

One Curriculum to Rule Them All

Experimental Evidence from Utah's Jordan District

Sam Lee, Gavin Hatch, Brigg Trendler

Department of Statistics, Brigham Young University

December 9, 2025

Introduction

Overview

- Elementary school curriculum
 - How much does curriculum influence student outcomes?
 - Traditional vs Curated
- Randomized crossover experiment
- Multivariate fixed effects and Bayesian random effects model
- Results: Curated rules them all!

Background



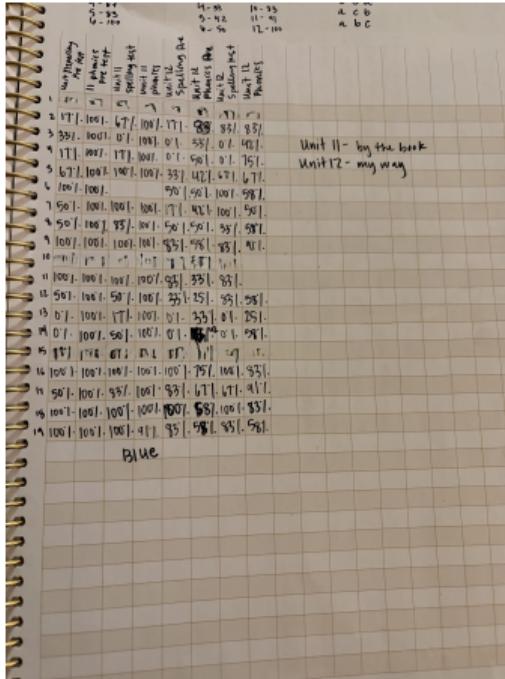
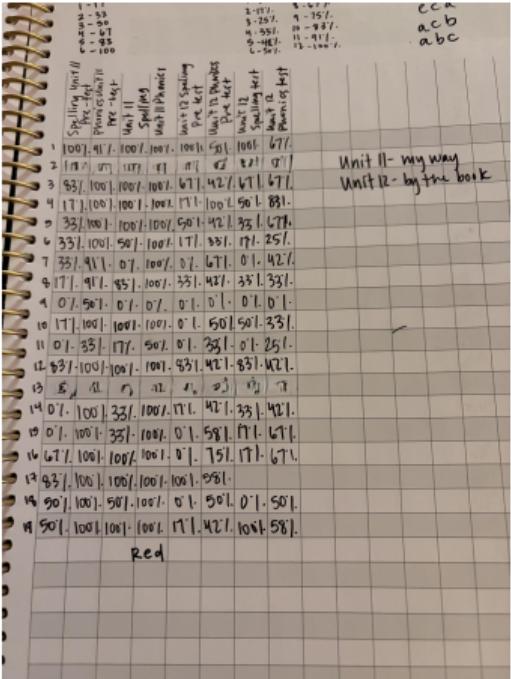
Experimental Design

Experimental Design

		Week 1	Week 2
		Curated	Traditional
Red Class 17 Students	Pre-Spelling	→	Post-Spelling
	Pre-Phonics	→	Post-Phonics
Blue Class 16 Students	Pre-Spelling	→	Post-Spelling
	Pre-Phonics	→	Post-Phonics

Data

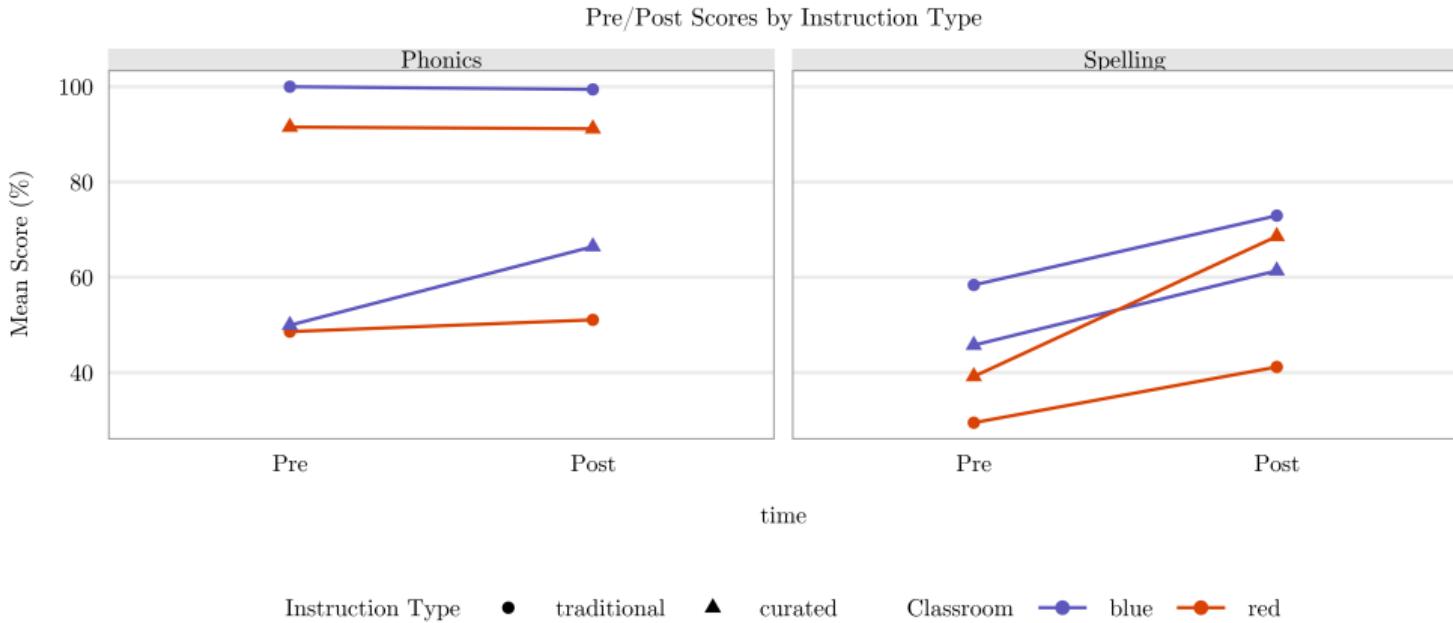
Data



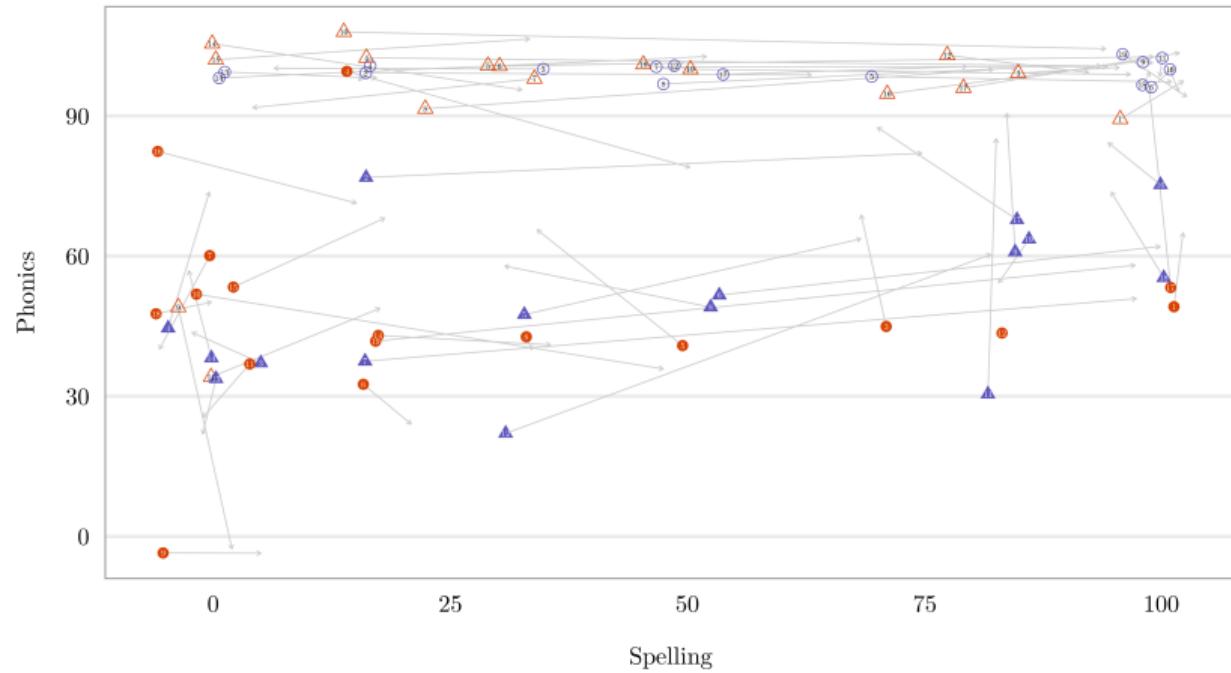
Data

- Bivariate response:
 - **Spelling and phonics test scores**
- Covariates:
 - **Curriculum type**
 - **Classroom** (Red or Blue)
 - **Week** (1 or 2)
 - **Age** (7 or 8)

Data (Aggregated)



Data (Disaggregated)



Methodology

Fixed Effect Model

- For student i from class c in week w who takes a test in period t ,

$$\mathbf{y}_{iwt} = \begin{pmatrix} \text{spelling}_{iwt} \\ \text{phonics}_{iwt} \end{pmatrix} \in \mathbb{R}^2,$$

$$\begin{aligned}\mathbf{y}_{iwt} = & \alpha_i + \lambda_w + \kappa_c + \mathbf{x}'_{iwt} \gamma \\ & + \mathbb{1}\{t = \text{post}\} \beta_1 + \mathbb{1}\{\text{instruction}_{cw} = \text{curated}\} \beta_2 \\ & + \mathbb{1}\{t = \text{post}\} \mathbb{1}\{\text{instruction}_{cw} = \text{curated}\} \beta_3 \\ & + \mathbf{u}_{iwt},\end{aligned}$$

where $\mathbf{u}_{iwt} \sim \mathcal{N}_2 \left(\mathbf{0}, \Sigma_u = \begin{bmatrix} \sigma_s^2 & \rho \\ \rho & \sigma_p^2 \end{bmatrix} \right)$

Estimators

- Standard maximum likelihood estimators hold:

$$\hat{\boldsymbol{B}} := (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y},$$

$$\hat{\Sigma}_u := \frac{1}{n} \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{B}} \right)' \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{B}} \right),$$

$$\text{Var} \left(\text{vec}(\hat{\boldsymbol{B}}) \right) := \boldsymbol{\Sigma}_u \otimes (\mathbf{X}' \mathbf{X})^{-1}.$$

Hypothesis Testing

- Does the curated curriculum improve scores over the week relative to the traditional curriculum?

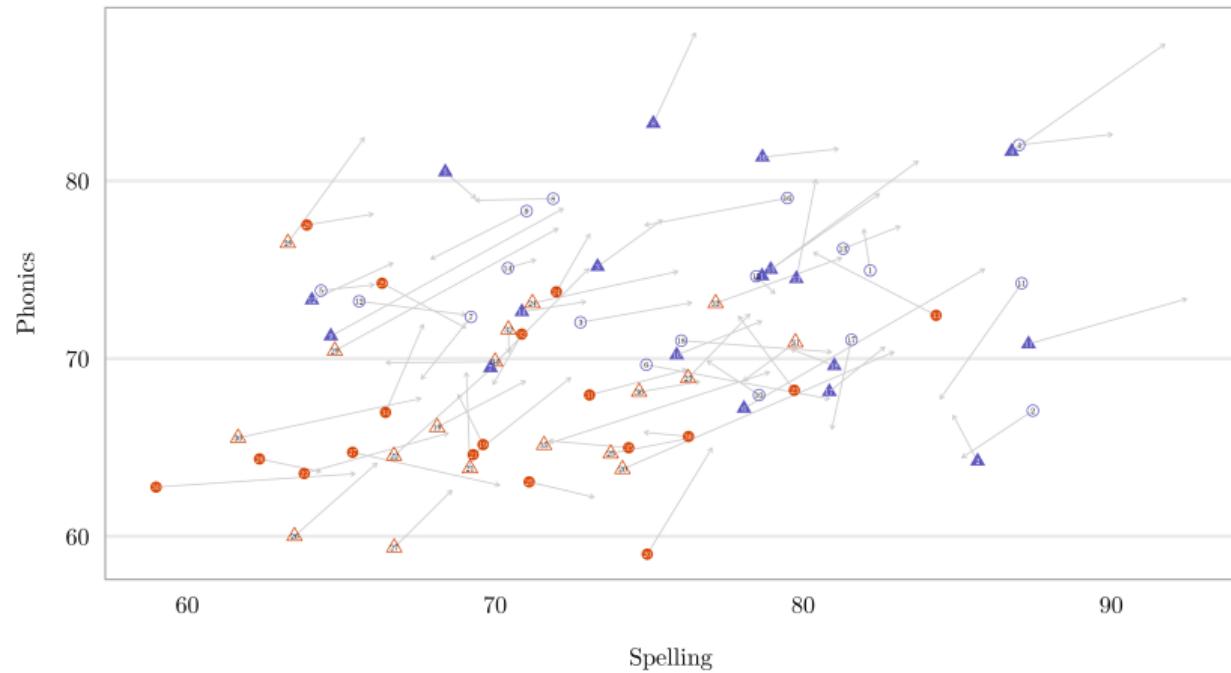
$$\mathcal{H}_0 : \mathbf{B}'\mathbf{c} = 0, \quad \mathbf{c} = (0 \ 0 \ \cdots \ 1)'$$

$$\mathcal{H}_1 : \mathbf{B}'\mathbf{c} \neq 0,$$

$$W := (\hat{\mathbf{B}}'\mathbf{c})' \left[(\mathbf{X}'\mathbf{X})_{kk}^{-1} \hat{\Sigma}_u \right]^{-1} (\hat{\mathbf{B}}'\mathbf{c}) \xrightarrow{d} \chi_2^2,$$

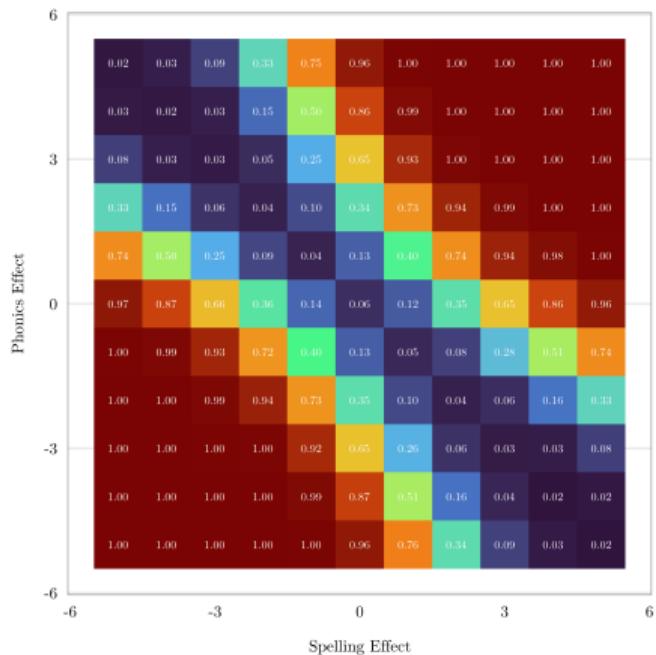
$$F := \frac{n - p - 1}{2(n - p)} W \sim F_{2, n-p-1}.$$

DGP

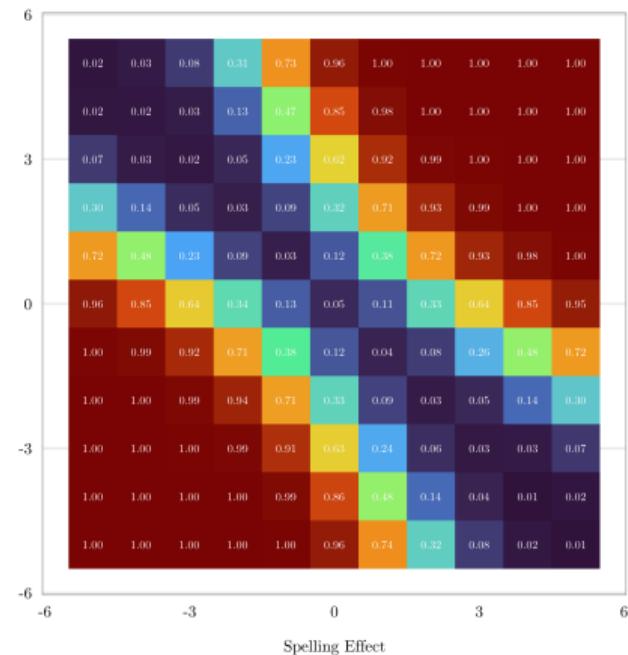


Rejection Rates

Wald test



F-test



Bayesian MLM

$$\begin{aligned}
 \mathbf{y}_{iwt} = & \mu_0 + \lambda_w + \kappa_c + \mathbf{x}'_{iwt} \boldsymbol{\gamma} \\
 & + \mathbf{z}'_{iwt} \mathbf{b}_i \\
 & + \mathbb{1}\{t = \text{post}\} \beta_1 + \mathbb{1}\{\text{instruction}_{cw} = \text{curated}\} \beta_2 \\
 & + \mathbb{1}\{t = \text{post}\} \mathbb{1}\{\text{instruction}_{cw} = \text{curated}\} \beta_3 \\
 & + \mathbf{u}_{iwt},
 \end{aligned}$$

where $\mathbf{u}_{iwt} \sim \mathcal{N}_2(\mathbf{0}, \Sigma_u = \begin{bmatrix} \sigma_s^2 & \rho \\ \rho & \sigma_p^2 \end{bmatrix})$,

$$\mathbf{z}_{iwt} = \mathbf{l}_2,$$

$\mathbf{b}_i \sim \mathcal{N}_2(\mathbf{0}, \Omega = \begin{bmatrix} \eta_s^2 & \eta_{sp} \\ \eta_{sp} & \eta_p^2 \end{bmatrix})$.

Bayesian Inference

- For each posterior draw $m = 1, \dots, M$,

$$\widehat{\Pr}(\beta_{3,\text{spelling}} > 0, \beta_{3,\text{phonics}} > 0 | \text{data}) = \frac{1}{M} \sum_{m=1}^M \mathbb{1}\left\{ \beta_{3,\text{spelling}}^{(m)} > 0, \beta_{3,\text{phonics}}^{(m)} > 0 \right\} \quad (1)$$

$$1 - \widehat{\Pr}(\beta_{3,\text{spelling}} < 0, \beta_{3,\text{phonics}} < 0 | \text{data}) = \dots \quad (2)$$

Results

Estimates

Panel A. Coefficient Summary

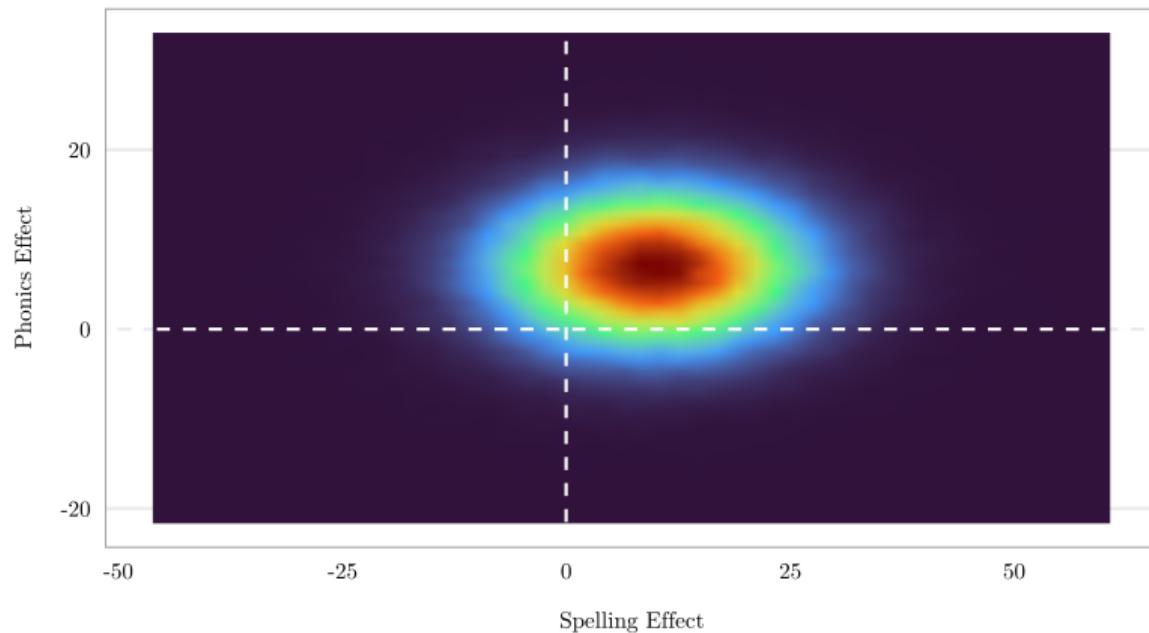
	Frequentist (FE model)		Bayesian (RE model)		
	Estimate	SE	Mean	Median	95% HPD
<i>Spelling</i>					
Age	22.92	9.19	17.37	17.38	[2.34, 32.40]
Class Red	-19.03	6.11	-17.40	-17.42	[-28.79, -5.90]
β_3	9.64	11.08	9.60	9.63	[-11.36, 30.83]
<i>Phonics</i>					
Age	16.54	4.72	13.69	13.69	[4.37, 22.65]
Class Red	-7.74	5.40	-8.00	-7.99	[-14.12, -1.88]
β_3	6.82	5.69	6.81	6.82	[-4.10, 17.92]

Inference

Panel B. Inference Summary

	Test	Statistic	p-value
<i>Frequentist</i>	Wald	$W = 1.44$	0.48
	F	$F = 0.72$	0.49
<i>Bayesian</i>	$\Pr(\beta_{3,\text{spell}} > 0, \beta_{3,\text{phon}} > 0 \text{data})$	0.72	
	$1 - \Pr(\beta_{3,\text{spell}} < 0, \beta_{3,\text{phon}} < 0 \text{data})$		0.98

Posterior Distribution of β_3



Conclusion

Conclusion

- Curated curriculum = better scores!
- Causality?
- Limitations
 - Small sample
 - Short horizon
 - Local results - population, teacher, etc.
 - Attenuation bias - lower bound!
- Future Work
 - Larger studies
 - Curriculum specifics

Thank You