

Using CBM-R for Planning and Program Evaluation in Three Tiers



Ben Ditkowsky, Ph.D.



technology so you can teach

Ben Silberglitt, Ph.D.

September, 2008

Midwest Leadership Summit II

Agenda

- CBM-R as a tool for decision-making
- Big ideas in CBM-R
 - Idea #1: Relevance of CBM-R scores
 - Idea #2: Relevance of progress on CBM-R
- BREAK
 - Idea #3: Linking to instruction using local percentiles
 - Idea #4: Linking to instruction using target scores
- Final thoughts about CBM
- Questions

How are assessment data used?

- Communicating individual student performance to parents, students, and teachers
- Communicating grade-level, building or district performance to the staff and the community
- Informing instructional decisions at all levels, from the individual to the district
- Evaluating effectiveness of programming
- Evaluating schools and districts for No Child Left Behind

Key Purposes of Assessment

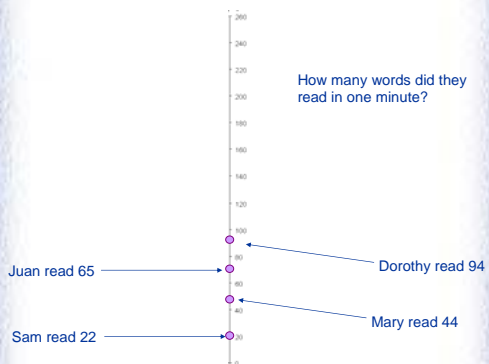
- Screening
- Diagnostic
- Progress Monitoring
- Outcomes

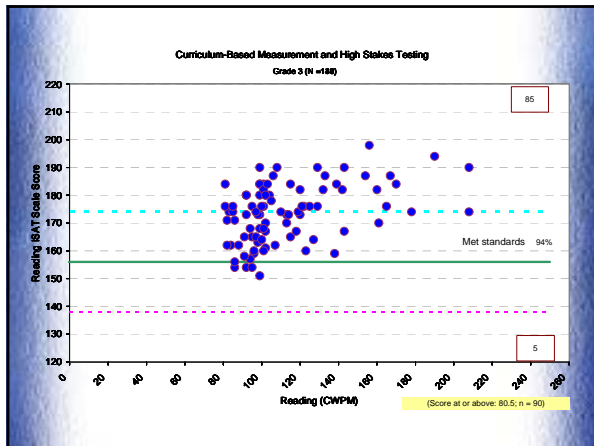
Why use local assessments, such as CBM?

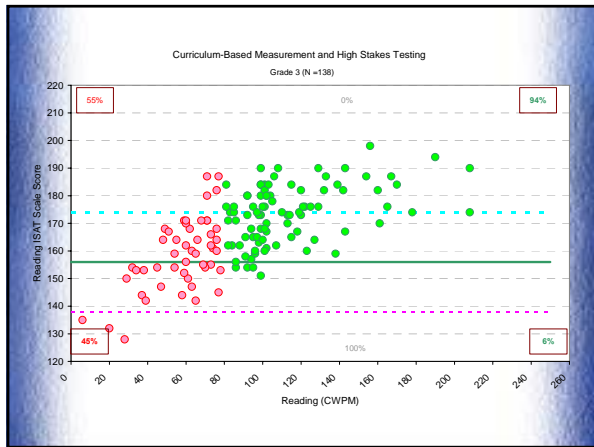
- State-mandated tests assess outcomes
- Local assessments allow us to:
 - Measure students earlier than 3rd grade
 - Monitor progress more frequently than once per year
 - Rely on multiple assessment tools for our information
 - Develop an integrated assessment system with benchmarks for performance, linked to a common outcome

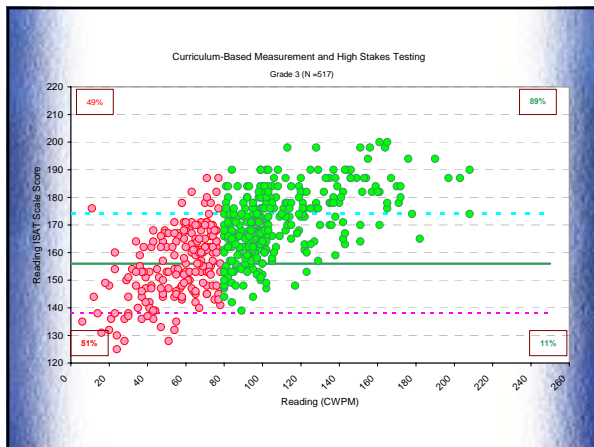
Question #1. Do scores on CBM Matter?

Source data AIMSWEB





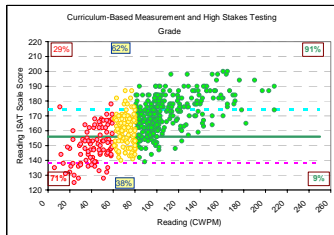




Curriculum-Based Measurement *is* a measure of general reading competence

- Validity coefficients for R-CBM with the Comprehension subtest of the SAT were .91 as compared with
 - Question Answering .82 , Recall .70, Cloze .72 (Fuchs, Fuchs & Maxwell, 1988)
- Validity coefficients for Text Fluency of Folk Tales with the Iowa Test of Basic Skills Comprehension was .83 (Jenkins, Fuchs, Espin, van den Broek & Deno, 2000)
- Fluency is causally related to reading comprehension (National Reading Panel -NICHHD, 2000)

Is fall Curriculum-Based Measurement related to state testing?



Additional research...

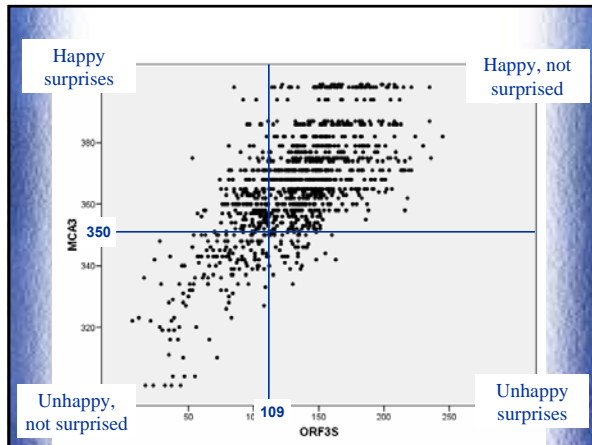
- Conducted at SCRED
- Explored CBM-R link to MCA-II-R
- Extended link to early literacy
- Developed system of target scores and a streamlined assessment plan

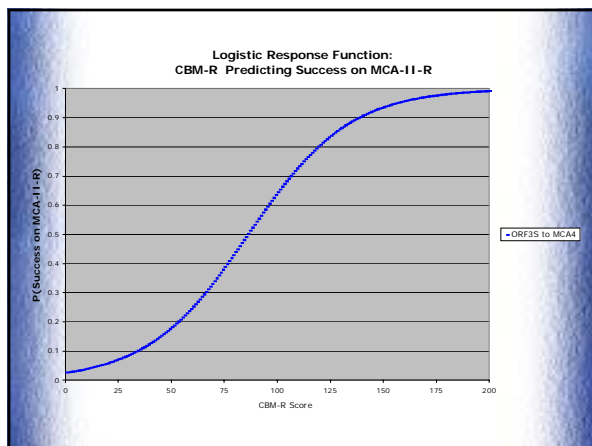
CBM-R & MCA-II-R Correlations

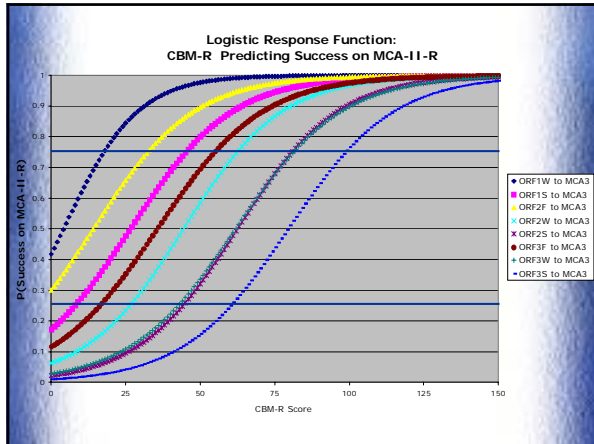
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7
r	.68	.61	.63	.64	.66
N	1108	1064	1065	1067	1121

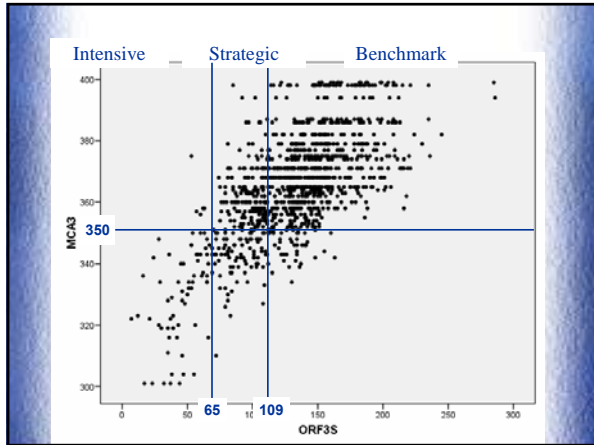
* - All correlations significant $p < .001$

* - All correlations are for Spring CBM-R tests with same-grade MCA-II-R (both 2006 & 2007 school years included)









Example: Target setting using CBM-R in a 3-tiered model

Grade	Tier	Fall	Winter	Spring
1	1		22	52
	2		0	19
2	1	43	72	90
	2	14	32	47
3	1	70	91	109
	2	24	50	65
4	1	95	114	127
	2	49	69	78
5	1	113	128	141
	2	69	83	100
6	1	135	152	166
	2	94	112	122

**Extending to 1st Grade Early Literacy:
Correlations with Spring 1st grade CBM-R**

	<u>NWF</u> Fall	<u>NWF</u> Win	<u>CBM- R</u> Win
r	.67	.62	.87
N	1336	1322	1552

* - All correlations significant $p < .01$

**Extending to Kindergarten Early Literacy:
Correlations with Spring 1st grade CBM-R**

	<u>PSF</u> Fall	<u>PSF</u> Win	<u>PSF</u> Spr	<u>LNF</u> Fall	<u>LSF</u> Nov	<u>LSF</u> Win	<u>LSF</u> Spr
r	.35	.41	.31	.52	.50	.61	.66
N	1260	1154	1181	429	1275	1343	1391

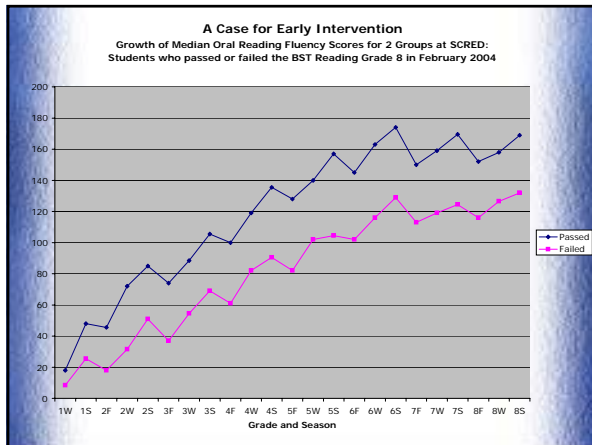
* - All correlations significant $p < .01$

Example: Target setting in Early Literacy

Kindergarten					
Measure	Tier	September	November	January	May
Letter Naming Fluency	1	9			
	2				
Letter Sound Fluency	1		10	21	41
	2			6	21
Phonemic Segmenting and Blending	1		Students must reach 12 words correct on the Blending and Segmenting tests to be considered phonemically aware		
	2				
Nonsense Word Fluency	1				35
	2				19

Example: Target setting in Early Literacy

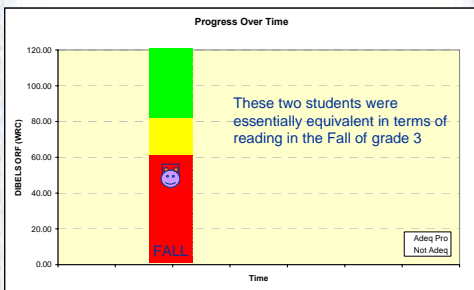
Grade 1					
Measure	Tier	September	November	January	May
Phonemic Segmenting and Blending	1	Students must reach 12 words correct on the Blending and Segmenting tests to be considered phonemically aware			
	2				
Nonsense Word Fluency	1	32	45	52	
	2	15	25	29	
Oral Reading Fluency	1			22	52
	2			0	19



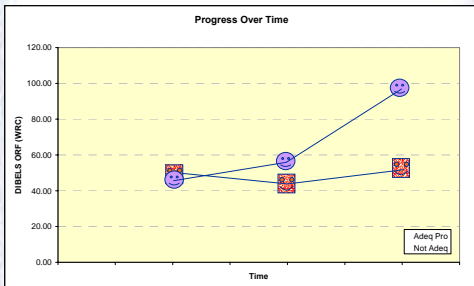
Big Idea # 1.
Scores on CBM are related to results of high-stakes testing

Question # 2.
**Is progress as measured by
CBM meaningful?**

**Consider two students from a different
district. Source data DIBELS
In Fall**



**Consider two students
Over the year (2005-06)**



Consider two students

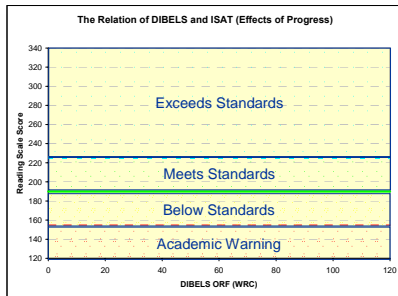
Over the year (2005-06)

- Both students began the year *Below Basic*
- In winter one student had not made any progress, one student had made minimal progress
- By Spring, one student had demonstrated an average gain of 1 word per week, the other demonstrated little gain all year

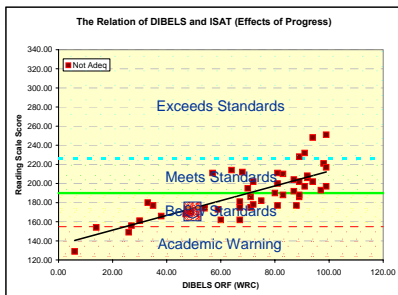
 Adequate

 Not Adequate

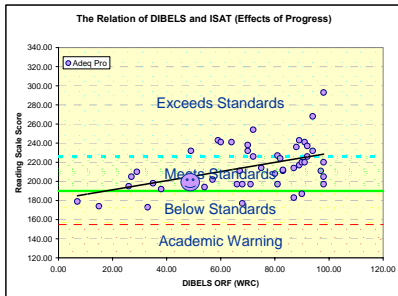
The Effect of Progress on DIBELS



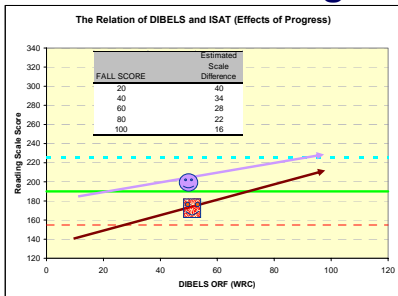
Fall DIBELS (NAP Group)



Fall DIBELS (AP Group)



The Effects of Progress



Big Idea #2.

Progress on CBM is related to improved outcomes in General reading

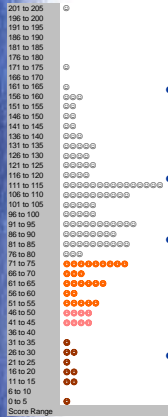
Given the same score in Fall, a student who gains an average of 1 word per week over the course of a year can be expected to score better on high stakes tests, that is, student improvement from Fall to Spring on DIBELS matters

Question: How can we use CBM to allocate resources?

The reality of schools

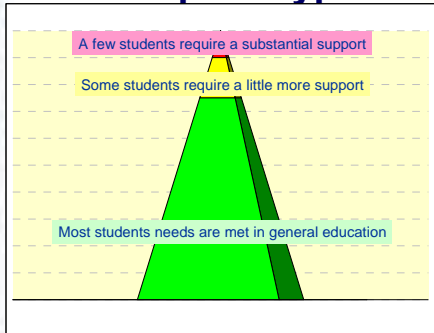
- “fair is not that everyone gets the same thing, fair means that everyone gets what they need,” at the same time...
- We do not have unlimited resources
- We have to make decisions regarding who gets more service

How did students score?



- If we align student scores according to performance...
- We see that scores distribute...
- Question: For how many students can we provide the most intensive services?
- Question: For how many students can we provide the supplemental services?

The prototype



Use local norms for resource allocation to match student needs with available resources.

Calculate Norms with Excel

- Percentiles indicate how many students score at or above a particular score

First Name	I.E.P.	CBM	Percentile
Adrian	1	7	0%
Cory	1	15	1%
Bridgette	1	16	1%
Dimitri	1	21	2%
Ekaterina	1	23	3%
Ashorina		27	4%
Kyle	1	32	4%
Nguyen	1	32	4%
Bryannalexis		39	6%
Aron		45	8%
Sokol	1	46	=PERCENTRANK(D:D,D12)
Saba		47	8%
Hana		49	9%
Appolokos		50	9%
Michael		50	9%
Christian		52	11%
Larh		54	11%
Tiana		55	12%
Simon		56	13%
Javeria		57	14%
Sabrina		58	14%
Katherine	1	59	15%
Gus		60	16%
Adin	1	62	18%
Azeba		65	17%
Shaville		66	18%
Hanna		66	18%
Ashley		67	19%
Melanie		69	20%
Kristen		72	21%
Joshua		73	21%
Miriam		74	22%
Amstar		75	23%
Emmett		75	23%
Silva		75	23%
Zoele		77	25%

Using percentiles keeps constant the number of students who are served.

Set up a standard protocol for intervention and problem solve when the standard is not sufficient

First Name	I.E.P.	CBM	Percentile
Adrian	1	7	0%
Sory	1	15	1%
Bridganae	1	16	1%
Brian	1	21	2%
Ekaterina	1	23	3%
Ashorena	1	27	4%
Kyle	1	32	4%
Ruben	1	32	4%
Bryanna/lexis	1	39	6%
Kron	1	45	6%
Sokol	1	46	7%
Saba	1	47	8%
Nana	1	49	9%
Apostolos	1	50	9%
Michael	1	50	9%
Christian	1	52	11%
Linh	1	54	11%
Tara	1	55	12%
Simon	1	56	13%
Javeria	1	57	14%
Sabrina	1	58	14%
Katherine	1	59	15%
Gus	1	60	16%
Adin	1	62	16%
Areeba	1	65	17%
Seville	1	66	18%
Hanna	1	66	18%
Ashley	1	67	19%
Melanie	1	69	20%
Kristen	1	72	21%
Joshua	1	73	21%
Miriam	1	74	22%
Anaar	1	75	23%
Ernest	1	76	23%

Big Idea #3.

We can use local percentile ranks to match our resources with our needs

Contexts for Decision-Making

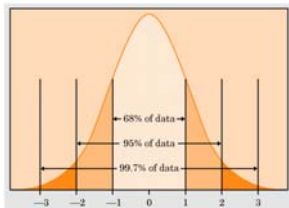
- Norm-Referenced
 - Compares a student’s score to some group of students
 - “Typical score” or “Typical growth rate”
 - Provides a context for judging performance relative to peers
- Criterion-Referenced
 - Compares a student’s performance to a pre-defined standard
 - “Target score” or “Target growth rate”
 - Provides a context for judging performance relative to an established benchmark

Contexts for Decision-Making

- Norm-Referenced
 - Josie, a 5th grader, got 32 items correct on the ITBS
 - *Is that a good score?*
 - The national average for 5th graders was 26
 - *Is that a big difference?*
 - The standard deviation was 6
 - 1 SD above the mean = 84th %ile (1 below = 16th %ile)
 - 2 SD above the mean = 97th %ile (2 below = 3rd %ile)

The normal curve

- A source of much confusion!



- Equal jumps in score \neq Equal jumps in percentile rank

Contexts for Decision-Making

- Criterion-Referenced
 - Josie, a 5th grader, got 32 items correct on the MCA
 - *Is that a good score?*
 - The proficiency standard for grade 5 is 26 items
 - *Is that a big difference?*
 - “Exceeds proficiency” standard is 31 items

Norm-Referenced Assessment

- Examples of types of scores:
 - Standard Scores
 - Percentile Ranks
 - Age and Grade Equivalents
- Who am I comparing to?
 - Norm Sample
 - User Norm
 - Local Norm

Considerations with Norms

- Recency
 - When were the norms gathered?
 - What's changed since then?
- Representativeness
 - What population do I want to compare my student to?
 - Is the norm group a representative sample of that population?
- Relevance
 - Is the norm group a comparison I care about?
 - Does success relative to the norm match my instructional goals for the student?

Criterion-Referenced Assessment

- Can be based on different types of scores
- Types of scores typically used:
 - Raw Scores
 - Scale Scores
 - RIT Scores
- What am I comparing to?
 - Target scores
 - Goal scores
 - Grade-level standard

Norm vs. Criterion-Referenced

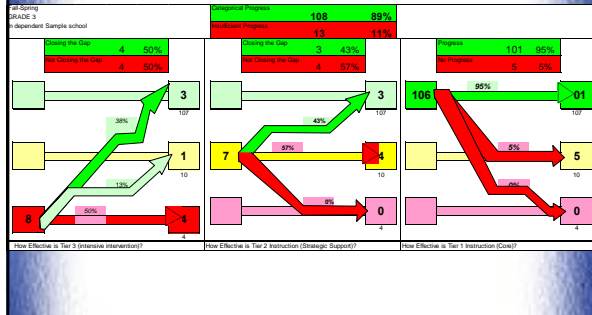
- Many tests now allow for both types of decisions (but typically involve a user norm)
 - NWEA / Scantron
 - State-mandated assessments
 - CBM / DIBELS

Question : How can we use CBM to evaluate progress for groups of students?

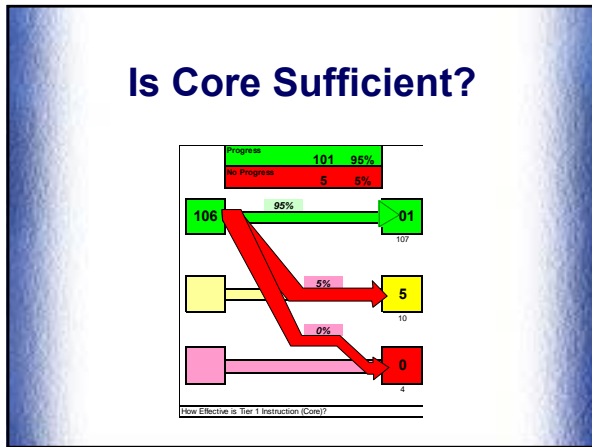
Summary of Adequate Progress

- Cut Scores for R-CBM and DORF have been derived for assessment 3 times per year.
 - When student are likely to **meet standards**, we might refer to them as **proficient**
 - When students are likely **not to met standards**, we might refer to them as **below basic**
 - There are some students whose performance is such that we do not know, we refer to them as questionable.
- We know what scores we expect in Fall, Winter and Spring.
- We can use this information to determine whether a school is making adequate progress, within a given tier, or as a whole.

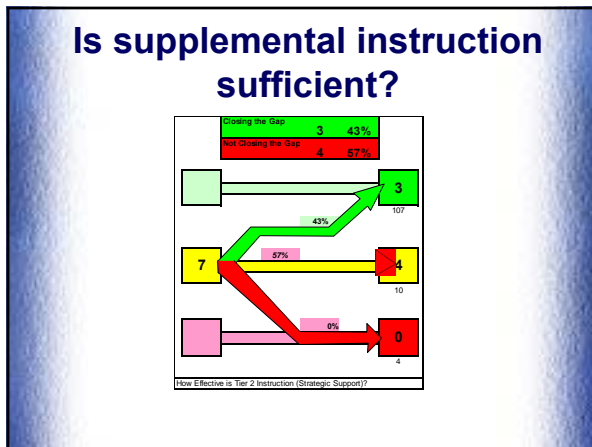
Evaluate Progress



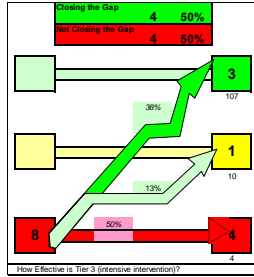
Is Core Sufficient?



Is supplemental instruction sufficient?



Is intensive intervention sufficient?

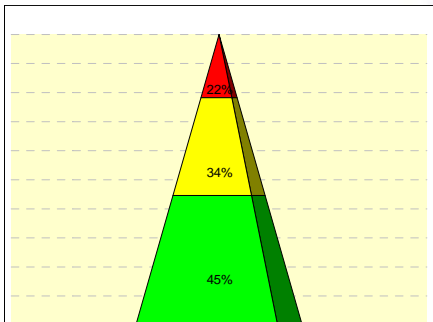


Where should Independent Sample School focus resources?

- About 13% of students were not proficient
 - 6% were questionable
 - 43% closed the gap
 - 7% were below basic
 - 50% were closing the gap

Where would you focus resources?

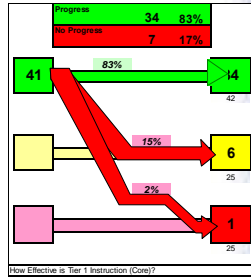
GW School a moderately needy school



At GW Elementary

Forty one (~45%) of students began on track.

Thirty four (83%) of them, remained on track.



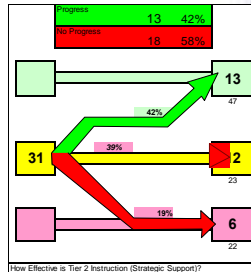
What are the implications?

At GW elementary

Thirty one (34%) of students entered with questionable performance.

Thirteen (42%) made enough progress where teachers could be confident they were back on track.

Nearly 20% of these students fell further behind?

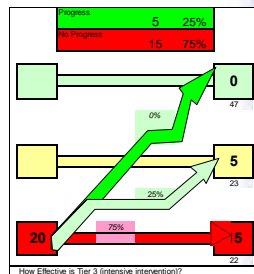


What are the implications?

At GW elementary

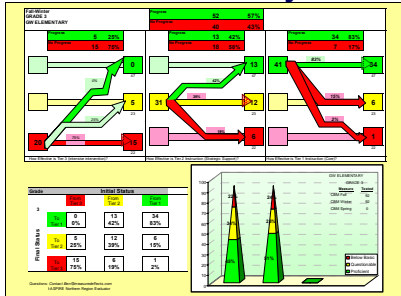
75% of students who began the year Below Basic maintained a status of Below Basic.

25% made progress.



What are the implications?

What should be the focus for GW Elementary?



Question: How can we use CBM to evaluate progress, and programs?

Answer:

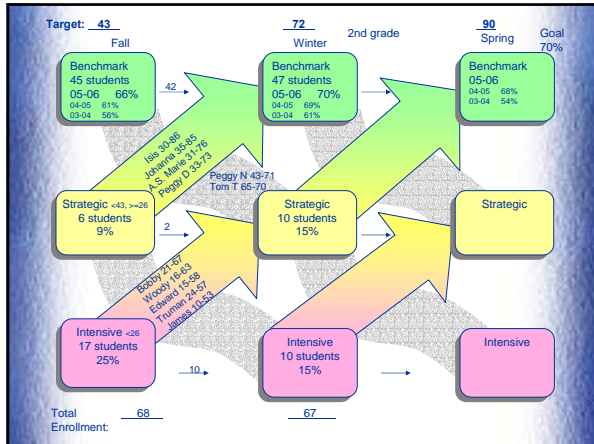
A Summary of Adequate Progress

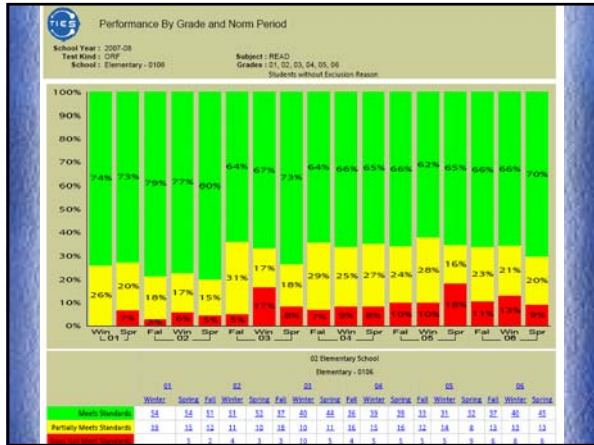
Big Idea #4.

We can disaggregate progress to examine student progress based on initial skill, instructional program, tier, etc...

Measuring Outcomes

- Evaluating performance at the group level
 - Grade / Group
 - Building
 - District
 - Growth and Status





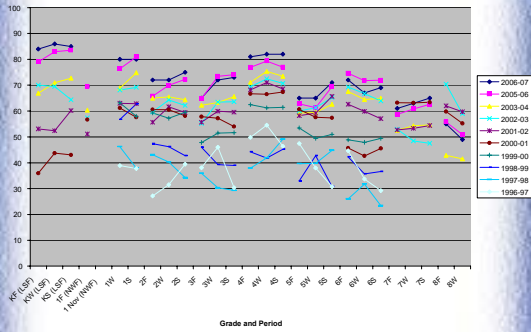
Instructional Recommendation	Participation in Core	Supplemental & Intervention Programs/Strategies	Supplemental & Intervention Program Delivery	Frequency of AllWeb progress monitoring	Determining Instructional Effectiveness
Benchmark: Who: All benchmark students When: Activities: Group Size:			Who: When: Time: Group Size:	3 times/year Fall goal: Winter goal: Spring goal:	Who: How Often: Criteria:
Strategic: Who: All strategic students When: Activities: Group Size:			Who: When: Time: Group Size:	Twice/month Fall goal: Be at or above goal line Winter goal: Be at or above goal line Spring goal:	Who: How Often: Criteria: QR# data

CSI Maps developed by Oregon Reading First Center

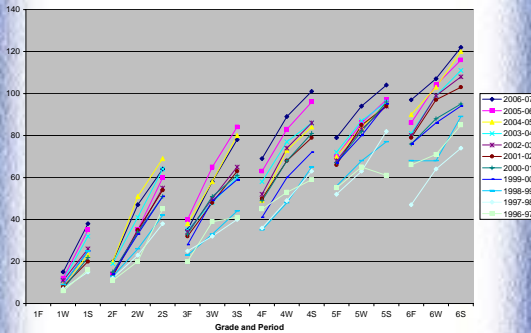
Instructional Recommendation	Participation in Core	Supplemental & Intervention Programs/ Strategies	Supplemental & Intervention Program Delivery	Frequency of AIMSweb progress monitoring	Determining Instructional Effectiveness
Intensive:	Who: All intensive students When: Activities: Group Size:		Who: When: Time: Group Size:	Once/week Fall goal: Be at or above goal line Winter goal: Be at or above goal line Spring goal:	Who: How Often: Criteria: ORF data

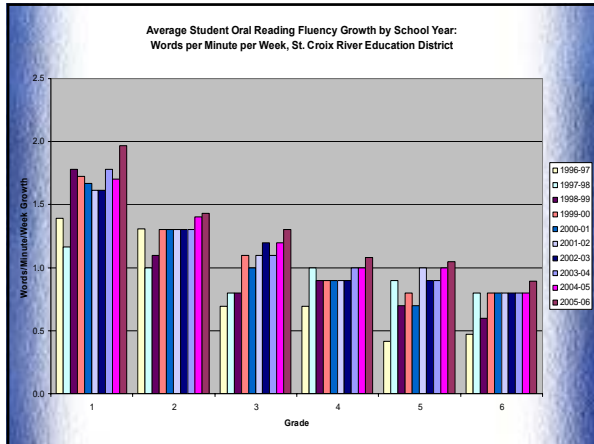
CSI Maps developed by Oregon Reading First Center

St. Croix River Education District - Percent Above Target (Reading)



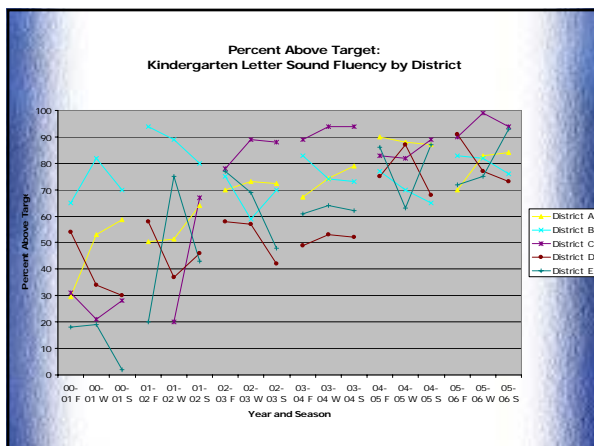
All-SCREd - Historical 10th percentile scores (ORF)

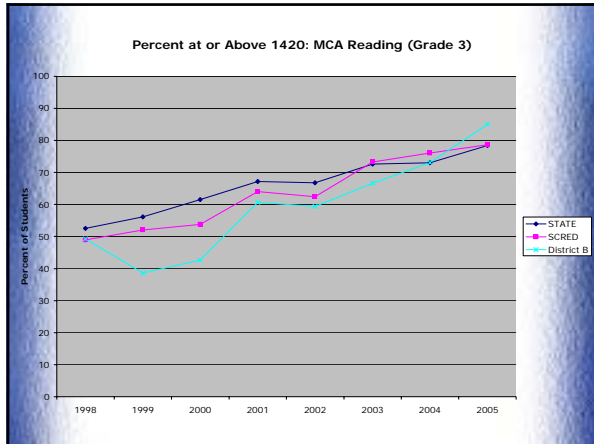


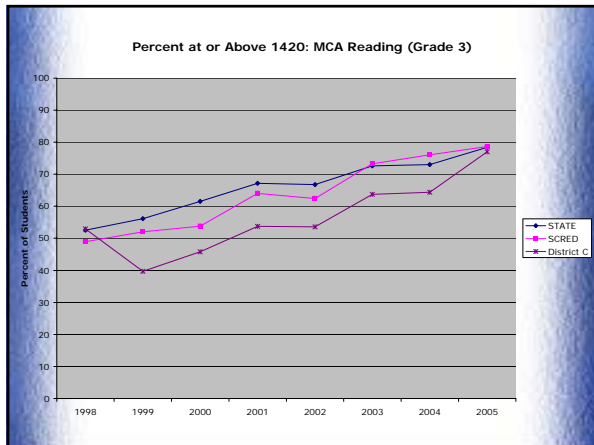


SCRED Reading Model's effect on student progress

- The average student growth in Oral Reading Fluency has improved at every grade level since the reading model's inception in 1996-97.
- Most grade levels saw an average improvement in student growth rates of 0.4 to 0.6 words per minute per week over this time period.
- A student who grows 0.5 words per minute per week faster than his or her peers will be about 20 words per minute higher at the end of a school year.
- In the spring of third grade in SCRED, a student who reads 107 words/minute has a 72% probability of reaching grade-level standards on the Minnesota Comprehensive Assessment in Reading. A student who reads 87 words/minute has only a 49% probability of reaching grade-level standards.







BIG IDEAS

- Scores on CBM are related to results of high-stakes testing
- Progress on CBM is related to improved outcomes in General reading
- We can use local percentile ranks to match our resources with our needs
- We can disaggregate progress to examine student progress based on initial skill, or by instructional program.

Some final thoughts about CBM & GOM

- Reliability
- SEM
- User norms
- Planning to maximize your quality of implementation

Reliability

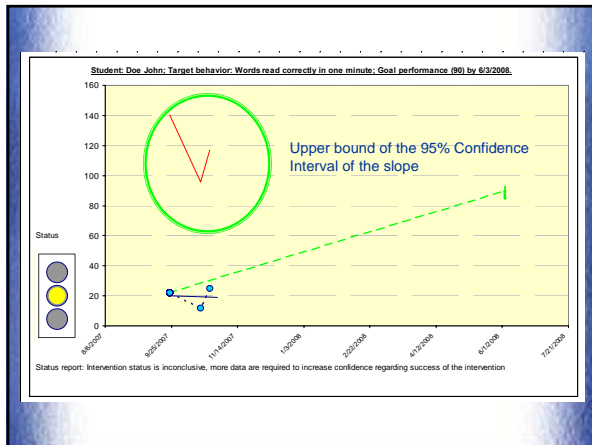
- The *consistency* of measurement
 - Across measurements
 - Across observers
 - Across items within an assessment
- For CBM, consider:
 - Fidelity of assessment checklist / check-in
 - Consistency of administration

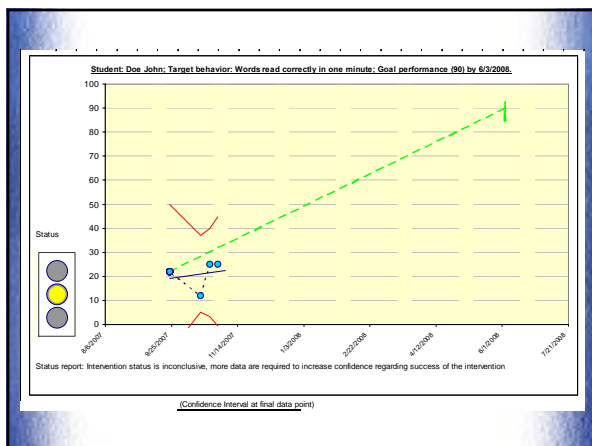
SEM

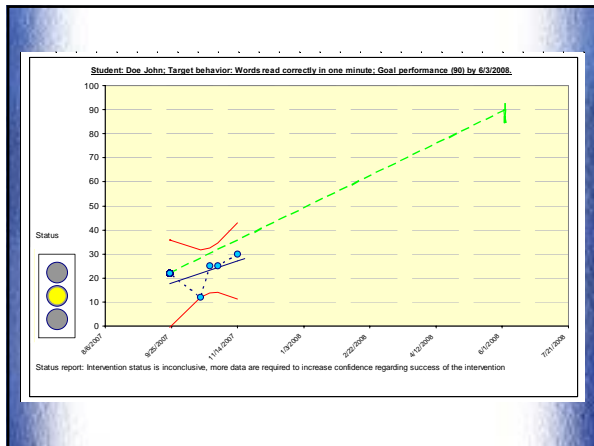
- Standard Error of Measurement
 - Typical range of differences between observed scores and true scores
 - Research has found SEM ranges for CBM-R from 8 – 12 WRCM (Christ & Silbergitt, 2007)
 - 68% confidence interval = +/-1 SEM
 - 95% confidence interval = +/-2 SEM
- Importance of multiple measures
- SEM for individual students vs. groups
- Resist temptation to make individual exceptions when analyzing groups!

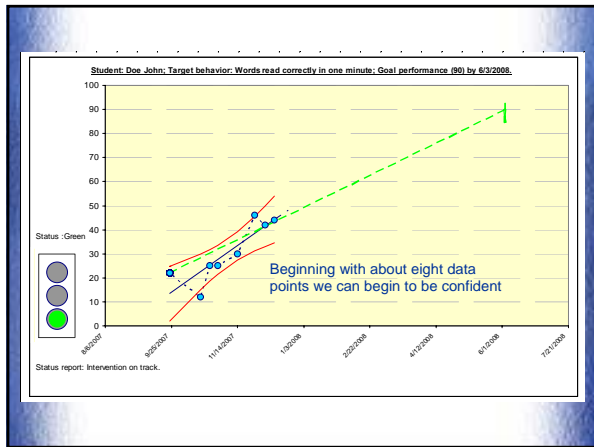
SEM and Progress Monitoring

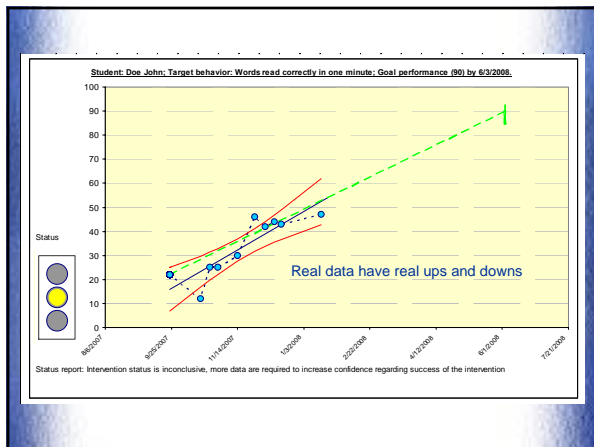
- Standard Error of Measurement is a real issue with progress monitoring and not just with CBM.
- The degree of confidence that we have in our decisions is important –
- How confident should we be?
- Let's look at a progress monitoring graph of a real student from last year.

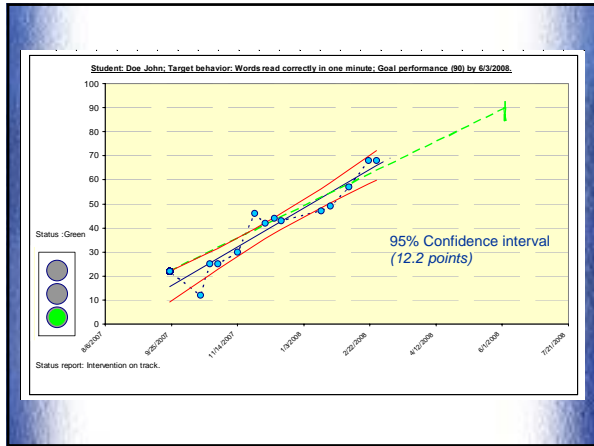


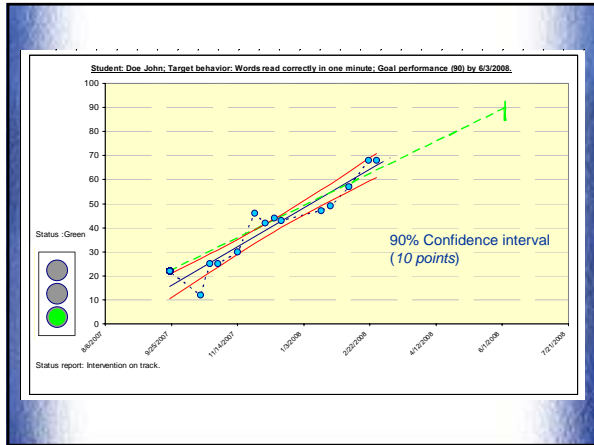


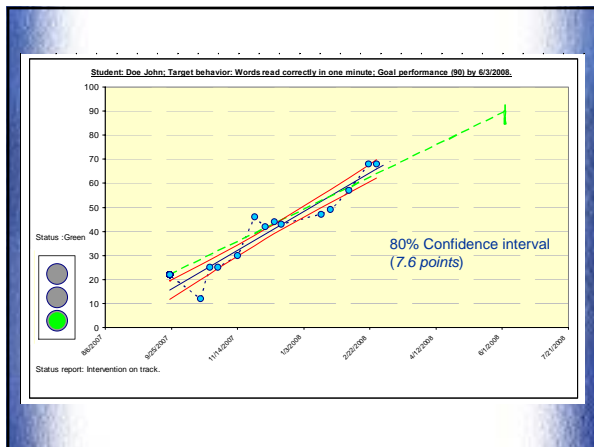


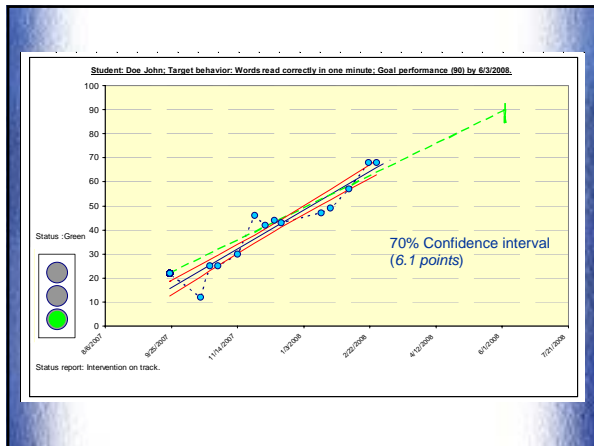


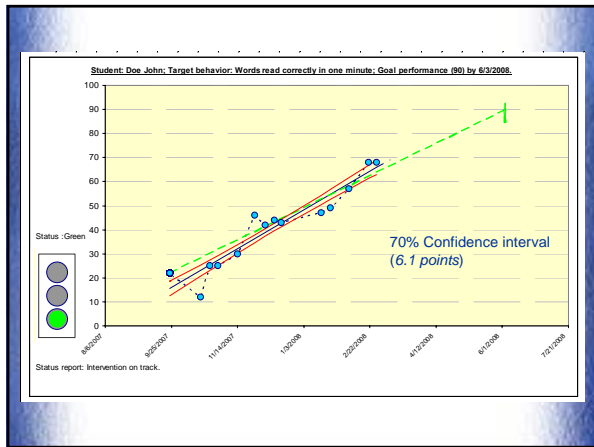












Being confident with progress monitoring

- ...is more than just projecting out the trend line
- ...typically requires far more than 3 data points
- ...requires informed decision-making

- You can download a template for progress monitoring with confidence from MeasuredEffects.com
- File name: **DBG_SL_free_V1.02.xls**

Measured Effects

Issues with “Off-the-shelf” tools for norms

- **User norms**
 - Example: NWEA norms changed from 2002 to 2005
 - Potential selection bias
- **Norms dependent on reliable and appropriate use of tool**
 - Example: Math Facts vs. M-CBM
 - Appropriate scoring?
 - Passage differences?
- **Tools display data in specific and proprietary ways**
 - Confusing to consumer
 - Requires access to and maintenance of multiple tools

Develop a plan for data management

- **Determine desired data elements, and gaps with current practices**
- **Identify critical staff and software**
- **Develop your plan**
 - Data collection and data processing timelines
 - Coordinate timelines with critical staff
 - Coordinate timelines with key events (i.e., team meetings, retreats, etc.)

Develop a plan for staff development

- **Accompany implementation with decisions about strategies for data use**
 - When will I use these data?
 - Which staff will be involved?
 - What decisions (changes) will be made as a result?
 - How will I continue to monitor the data to see if the change was effective?
- **Accompany training on assessment & software use with training on data-based decision making**
 - “Training before clicking”
 - Models for implementation
 - “Actionable” data means take action in response

Questions?

Ben Ditkowsky

Ben@MeasuredEffects.com

<http://measuredeffects.com>

Ben Silbergliitt

Ben.Silbergliitt@ties.k12.mn.us

<http://www.ties.k12.mn.us/DDDM.html>
