A Comparison of Growth via Curriculum-Based Measurement and Computer-Adaptive Tests

NASP 2023 Annual Conference
Denver, CO

Emily R. Forcht, M.Ed. & Ethan R. Van Norman, Ph.D.

Lehigh University

Conflict of Interest Statement

- Data were collected and maintained by Renaissance Learning (RL), the developer of the assessments in this study.
- Ms. Forcht completed this research as part of a sponsored research agreement in the role of a graduate research assistantship between RL and Lehigh University
- Dr. Van Norman received no financial compensation for this project and has not previously received financial compensation from RL
- RL reviewed but did not edit or author the results of this study, per the conditions of the sponsored research agreement
- The office of sponsored research projects and the IRB at Lehigh University continually review this relationship for ethical compliance and potential conflicts of interest

Learner Objectives

- Evaluate the extent to which CBM and CAT capture growth in unique skills in reading
- Determine which progress monitoring tool, CBM or CAT, is more appropriate for different scenarios
- Apply more informed decisions during the progress monitoring process in their practice

NASP Domains

Practice Model Domains

- Domain 1: Data-Based Decision Making and Accountability
- Domain 9: Research and Evidence-Based Practice

LEHIGH UNIVERSITY

Background

Progress Monitoring

- In multi-tiered system of support (MTSS), students are identified as at risk for academic difficulties and receive supplemental intervention
- Educators collect performance data to monitor effects of supplemental intervention → progress monitoring
- Curriculum-based measurement (CBM): most common PM tool (Deno, 1985)
- Computer-adaptive tests (CATs) has emerged another option for monitoring

CBM as Progress Monitoring Tool

Advantages

- Easy to administer, easy to score, and assess a variety of grade-level skills (Deno, 2003)
  - CBM of oral reading (CBM-R) = Passage Oral Reading (Renaissance Learning, 2021)
- Probes are short (1-3 minutes)
- Predictive of broader academic skills (Shinn, 2007)

Disadvantages

- Can be an unreliable indicator of true growth in oral reading fluency
- High levels of residual, or error, have been associated → translates to less reliable estimates of growth (Christ, 2006)

LEHIGH UNIVERSITY

Computer Adaptive Tests

- Examinees receive a unique version of the test during each administration (Meijer & Nering, 1999)
- CATs can accurately estimate ability with fewer items than traditional fixed item tests (Wang & Shin, 2010)
- Used in clinical and educational settings
  - Several CATs to measure reading skills
- Star Reading (SR), reading CAT developed by Renaissance Learning, is the focus of the current study

LEHIGH UNIVERSITY

Comparison of CATs and CBM

CATs

- Questions selected in real-time based upon responses
- Item Response Theory (IRT; Carlson, 1994)
- Item-level information
- Broad number of literacy skills

CBM

- Fixed item forms
- Classical Test Theory (CTT)
- Test-level information
- Oral reading fluency skills

LEHIGH UNIVERSITY

Purpose and Research Questions

To determine whether CBM and CAT yield distinct growth trajectories in reading skills across a school year.

Research Questions:

1. To what degree does growth, on average, measured concurrently via CBM-R and SR differ, across a school year?
2. To what degree do the assessments differ in their capacity to capture meaningful variability in growth between students?
3. To what degree does the magnitude of residual variance, or error, differ between assessments?

LEHIGH UNIVERSITY

Method & Analysis

LEHIGH UNIVERSITY

Method

Participants

- Total of 3,192 students
  - Grade 1 (n = 298), Grade 2 (n = 1149), Grade 3 (n = 1,062), Grade 4 (n = 462), and Grade 5 (n = 221)
- 398 schools in 41 states
- Student-level demographic info largely un-reported

Measures

- Passage Oral Reading (CBM-R; Renaissance Learning, 2021)
- Star Reading (SR; Renaissance Learning 2022)

LEHIGH UNIVERSITY

Data Analysis

- Fit Separate Multilevel Models
  - Fixed and Random Effects for CBM-R
  - Fixed and Random Effects for Star Reading
- Problematic
  - Vastly Different Scales
  - Between Measure Outcomes Correlated
  - Straightforward Significance Tests?
- Multivariate Multilevel Growth Modeling
  - Standardize Outcomes
  - Simultaneously Model Fixed and Random Effects for Both Outcomes
  - Explicitly Model and Evaluate Dependencies
- Bayesian Framework
  - Compare Magnitude and Direction of Differences without NHST
  - Leverage Prior Information to Increase Computational Efficiency

Typical Regression

```
Call:
lm(formula = sales ~ ., data = adv_training)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.6331 -0.8971  0.2283  1.1971  2.9630 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 2.707724   0.354788   7.632 2.05e-12 ***
TV          0.046521   0.001561  29.801  < 2e-16 ***
radio       0.188857   0.009423  20.041  < 2e-16 ***
newspaper   0.001619   0.006255   0.259    0.796    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 1.677 on 158 degrees of freedor
Multiple R-squared:  0.8967,	Adjusted R-squared:  0.
F-statistic: 457.1 on 3 and 158 DF,  p-value: < 2.2e-16
```

- Parameter Estimates
- Standard Error
- *p* value

- Wald Test

Typical Regression

```
Call:
lm(formula = sales ~ ., data = adv_training)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.6331 -0.8971  0.2283  1.1971  2.9630 

Coefficients:
            Estimate td. Error t value Pr(>|t|)    
(Intercept) 2.707724  0.354788   7.632 2.05e-12 ***
TV          0.046521  0.001561  29.801  < 2e-16 ***
radio       0.188857  0.009423  20.041  < 2e-16 ***
newspaper   0.001619  0.006255   0.259    0.796    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 1.677 on 158 degrees of freedor
Multiple R-squared:  0.8967,	Adjusted R-squared:  0.8
F-statistic: 457.1 on 3 and 158 DF,  p-value: < 2.2e-16
```

- Parameter Estimates
- Standard Error
- *p* value

- Wald Test

Bayesian Regression

$$P(\beta|y, X) = \frac{P(y|\beta, X) * P(\beta|X)}{P(y|X)}$$

Bayesian Regression

- Probability of parameter β

$$P(\beta|y,X) = \frac{P(y|\beta,X) * P(\beta|X)}{P(y|X)}$$

Bayesian Regression

- Probability of parameter β
- Given the observed data

$$P(\beta|y, X) = \frac{P(y|\beta, X) * P(\beta|X)}{P(y|X)}$$

Bayesian Regression

- Probability of parameter β
- Given the observed data

$$P(\beta|y,X) = \frac{P(y|\beta,X) * P(\beta|X)}{P(y|X)}$$

- Likelihood

Bayesian Regression

- Probability of parameter β
- Given the observed data

$$P(\beta|y,X) = \frac{P(y|\beta,X) * P(\beta|X)}{P(y|X)}$$

- Likelihood x Prior Assumptions

Bayesian Regression

$$P(\beta|y,X) = \frac{P(y|\beta,X) * P(\beta|X)}{P(y|X)}$$

- Probability of parameter β
- Given the observed data
- Likelihood x Prior Assumptions
- Divided by a regularizing constant

Bayesian Regression

$$P(\beta|y, X) = \frac{P(y|\beta, X) * P(\beta|X)}{P(y|X)}$$

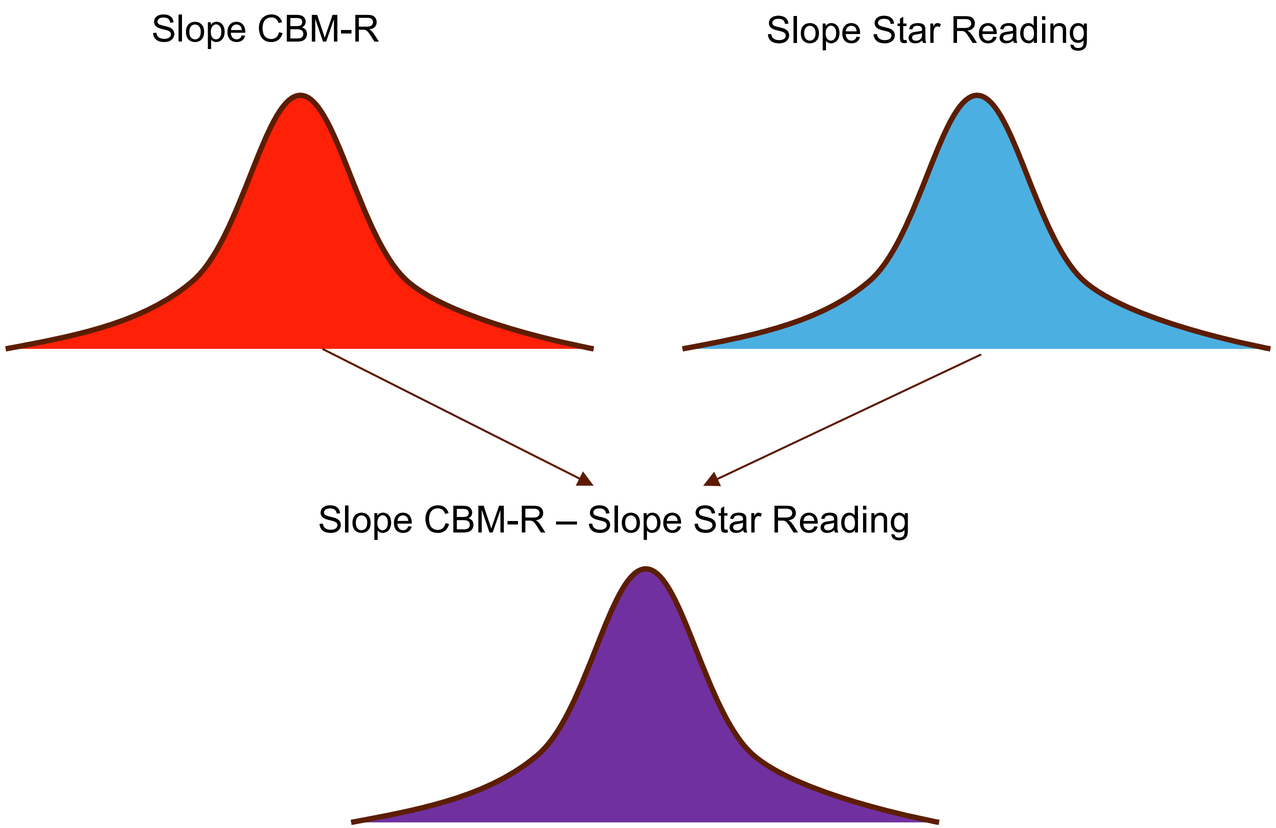
- Not a single value
- Posterior Distribution of *possible* Parameter Values

Bayesian Regression

$$P(\beta|y,X) = \frac{P(y|\beta,X) * P(\beta|X)}{P(y|X)}$$

Use Simulation to Sample from the Posterior

Create *Posterior Distributions* of each Parameter in the Model



Does the Distribution overlap with 0?

Make *probabilistic* statements about differences

LEHIGH UNIVERSITY

Data Analysis

- RQ#1: Differences between fixed effects for growth from each measure for each grade
- RQ #2: Differences in random effects for slope terms from each measure for each grade
- RQ #3: Differences residual variance between each measure for each grade
- Did any part of the 95% CI of the Posterior Distributions of Differences overlap with 0?
- Was the mean of the distribution + or -?

LEHIGH UNIVERSITY

Results

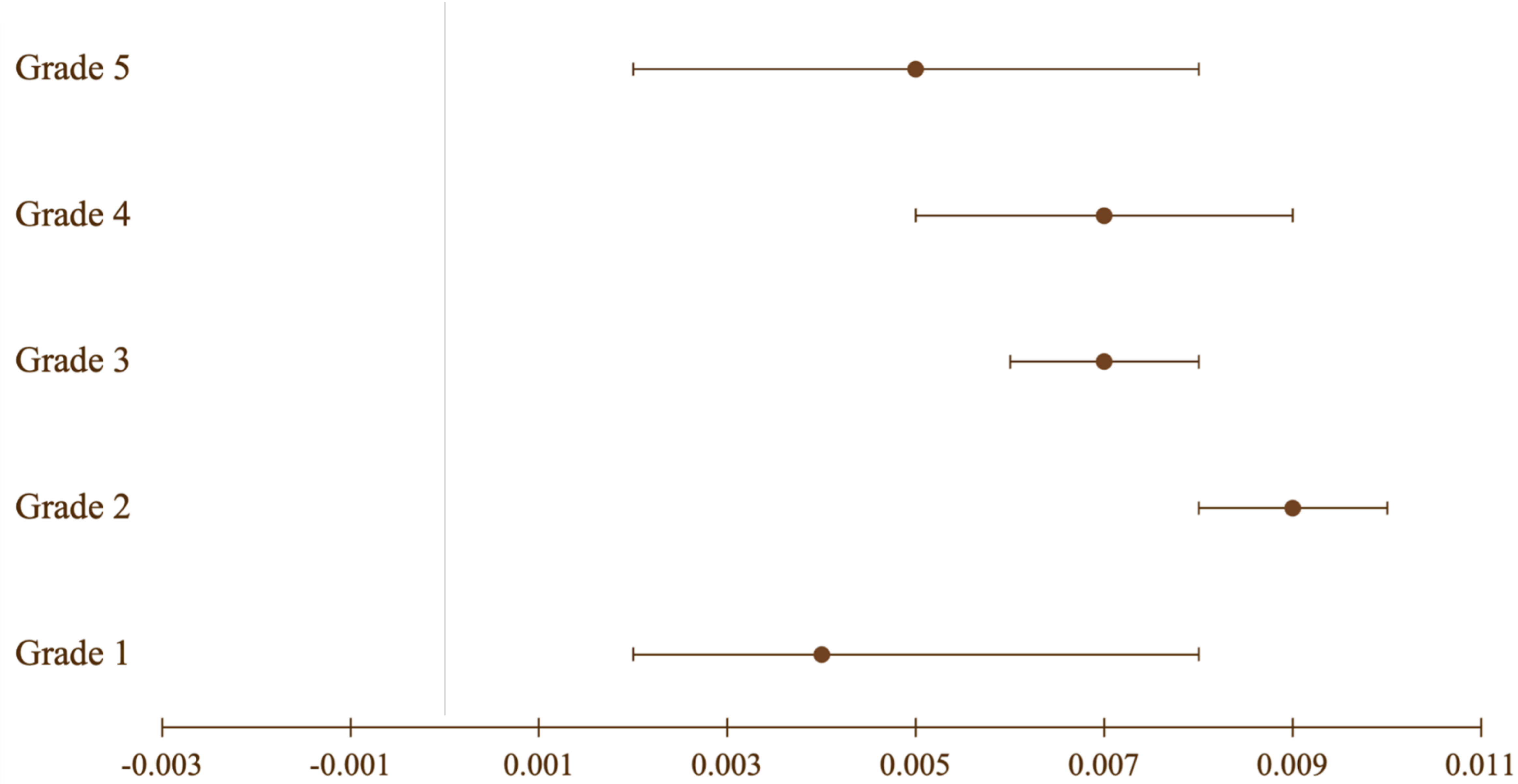
Univariate Hierarchical Growth Models

CBM	Grade One	Grade Two	Grade Three	Grade Four	Grade Five
Fixed	B (*SE*)	B (*SE*)	B (*SE*)	B (*SE*)	B (*SE*)
Intercept	15.41 (1.46)	41.12 (0.95)	71.70 (1.07)	82.58 (1.48)	96.17 (2.17)
Slope	1.19 (0.04)	1.10 (0.02)	0.84 (0.02)	0.72 (0.03)	0.58 (0.05)
Random	*SD*	*SD*	*SD*	*SD*	*SD*
Intercept	22.68	30.40	33.07	28.31	28.73
Slope	0.59	0.49	0.41	0.36	0.29
Residual	9.90	11.34	13.09	17.31	16.63
Correlation	*r*	*r*	*r*	*r*	*r*
Intercept, Slope	.28	.01	-.05	.04	-.20

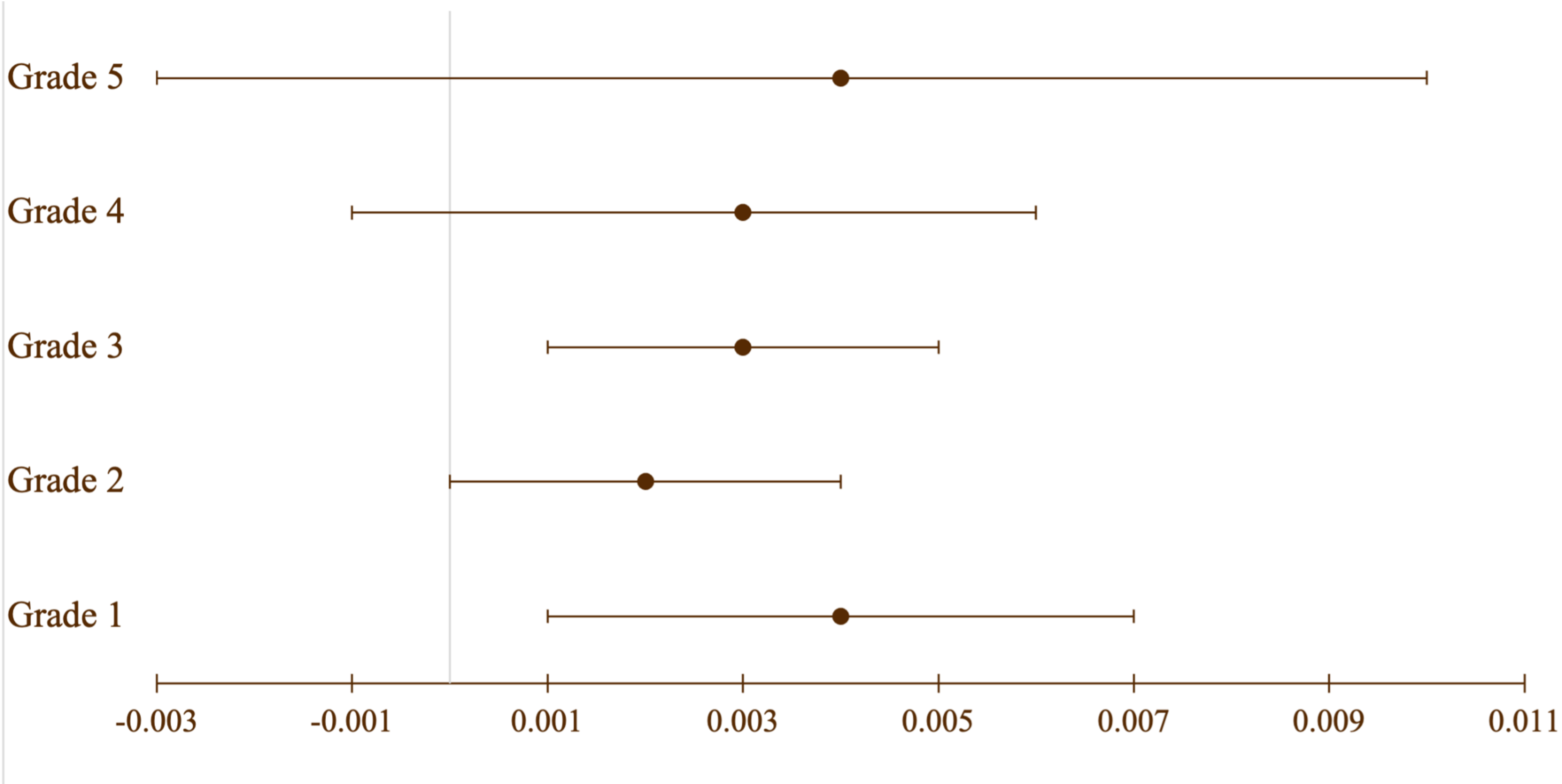
SR	Grade One	Grade Two	Grade Three	Grade Four	Grade Five
Fixed	B (*SE*)	B (*SE*)	B (*SE*)	B (*SE*)	B (*SE*)
Intercept	719.13 (5.32)	820.58 (2.58)	895.01 (2.45)	931.82 (3.33)	970.90 (4.98)
Slope	3.38 (0.16)	2.03 (0.06)	1.34 (0.05)	0.96 (0.07)	0.99 (0.09)
Random	*SD*	*SD*	*SD*	*SD*	*SD*
Intercept	70.32	78.11	73.08	63.86	66.94
Slope	1.38	1.11	0.70	0.60	0.33
Residual	57.07	42.59	37.17	35.98	35.84
Correlation	*r*	*r*	*r*	*r*	*r*
Intercept, Slope	-.09	-.30	-.34	-.29	-.61

LEHIGH UNIVERSITY

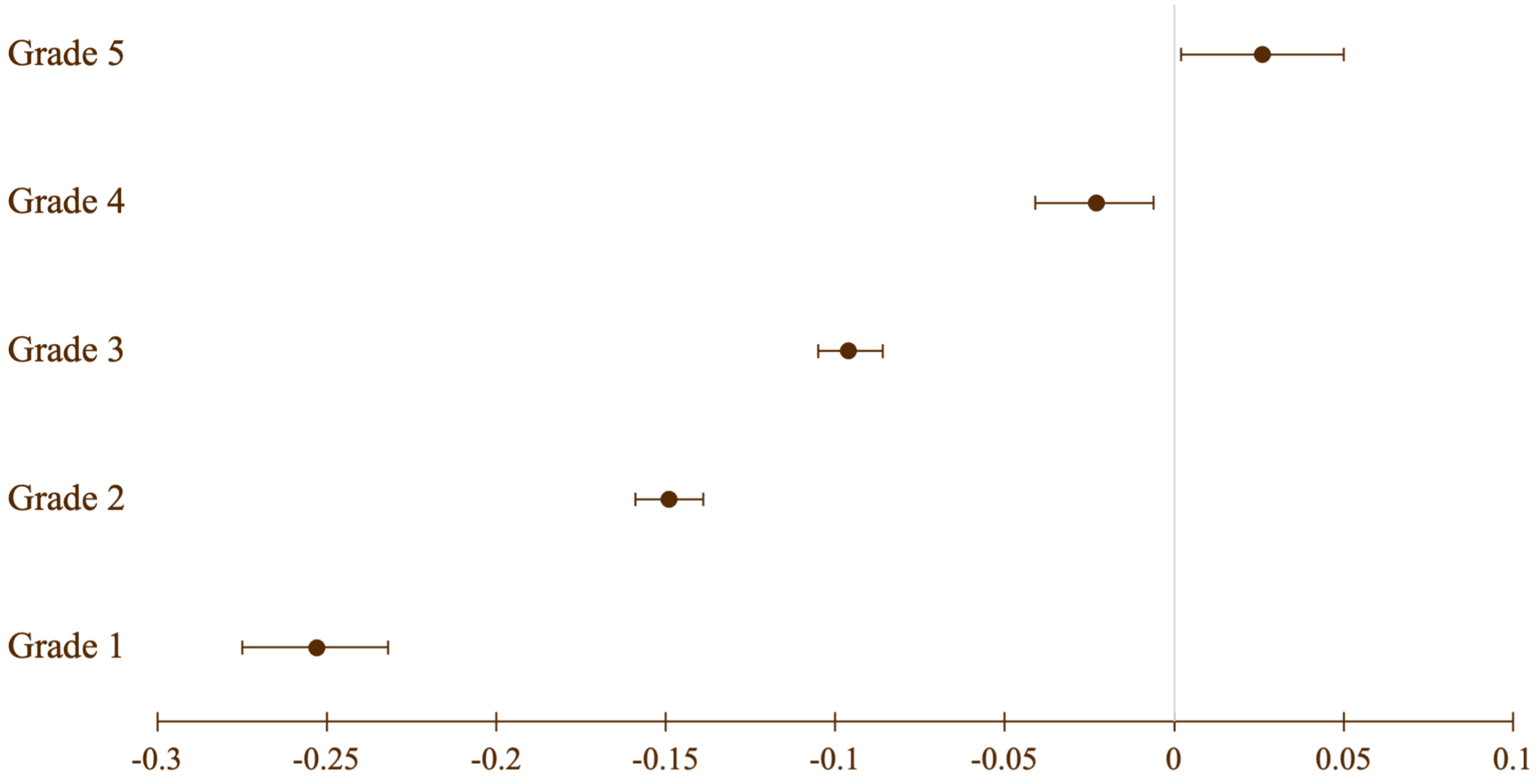
RQ #1: Average Growth



RQ #2: Between-Student Variability



RQ #3: Residual



Discussion

Discussion

The purpose of this presentation is to determine whether CBM and CAT yield distinct growth trajectories in reading skills across a school year.

Research Questions:

1. To what degree does growth, on average, measured concurrently via CBM-R and SR differ, across a school year?
2. To what degree do the assessments differ in their capacity to capture meaningful variability in growth between students?
3. To what degree does the magnitude of residual variance, or error, differ between assessments?

LEHIGH UNIVERSITY

Discussion

- RQ#1: The average rate of growth observed via CBM-R and CAT across grade levels was highly similar.
- RQ#2: The magnitude of between-student variability in growth was also highly similar.
- RQ#3: The most noticeable differences in progress monitoring outcomes occurred when comparing the magnitude of error, or residual variance.
  - Grades 1-4: SR > CBM-R
  - Grade 5: CBM-R > SR
  - Why?

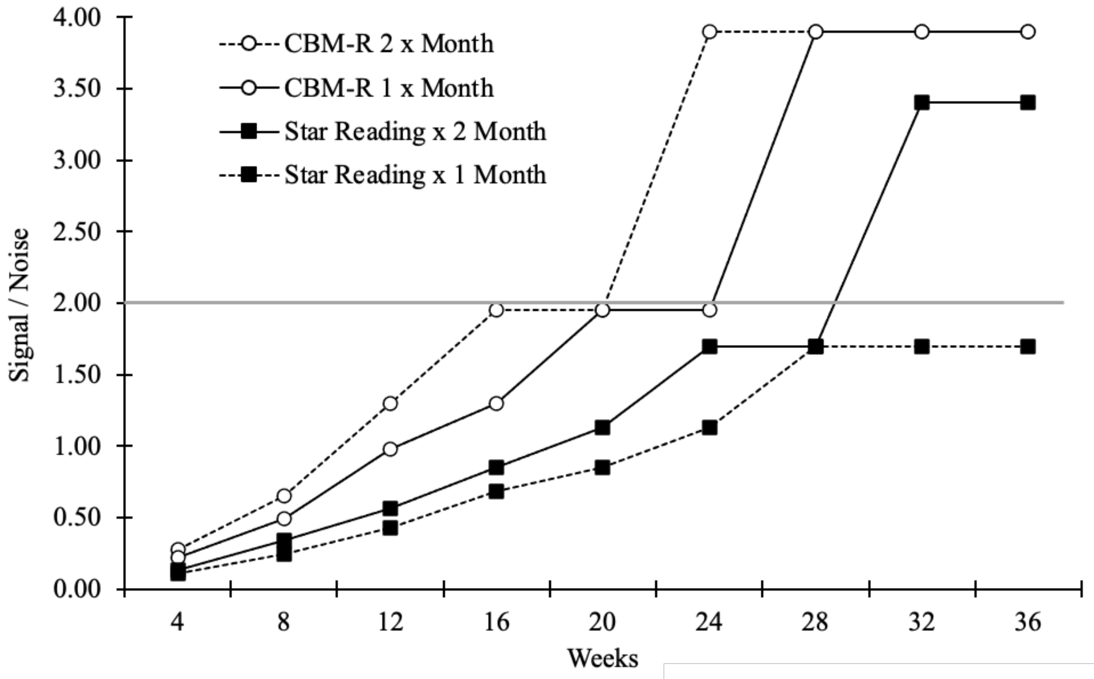
LEHIGH UNIVERSITY

Signal-to-Noise Ratio

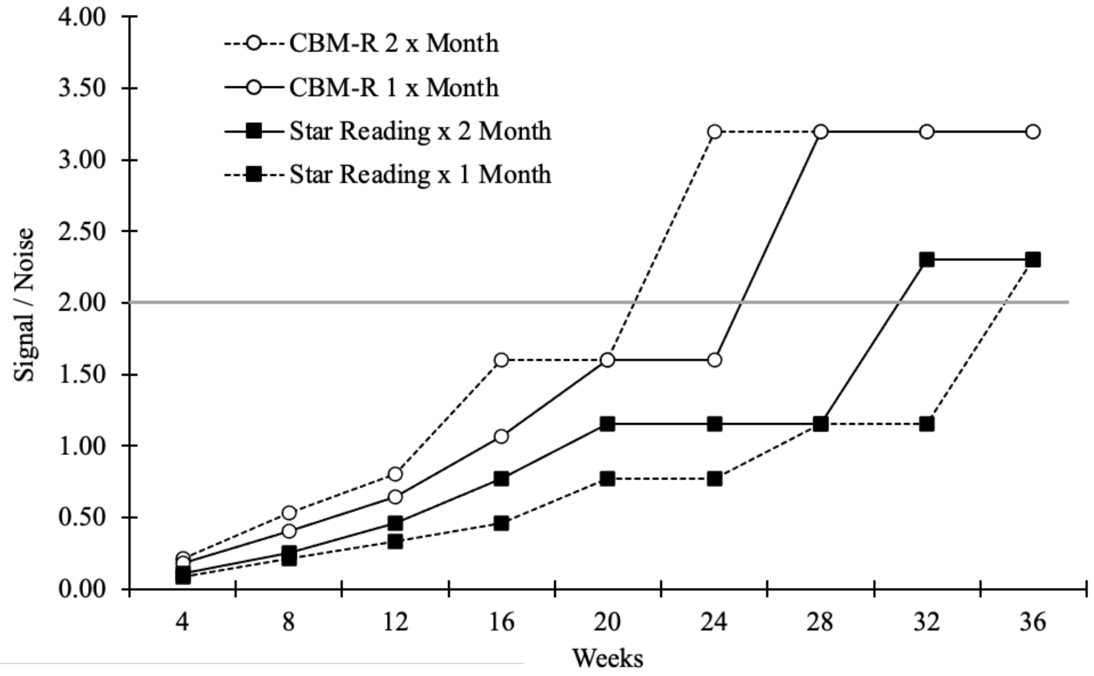
- One way to contextualize typical growth on assessments
- Signal = average rate of improvement
- Noise = amount of residual variance
- Recommendation = 2:1
  - Growth at least twice as large as error (Christ et al., 2013b)

LEHIGH UNIVERSITY

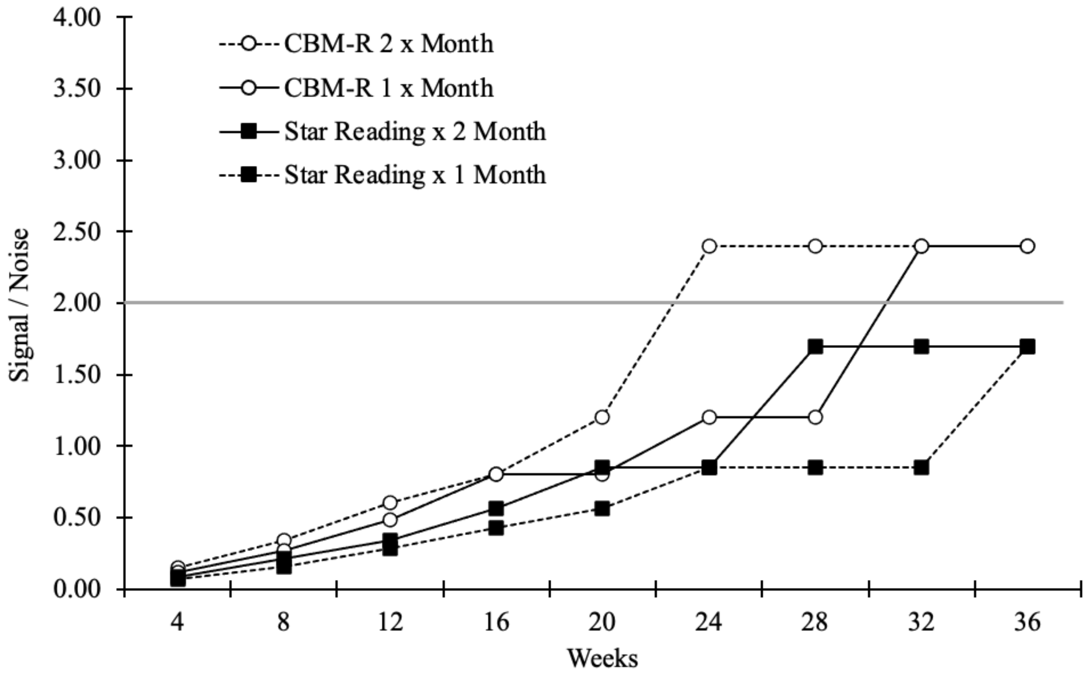
Grade 1



Grade 2

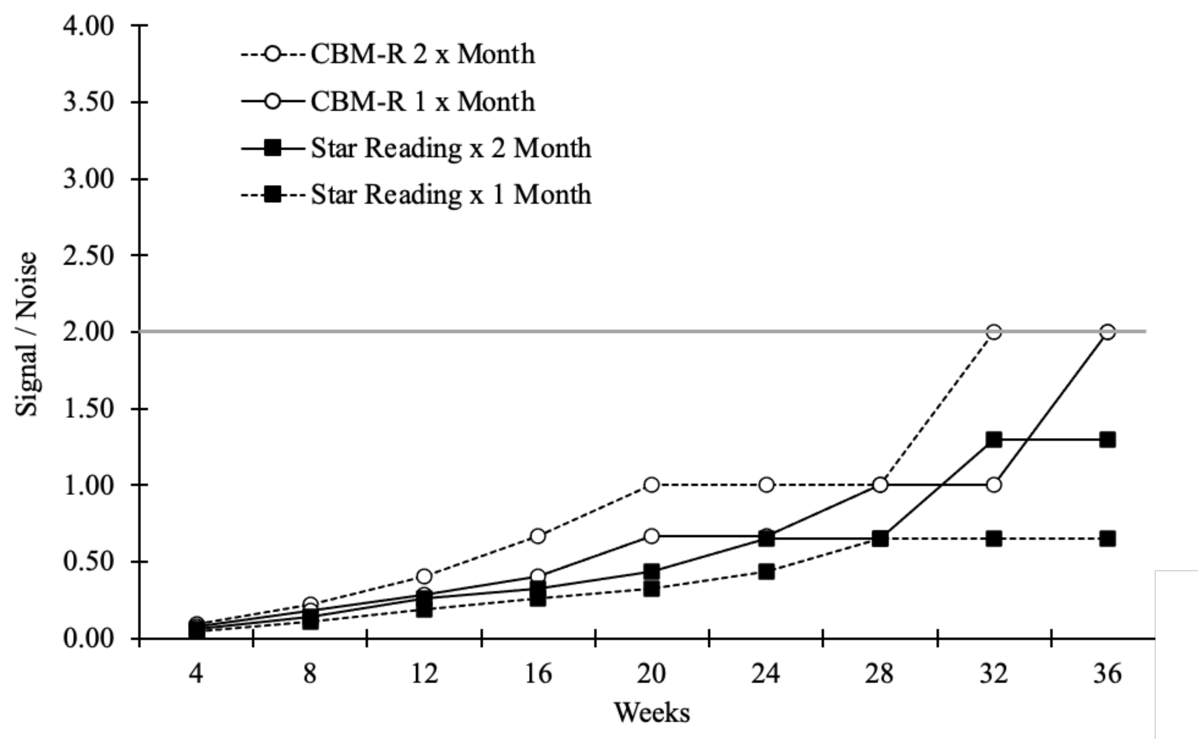


Grade 3

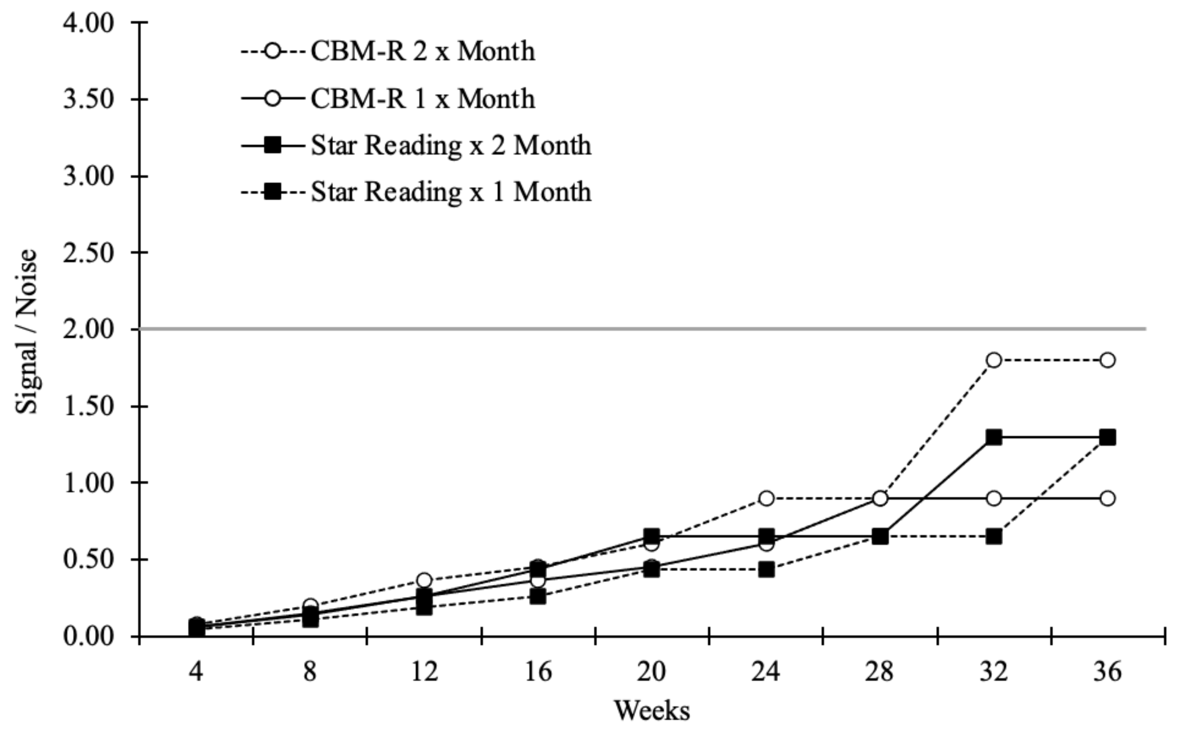


LEHIGH UNIVERSITY

Grade 4



Grade 5



Discussion – Why?

CBM-R

- Measures oral reading rate
- Single skill → less unwanted bounce or error associated between time points
- Suitable for monitoring progress towards yearlong goals

SR

- Measures any number of broader, more complex skills
- Each test is individualized to each test-taker → unwanted bounce or error may be likely
- Suitable for monitoring progress across multiple years

LEHIGH UNIVERSITY

Implications – CBM or CAT?

CBM-R

- Evaluate instructional effects within a single-school year
- Monitoring younger students, grades 1-3
  - Or older students with consistent reading difficulties
- Instruction/intervention = early reading skills and building fluent reading

SR

- Assess general achievement over several years
  - Compare student performance across grade levels
- Monitoring older students, grades 4-5
- Instruction/intervention = building comprehension

LEHIGH UNIVERSITY

Limitations & Future Directions

- Only one type of CBM-R and reading CAT were evaluated
- Data collected infrequently across an entire school year
- Limited demographic information available at student-level
- We assumed monotonic linear growth across the school year
  - Future research → quadratic growth
- No access to if students received any supplemental support

References

Carlson, R. D. (1994). Computer adaptive testing: A shift in the evaluation paradigm. *Journal of Educational Technology Systems, 22*(3), 213-224. https://doi.org/10.2190/QP36-WV9L-ATT3-FYWK

Christ, T. J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review, 35*(1), 128-133. https://doi.org/10.1080/02796015.2006.12088006

Christ, T. J., Zopluoglu, C., Monaghen, B. D., & Van Norman, E. R. (2013b). Curriculum-based measurement of oral reading: Multi-study evaluation of schedule, duration, and dataset quality on progress monitoring outcomes. *Journal of School Psychology, 51*(1), 19-57. https://doi.org/10.1016/j.jsp.2012.11.001

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219-232. https://doi.org/10.1177/001440298505200303

Deno, S. L. (2003). Curriculum-based measures: Development and perspectives. *Assessment for Effective Intervention, 28*(3-4), 3-12. https://doi.org/10.1177/073724770302800302

Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement, 23*(3), 187-194. https://doi.org/10.1177/01466219922031310

Renaissance Learning. (2021). Star CBM Reading Technical Manual.

Renaissance Learning. (2022). Star Assessments™ for Reading Technical Manual. https://help.renaissance.com/US/PDF/SR/SRRPTechnicalManual.pdf

Shinn, M. R. (2007). Identifying students at risk, monitoring performance, and determining eligibility within response to intervention: Research on educational need and benefit from academic intervention. *School Psychology Review, 36*(4), 601-617. https://doi.org/10.1080/02796015.2007.12087920

Wang, H., & Shin, C. D. (2010). Comparability of computerized adaptive and paper-pencil tests. *Test, Measurement and Research Service Bulletin, 13*, 1-7.

LEHIGH UNIVERSITY

Questions?

Thank you for listening!

Emily Forcht erf220@lehigh.edu

Ethan Van Norman erv418@lehigh.edu

QuALITY Lab: https://wordpress.lehigh.edu/quality/