%	0%	14	4
		32	13	1	2	78%	56%		33		24	48	27	22	15	16	10	43%	0%	10	1
		19	2	1	1	33%	44%		12		45	51	30	32	12	20	11	71%	14%	18	11
		9	6	1	1	22%	33%	15	13	31	22	10	11	23	3	13	8	14%	0%	2	0
		20	8	1	1	56%	44%		29		32	42	25	27	22	12	6	43%	0%	4	2
		0	4	0	1	11%	44%	13	0	0	0	4	0	12	4	5	3	0%	0%	0	0
		9	5	1	1	11%	56%	33	12	25	45	34	20	28	21	15	11	0%	0%	5	0
		96	65	3	3	78%	78%		35		22	77	47	37	24	18	15	71%	29%	13	14
		21		1		67%						51		41		19		57%		8	
		33	10	1	1	89%	11%		23		54	31	16	29	21	12	3	29%	0%	12	8
HOMEROOM		20	8	1	1	44%	37%	23	24	19	29	39	24	28	15	15	9	34%	4%	10	4
SCHOOL		19	8	1	1	51%	40%	26	32	24	28	38	23	25	14	14	10	26%	5%	7	3
DISTRICT		30	10	2	1	66%	46%	27	29	25	26	40	28	29	21	16	12	37%	17%	8	3

Other Considerations in screening/benchmarking

- Generalizability of the screener data/ benchmarking studies to your population
- Efficiency of the screening tool(s)
- Time of screening per student and per teacher
- Language of the screener and accommodations
- Can the measures be copied, adapted
- Cost of the screener per student or per site license
- Training needed for the instrument and training cost
- Scores available through the screener (e.g., national percentiles)
- How often the screener can be given