On the Importance of Integrated Psychometrics and Multilevel Impact Estimation in Multi-Site RCTs: Lessons Learned from CATS

Li Cai, Kilchan Choi, Greg Chung, Eva Baker
CRESST

CATS Conference
Redondo Beach, CA
April 2014

cats
center for advanced technology in schools

CRESST

Statistical Routines in Multi-Site RCTs

- Random assignment, measurement at pretest and posttest.
- Standard off-the-shelf outcomes (e.g., commercially developed instruments, end-of-grade assessments).
- Or, researcher developed outcome measures.
- Classical or more modern psychometric analyses to gather technical quality information.
- Outcome scores are computed and hierarchical linear models fitted for estimating the impact of the intervention on the outcome.

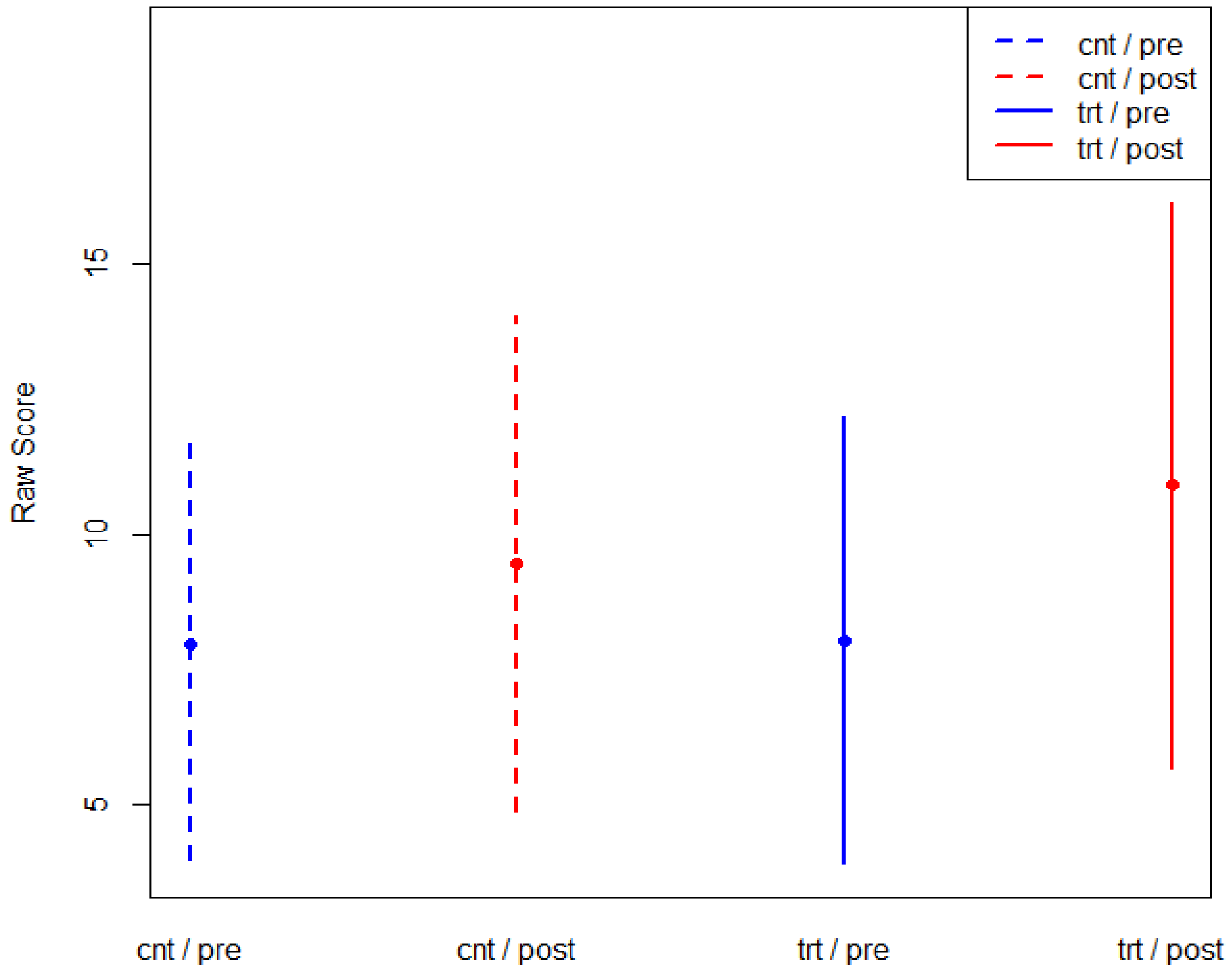
What the Routine Ignores

- The dependence between the outcome constructs at each occasion due to a longitudinal design.
- The item-level residual dependence due to repeated (pre-post) exposure to the same set of measures.
- The practical implausibility of assuming full exchangeability of subjects across treatment and control conditions.
- The obvious dependence of individuals due to their nesting in sites.

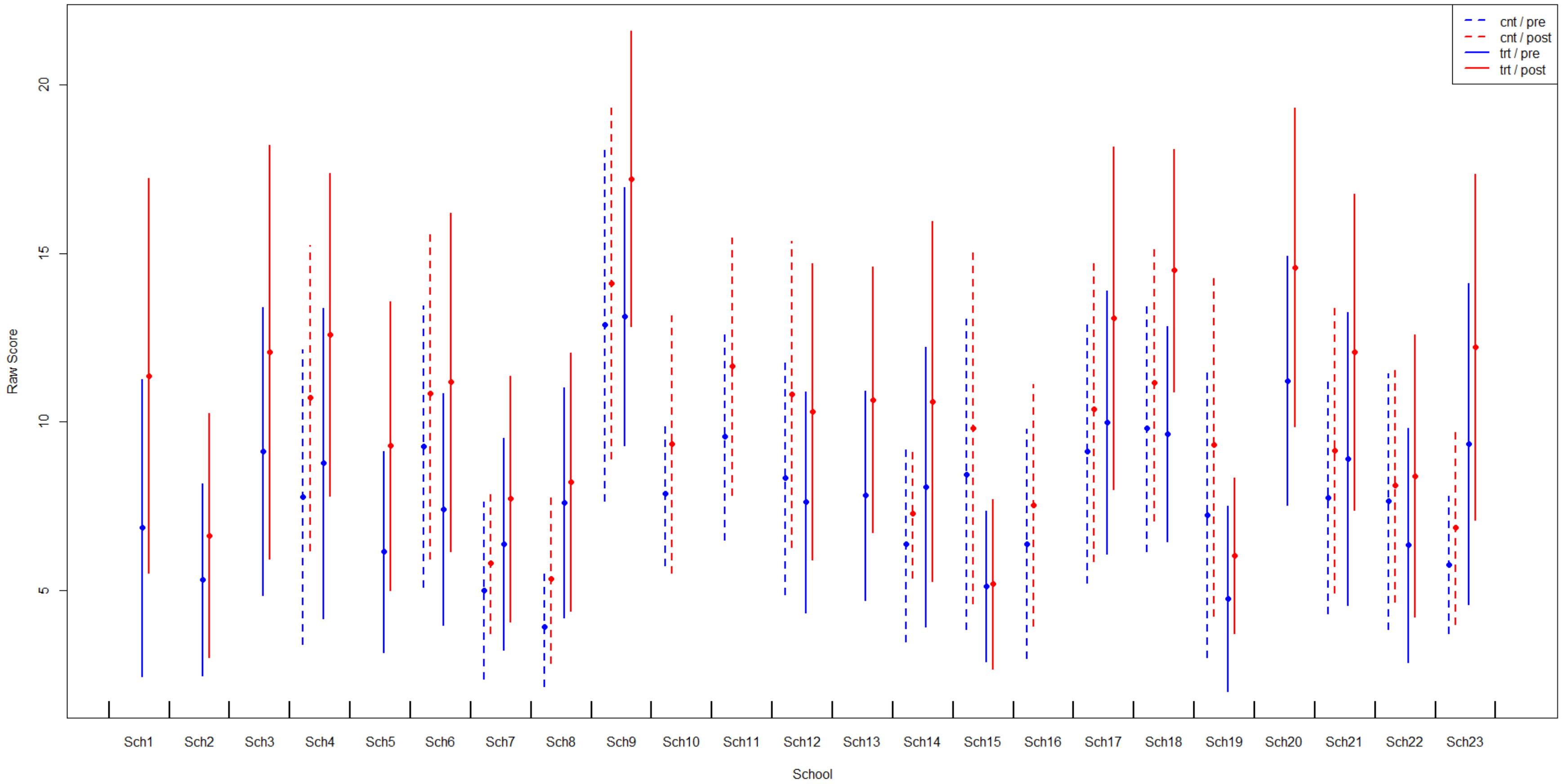
CATS

- Detailed knowledge specifications for assessment development, game development, and PD.
- Intervention classrooms played games focusing on rational number and fraction concepts.
- Comparison classrooms played games focusing on solving equations.
- The main outcome is developed to principally measure rational number and fraction knowledge.
- Within site randomization led to 30 intervention classrooms, and 29 comparison classrooms.
- Total efficacy sample: ~1500 students, 9 districts, 24 schools.

Raw Scores


cnt / pre
cnt / post
trt / pre
trt / post
Raw Score
5
10
15
cnt / pre
cnt / post
trt / pre
trt / post

Raw Scores (by Site)



Multilevel Impact Models

M1:

$$Y_{ij} = \beta_{0j} + \beta_{1j} * Trt_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$
$$\beta_{0j} = \gamma_{00} + u_{0j} \quad u_{0j} \sim N(0, \tau_{00})$$
$$\beta_{1j} = \gamma_{10} + u_{1j} \quad u_{1j} \sim N(0, \tau_{11})$$

M2:

$$Y_{ij} = \beta_{0j} + \beta_{1j} * Trt_{ij} + \beta_{2j} * Pre_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$
$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad u_{0j} \sim N(0, \tau_{00})$$
$$\beta_{1j} = \gamma_{10} + u_{1j}, \quad u_{1j} \sim N(0, \tau_{11})$$
$$\beta_{2j} = \gamma_{20} + u_{2j}, \quad u_{2j} \sim N(0, \tau_{22})$$

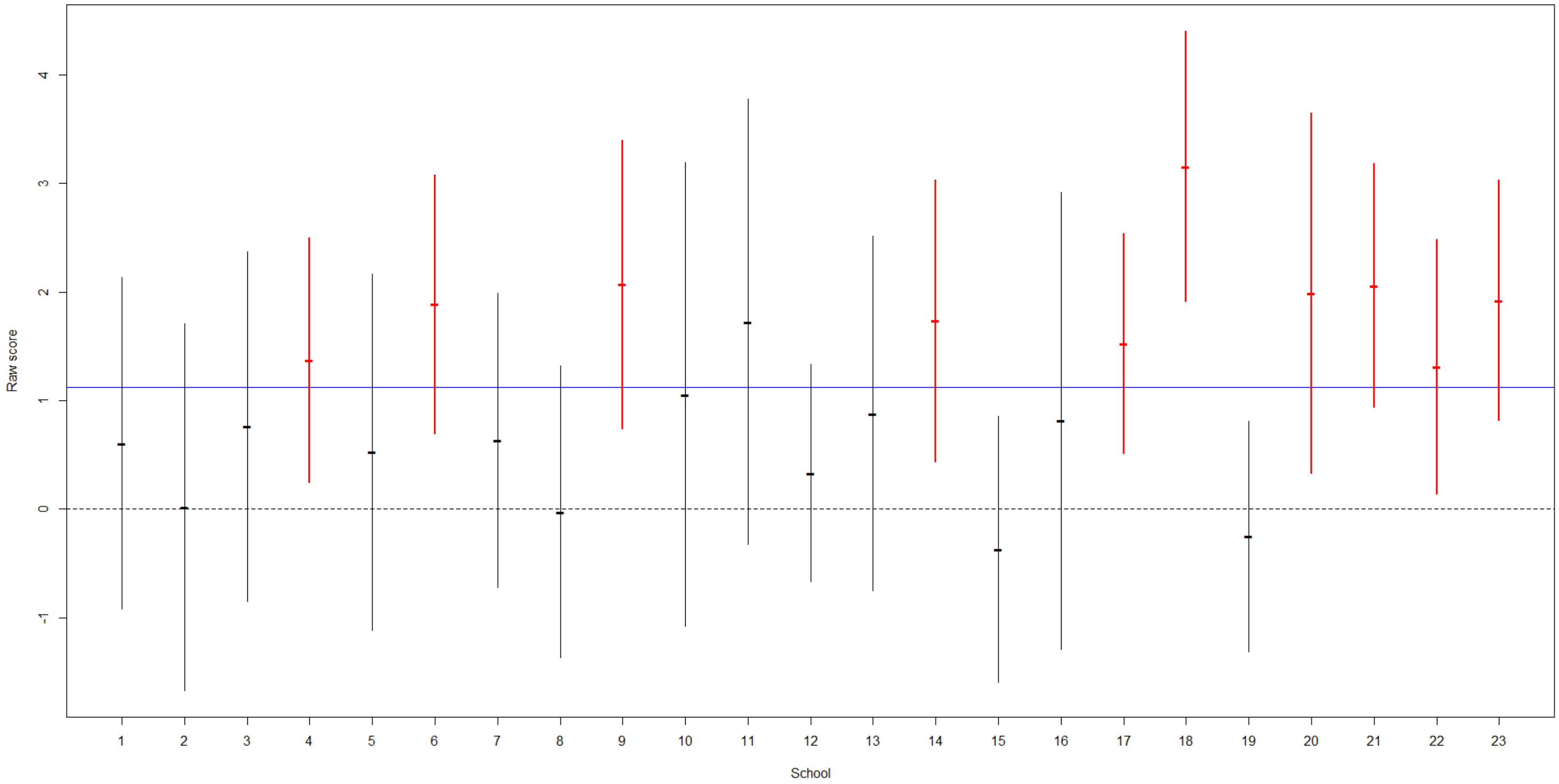
Impact Estimates with Raw Scores

	Model 1: Unconditional			Model 2: Pretest as covariate		
Fixed Effects	Estimate	*SE*	*p*-value	Estimate	*SE*	*p*-value
Intercept (γ_{00})	10.062	0.489	< .0001	10.334	0.486	< .0001
Trt (γ_{10})	1.388	0.636	.029	1.119	0.332	.001
Pretest(γ_{20})				0.9304	0.026	< .0001
Variance Components	Estimate	*SE*	*p*-value	Estimate	*SE*	*p*-value
Level-1 (σ^2)	19.914	0.728	< .0001	7.919	0.298	< .0001
Intercept (τ_{00})	4.723	1.634	.002	5.178	1.646	.001
Trt (τ_{11})	5.669	2.436	.010	1.307	0.686	.029
Pretest (τ_{22})				0.005	0.298	< .0001

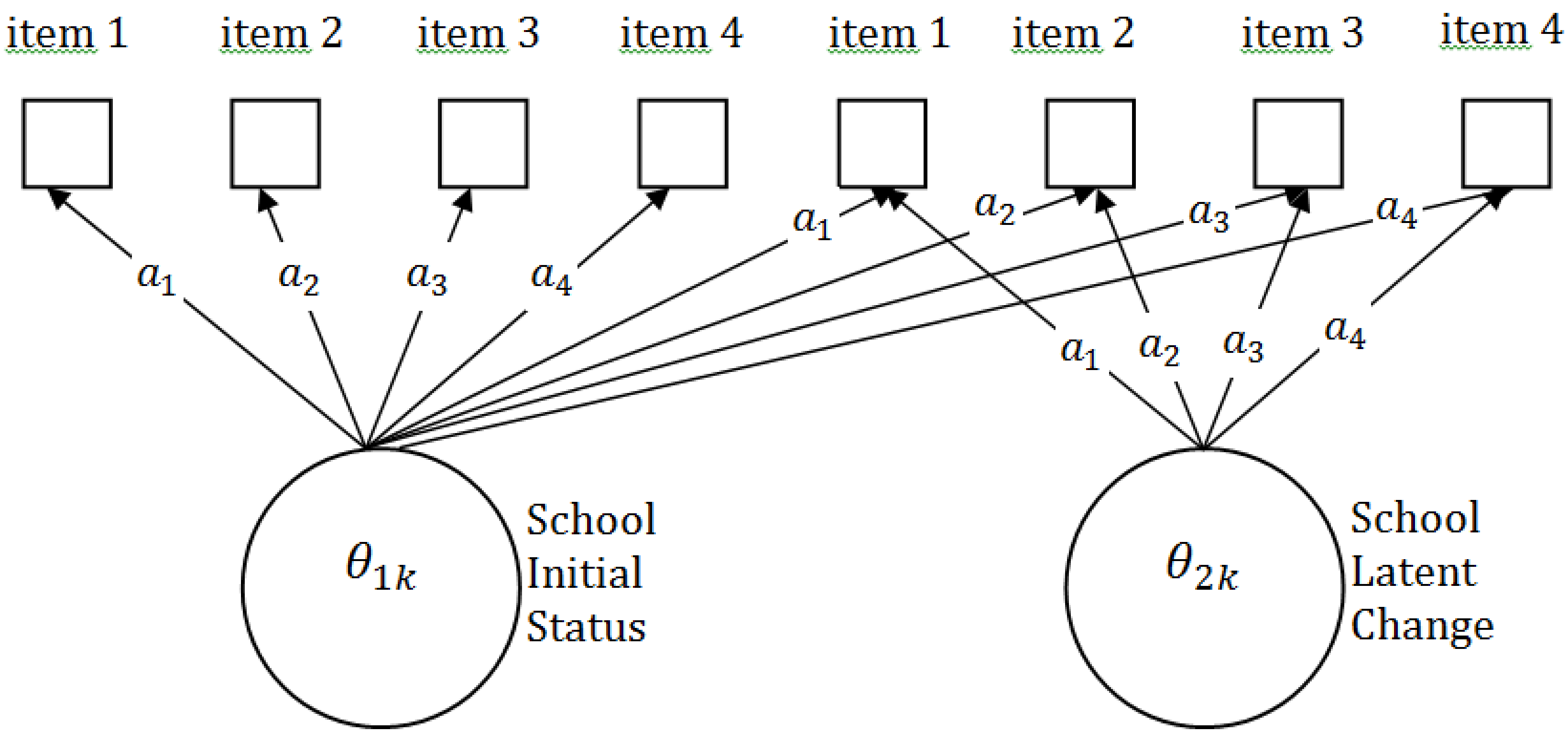
Impact Estimates with Raw Scores

	Model 1: Unconditional			Model 2: Pretest as covariate		
Fixed Effects	Estimate	*SE*	*p*-value	Estimate	*SE*	*p*-value
Intercept (γ_{00})	10.062	0.489	< .0001	10.334	0.486	< .0001
Trt (γ_{10})	1.388	0.636	.029	1.119	0.332	.001
Pretest(γ_{20})				0.9304	0.026	< .0001
Variance Components	Estimate	*SE*	*p*-value	Estimate	*SE*	*p*-value
Level-1 (σ^2)	19.914	0.728	< .0001	7.919	0.298	< .0001
Intercept (τ_{00})	4.723	1.634	.002	5.178	1.646	.001
Trt (τ_{11})	5.669	2.436	.010	1.307	0.686	.029
Pretest (τ_{22})				0.005	0.298	< .0001

Empirical Bayes Impact Estimates

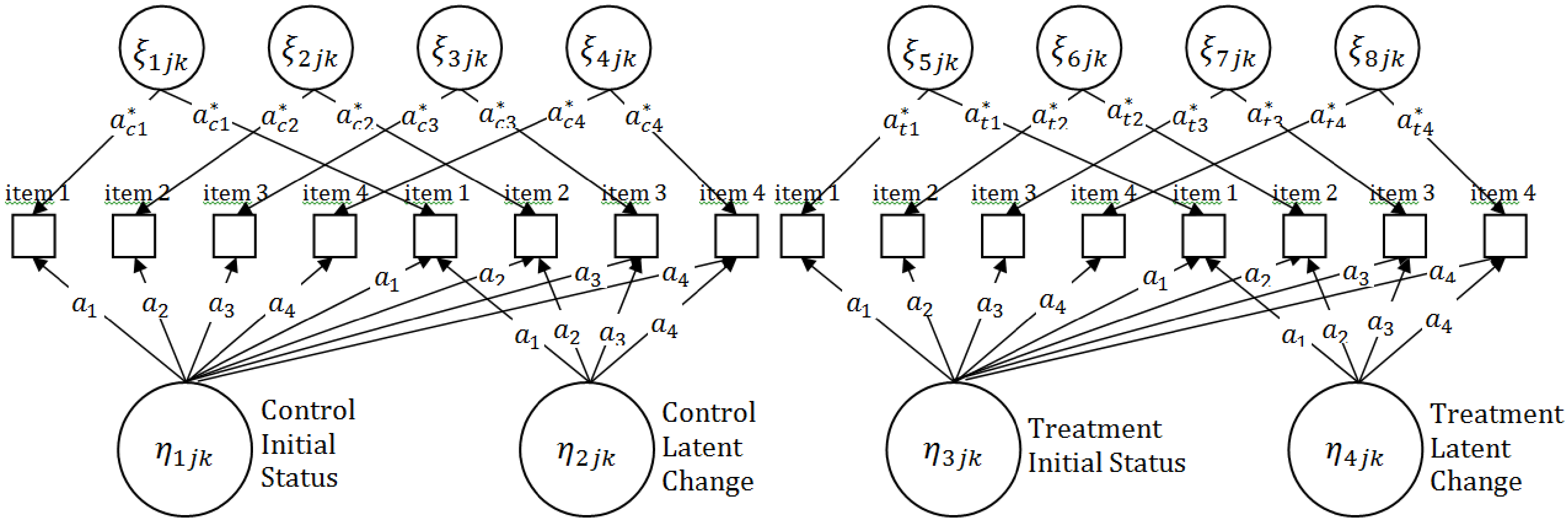


Multilevel Two-Tier (MTT) Model Example



Between-School (Site-Level) Model

Multilevel Two-Tier (MTT) Model Example



Within-School (Student-Level) Model

Item Level Formulation

Pretest Control:

$$a_i(\theta_{1k}+\eta_{1jk}) + a_{ci}^{*}\xi_{ijk}$$

Posttest Control:

$$a_i[(\theta_{1k}+\eta_{1jk}) + (\theta_{2k} + \eta_{2jk})] + a_{ci}^{*}\xi_{ijk}$$

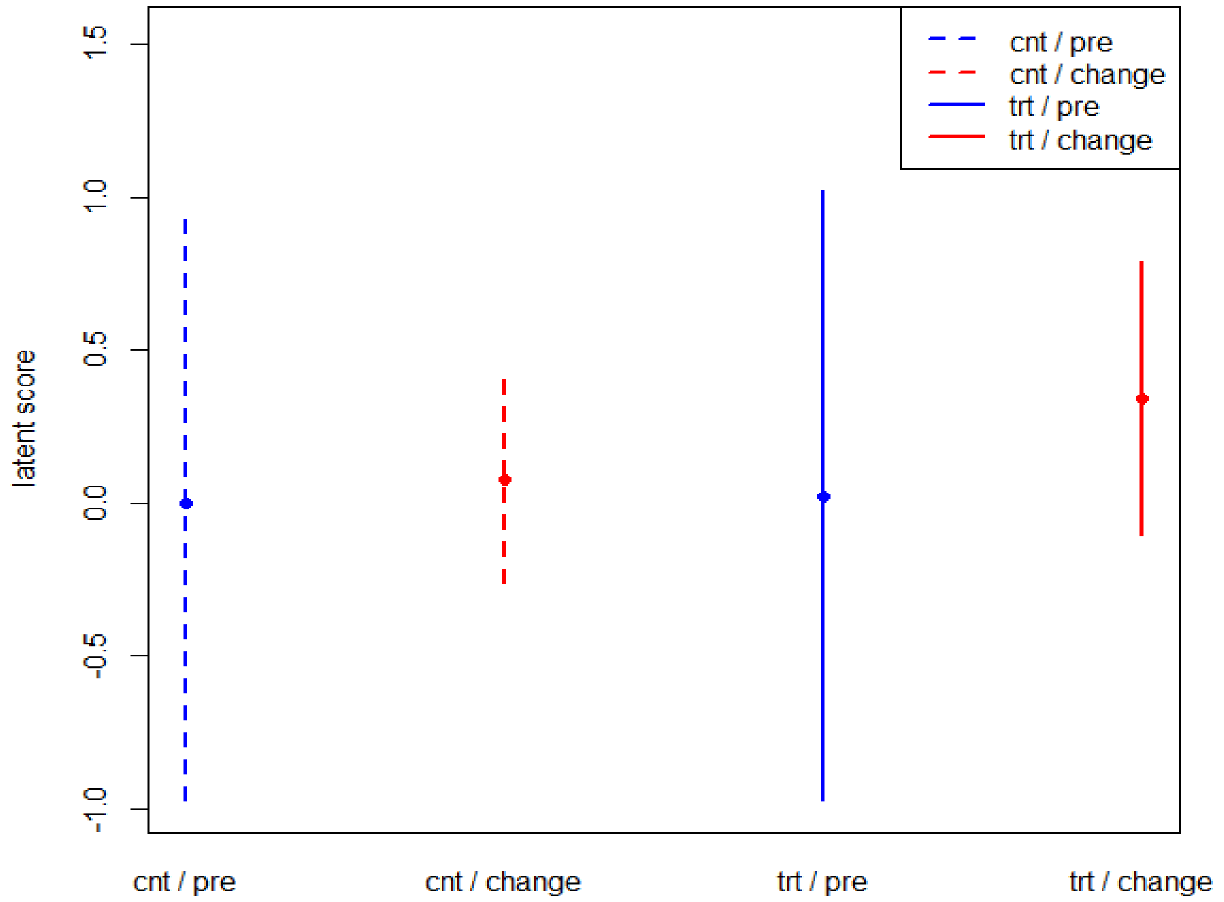
Pretest Treatment:

$$a_i(\theta_{1k}+\eta_{3jk}) + a_{ti}^{*}\xi_{ijk}$$

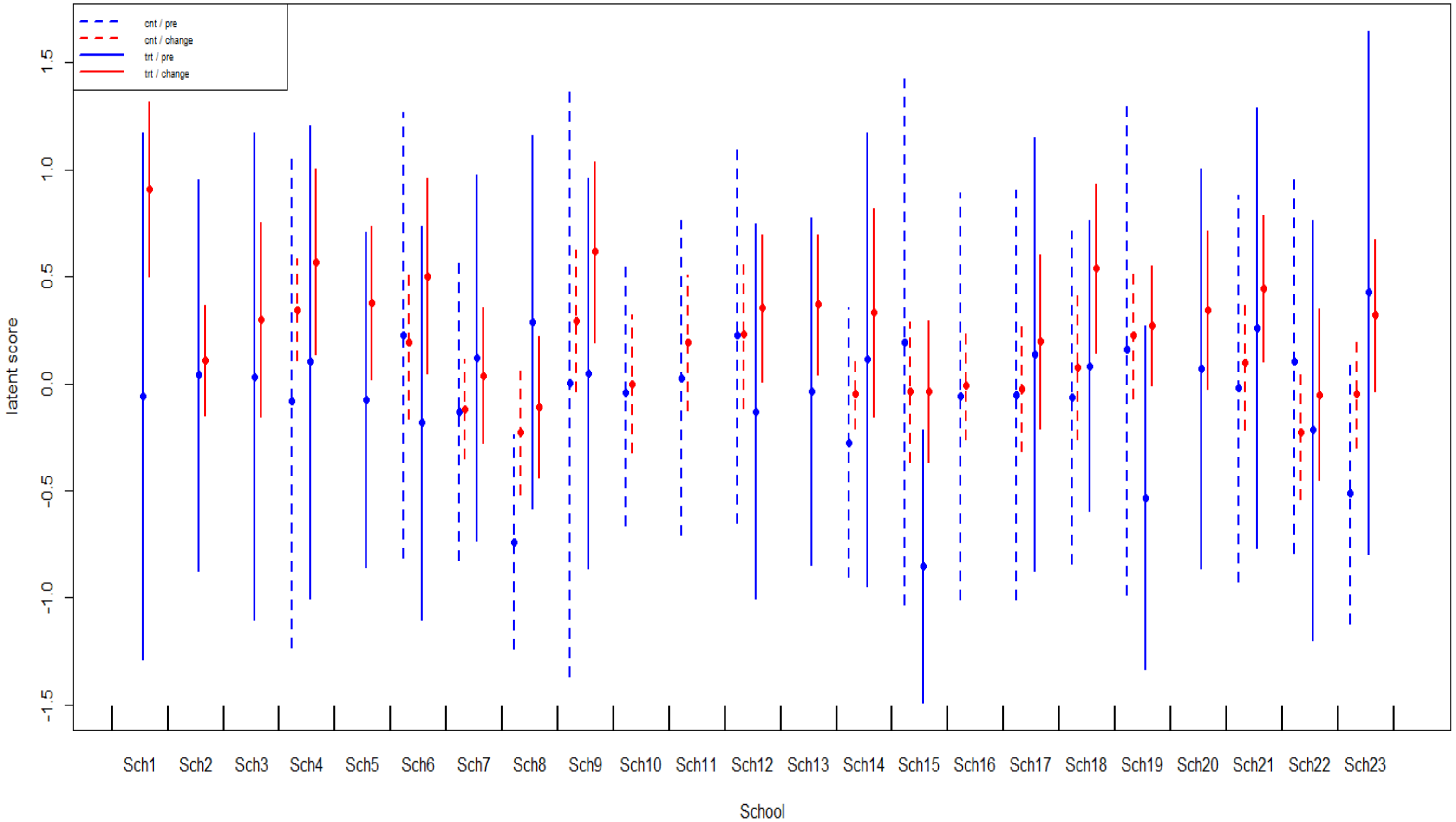
Posttest Treatment:

$$a_i[(\theta_{1k}+\eta_{3jk}) + (\theta_{2k} + \eta_{4jk})] + a_{ti}^{*}\xi_{ijk}$$

Latent Variable Pre and Change Scores



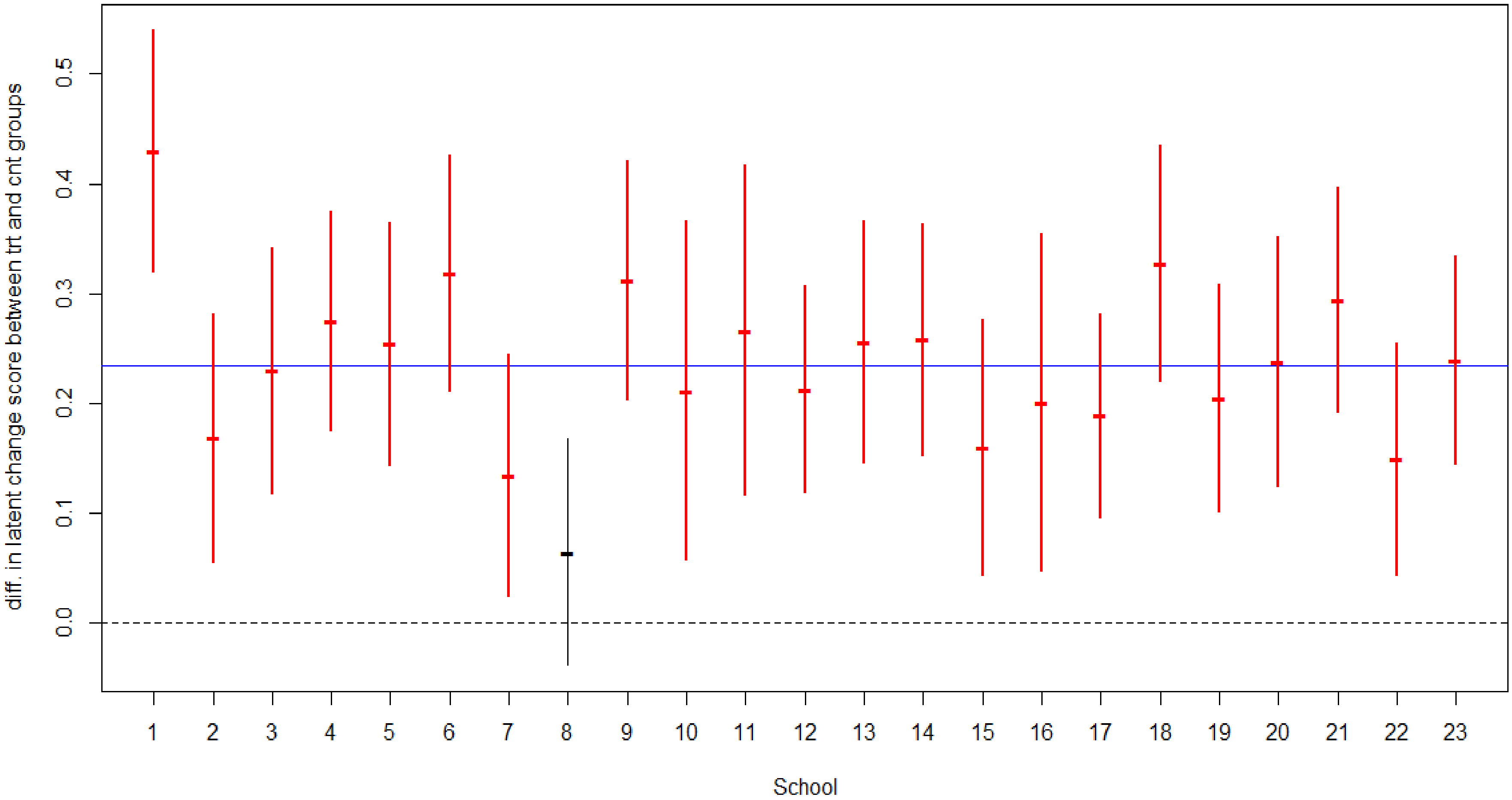
Latent Variable Pre and Change Scores (by Site)



Impact Estimates with Latent Change Scores

	Model 1: Unconditional			Model 2: Pretest as covariate		
Fixed Effects	Estimate	*SE*	*p*-value	Estimate	*SE*	*p*-value
Intercept (γ_{00})	0.199	0.042	.0001	0.207	0.042	< .0001
Trt (γ_{10})	0.243	0.035	< .0001	0.236	0.029	< .0001
Pretest(γ_{20})				0.107	0.009	< .0001
Variance Components	Estimate	*SE*	*p*-value	Estimate	*SE*	*p*-value
Level-1 (σ^2)	0.119	0.012	< .0001	0.110	0.004	< .0001
Intercept (τ_{00})	0.038	0.008	.001	0.037	0.012	.001
Trt (τ_{11})	0.013	0.004	.049	0.008	0.006	.094
Pretest (τ_{22})				0.000	0.298	.423

Empirical Bayes Impact Estimates



Effect Size Estimates

Cohen's δ = γ_{10}/ SD of outcome

Hedges' ES = γ_{10}/ sqrt $(\sigma^2 + \tau_{00} + \tau_{11})$

Conditional ES = γ_{10}/ sqrt (σ^2)

Outcome scores	Est. (γ_{10})	Posttest SD	Lev-1 var(σ^2)	Lev-2 Int var(τ_{00})	Lev-2 Trt var(τ_{11})	Hedges' ES	Cond. ES	Cohen's δ
Raw Score	1.139	4.986	19.914	4.723	5.669	0.207	0.405	0.228
Latent Change	0.236	0.419	0.110	0.037	0.009	0.615	0.705	0.579

Summary

- Integration of domain understanding, design of intervention, design of assessment, psychometric analysis, and impact modeling.
- Full exchangeability vs. conditional exchangeability.
- Psychometric models should reflect design constraints.
- Improved measurement => increased sensitivity.

center for advanced technology in schools

lcai@ucla.edu

National Center for Research
on Evaluation, Standards, & Student Testing

UCLA | Graduate School of Education & Information Studies

VISIT US ON THE WEB
cresst.org