CBM-R Progress Monitoring Outcomes:
Reliability and Validity of Slope Estimates Compared to 3, 4, and 5 Point Decision Rules

Reliability and Validity of Slope Estimates Compared to 3, 4, and 5 Point Decision Rules

Theodore J. Christ, PhD
(tchrist@umn.edu)
With Amanda Pike & Barbara Monaghen
University of Minnesota

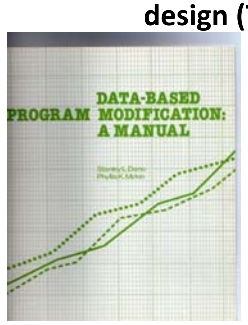
Our Plan & Whose in the Room

- Do you use progress monitoring data?
- Do you consider the data "noisy"?
- Do you have trouble estimating trends?

Curriculum Based Measurement (CBM)

- Assessment of academic skills that emphasizes direct measurement of academic behaviors
- Characterized by frequent measurement and short duration
- Rate of change over time is compared to an expected rate of progress
  - Estimates of Growth
  - Data Point Decision Rules

Enter time series data & the single case design (The early 1970s)



Data-Based Program Modification: A Manual (Deno & Mirkin, 1977)

1) As the present time we are unable to prescribe specific and effective changes to instruction for individual pupils with certainty. Therefore, changes in instructional programs which are arranged for an individual child can be treated only as hypotheses which must be empirically tested before a decision can be made on whether they are effective for that child (p. 11).

University of Minnesota
Driven to Discover
RIPS Research Institute for Problem Solving

Data-Based Program Modification: A Manual (Deno & Mirkin, 1977)

2) Time series research designs are uniquely appropriate for testing instructional reforms (hypotheses) which are intended to improve individual performance (p. 11).

3) Special education is an intervention system, created to produce reforms in the educational programs of selected individuals, which can (and, now, with due process requirements, must) be empirically tested (p. 13).

University of Minnesota
Driven to Discover
RIPS Research Institute for Problem Solving

Data-Based Program Modification: A Manual (Deno & Mirkin, 1977)

4) To apply time series designs to (special) educational reforms we need to specify the data representing the "vital signs" of educational development which can be routinely (frequently) obtained in and out of school (p. 14).

University of Minnesota
Driven to Discover
RIPS Research Institute for Problem Solving

Data-Based Program Modification: A Manual (Deno & Mirkin, 1977)

5) Testing program modifications (reforms) requires well-trained professionals capable of using time series data analysis to draw valid conclusions about program effects (p. 15).

University of Minnesota
Driven to Discover
RIPS Research Institute for Problem Solving

Evaluate Growth

- Compare to Expected Rate of Growth
  - Goal/Aimline Decision Rules
- Estimate Trend/Slope
  - Ordinary Least Squares Regression (OLS)
  - Split middle
  - Quarter Intersect
  - Split Middle

Data Point Decision Rules

- Alternative method to evaluating student progress

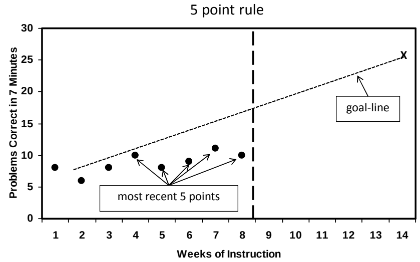
If (**3,4,5**) most recent data points are:

- ABOVE the line keep current intervention, ⟶ increase goal
- BELOW the line keep current goal, ⟶ modify instruction
- Neutral – points are above and below the line ⟶ maintain current goal and instruction

Decision Rules

National Center for Progress Monitoring

5 point rule

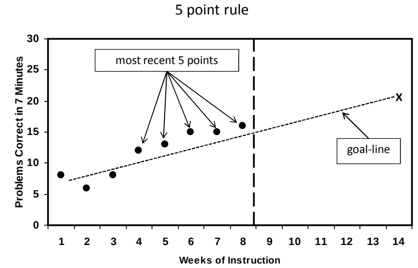


Slide from

Decision Rules

National Center for Progress Monitoring

5 point rule



Slide from

Use of Slope Estimates

SOMETHING ABOUT SLOPE ESTIMATES

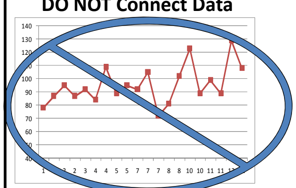
Progress Monitoring

- How many data points do I need to collect?
  - 10 Data Points (professional literature)
    - Shinn, M. R. (2002). Best practices in using curriculum-based measurement in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology. (pp. 671-698). Bethesda, MD: National Association of School Psychologists.*
  - **10 Weeks with 2 data points per week**
    - Christ, T. J. (2006). Short term estimates of growth using curriculum-based measurement of oral reading fluency: Estimates of standard error of the slope to construct confidence intervals. *School Psychology Review, 35(1), 128-133.*
    - Also
      - Good, R. H., & Shinn, M. R. (1990). Forecasting accuracy of slope estimates for reading curriculum-based measurement: Empirical evidence. *Behavioral Assessment, 12(2), 179-193.*
      - Shinn, M. R., Good, R. H., & Stein, S. (1989). Summarizing trend in student achievement: A comparison of methods. *School Psychology Review, 18(3), 356-370.*

University of Minnesota
Driven to Discover℠
RIPS Research Institute for Problem Solving

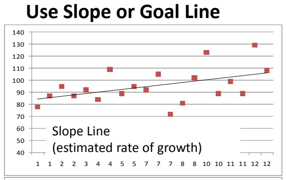
Estimating Progress

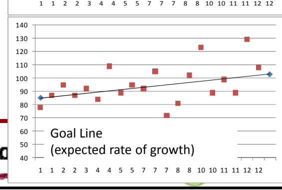
DO NOT Connect Data



Each assessment is an **ESTIMATE** of reading achievement

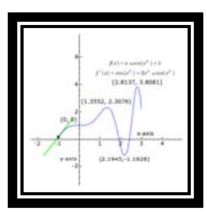
Use Slope or Goal Line



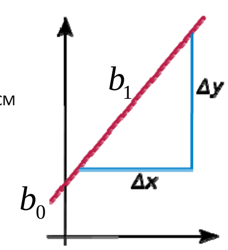


University of Minnesota
Driven to Discover℠

What is slope?



50 WRCM

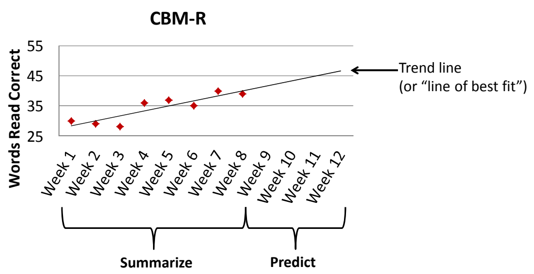


$$b_1 = \frac{\Delta y}{\Delta x}$$

10 weeks

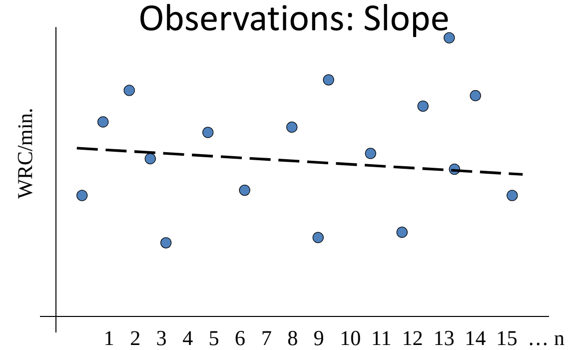
University of Minnesota
Driven to Discover℠
RIPS Research Institute for Problem Solving

Uses of Estimates of Growth



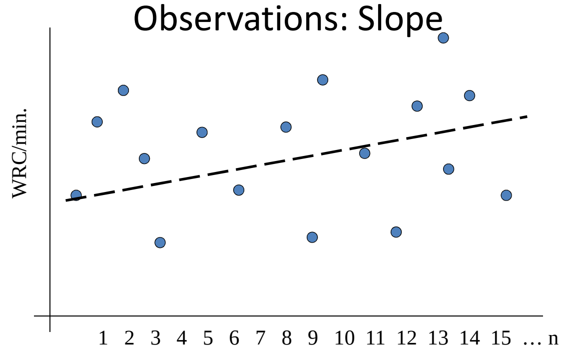
University of Minnesota
Driven to Discover
RIPS

Observations: Slope



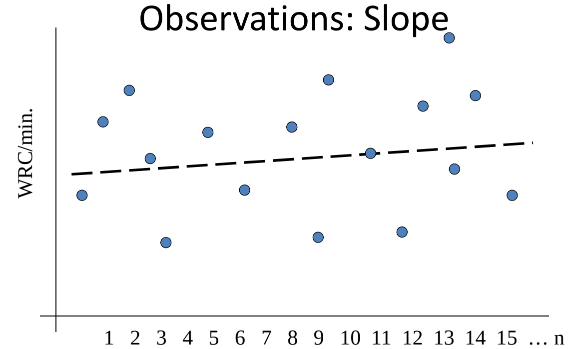
University of Minnesota
Driven to Discover
RIPS

Observations: Slope



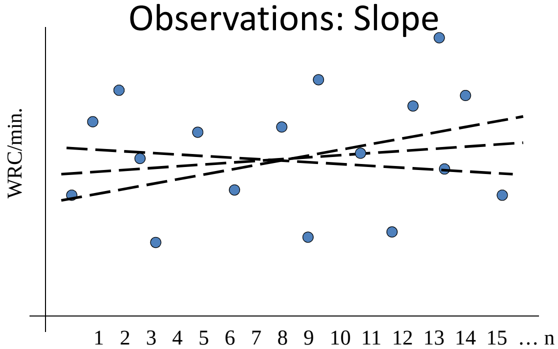
University of Minnesota
Driven to Discover
RIPS

Observations: Slope

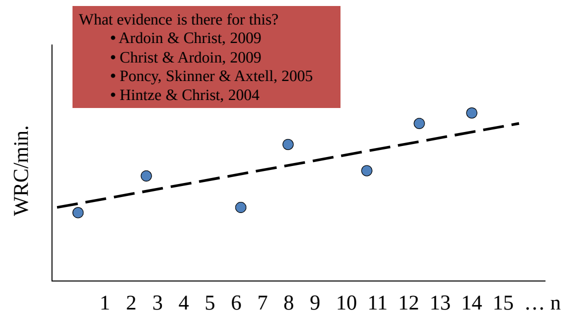


University of Minnesota
Driven to Discover
RIPS

Observations: Slope



Observations: Slope

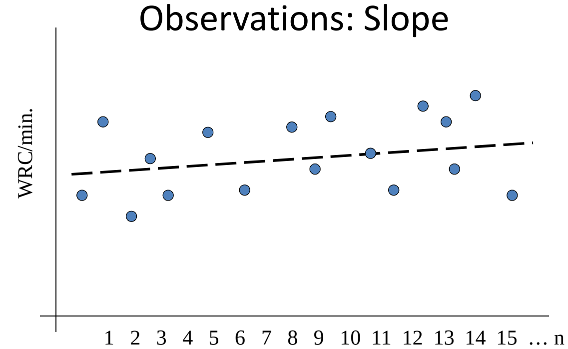


Ordinary Least Squares (OLS) Regression

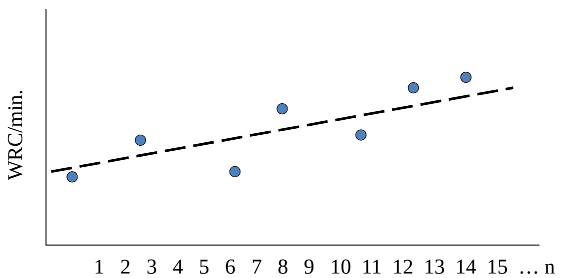
- OLS fits the data set with a monotonic linear model so that the residual is minimized
    - i.e. residual is the difference between point predictions and observed scores
  - The residual is often conceptualized
    - as error in model or measurement
    - variability not explained in the model (linear growth), such as antecedent or environmental conditions

Shapiro (2004); Shinn, Good, & Stein (1989)

Observations: Slope

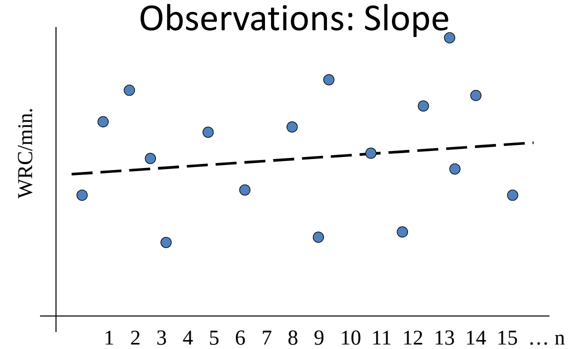


Observations: Slope

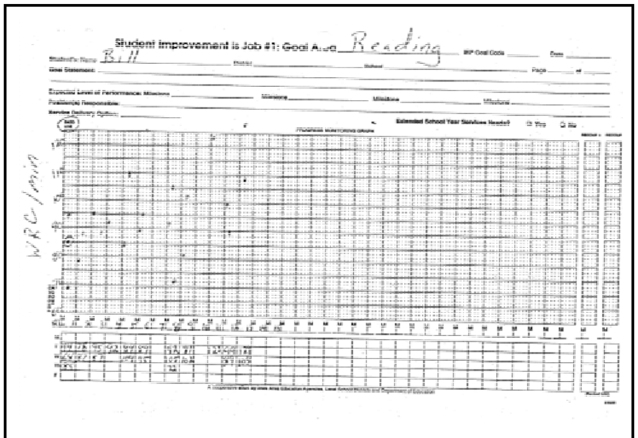


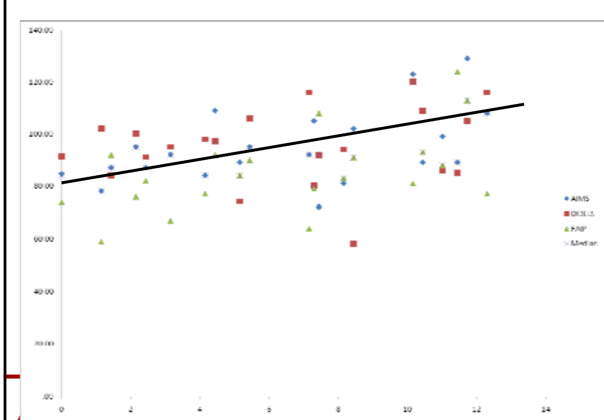
University of Minnesota
Driven to Discover
RIPS Research Institute for Problem Solving

Observations: Slope



University of Minnesota
Driven to Discover
RIPS Research Institute for Problem Solving



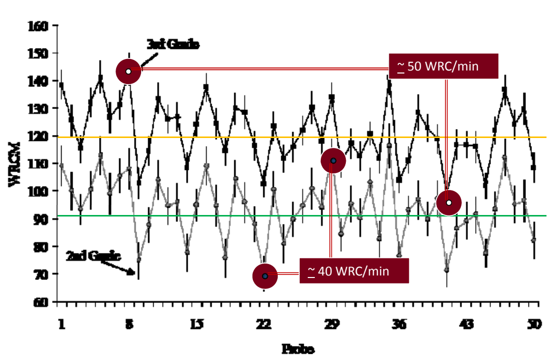


Driven to Discover

MEASUREMENT ERROR

Residual (Measurement) Error

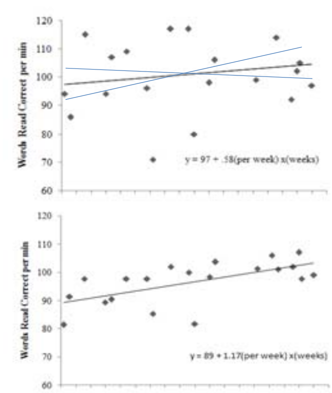
- A variety of influences impact CBM-R outcomes including:
  – The characteristics of the administrator (Derr-Minneci & Shapiro, 1992)
  – Setting or location (Derr-Minneci, 1990)
  – Delivery of directions (Colon & Kranzler, 2006)
  – Instrumentation (Poncy, Skinner, & Axtell, 2005; Christ & Ardoin, 2009)
- Together these influences result in **measurement error – or residual in growth models**



Christ, T. J., & Ardoin, S. P. (2009). Curriculum-based measurement of oral reading: Passage equivalence and probe-set development. *Journal of School Psychology*, 47, 55-75.

Define the Dependent Variables

- **Intercept** – intersection of line of best fit & y-axis
- **Slope** - weekly growth per 7 days
- Standard error of the slope (**SEb**) – relative imprecision of slope for weekly growth
- Standard error of the estimate (**SEE**) – relative impression of point predictions



Psychometrics

- Standard Error of Measurement (SEM)
  - amount an observed score is likely to fluctuate
  - Inversely proportional to the reliability of a test

$$SEM = SD_x \sqrt{1 - r_{xx}}$$

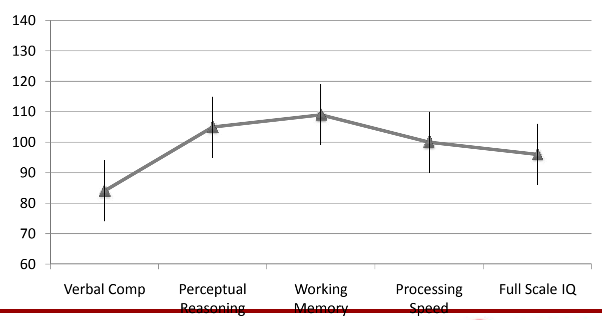
(we have successfully developed generic estimates of *SEM* within CTT)

Psychometrics

- Confidence Interval
  - band or range around the observed score in which the true score is most likely to fall
  - CI ~ N(0,SEM)
    - **CI(68%) = 1.00 * SEM +/- *x***
    - CI(85%) = 1.44 * SEM +/- *x*
    - CI(90%) = 1.65 * SEM +/- *x*
    - **CI(95%) = 1.96 * SEM +/- *x***
    - CI(99%) = 2.58 * SEM +/- *x*

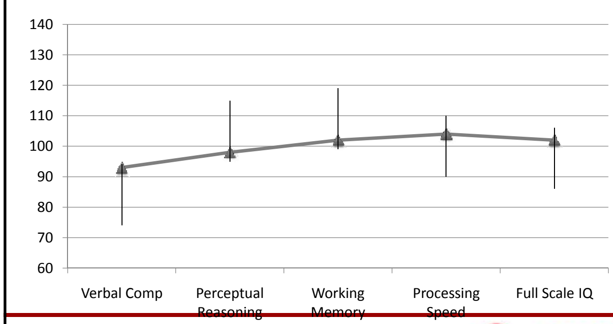
University of Minnesota
Driven to Discover
RIPS

Self-Referenced: Profile Analysis



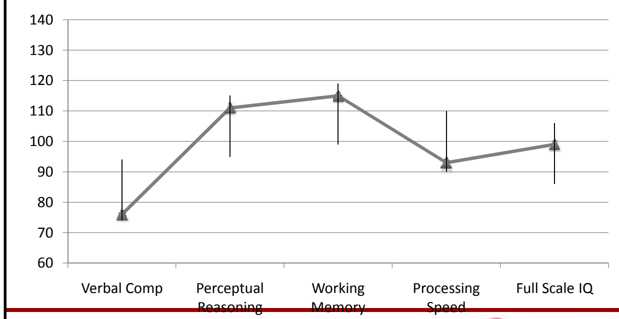
University of Minnesota
Driven to Discover
RIPS

Self-Referenced: Profile Analysis



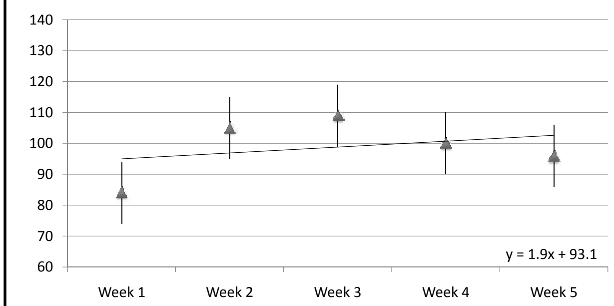
University of Minnesota
Driven to Discover
RIPS

Self-Referenced: Profile Analysis

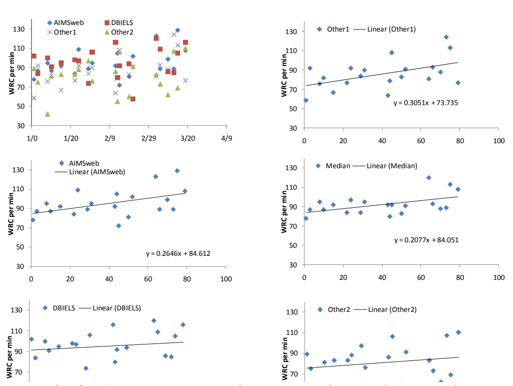


UNIVERSITY OF MINNESOTA
Driven to Discover℠
RIPS
Research Institute for Problem Solving

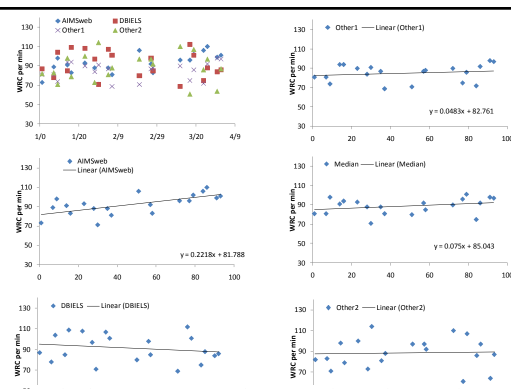
Self-Referenced: Progress Monitor



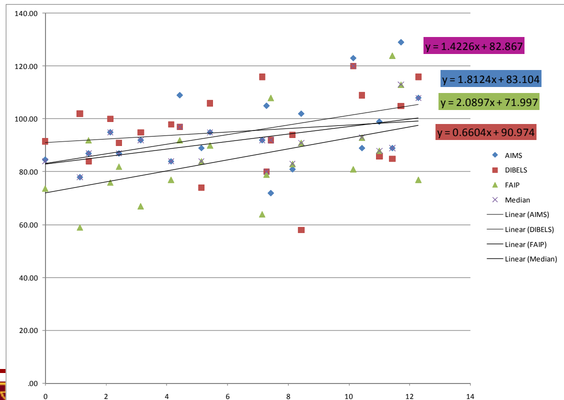
UNIVERSITY OF MINNESOTA
Driven to Discover℠
RIPS
Research Institute for Problem Solving

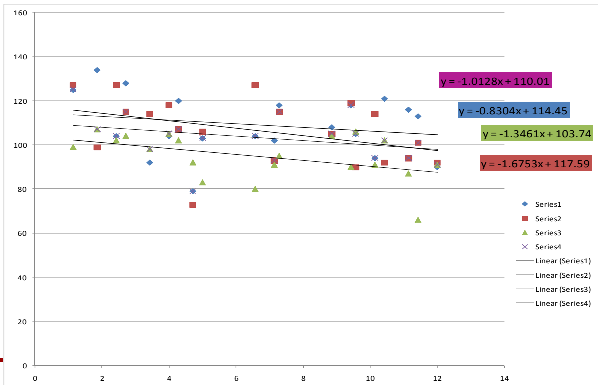


Week Calcuations	AIMSweb	DBIELS	Other1	Other2	Median
Intercept	84.61	91.52	73.74	75.61	84.05
Slope	1.85	0.67	2.14	0.90	1.45
Slope SE	3.51	4.20	3.86	4.86	2.58
SEE	12.87	15.37	14.14	17.81	9.45



Week Calcuations	AIMSweb	DBIELS	Other1	Other2	Median
Intercept	82.07	94.86	83.06	87.48	85.18
Slope	1.53	-0.54	0.30	0.13	0.51
Slope SE	1.87	3.03	2.03	3.37	1.83
SEE	8.28	13.42	9.01	14.96	8.12





Comparing Methods of Slope Estimates

Shinn, Good, and Stein (1989)

- Compared predominate methods at the time
  - Ordinary least squares (OLS) regression
  - split-middle (SM) procedures
- Although equivalent accuracy was observed, **OLS provided significant increases in precision**

Comparing Methods of Slope Estimates

Good and Shinn (1990)

- Used moving root mean squared error (RMSE) and moving mean error (ME) to evaluate the forecasting accuracy of OLS and SM slope estimation
  - **RMSE** is an estimate of precision, where small values indicate more accurate predictions
  - **ME** is an estimate of bias, or the systematic over- or under-prediction of future performance

Comparing Methods of Slope Estimates

Estimation Method	Forecast Horizon: 2 Weeks	4 Weeks	6 Weeks
Interpret Similar to Standard Error of the Estimate (SEE)			
Ordinary Least Squares			
10 Points	15.97 (6.68)	21.30 (8.72)	27.19 (12.76)
20 Points	11.42 (4.30)	13.33 (4.58)	15.71 (5.77)
30 Points	10.69 (4.12)	12.00 (4.50)	14.70 (5.68)
Split Middle			
10 Points	21.97 (10.68)	30.67 (15.43)	40.94 (20.47)
20 Points	13.86 (5.15)	17.52 (6.41)	21.01 (7.27)
30 Points	13.73 (4.91)	15.75 (5.29)	18.10 (5.79)

OLS consistently yielded superior estimates for all number of time points and forecast-horizon combinations

Good & Shinn (1990)

Knowns and Unknowns

Known

- OLS is more precise than Split Middle
- Neither does a good job with few data points

Unknown

- But how data points do we need to estimate growth and predict future performance?
- How do decision rules (3, 4, 5) compare?

Instrumentation, Conditions, and Number of Data Points

- Christ (2006)
  - Literature review conducted to obtain the range of SEE for CBM-R
    - Optimal – 2, 4, 6
    - Moderate – 8, 10, 12
    - Poor – 14, 16, 18
  - Number of data points & observation occasions
    - 2 through 15

Standard Error of the Slope Estimate (*SEb*) by Progress Monitoring Duration in Weeks

SEE[c]	Weeks of Progress Monitoring[a,b]: 2	3	4	5	6	7	8	9	10	11	12	13	14	15
Optimal[d]														
2	1.84	0.97	0.62	0.44	0.34	0.27	0.22	0.18	0.16	0.13	0.12	0.10	0.09	0.08
4	3.68	1.93	1.24	0.88	0.67	0.53	0.43	0.36	0.31	0.27	0.24	0.21	0.19	0.17
6	5.51	2.90	1.86	1.33	1.01	0.80	0.65	0.55	0.47	0.40	0.35	0.31	0.28	0.25
Moderate[d]														
8	7.35	3.87	2.48	1.77	1.34	1.06	0.87	0.73	0.62	0.54	0.47	0.42	0.37	0.34
10	9.19	4.48	3.11	2.21	1.68	1.33	1.09	0.91	0.78	0.67	0.59	0.52	0.47	0.42
12	11.03	5.80	3.73	2.65	2.01	1.59	1.30	1.09	0.93	0.81	0.71	0.63	0.56	0.51
Poor[d]														
14	12.87	6.77	4.35	3.10	2.35	1.86	1.52	1.27	1.09	0.94	0.83	0.73	0.66	0.59
16	14.71	7.74	4.97	3.54	2.68	2.13	1.74	1.46	1.24	1.08	0.94	0.84	0.75	0.68
18	16.54	8.71	5.59	3.98	3.02	2.39	1.96	1.64	1.40	1.21	1.06	0.94	0.84	0.76

[a] Estimates based on the assumption that two data points are collected weekly.
[b] Reported for weekly standard errors of the slope [$SEb = SEE/(SD_d * \sqrt{n})$].
[c] *SEE* is the average standard deviation in words read correctly per minute from predicted CBM-R performances along the line of best fit.
[d] Qualitative descriptor for measurement conditions based on author's subjective evaluation.

Smaller magnitudes of *SEb* exist under "optimal" conditions and longer progress monitoring duration

Christ (2006)

Standard Error Associated with Progress Monitoring

Ardoin and Christ (2009)

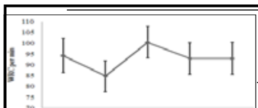
- evaluate progress monitoring outcomes for
  - **AIMSweb, DIBELS, FAIP** probe-sets
- Attending to
  - Intercept,
  - weekly estimates of growth (slope),
  - standard error of slope (SEb), &
  - standard error of estimate (SEE)

Ardoin, S. P., & Christ, T. J. (2009). Curriculum based measurement of oral reading: Standard errors associated with progress monitoring outcomes from DIBELS, AIMSweb, and an experimental passage set. *School Psychology Review, 38(2), 266-283.*

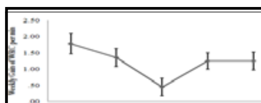
Table 1
Descriptive Statistics for Progress Monitoring Outcomes Across Three Passage Sets: Intercept, Weekly Growth, and Standard Error

Statistic	Passage Set	*M*	*SD*	Skewness	Kurtosis
Intercept					
	FAIP-R	84.81	29.92	0.01	1.00
	AIMSweb	94.29	33.14	0.06	1.21
	DIBELS	100.40	30.62	0.12	1.17
Slope					
	FAIP-R	1.36	1.15	0.48	0.71
	AIMSweb	1.79	1.35	0.52	0.62
	DIBELS	0.45	1.12	−0.13	0.09
SEb					
	FAIP-R	0.64	0.20	1.72	5.29
	AIMSweb	0.71	0.23	1.17	2.44
	DIBELS	0.91	0.29	0.91	1.03
SEE					
	FAIP-R	10.68	2.96	1.34	3.04
	AIMSweb	11.89	3.29	1.14	2.78
	DIBELS	15.26	4.27	0.86	0.84

Note. FAIP-R = Formative Assessment Instrumentation and Procedures for Reading; DIBELS = Dynamic Indicators of Basic Early Literacy Skills; SEb = standard error of slopes; SEE = standard error of the estimate. *n* (progress monitoring data set) = 68 within each probe set with a mean (*SD*) of 19.91 (0.41) number of observations per data set across a mean (*SD*) of 11.66 (0.77) weeks.

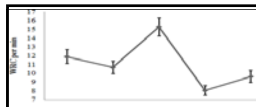


		Marginal Mean		Pairwise Comparisons[a] FAIP	DIBELS	Mean	Median
				85	*100*	*93*	*93*
Stat	Probe	Mean	Standard Error	$M_r - M_c$	$M_r - M_c$	$M_r - M_c$	$M_r - M_c$
Intercept							
	AIMS	94	4	9.5**	-6.1**	1.3*	1.3*
	FAIP	85	4		-15.6**	-8.2**	-8.2**
	DIBELS	100	4			7.4**	7.4**
	Mean	93	4				-0.1
	Median	93	4				



		Marginal Mean		Pairwise Comparisons[a] FAIP	DIBELS	Mean	Median
				1.36	*.45*	*1.25*	*1.25*
Stat	Probe	Mean	Standard Error	$M_r - M_c$	$M_r - M_c$	$M_r - M_c$	$M_r - M_c$
Slope							
	AIMS	1.79	.16	.43**	1.34**	.54**	.54**
	FAIP	1.36	.14		.91**	.11	.11
	DIBELS	.45	.14			-.80**	-.80**
	Mean	1.25	.13				.00
	Median	1.25	.14				

Stat	Probe	Mean	Standard Error	Pairwise Comparisons [a] FAIP	DIBELS	Mean	Median
		Marginal Mean		.64	.91	.50	.58
				$M_r - M_c$	$M_r - M_c$	$M_r - M_c$	$M_r - M_c$
SE(Slope)							
	AIMS	.71	.03	0.08**	-.20**	.22**	.14**
	FAIP	.64	.02		-.28**	.14**	.06**
	DIBELS	.91	.04			.42**	.34**
	Mean	.50	.02				-.08**
	Median	.58	.02				



Stat	Probe	Mean	Standard Error	Pairwise Comparisons [a] FAIP	DIBELS	Mean	Median
		Marginal Mean		10.68	15.26	8.03	9.63
				$M_r - M_c$	$M_r - M_c$	$M_r - M_c$	$M_r - M_c$
SE(estimate)							
	AIMS	11.89	.40	1.21**	-3.3**	3.86**	2.26**
	FAIP	10.68	.36		-4.6**	2.65**	1.05**
	DIBELS	15.26	.52			7.22**	5.63**
	Mean	8.03	.28				-1.6**
	Median	9.63	.34				

Where We Are Today..

Given the literature we can conclude:

- **Longer progress monitoring** durations reduce magnitude of Standard Error of the Slope Estimate (SEb)
- Smaller magnitudes of Standard Error of the Estimate (SEE) are likely to occur when we **optimize measurement conditions**
  - Standardize conditions for administration
  - Select optimal (equivalent) CBM-R passages
  - Collect multiple CBM-R on each occasion

Baseline

- Jenkins, Zumeta, Dupree, and Johnson (2005)
  - When four CBM-R passages were administered at baseline a 6-8 word standard error (SE) for the mean was found
  - The average 10 week gain was 11-13 words
  - **SE at baseline is more than half the average reading gain**
  - Conclusion: more than 3 or 4 passages may be required to confidently establish a baseline
- However, **essentially no procedures have been empirically supported for establishing a good baseline**

Big Questions…

- How do alternate data collection **schedules** and **durations** impact progress monitoring outcomes?
  - One CBMs-R weekly
  - 3 CBMs-R weekly
  - pre-post
- So, should we use decision rules instead?

50 Years of Research

- Ardoin et al. (in submission)
  - Identified research that evaluated the degree to which CBM-R decision rule(s) lead to accurate decision regarding the impact of intervention on individual students' growth.
  - Summarized published recommendations.
  - CBM-R instructional manuals and chapters related to problem solving and CBM were reviewed.

Ardoin, S. P., Morena, L. S., Christ, T. J., Cormier, D. C., & Klingbeil, D. A. (2010). *Exploring the evidence behind Curriculum Based Measurement of oral Reading (CBM-R) decision rules. Manuscript submitted for publication.*

University of Minnesota Driven to Discover℠ RIPS Research Institute for Problem Solving

Summary of Findings From a Review of 103 Documents

	Decision rule discussed[1]				Trend line discussed						Minimum # of data points[3]						
Document Type	Data points	Trend line	Both		OLS	SM	Quarter intersect	Tukey	None		3	4	6	7	8	9	10
Empirical	4	25	3		22	4	4	1	2		1	3	1	7	3	0	2
Manual	8	9	18		10	4	2	2	12		4	5	7	4	8	2	3

Conclusions:
Almost no psychometric and empirical evaluation of decision rules.
Very little on trend estimates.

University of Minnesota Driven to Discover℠ RIPS Research Institute for Problem Solving

Road Blocks to Data Collection

- True growth is unknown.
  - Jenkins, Graff, and Miglioretti (2009) used 29 CBM-R scores as a proxy for "true" growth estimate
  - With actual data, can only estimate "true" score
- Multiple factors create large numbers of interactions.
  - Datasets with very large N's are required

University of Minnesota Driven to Discover℠ RIPS Research Institute for Problem Solving

Reliability and Validity

GROWTH ESTIMATES VERSUS DATA POINT DECISION RULES

Research Question

- What is the relationship between true slopes and observed slopes derived from **one CBM-R per week** for
  – very good to very poor progress monitoring conditions
  – with very few (6) to very many weeks of data

University of Minnesota Driven to Discover
RIPS Research Institute for Problem Solving

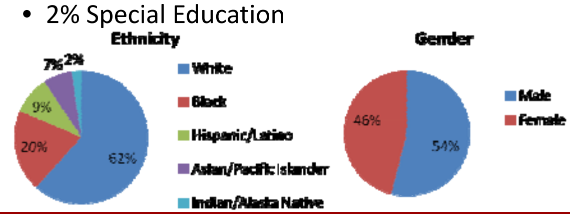
Method

- **Sample**
  – Derive model parameters from large high quality progress monitoring dataset
    - Intercept (*M, SD*)
    - Slope (*M, SD*)
    - Residual
- **Simulate**
  – Simulate true and observed slope estimates

University of Minnesota Driven to Discover
RIPS Research Institute for Problem Solving

Minnesota Reading Core – Progress Monitoring Data Set

- 2nd grade (N=1517)
- 3rd grade (N=1561)
- 2% Special Education



University of Minnesota Driven to Discover
RIPS Research Institute for Problem Solving

MN Reading Core **AIMSweb** Data & Simulation Parameters

	N	β_0 *M (SE)*	*SD*(b_{0i}) *M*	β_1 *M (SE)*	*SD*(b_{1i}) *M*	*Residual* (σ_ε) *M*	Number of Time/Data Points Per case *range*
Datasets Grade 2	1517	31	11	1.6	0.66	10	3-31
Datasets Grade 3	1561	56	15	1.4	0.61	11	3-31
Simulation	**9000[a] (300 x 30)**	**40**	**12**	**1.5**	**0.63**	**5 to 20**	**6,8,10…20**

UNIVERSITY OF MINNESOTA Driven to Discover℠ — RIPS Research Institute for Problem Solving

Method

- **Independent Variables**
  - Residual (measurement) error (akin to SEE)
    - 20 very poor
    - 15 poor (old DIBELS)
    - 10 good (AIMSweb)
    - 5 very good (FAIP, we hope)
  - Number of Observations
    - 6, 8, 10 … 20 observation occasions
    - Assume weekly progress monitoring

UNIVERSITY OF MINNESOTA Driven to Discover℠ — RIPS Research Institute for Problem Solving

Methods

- **Dependent Variables**
  - Reliability
    - Squared correlation of true and observed slope estimates (this is also validity)
    - Corrected split half reliability
  - Root Mean Square Error (RMSE)
    - RMSE for all pairs of true and observed slope estimates

UNIVERSITY OF MINNESOTA Driven to Discover℠ — RIPS Research Institute for Problem Solving

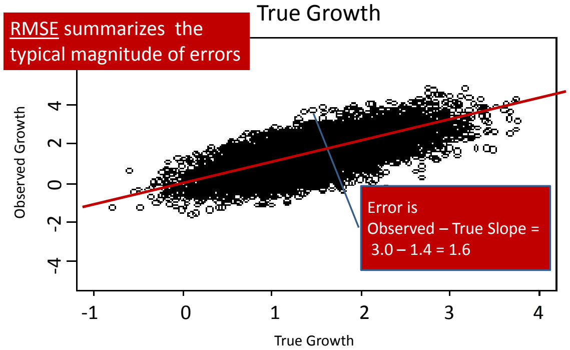
Root Mean Square Error (RMSE)

- Index of precision for observed estimates of growth
  - Smaller magnitudes RMSE correspond with more precise estimates (closer to true slopes)
- RMSE $\underset{\sim}{\ }$ Standard error of slope (SEb)
  - Can be used to construct confident intervals.

Christ, Zopluoglu, & Long (in submission)

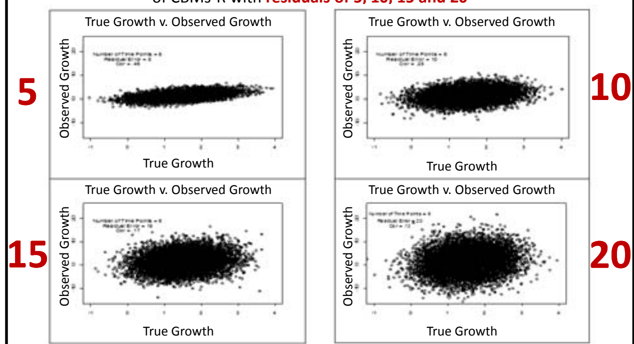
UNIVERSITY OF MINNESOTA Driven to Discover℠ — RIPS Research Institute for Problem Solving

Scatter plot of OLS Observed Estimates and True Growth



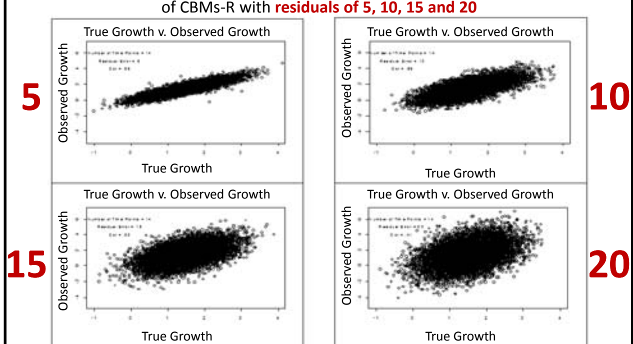
UNIVERSITY OF MINNESOTA
Driven to Discover℠
RIPS Research Institute for Problem Solving

Scatter plot of OLS Observed Estimates and True Growth for **6 Observations** of CBMs-R with **residuals of 5, 10, 15 and 20**



UNIVERSITY OF MINNESOTA
Driven to Discover℠
RIPS Research Institute for Problem Solving

Scatter plot of OLS Observed Estimates and True Growth for **14 Observations** of CBMs-R with **residuals of 5, 10, 15 and 20**



UNIVERSITY OF MINNESOTA
Driven to Discover℠
RIPS Research Institute for Problem Solving

Reliability /Validity: Squared Correlations (True & Observed)

Dataset Quality	Very Good			Good			Poor			Very Poor		
Residual (ε_{ij}) =	5			10			15			20		
Methods of Growth Estimation	OLS	*Mdn*	*M*	OLS	*Mdn*	*M*	OLS	*Mdn*	*M*	OLS	*Mdn*	*M*
6	.21	.15	.13	.06	.04	.05	.03	.02	.02	.02	.01	.01
8	.38	.33	.29	.14	.10	.12	.06	.04	.05	.04	.02	.03
10	.56	.55	.44	.25	.18	.23	.14	.09	.13	.07	.05	.06
12	.69	.69	.58	.35	.26	.32	.20	.14	.18	.12	.08	.10
14	.78	.79	.68	.48	.34	.42	.28	.18	.23	.17	.10	.14
16	.83	.84	.73	.57	.41	.49	.38	.26	.32	.23	.15	.18
18	.88	.88	.78	.65	.48	.56	.45	.30	.36	.32	.20	.25
20	.91	.92	.83	.72	.56	.63	.53	.36	.43	.40	.25	.30

UNIVERSITY OF MINNESOTA
Driven to Discover℠
RIPS Research Institute for Problem Solving

Reliability: Corrected Split Half

Dataset Quality	Very Good			Good			Poor			Very Poor		
Residual (ε_{ij}) =	5			10			15			20		
Methods of Growth Estimation	OLS	*Mdn*	*M*	OLS	*Mdn*	*M*	OLS	*Mdn*	*M*	OLS	*Mdn*	*M*
Number of Time/Data Points												
6	.28	.14†	.19†	.26	.04†	.06†	.24	.02†	.02†	.23	.01†	.01†
8	.38	.32†	.39†	.26	.11†	.14†	.24	.05†	.07†	.24	.03†	.04†
10	.53	.46†	.54†	.31	.19†	.24†	.24	.10†	.13†	.24	.06†	.07†
12	.66	.50	.60	.35	.26	.32	.29	.26	.28	.24	.24	.25
14	.76	.65	.74	.45	.36	.43	.33	.28	.30	.28	.24	.26
16	.83	.74	.81	.54	.43	.52	.36	.31	.35	.28	.26	.28
18	.88	.81	.86	.64	.52	.61	.46	.36	.43	.32	.30	.31
20	.91	.86	.90	.71	.61	.69	.52	.43	.50	.38	.31	.36

UNIVERSITY OF MINNESOTA
Driven to Discover™
RIPS Research Institute for Problem Solving

RMSE

Table 5.

Root Mean Square Error for True and Observed Estimates of Growth

Dataset Quality	Very Good			Good			Poor			Very Poor		
Residual (ε_{ij}) =	5			10			15			20		
Methods of Growth Estimation	OLS	Mdn	M	OLS	Mdn	M	OLS	Mdn	M	OLS	Mdn	M
Number of Time/Data Points												
6	1.19	1.62	1.37	2.38	3.16	2.72	3.58	4.72	4.07	4.73	6.24	5.37
8	.77	.95	.81	1.55	1.90	1.64	2.27	2.81	2.41	3.06	3.77	3.24
10	.54	.68	.58	1.09	1.33	1.16	1.62	2.00	1.70	2.20	2.70	2.34
12	.41	.52	.44	.83	1.05	.89	1.25	1.58	1.35	1.65	2.09	1.79
14	.32	.42	.36	.65	.86	.73	.98	1.29	1.11	1.32	1.72	1.47
16	.27	.37	.31	.53	.72	.62	.80	1.08	.93	1.08	1.44	1.24
18	.23	.32	.27	.45	.63	.54	.68	.95	.82	.89	1.24	1.06
20	.19	.28	.24	.38	.55	.48	.58	.83	.72	.77	1.11	.96

Note. Standard error of estimated values < .05. OLS – Ordinary Least Square, Mdn – moving median, M – moving mean

Christ, Zopluoglu, & Long (in submission)

UNIVERSITY OF MINNESOTA
Driven to Discover™
RIPS Research Institute for Problem Solving

Conclusions

- Important to **optimize conditions & passages**
- OLS estimates based on fewer than **14 data points are likely to be insufficient**
  - Reliabilities with 14 data points were.76 (very good), .46 (good), .33 (poor), .28 (very poor)
- **OLS slightly better than moving mean**
  - Cost benefit analysis how to allocate resources
  - Pre-post assessments might be superior (but we don't know that for sure; see Jenkins, 2009)

UNIVERSITY OF MINNESOTA
Driven to Discover™
RIPS Research Institute for Problem Solving

Research Question

- **Observed:** Moving mean did pretty well in previous study.
- **Question**: What is the relationship between true slopes and observed slopes derived **pre-post datasets**
  - very good to very poor progress monitoring conditions
  - with very few weeks (6 weeks) of intervention to very many weeks of intervention (20 weeks)

UNIVERSITY OF MINNESOTA
Driven to Discover™
RIPS Research Institute for Problem Solving

Method

- **Independent Variables**
  - Residual (measurement) error
    - 20 very poor
    - 15 poor (old DIBELS)
    - 10 good (AIMSweb)
    - 5 very good (FAIP, we hope)
  - Number of Observations and Weeks of Intervention
    - 3 observation on each of two occasions
    - Pre-post assessments 6 to 20 weeks apart

Reliability for Weekly & PP

Table 4.

Split-Half Reliability of Observed Estimates of Growth

Weekly Progress Monitoring

Dataset Quality	Very Good			Good			Poor			Very Poor		
Residual (ε_{ij}) =	5			10			15			20		
Methods of Growth Estimation	OLS	*Mdn*	*M*	OLS	*Mdn*	*M*	OLS	*Mdn*	*M*	OLS	*Mdn*	*M*
Number of Time/Data Points												
6	.28	.14†	.19†	.26	.04†	.06†	.24	.02†	.02†	.23	.01†	.01†
8	.38	.32†	.39†	.26	.11†	.14†	.24	.05†	.07†	.24	.03†	.04†
10	.53	.46†	.54†	.31	.19†	.24†	.24	.10†	.13†	.24	.06†	.07†
12	.66	.50	.60	.35	.26	.32	.29	.26	.28	.24	.24	.25
14	.76	.65	.74	.45	.36	.43	.33	.28	.30	.28	.24	.26
16	.83	.74	.81	.54	.43	.52	.36	.31	.35	.28	.26	.28
18	.88	.81	.86	.64	.52	.61	.46	.36	.43	.32	.30	.31
20	.91	.86	.90	.71	.61	.69	.52	.43	.50	.38	.31	.36

Note. Split half reliability estimates corrected with Spearman Brown. Standard error of estimated values < .16. OLS – Ordinary Least

Pre-Post Monitoring

Dataset Quality	Very Good		Good		Poor		Very Poor	
Residual Error (ε_{ij}) =	5		10		15		20	
Methods of Growth Estimation	PP-Mdn	PP-M	PP-Mdn	PP-M	PP-Mdn	PP-M	PP-Mdn	PP-M
Number of Time/Data Points								
6	.27	.30	.08	.09	.04	.04	.02	.02
8	.40	.43	.15	.17	.07	.08	.04	.05
10	.52	.57	.22	.24	.12	.13	.07	.08
12	.63	.67	.29	.31	.16	.18	.10	.11
14	.70	.73	.37	.40	.21	.24	.13	.14
16	.76	.79	.44	.48	.27	.29	.16	.18
18	.81	.84	.49	.53	.31	.35	.20	.22
20	.82	.84	.56	.60	.36	.40	.24	.27

Reliability for Weekly & PP

Pre-Post Monitoring

Dataset Quality	Very Good		Good		Poor		Very Poor	
Residual Error (ε_{ij}) =	5		10		15		20	
Methods of Growth Estimation	PP-Mdn	PP-M	PP-Mdn	PP-M	PP-Mdn	PP-M	PP-Mdn	PP-M
Number of Time/Data Points								
6	.27	.30	.08	.09	.04	.04	.02	.02
8	.40	.43	.15	.17	.07	.08	.04	.05
10	.52	.57	.22	.24	.12	.13	.07	.08
12	.63	.67	.29	.31	.16	.18	.10	.11
14	.70	.73	.37	.40	.21	.24	.13	.14
16	.76	.79	.44	.48	.27	.29	.16	.18
18	.81	.84	.49	.53	.31	.35	.20	.22
20	.82	.84	.56	.60	.36	.40	.24	.27

Research Question

- What is the relationship between true slopes and observed slopes derived from **three CBMs-R per week** for
  - very good to very poor progress monitoring conditions
  - with very few (6) to very many weeks (20) of data

RMSE for Weekly & PP

Table 5.

Root Mean Square Error for True and Observed Estimates of Growth

Weekly Progress Monitoring

Dataset Quality	Very Good			Good			Poor			Very Poor		
Residual (ε_{ij}) =	5			10			15			20		
Methods of Growth Estimation	OLS	Mdn	M	OLS	Mdn	M	OLS	Mdn	M	OLS	Mdn	M
Number of Time/Data Points												
6	1.19	1.62	1.37	2.38	3.16	2.72	3.58	4.72	4.07	4.73	6.24	5.37
8	.77	.95	.81	1.55	1.90	1.64	2.27	2.81	2.41	3.06	3.77	3.24
10	.54	.68	.58	1.09	1.33	1.16	1.62	2.00	1.70	2.20	2.70	2.34
12	.41	.52	.44	.83	1.05	.89	1.25	1.58	1.35	1.65	2.09	1.79
14	.32	.42	.36	.65	.86	.73	.98	1.29	1.11	1.32	1.72	1.47
16	.27	.37	.31	.53	.72	.62	.80	1.08	.93	1.08	1.44	1.24
18	.23	.32	.27	.45	.63	.54	.68	.95	.82	.89	1.24	1.06
20	.19	.28	.24	.38	.55	.48	.58	.83	.72	.77	1.11	.96

Note. Standard error of estimated values < .05. OLS – Ordinary Least Square, Mdn – moving median, M – moving mean

Pre-Post Monitoring

Dataset Quality	Very Good		Good		Poor		Very Poor	
Residual Error (ε_{ij}) =	5		10		15		20	
Methods of Growth Estimation	PP-Mdn	PP-M	PP-Mdn	PP-M	PP-Mdn	PP-M	PP-Mdn	PP-M
Number of Time/Data Points								
6	1.09	1.02	2.20	2.05	3.30	3.07	4.44	4.13
8	0.79	0.74	1.56	1.46	2.35	2.18	3.15	2.94
10	0.62	0.57	1.22	1.14	1.83	1.70	2.44	2.28
12	0.50	0.47	1.01	0.95	1.50	1.41	1.99	1.86
14	0.42	0.39	0.85	0.79	1.28	1.18	1.71	1.59
16	0.37	0.34	0.73	0.68	1.08	1.01	1.46	1.37
18	0.32	0.30	0.64	0.60	0.97	0.90	1.31	1.22
20	0.29	0.27	0.58	0.54	0.87	0.81	1.16	1.08

Split Half Reliability
Summery of 3 Studies

Dataset Quality	Very Good			Good		
Residual (ε_{ij}) =	5			10		
	1 CBM	3 CBM	3 CBM	1 CBM	3 CBM	3 CBM
Methods of Growth Estimation	OLS	OLS	PP-*M*	OLS	OLS	PP-*M*
Number of Time/Data Points						
6	0.28	0.33	0.30	0.26	0.26	0.09
8	0.38	0.53	0.43	0.26	0.29	0.17
10	0.53	0.68	0.57	0.31	0.38	0.24
12	0.66	0.80	0.67	0.35	0.50	0.31
14	0.76	0.87	0.73	0.45	0.60	0.40
16	0.83	0.91	0.79	0.54	0.70	0.48
18	0.88	0.93	0.84	0.64	0.77	0.53
20	0.91	0.95	0.84	0.71	0.83	0.60

University of Minnesota
Driven to Discover™
RIPS Research Institute for Problem Solving

RMSE:
Summary of 3 Studies

Dataset Quality	Very Good			Good		
Residual (ε_{ij}) =	5			10		
	1 CBM	3 CBM	3 CBM	1 CBM	3 CBM	3 CBM
Methods of Growth Estimation	OLS	OLS	PP-*M*	OLS	OLS	PP-*M*
Number of Time/Data Points						
6	1.2	0.9	1.0	2.4	1.7	2.1
8	0.8	0.6	0.7	1.6	1.1	1.5
10	0.5	0.4	0.6	1.1	0.8	1.1
12	0.4	0.3	0.5	0.8	0.6	1.0
14	0.3	0.2	0.4	0.7	0.5	0.8
16	0.3	0.2	0.3	0.5	0.4	0.7
18	0.2	0.2	0.3	0.5	0.3	0.6
20	0.2	0.1	0.3	0.4	0.3	0.5

University of Minnesota
Driven to Discover™
RIPS Research Institute for Problem Solving

So what is next?

- It is difficult to establish a reliable and precise estimates of slope.
- Lots of data are necessary
- Do decision rules function better?

University of Minnesota
Driven to Discover™
RIPS Research Institute for Problem Solving

Challenges

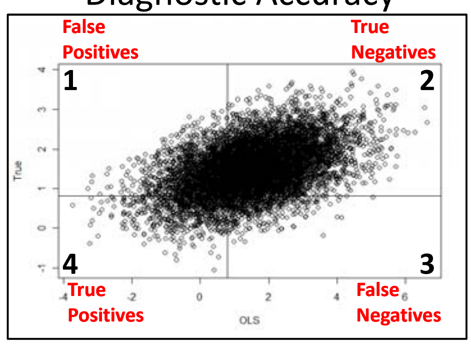
- We need an alternate approach to analysis.
  - No model fit estimates, like SEb or SEE, for decision rules
  - This delayed – and maybe precluded – such examination of decision rules

DIAGNOSTIC ACCURACY & ROC ANALYSIS

Diagnostic Accuracy

- Estimation of the proportion accurate (true) and inaccurate (false) classifications
- Four possible outcomes:
  - True Positive (TP)
  - True Negative (TN)
  - False Positive (FP)
  - False Negative (FN)

Diagnostic Accuracy

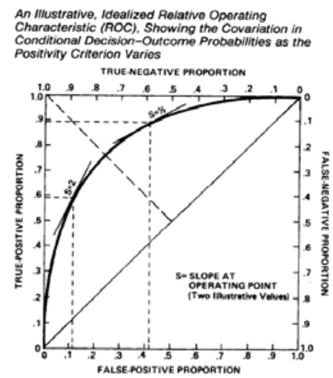


Classifications Example

- Melanoma testing
  - TP – positive test in which the individual truly ***has*** the ailment
  - TN – negative test in which the individual is truly ***free*** of the ailment
  - FP – positive test in which the individual is truly ***free*** of the ailment
  - FN – negative test in which the individual truly ***has*** the ailment

The ROC Curve

- Graphical depiction of the proportions of TP, TN, FP and FN

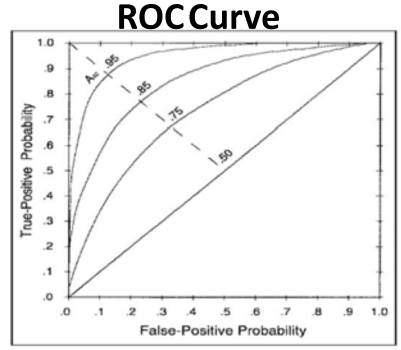


Swets (1992)

ROC Terms

- Sensitivity (True Positive Rate)
- Specificity (True Negative Rate)
- Positive Predictive Power
- Negative Predictive Power
- Kappa
- Area Under the Curve (AUC or A)

ROC Curve



If area under the curve is .5 or less, it means we do not predict more than chance

Area Under the Curve (AUC)

- Values always between 0.00 and 1.00
  - Need to be greater than 0.50
- Value classifications:
  - Excellent: 0.90 to 1.00 → **High – stakes Decisions**
  - Good: 0.80 to 0.89 → **Moderate – stakes Decisions**
  - Fair: 0.70 to 0.79 → **Low – stakes Decisions**
  - Poor: 0.60 to 0.69

University of Minnesota — Driven to Discover
RIPS Research Institute for Problem Solving

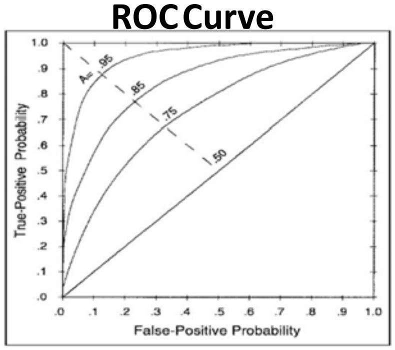
Diagnostic Accuracy Standards
National Center for RTI (2010)

- Used to evaluate benchmark and screening measures
  - *Convincing evidence* – **AUC of 0.85**
  - *Somewhat convincing evidence* – **AUC of 0.75**

  - Must use sufficient methods to establish decision thresholds in order to effectively identify students at-risk (e.g., 20th percentile)

University of Minnesota — Driven to Discover
RIPS Research Institute for Problem Solving

ROC Curve

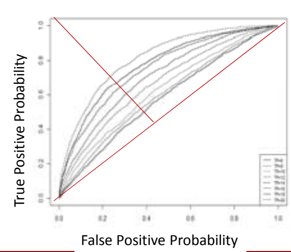


If area under the curve is .5 or less, it means we do not predict more than chance

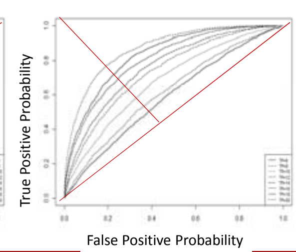
University of Minnesota — Driven to Discover
RIPS Research Institute for Problem Solving

Receiver Operator Curves (ROC)

Very Poor (Residuals = 20)

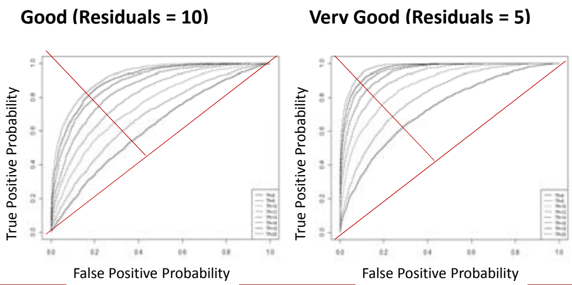


Poor (Residuals = 15)



University of Minnesota — Driven to Discover
RIPS Research Institute for Problem Solving

Receiver Operator Curves (ROC)



Area Under the Curve (AUC)

Dataset Quality / Residual (ε_{ij}) =	Very Good 5			Good 10			Poor 15			Very Poor 20		
Methods of Growth Estimation	3 Pnt	5 Pnt	OLS	3 Pnt	5 Pnt	OLS	3 Pnt	5 Pnt	OLS	3 Pnt	5 Pnt	OLS
Number of Time/Data Points												
6	.61	.57	.75	.55	.53	.63	.53	.52	.59	.53	.52	.58
8	.63	.59	.81	.57	.55	.70	.54	.53	.63	.54	.53	.61
10	.67	.62	.88	.60	.57	.76	.57	.54	.68	.55	.54	.63
12	.69	.66	.92	.61	.60	.81	.57	.56	.73	.56	.55	.67
14	.71	.68	.95	.62	.60	.85	.58	.57	.77	.57	.56	.72
16	.73	.71	.96	.65	.62	.89	.61	.60	.82	.56	.55	.76
18	.74	.74	.97	.65	.65	.91	.61	.61	.84	.58	.58	.78
20	.76	.77	.98	.67	.66	.93	.63	.62	.88	.61	.60	.82

University of Minnesota
Driven to Discover™
RIPS Research Institute for Problem Solving

Conclusion

- Diagnostic accuracy of decision rules are worse than OLS
- Collect 3 CBMs-R per occasion for 14 Weeks
  - or collect a minimum of 14 CBMs-R for an abbreviated duration
  - Use OLS and plot all data point (don't take the mean or median if multiple are administered within a day)
- Yet to evaluate Jenkins (2009) recommendations for intermittent progress monitoring

University of Minnesota
Driven to Discover™
RIPS Research Institute for Problem Solving

Challenges

- No other research like this for progress monitoring practices
- Difficult to convey information
  - Very technical
  - Finding the perfect balance
- Implications
  - May change "Best Practices" for progress monitoring

University of Minnesota
Driven to Discover™
RIPS Research Institute for Problem Solving

Where We Are Going…

- More research (Simulation and Actual Data) is necessary
- What proportion of the time are we making judgments that are wrong?
  - Inform instruction/intervention
- Optimal progress monitoring practices

OTHER PROJECTS

University of Minnesota
Driven to Discover
RIPS

CBAS

- Computer-Based Assessment System
- Online database
  - Reading
    - Over 600 items with parameters
    - Computer Adaptive
  - Math
    - Developing Items
    - Begin data collection this Spring

http://www.cehd.umn.edu/EdPsych/C-BAS-R

University of Minnesota
Driven to Discover
RIPS

FAIP

- Formative Assessment Instrumentation and Procedures
- Optimal set of CBM-R progress monitoring materials
- Online system (recent data collection)
  - Audio recording
  - Error coding
  - Inter-rater reliability

FAIP

http://www.cehd.umn.edu/EdPsych/FAIP-R/

University of Minnesota
Driven to Discover
RIPS

References

Ardoin, S. P., & Christ, T. J. (2009). Curriculum based measurement of oral reading: Standard errors associated with progress monitoring outcomes from DIBELS, AIMSweb, and an experimental passage set. School Psychology Review, 38(2), 266-283.

Christ, T. J. (2006). Short term estimates of growth using curriculum-based measurement of oral reading fluency: Estimates of standard error of the slope to construct confidence intervals. School Psychology Review, 35(1), 128-133.

Christ, T. J., Zopluoglu, C., & Long, J. (in submission). Curriculum Based Measurement of oral Reading (CBM-R): Evaluation of progress monitoring outcomes and trend line estimates. University of Minnesota, Minneapolis, MN.

Colon, E. P., & Kranzler, J. H. (2006). Effect of instructions on curriculum-based measurement of reading. *Journal of Psychoeducational Assessment, 24*(4), 318-328.

Derr-Minneci, T. F. (1990). A behavioral evaluation of curriculum-based assessment for reading: Tester, setting, and task demand effects on high- vs. average- vs. low-level readers. Dissertation Abstracts International, 51(5-B), 2669.

Derr-Minneci, T. F., & Shapiro, E. S. (1992). Validating curriculum-based measurement in reading from a behavioral perspective. *School Psychology Quarterly, 7*, 2-16.

Good, R. H., & Shinn, M. R. (1990). Forecasting accuracy of slope estimates for reading curriculum-based measurement: Empirical evidence. Behavioral Assessment, 12(2), 179-193.

Jenkins, J. R., Graff, J. J., & Miglioretti, D. L. (2009). Estimating reading growth using intermittent CBM progress monitoring. *Exceptional Children, 75(2), 151-163.*

National Center for Response to Intervention (2010). http://www.rti4success.org/

Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2005). An investigation of the reliability and standard error of measurement of words read correctly per minute using curriculum-based measurement. *Journal of Psychoeducational Assessment, 23*(4), 326-338.

Shapiro, E. S. (2004). *Academic skills problems: Direct assessment and intervention* (3rd ed.). New York, NY: The Guildford Press.

Shinn, M. R., Good, R. H., & Stein, S. (1989). Summarizing trend in student achievement: A comparison of methods. *School Psychology Review, 18*(3), 356-370.

Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist, 47*(4), 522-532.

UNIVERSITY OF MINNESOTA
Driven to Discover℠

RIPS
Research Institute for Problem Solving