COULD YOUR DIBELS, AIMSWEB, OR STAR PREDICTION MODELS BE INACCURATE?

Improving Your Prediction Modeling within an RTI Framework

INTRODUCTIONS

- Timothy J. Runge, PhD, NCSP
  - Indiana University of Pennsylvania
  - trunge@iup.edu
- Ellie Edman, DEd, NCSP
  - Lincoln Intermediate Unit
  - ewedman@iu12.org
- Matthew Ferchalk, DEd, NCSP
  - Northern Lebanon School District
  - mferchalk@norleb.k12.pa.us

2

OUTLINE

1. Universal Screening and Basic Prediction Statistics
2. Suggestions for Improving Prediction Accuracy
3. Case Study Illustration

3

UNIVERSAL SCREENING

- Universal screening is a systematic assessment of the students on an academic indicator (Deno, 2003)
- They correspond to Tier 1 within an RtI model and should assess all students in a grade level
- They are typically assessed three times per year in the fall, winter, and spring of the school year (Hughes & Dexter, 2011)
- Ideally, universal screening assessments are cost-effective and quick and easy to administer

4

UNIVERSAL SCREENING

- Universal screening measures should be:
  - closely linked to state or national standards;
  - able to be administered to large groups of students in an efficient manner;
  - administered frequently and sensitive to small amounts of change so that student growth can be effectively assessed;
  - user-friendly and efficiently organized to show student progress on each individual skill assessed (Kovaleski & Pedersen, 2008);
  - Curriculum Based Measures such as DIBELS, AIMSweb, or STAR meet all of the criteria for a good universal screening assessment.

5

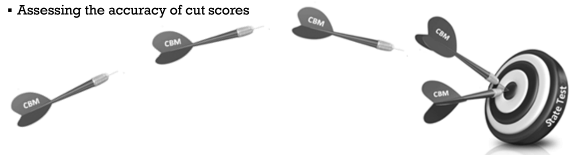
UNIVERSAL SCREENING

- Universal screening data can be used to:
  - make adjustments that are needed to improve the core curriculum (Ikeda, Nessen, & Witt, 2008);
  - determine if supplemental interventions are necessary;
  - highlight the pervasiveness and severity of academic problems;
  - judge the effectiveness of both the core curriculum and intervention supports.

6

ACCURACY OF UNIVERSAL SCREENING ASSESSMENTS

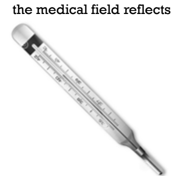
- CBM must be tasks that are highly predictive of overall academic outcomes (Deno, 2003; Ikeda, Neessen & Witt, 2008)
- Numerous studies that show a high correlation between CBM and general reading outcomes (Deno, Mirkin, & Chiang, 1982; Hintze & Silberglitt, 2005; Hosp & Fuchs, 2005; Jenkins & Jewell, 1993; Shinn et al., 1992)
- Accuracy is relative to its use and the criterion you have for success
- Assessing the accuracy of cut scores



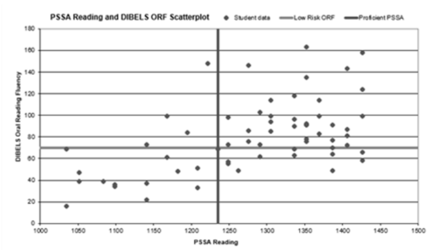
7

ACCURACY STATISTICS

- Interpretations of accuracy statistics are derived from a medical model approach indicating a positive or negative result (Silberglitt, 2008; VanDerHayden, 2011)
- A Positive result indicates the presence of an academic problem as positive test in the medical field reflects the presence of a disease



8



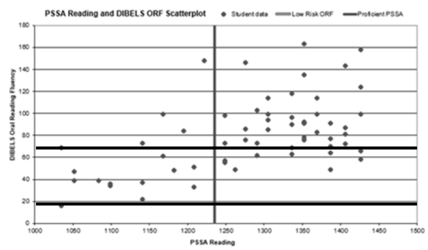
MEDICAL MODEL FOR INDICATOR ACCURACY

True Positives

True Negatives

False Positives

False Negatives



ACCURACY STATISTICS

Sensitivity

Specificity

Positive Predictive Power

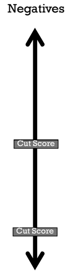
Negative Predictive Power

Each measures the balance between Positives and Negatives in different ways.

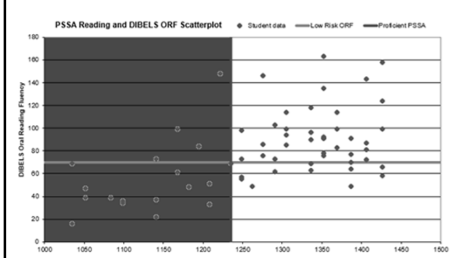
ACCURACY STATISTICS

- The cut score determines the percentage of false positives and false negatives
  - Reduce cut score and false positives decrease while false negatives increase.
  - Increase the cut score and false negatives decrease while false positives increase.

False Negatives



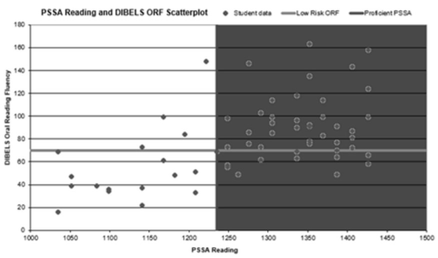
False Positives



SENSITIVITY

Sensitivity refers to the percentage of students who were correctly predicted to fail the outcome measure out of all of the students who actually failed the outcome (Silberglitt, 2008).

Failed DIBELS & PSSA/ Total number who failed PSSA (Cogan-Ferchalk, 2013)

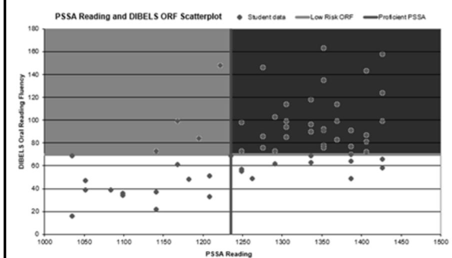


SPECIFICITY

Specificity refers to the percentage of students who were accurately predicted to pass the outcome out of all of the students who passed the outcome measure (Silberglitt, 2008).

Passed DIBELS & PSSA/ Total number who passed PSSA (Cogan-Ferchalk, 2013)

13

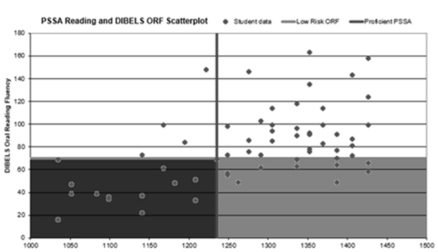


NEGATIVE PREDICTIVE POWER

Negative Predictive Power (NPP) refers to the proportion of students that were predicted to pass who actually passed the outcome measure (Silberglitt, 2008).

Passed DIBELS & PSSA/ Total number who passed DIBELS (Cogan-Ferchalk, 2013)

15



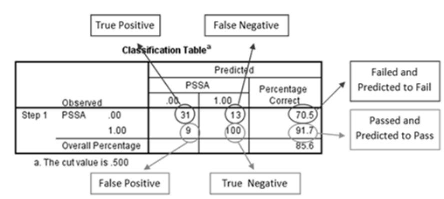
POSITIVE PREDICTIVE POWER

Positive predictive power (PPP) indicates the proportion of students that were predicted to fail and actually failed the outcome measure (Silberglitt, 2008).

Failed DIBELS & PSSA/ Total number who failed DIBELS (Cogan-Ferchalk, 2013)

14

ACCURACY IN SPSS



16

OUTLINE

1. Universal Screening and Basic Prediction Statistics
2. Suggestions for Improving Prediction Accuracy
3. Case Study Illustration

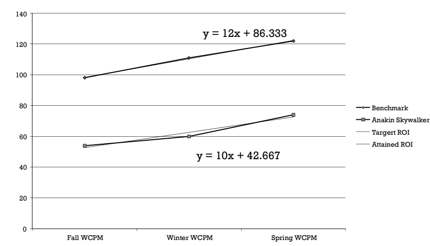
17

LIMITATIONS OF NATIONALLY-DERIVED CUT SCORES

- Developed based on predicting (typically) 40th percentile rank on an external criterion <u>other than PSSA</u>
- Publishers balance sensitivity, specificity, and other statistics to maximize the number of true positives and true negatives
- Lack of consensus about whether 40th PR is appropriate threshold
- Systems-level feedback is stronger when using local norms
  - Dedicated professional development in phonemic awareness should translate to improved local norms over multiple-years

18

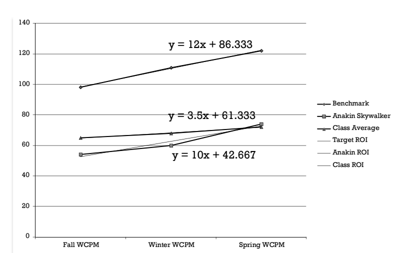
HOW IS ANAKIN DOING (6TH GRADE)?



Describe growth in these data. Acceptable?

19

HOW IS ANAKIN DOING (6TH GRADE)?



Now describe growth in these data. Acceptable?

20

NATIONAL V. LOCAL NORMS – 3RD GRADE AIMSWEB SPELLING (CLS)

	Fall	Winter	Spring	ROI
National Benchmarks	83	92	100	0.447 / wk
Dylan	79	91	98	.528 / wk

How is Dylan performing?

21

NATIONAL V. LOCAL NORMS – 3RD GRADE AIMSWEB SPELLING (CLS)

	Fall	Winter	Spring	ROI
National Benchmarks	83	92	100	0.447 / wk
Dylan	79	91	98	0.528 / wk
Local Benchmarks	95	106	126	0.861 / wk

How is Dylan performing now?

22

LIMITATIONS OF NATIONALLY-DERIVED CUT SCORES (CONTINUED)

- District demographics may not match the demographics of the normative group (Steward & Silberglitt, 2008)
  - Local norms may reduce bias in decision-making (Shinn, 1988)
  - Comparison group easier to explain to teachers and parents
  - Facilitate intra-classroom comparisons (versus national comparisons) (Shinn, Tindal, & Spira, 1997)
- The external criterion may or may not be as challenging as the PSSA (Kingsbury, Olson, Cronin, Hauser, & Houser, 2003)
- Fundamentally, we want to know how our students will do on the PSSA, not something like the GRADE, GMADE, or SAT

23

LIMITATIONS OF SENSITIVITY & SPECIFICITY

- Remember, why do we screen students?
  - To determine their level of **risk** for an outcome (e.g., failing the PSSA; significant behavioral challenges)
  - Sensitivity and specificity don't give us an indication of **risk** for an outcome
    - 80% Sensitivity = 80% of all students who fail PSSA earned that score or below
    - Tells us the proportion of students who achieved an outcome
    - Does NOT tell us probability that a student with that score will achieve an outcome (e.g., fail PSSA)
  - Educators typically want to know: "How at-risk is this student for failing the PSSA?"
  - Sensitivity and specificity will **not** give you this information

VanDerHeyden (2011)

24

A NEW STATISTIC: LIKELIHOOD RATIO

- A **Likelihood Ratio** (LR) is the probability that a particular outcome will occur given a particular screening score

- Examples:

Strep test screeners are use to diagnose the probability that someone has strep throat without resorting to costly, invasive, and often unnecessary tests

Battery of tests to determine if someone has appendicitis

POSITIVE LIKELIHOOD RATIO (LR+)

- The probability (likelihood) of a positive outcome (true positive) given a particular screening score

- In the context of medicine: the probability that a positive screen means disease is present

- In context of education: the probability that a positive screen means the student will fail the PSSA (or other true positive)

CALCULATING AND INTERPRETING LR+

$$LR+ = \frac{\text{Sensitivity}}{(\text{1- Specificity})}$$

- Purpose: to **rule in** the probability of a disorder (or failing PSSA)

- Interpretative Guidelines:
  - 1-2 minimum probability
  - 2-5 small probability
  - 5-10 moderate probability
  - >10 large and conclusive probability

Office of Medical Education Research and Development, Michigan State University (n.d.)

NEGATIVE LIKELIHOOD RATIO (LR-)

- The probability (likelihood) of a negative outcome (true negative) given a particular screening score

- In the context of medicine: the probability that a negative screen means disease is **NOT** present

- In context of education: the probability that a negative screen (scoring above the threshold) means the student will **PASS** the PSSA (or other true negative)

CALCULATING AND INTERPRETING LR-

$$LR- = \frac{(1 - \text{Sensitivity})}{\text{Specificity}}$$

- Purpose: to **rule out** the probability of a disorder (or Passing PSSA)
- Interpretative Guidelines:
  - 0.5 – 1.0 minimum probability
  - 0.2 – 0.5 moderate probability
  - 0.0 – 0.2 large probability

Office of Medical Education Research and Development, Michigan State University (n.d.)

SOME OTHER STATISTICS TO CONSIDER

- Post-test probabilities "are used to estimate the probability that an individual has a condition given a positive test result of does not have a condition given a negative test result" (VanDerHeyden, 2011, p. 345) in the context of prevalence rates of the condition
- If we know the prevalence rate, then the post-test probabilities inform us about how much better (or worse) the screening is at predicting an outcome
- Example: if we know 15% of all students have an SLD in reading, then a quality screening instrument should be more likely to predict an SLD in reading than simply guessing 15 in 100 students is SLD in reading
- Prevalence rates influence the value of post-test probabilities; higher prevalence rates means post-test probabilities are less valuable

POST-TEST PROBABILITIES

- Positive Post-test Probability = probability that a positive screen (e.g., scoring at or below a threshold) will accurately predict a positive outcome (e.g., failing PSSA)
- Negative Post-test Probability = probability that a negative screen (e.g., scoring above a threshold) will accurately predict a negative outcome (e.g., passing PSSA)
- In most contexts, prevalence rates skew interpretability of Positive Post-test Probabilities and Negative Post-test Probabilities
- VanDerHeyden (2011) advocated against Post-test Probabilities when predicting failing / passing high-stakes tests

EXAMPLE OF LR+, LR-

- 3rd Grade Fall STAR Reading predicting Failing PSSA
- Balance of LR+ and LR-

Positive if Greater Than or Equal to	Sensitivity (% True Positives)	1 - Specificity (False Positive Rate; Type 1 Error)	Specificity (% True Negatives)	1-Sensitivity (False Negative Rate; Type II error)	Positive Likelihood Ratio (probability that the score predicts FAILING PSSA)	Negative Likelihood Ratio (Probability that the score predicts PASSING PSSA)
184.50	.200	0.000	1.000	0.800	#DIV/0!	0.8
191.00	.217	0.000	1.000	0.783	#DIV/0!	0.783333
200.00	.233	0.000	1.000	0.767	#DIV/0!	0.766667
212.00	.250	0.000	1.000	0.750	#DIV/0!	0.75
222.00	.267	0.000	1.000	0.733	#DIV/0!	0.733333
224.50	.267	.004	0.996	0.733	61.06667	0.73655
230.50	.283	.004	0.996	0.717	64.88333	0.71981
238.00	.300	.004	0.996	0.700	68.7	0.70307
247.50	.317	.004	0.996	0.683	72.51667	0.68633
258.00	.333	.004	0.996	0.667	76.33333	0.669591
262.00	.350	.004	0.996	0.650	80.15	0.652851
265.50	.367	.004	0.996	0.633	83.96667	0.636111
271.00	.383	.004	0.996	0.617	87.78333	0.619371
281.00	.400	.004	0.996	0.600	91.6	0.602632
290.50	.417	.004	0.996	0.583	95.41667	0.585892
298.00	.417	.009	0.991	0.583	47.70833	0.588473
303.50	.417	.013	0.987	0.583	31.80556	0.591077
304.50	.433	.013	0.987	0.567	33.07778	0.574189
309.50	.433	.026	0.974	0.567	16.53889	0.581913
315.50	.450	.026	0.974	0.550	17.175	0.564798
318.00	.450	.031	0.969	0.550	14.72143	0.567342
320.00	.450	.035	0.965	0.550	12.88125	0.56991
321.50	.467	.035	0.965	0.533	13.35833	0.55264

HOW TO CALCULATE LR+ AND LR-

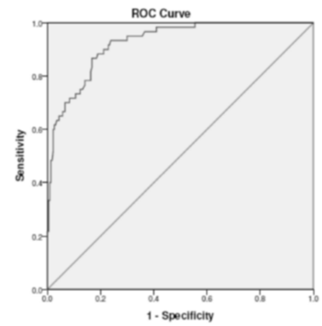
- Summary on Runge's website (http://www.iup.edu/page.aspx?id=124318)
- Resources needed:
  - Local benchmarking data and PSSA performance dichotomized for at least one complete cohort
  - Microsoft™ Excel (or something comparable)
  - SPSS (or something comparable)
  - Some time
  - Some wine (or beer, soda, coffee, snacks)
  - Patience!

33

(ABBREVIATED) STEPS TO CALCULATING LR+ AND LR-

1. Make sure data set is completely clean
2. Make sure PSSA performance is dichotomized
3. Perform Receiver Operating Curve Analysis with lower scores predicting failing PSSA
   - https://www.youtube.com/watch?v=_2zN2a3MgmU
4. ROC Results for 3rd Grade Fall STAR Reading on next slide

34



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s): Fall STAR SS

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.925	.017	.000	.892	.959

The test result variable(s): Fall STAR SS has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

35

Coordinates of the Curve

Test Result Variable(s): Fall STAR SS

Positive if Less Than or Equal To[a]	Sensitivity	1 - Specificity
70.00	.000	.000
72.50	.017	.000
75.00	.033	.000
80.00	.050	.000
86.00	.067	.000
92.00	.083	.000
97.50	.100	.000
101.00	.117	.000
105.50	.133	.000
110.50	.150	.000
114.00	.167	.000
116.50	.183	.000
119.00	.200	.000
125.50	.217	.000
132.00	.217	.004

5. Highlight these three columns and copy / paste into Excel spreadsheet on website

36

Positive if Greater Than or Equal to	Sensitivity (% True Positives)	1 - Specificity (False Positive Rate; Type 1 Error)	Specificity (% True Negatives)	1-Sensitivity (False Negative Rate; Type II error)	Positive Likelihood Ratio (probability that the score predicts FAILING PSSA)	Negative Likelihood Ratio (Probability that the score predicts PASSING PSSA)
78.00	0.000	0.000	1.000	1.000		1
80.50	.017	0.000	1.000	0.983	#DIV/0!	0.983333
83.50	.033	0.000	1.000	0.967	#DIV/0!	0.966667
87.00	.050	0.000	1.000	0.950	#DIV/0!	0.95
91.00	.067	0.000	1.000	0.933	#DIV/0!	0.933333
97.00	.083	0.000	1.000	0.917	#DIV/0!	0.916667
103.50	.100	0.000	1.000	0.900	#DIV/0!	0.9
107.50	.117	0.000	1.000	0.883	#DIV/0!	0.883333
127.00	.133	0.000	1.000	0.867	#DIV/0!	0.866667
158.50	.150	0.000	1.000	0.850	#DIV/0!	0.85
175.00	.167	0.000	1.000	0.833	#DIV/0!	0.833333
180.50	.183	0.000	1.000	0.817	#DIV/0!	0.816667
184.50	.200	0.000	1.000	0.800	#DIV/0!	0.8
191.00	.217	0.000	1.000	0.783	#DIV/0!	0.783333
200.00	.233	0.000	1.000	0.767	#DIV/0!	0.766667
212.00	.250	0.000	1.000	0.750	#DIV/0!	0.75
222.00	.267	0.000	1.000	0.733	#DIV/0!	0.733333
224.50	.267	.004	0.996	0.733	61.06667	0.73655
230.50	.283	.004	0.996	0.717	64.88333	0.71981
238.00	.300	.004	0.996	0.700	68.7	0.70307
247.50	.317	.004	0.996	0.683	72.51667	0.68633
258.00	.333	.004	0.996	0.667	76.33333	0.669591
262.00	.350	.004	0.996	0.650	80.15	0.652851
265.50	.367	.004	0.996	0.633	83.96667	0.636111

6. LR+ and LR- should automatically populate

OUTLINE

1. Universal Screening and Basic Prediction Statistics
2. Suggestions for Improving Prediction Accuracy
3. Case Study Illustration

COMPARISON OF DIFFERENT CUT SCORES

Side-By-Side Comparison of Nationally- and Locally-Derived Winter STAR Cut Scores

STAR SS	Sensitivity	Specificity	LR+	LR-	Difference in Number of Students: School A	School B	School C	School D
Nationally-Derived								
352	62%	92%	8.06	0.40	n/a	n/a	n/a	n/a
Locally-Derived								
387	82%	82%	4.45	0.22	10	10	7	9
379	78%	84%	4.72	0.26	7	9	7	9
366	75%	87%	5.92	0.29	4	6	4	4
359	70%	91%	7.63	0.33	1	3	1	1
351	63%	92%	8.06	0.40	0	0	0	0

Note. SS = Standard Score; LR+ = Positive Likelihood Ratio; LR- = Negative Likelihood Ratio

APPLYING LOCAL NORMS

Site

- Rural Pennsylvania district
- 3500 total students – 87% white; 13% non-white
- 4 elementary schools – K-6 configuration

Universal screenings – STAR Reading and STAR Math

Benchmark: 40th percentile and above
On Watch: 25th – 39th percentile
Intervention: 10th – 24th percentile
Urgent Intervention: 1st – 9th percentile

WHY LOCAL CUT SCORES FOR THIS DISTRICT?

National cut scores are too lenient – not identifying students who go on to fail the PSSA

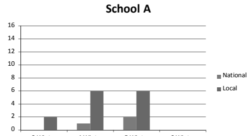
4th Grade Reading – 2012/2013 Academic Year

		PSSA_Descriptor					Total
			Advanced	Proficient	Basic	Below Basic	
Fall_STAR_Descriptor		4	1	10	2	2	19
	At/Above Benchmark	2	67	95	30	4	198
	On Watch	3	0	13	15	4	35
	Intervention	0	0	1	12	6	19
	Urgent Intervention	3	0	0	3	17	23
Total		12	68	119	62	33	294

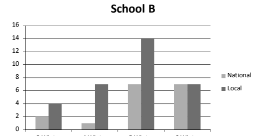
41

ON WATCH – FAILED PSSA READING (13-14)

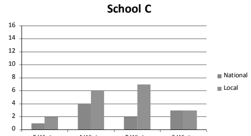
School A



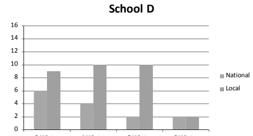
School B



School C



School D



42

CHOOSING LOCAL CUT SCORES

- Resource allocation – cannot identify unmanageable number of students in need of intervention
- Balance of technical qualities (LR+ and LR-)
- Important to correctly identify students who are likely to fail the PSSA (true positive) in order to provide additional supports

STAR 40th Percentile	Cut Scores	LR+	LR-	Difference			
				A	B	C	D
310	317	5.18	0.16	2	0	2	2
(LR+ ~4.9;	314	4.98	0.20	1	0	2	2
LR- ~0.26)	286	6.08	0.29	-4	-2	-4	-1

43

LOCAL VS NATIONAL CUT SCORES - READING

	Cut Score	LR+	LR-
3rd Grade			
National	310	4.9	0.26
Local	314	5.0	0.20
4th Grade			
National	402	7.8	0.41
Local	435	5.1	0.30
5th Grade			
National	479	9.5	0.42
Local	553	5.1	0.17
6th Grade			
National	573	5.5	0.44
Local	598	5.1	0.31

National cut scores identifying fewer students as "at risk," but with higher accuracy

The national cut scores were not catching enough students in need of intervention

44

APPLYING LOCAL CUT SCORES - STAR READING

Current school year (2014-2015)

Grade	National Cut Score	Local Cut Score	Difference in On Watch Students – By School			
			A	B	C	D
3	319	314	0	0	-1	-2
4	415	435	0	6	2	1
5	514	553	7	8	6	9
6	614	598	-1	-1	-6	-3

*Note – STAR adjusted Reading cut scores for 2014-15 (higher than 13-14)

SO, HOW DO WE USE THE CUT SCORES?

- Consider **both** national benchmark cut score and selected locally-derived cut scores that maximize LR+ and LR-
- Have conversations about the students that fall between the nationally- and locally-derived cut scores
- What other data do you have to help determine risk level of students that fall between those two cut scores?
- Considerations when making decisions about those students

CHANGE IN NUMBER REACHING BENCHMARK

STAR Reading Universal Screening Fall 2014

Grade	A		B		C		D	
	National	Local	National	Local	National	Local	National	Local
4	43/60 (72%)	43/60 (72%)	54/75 (72%)	48/75 (64%)	42/61 (69%)	39/61 (64%)	50/80 (63%)	49/80 (61%)
5	56/81 (69%)	48/81 (59%)	40/70 (57%)	32/70 (46%)	52/81 (64%)	46/81 (57%)	46/69 (67%)	37/69 (54%)

2014 PSSA Reading results –

3rd, 4th, and 6th grades at 74% proficiency
5th grade at 65% proficiency

TIER 2 READING INTERVENTION GROUPS

- Tier 2 Reading Intervention Groups by School

Grade	A		B		C		D	
	On Watch Tier 2	Total Tier 2	On Watch Tier 2	Total Tier 2	On Watch Tier 2	Total Tier 2	On Watch Tier 2	Total Tier 2
3	3	6	6	24	4	15	3	9
4	2	7	3	12	5	13	6	14
5	6	18	18	29	6	13	2	10
6	4	9	5	19	4	13	2	8

*Includes Tier 2 groups taught by intervention specialists, reading specialists, instructional aides, and classroom teachers

DATA TEAM CONVERSATIONS - READING

- Each building screened for Title 1 eligibility before data team meetings
  - Took several assessment scores into account
- On Watch students who did not qualify for Title 1 services
  - Should they be placed into a Tier 2 intervention?
  - Should they be supported within the classroom?
  - How does Johnny's performance on assessments X, Y, and Z support or contradict STAR results?
- Who collects additional data? It depends on the building.
  - Classroom teacher
  - Interventionist, reading specialist

49

DATA TEAM CONVERSATIONS - READING

What support will be given to On Watch Students?

- Support them in the classroom
  - Differentiated instruction
  - Fluid groups based on skill deficits on Study Island
  - Continue to PM to watch for lack of progress
- Provide Tier 2 intervention
  - Collect additional assessment data
  - Progress monitor for several weeks and examine trend line
  - Provide targeted Tier 2 intervention for all students who were On Watch (with supporting classroom performance)
    - This called for reallocation of instructional aides, push-in intervention, shuffling of benchmark students between remaining grade level teachers
    - On Watch students who did not fit into another Tier 2 intervention were placed in an isolated Tier 2 in certain buildings/grade levels

Additional data: STAR diagnostic reports, Study Island, ORF, CORE (decoding, vocabulary)

50

LOCAL VS NATIONAL CUT SCORES - MATH

	Cut Score	LR+	LR-
3rd Grade			
National	479	7.6	0.30
Local	497	5.5	0.20
4th Grade			
National	563	13.2	0.34
Local	591	4.9	0.27
5th Grade			
National	628	11.9	0.46
Local	678	5.07	0.17
6th Grade			
National	694	8.5	0.30
Local	710	5.3	0.20

National norms had high LR+, particularly at 4th and 5th grades

Local cut scores had biggest impact in 5th grade

51

APPLYING LOCAL CUT SCORES – STAR MATH

Current School Year (2014-2015)

Grade	National Cut Score	Local Cut Score	Difference in On Watch Students – By School			
			A	B	C	D
3	479	497	4	3	2	8
4	563	591	1	7	5	5
5	628	678	15	12	13	12
6	694	710	3	4	2	3

52

CHANGE IN NUMBER REACHING BENCHMARK - MATH

- STAR Math Universal Screening Fall 2014

Grade	A		B		C		D	
	National	Local	National	Local	National	Local	National	Local
3	57/66 (86%)	53/66 (80%)	65/85 (76%)	62/85 (73%)	50/59 (85%)	48/59 (81%)	67/84 (80%)	59/84 (70%)
4	50/60 (83%)	49/60 (82%)	61/75 (81%)	54/75 (72%)	52/61 (85%)	47/61 (77%)	63/78 (81%)	58/78 (74%)
5	66/80 (83%)	51/80 (64%)	55/70 (79%)	43/70 (61%)	73/81 (90%)	60/81 (74%)	49/68 (72%)	37/68 (54%)
6	50/61 (82%)	47/61 (77%)	59/80 (74%)	55/80 (69%)	59/75 (79%)	57/75 (76%)	47/62 (76%)	44/62 (71%)

PSSA Math Proficiency Levels:

2013: 3rd – 84%; 4th – 81%; 5th – 75%; 6th – 77%
2014: 3rd – 76%; 4th – 76%; 5th – 76%; 6th – 85%

DATA TEAM CONVERSATIONS - MATH

- There is no structured math intervention in daily schedule
- How do we support students in need of math intervention?
  - Provide intervention within or between classrooms during one or two days of the 6-day cycle
  - Differentiated small groups in the classroom
- Data collection - ongoing to inform instructional planning
  - Pre-testing and post-testing with EM
  - Study Island assessments
  - easyCBM

HOW DID LOCAL CUT SCORES IMPACT DECISIONS?

- On Watch does not = safe
  - Tend to treat this category as "bubble kids" who are not in need of targeted intervention
  - In reality, they are below benchmark and should be considered at risk to fail the PSSA
- Some grade levels embraced this shift in mindset
  - Providing math intervention on off days for reading intervention
  - Organizing additional Tier 2 reading intervention groups to include On Watch students
    - These students would previously have been in a large Tier 1 group
  - Use of Study Island assessments to inform intervention and small group instruction
- Overall understanding that we need to be providing more support for On Watch students in the classroom and through structured intervention
  - Still necessary to collect additional information to confirm or refute screening results

OUTLINE

1. Universal Screening and Basic Prediction Statistics
2. Suggestions for Improving Prediction Accuracy
3. Case Study Illustration

57

REFERENCES

Cogan-Ferchalk, J. R. (2013). *The utility of the Modified Checklist for Autism in Toddlers in a preschool-age special education sample*. Indiana University of Pennsylvania, Indiana, PA.

Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education*, *37*, 137-161.

Deno, S. L., Mirkin, P., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children*, *49*, 36-45.

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review*, *34*, 372-386.

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review*, *34*, 9-26.

Hughes, C. A., & Dexter, D. D. (2011). Response to intervention: A research-based summary. *Theory Into Practice*, *50*(1), 4-11. doi:10.1080/00405841.2011.534909

Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children*, *59*, 421-432.

Kingsbury, G. G., Olson, A., Cronin, J., Hauser, C., & Houser, R. (2003). *The state of state standards: Research investigating proficiency levels in fourteen states* (Technical Report). Lake Oswego, OR: Northwest Evaluation Association.

Kovaleski, J. F., & Pedersen, J. A. (2008). Best practices in data analysis teaming. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 115-130). Bethesda, MD: National Association of School Psychologists.

58

REFERENCES

Ikeda, M. J., Neessen, E., & Witt, J. C. (2008). Best practices in universal screening. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 103-114). Bethesda, MD: National Association of School Psychologists.

Office of Medical Education Research and Development, Michigan State University. (n.d.). *Likelihood ratios part 1: Introduction*. Retrieved from: http://omerad.msu.edu/ebm/Diagnosis/Diagnosis6.html

Shinn, M. R. (1988). Development of curriculum-based local norms for use in special education decision making. *School Psychology Review, 17*, 61-80.

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, *21*, 459.

Shinn, M. R., Tindal, G., & Spira, D. (1997). Special education referrals as an index of teacher tolerance: Are teachers imperfect tests? *Exceptional Children, 54*, 32-40.

Silberglitt, B. (2008). Best practices in using technology for data-based decision making. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 1869-1884). Bethesda, MD: National Association of School Psychologists.

Stewart, L. H., & Silberglitt, B. (2008). Best practices in developing academic local norms. In A. Thomas & J. Grimes (Eds), *Best practices in school psychology, volume 5* (pp. 225-242). Bethesda, MD: National Association of School Psychologists.

VanDerHeyden, A. M. (2011). Technical adequacy of response to intervention decisions. *Exceptional Children, 77*, 335-350.

59