State Based Intraclass Correlation Values for Planning Group-Randomized Trials in Education: Within and Between District Estimates

Larry V. Hedges
Northwestern
E. C. Hedberg (hedberg-eric@norc.org)
NORC at the University of Chicago

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D110032, NORC at The University of Chicago. The opinions expressed are those of the authors and do not necessarily represent the views of the Institute or the U.S. Department of Education.
Education evaluation design

• Cluster randomized trial
  – Randomize entire schools into treatment or control

• Issues of cost:
  – How many schools do we need to detect an effect?
  – How many students within those schools?

• Power analysis and optimal design require knowledge about the variance decomposition of academic achievement
  – This is summarized in the intraclass correlation statistic \(ICC\) or \(\rho\)
Districts

- Few districts typically employed by researchers
- Tend to be local
A noncentrality parameter

\[
\sqrt{\frac{nm}{2}} \sqrt{\frac{1}{1+\left(n-1\right)}} \frac{1}{R_1^2 + \left(nR_2^2 + R_1^2\right)}
\]

Effect size

Sample size

Design effect

Adjustment to the design effect based on variance explained by controls
The Intraclass Correlation (ICC)

- The proportion of the total variation that occurs at a specific level of analysis (except level 1)
- For level \( k \), random effect \( \zeta \)

\[
\frac{\text{Variance at level } k}{\text{Total variance}} = \frac{\text{var}(\zeta)}{\text{var}(\zeta) + \cdots + \text{var}()} 
\]

- Typically, we deal with 2-, 3-, or 4-level models in education
- Indicates the within-cluster similarity
ICCs in power calculations for different designs

- 2-level cluster randomized design (Hedges & Rhoads 2009)
  \[ \lambda = \delta \sqrt{\frac{mn}{2}} \sqrt{\frac{1}{1 + n - 1 \rho_2 - \left[ R_1^2 + nR_2^2 - R_1^2 \rho_2 \right]}} \]

- 2-level block randomized design (Hedges & Rhoads 2009)
  \[ \lambda = \delta \sqrt{\frac{mn}{2}} \sqrt{\frac{mn/2}{1 + \left( \frac{n \sigma_{T2}^2}{2 \sigma_2^2} - 1 \right) \rho_2 - \left[ R_1^2 + \left( \frac{n \sigma_{T2}^2}{2 \sigma_2^2} R_{T2}^2 - R_1^2 \right) \rho_2 \right]}} \]

- 3-level cluster randomized design (Konstantopoulos, 2008)
  \[ \lambda = \delta \sqrt{\frac{mpn}{2}} \sqrt{\frac{1}{1 - R_1^2 + pn 1 - R_3^2 - 1 - R_1^2 \rho_3 + n 1 - R_2^2 - 1 - R_1^2 \rho_2}} \]

- 3-level block randomized design (Konstantopoulos, 2008)
  \[ \lambda = \delta \sqrt{\frac{mpn}{2}} \sqrt{\frac{1}{1 - R_1^2 + \left( \frac{pn \sigma_{T3}^2}{\sigma_3^2} 1 - R_3^2 - 1 - R_1^2 \right) \rho_3 + n 1 - R_2^2 - 1 - R_1^2 \rho_2}} \]
Previous work on ICC reference values in education

- **Previous experiments**
  - Limited in scope and applicability

- **Hedges & Hedberg (2007)**
  - 2-level ICCs using national surveys
  - Students nested within schools
  - Grade 3 Reading ICC = 0.271

- **Bloom, Richburg-Hayes, & Black (2007)**
  - Within district ICCs were smaller than Hedges & Hedberg’s national 2-level ICCs
  - Grade 3 Reading within-district average ICC = 0.184
Districts

• Few districts typically employed by researchers
• Tend to be local
• National level ICCs may not be appropriate
  – Masks local variation in ICC values
  – Incorporate between-district variation that can be modeled away with fixed effects
Present study

• State-specific Design Parameters for Designing Better Evaluation Studies
  – Use SLDS to examine the variance structure of academic achievement in participating states
  – Produce ICCs and $R^2$ estimates
  – Examine 2-, & 3-level models
  – Particular attention to within-district, school-level, ICCs
  – Examine models in variety of states
State recruitment

• 7 states agreed to be part of the study
  – Arkansas
  – Arizona
  – Florida
  – Kentucky
  – Massachusetts
  – North Carolina
  – Wisconsin
Methods

• Estimated parameters
  – Unconditional ICCs
  – Standard Errors of ICCs
    • Derived from the estimated variance of the variance components (Hedges, Hedberg, & Kyper, 2012)
    • Stata software written by Hedberg for computations
  – $R^2$ values based on models with covariates
Methods

• Models
  – Unconditional models include no covariates
  – Academic achievement (AA) models include pretest and pretest group means
  – Demographic models (D) include race/ethnicity indicators (American Indian, Asian, Black, Hispanic), English learner status, Free/Reduced lunch status, and group means of these indicators
  – Academic achievement and demographic models (AAD) include pretest and demographic indicators

• Disabled students and charter schools removed
• Student samples within 5% of expected CCD counts
## Grade Coverage

<table>
<thead>
<tr>
<th>Grade</th>
<th>AR</th>
<th>AZ</th>
<th>FL</th>
<th>KY</th>
<th>MA</th>
<th>NC</th>
<th>WI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>MR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>MR</td>
<td>MR</td>
<td></td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
</tr>
<tr>
<td>4</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
</tr>
<tr>
<td>5</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
</tr>
<tr>
<td>6</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
</tr>
<tr>
<td>7</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
</tr>
<tr>
<td>8</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
<td>MR</td>
</tr>
<tr>
<td>9</td>
<td>MR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>MR</td>
<td>MR</td>
<td>R</td>
<td>MR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M</td>
</tr>
</tbody>
</table>
Students (level-1) nested within schools (level-2)

2-LEVEL MODELS
2-level model

• Unconditional model

\[ Y_{ij} = \beta_0 + \beta_j + \epsilon_{ij} \]

• Models with covariates produce \( \phi_j^* \) and \( \epsilon_{ij}^* \)

• Unconditional ICC estimate

\[ R^2 = \frac{\text{var}(f_j)}{\text{var}(f_j) + \text{var}(\epsilon_{ij})} \]

• \( R^2 \) estimates

\[ R^2_1 = \frac{\text{var}(\epsilon_{ij}) \text{var}(\epsilon_{ij}^*)}{\text{var}(\epsilon_{ij}) \text{var}(f_j)}, R^2_2 = \frac{\text{var}(f_j) \text{var}(\epsilon_{ij}^*)}{\text{var}(f_j) \text{var}(\epsilon_{ij})} \]
ICCs from two-level models

- Somewhat consistent in lower grades
- More inconsistent in higher grades
- Larger for higher grades (on average)
Math $\rho$ (two-level)

$\rho$ for math

Grade 3

Grade 4

Grade 5

Grade 6

Grade 7

Grade 8

State mean = 0.180

State mean = 0.175

State mean = 0.187

State mean = 0.200

State mean = 0.220

State mean = 0.223

*Nation based on surveys, 3rd grade not available for FL model is $Y_{ij} = \mu_{ij} + \phi_j + \epsilon_{ij}$, $\rho = \text{var}(\phi_j)/[\text{var}(\phi_j)+\text{var}(\epsilon_{ij})]$
Reading $\rho$ (two-level)

$\rho$ for reading

Grade 3

Grade 4

Grade 5

Grade 6

Grade 7

Grade 8

State mean = 0.156

State mean = 0.170

State mean = 0.164

State mean = 0.170

State mean = 0.185

State mean = 0.191

*Nation based on surveys, 3rd grade not available for FL
model is $Y_{ij} = \mu_{0j} + \phi_j + \varepsilon_{ij}$, $\rho = \text{var}(\phi_j)/[\text{var}(\phi_j)+\text{var}(\varepsilon_{ij})]$
Math $R^2_2$ (two-level)

$R^2_2$ for math

Grade 3

Grade 4

Grade 5

Grade 6

Grade 7

Grade 8

3rd grade not available for FL
model is $Y_{ij} = \mu_{0j} + \sum_{p} \gamma_p W_{ij,p} + \sum_{q} \beta_q X_{ij,q} + \phi_j + \epsilon_{ij}$. 

Mean: AA=0.51, D=0.56 AAD=0.57

Mean: AA=0.74, D=0.59 AAD=0.76

Mean: AA=0.76, D=0.59 AAD=0.78

Mean: AA=0.73, D=0.61 AAD=0.76

Mean: AA=0.76, D=0.68 AAD=0.81

Mean: AA=0.88, D=0.66 AAD=0.88
Reading $R^2_2$ (two-level)

$R^2_2$ for reading

Grade 3

Grade 4

Grade 5

Grade 6

Grade 7

Grade 8

Mean: AA=0.72, D=0.68 AAD=0.77

Mean: AA=0.83, D=0.71 AAD=0.84

Mean: AA=0.85, D=0.72 AAD=0.86

Mean: AA=0.84, D=0.73 AAD=0.86

Mean: AA=0.83, D=0.77 AAD=0.88

Mean: AA=0.90, D=0.74 AAD=0.91

3rd grade not available for FL

model is $Y_{ij} = \mu_{0j} + \sum_p \gamma_p W_{i,p} + \sum_q \beta_{q} X_{i,j,q} + \phi_j + \epsilon_{ij}$
Students (level-1) nested within schools (level-2), nested within districts (level-3)

3-LEVEL MODELS
3-level model

- Unconditional model
  \[ Y_{ijk} = \beta_{0jk} + \beta_j + \beta_{ijk} \]

- Models with covariates produce \( \gamma_k, \phi_{jk}, \text{ and } \epsilon_{ijk} \)

- Unconditional ICC estimates
  \[ R^2 = \frac{\text{var}(\gamma_k)}{\text{var}(\gamma_k) + \text{var}(\phi_{jk}) + \text{var}(\epsilon_{ijk})} \]

- \( R^2 \) estimates
  \[ R_1^2 = \frac{\text{var}(\gamma_k) - \text{var}(\gamma_k^*)}{\text{var}(\gamma_k)}, R_2^2 = \frac{\text{var}(\phi_{jk}) - \text{var}(\phi_{jk}^*)}{\text{var}(\phi_{jk})}, R_3^2 = \frac{\text{var}(\epsilon_{ijk}) - \text{var}(\epsilon_{ijk}^*)}{\text{var}(\epsilon_{ijk})} \]
Within district ICC estimates

- $\rho_2$ provides estimate of the 2-level ICC within a district
  - Designs within a single district
  - Designs using district fixed effects
- Much more consistent
- Smaller ($\approx 0.1$, with exceptions)
4th Grade Math \( \rho \) and \( \rho_2 \)
Math $\rho_2$ (three-level)

$\rho_2$ for math

Grade 3
State mean = 0.108

Grade 4
State mean = 0.109

Grade 5
State mean = 0.115

Grade 6
State mean = 0.126

Grade 7
State mean = 0.148

Grade 8
State mean = 0.156

3rd grade not available for FL

model is $Y_{ijk} = \mu_{ijk} + \zeta_k + \phi_{jk} + \epsilon_{ijk}$, $\rho_2 = \text{var}(\phi_{jk})/\left[\text{var}(\zeta_k) + \text{var}(\phi_{jk}) + \text{var}(\epsilon_{ijk})\right]$
Reading $\rho_2$ (three-level)

$\rho_2$ for reading

Grade 3

State mean = 0.089

Grade 4

State mean = 0.102

Grade 5

State mean = 0.096

Grade 6

State mean = 0.100

Grade 7

State mean = 0.113

Grade 8

State mean = 0.128

3rd grade not available for FL

model is $Y_{ijk} = \mu_{ijk} + \zeta_k + \phi_{jk} + \varepsilon_{ijk}$, $\rho_2 = \frac{\text{var}(\phi_{jk})}{\text{var}(\zeta_k) + \text{var}(\phi_{jk}) + \text{var}(\varepsilon_{ijk})}$
Within district *ICC* estimates

• $\rho_2$ is much more consistent across states  
  – FL and NC are different, however
• Attempt to explain why
• FL has very few, large, districts
Math $\rho_2 = \ln(n_k)$?

$\rho_2$ for math

Grade 4

Grade 5

Grade 6

Grade 7

Grade 8

$R^2 = 0.55$, $R^2_{\text{NOPL}} = 0.22$

$R^2 = 0.88$, $R^2_{\text{NOPL}} = 0.62$

$R^2 = 0.98$, $R^2_{\text{NOPL}} = 0.79$

$R^2 = 0.99$, $R^2_{\text{NOPL}} = 0.85$

$R^2 = 1.00$, $R^2_{\text{NOPL}} = 0.99$

$\ln(\text{Students/Districts})$

solid line = $x\beta$, dashed line = $x\beta$ with FL omitted
Reading $\rho_2 = \ln(n_k)$?

$\rho_2$ for reading

Grade 4

$R^2 = 0.55, R^2_{\text{NOPL}} = 0.12$

Grade 5

$R^2 = 0.73, R^2_{\text{NOPL}} = 0.53$

Grade 6

$R^2 = 0.90, R^2_{\text{NOPL}} = 0.28$

Grade 7

$R^2 = 0.86, R^2_{\text{NOPL}} = 0.39$

Grade 8

$R^2 = 0.98, R^2_{\text{NOPL}} = 0.82$

$\ln(\text{Students/Districts})$

solid line = $x\beta$, dashed line = $x\beta$ with FL omitted
Within district ICC estimates

• $\rho^2$ is much more consistent across states
  – FL and NC are different, however
• Attempt to explain why
• FL has very few, large, districts
• Grade 7 reading does not fit model, otherwise: plausible
Between district *ICC* estimates

- *Also somewhat consistent, but different states (AZ and MA) stand out*
Math $\rho_3$ (three-level)

$\rho_3$ for math

Grade 3

Grade 4

Grade 5

Grade 6

Grade 7

Grade 8

3rd grade not available for FL

model is $Y_{ijk} = \mu_{ijk} + \zeta_k + \phi_{jk} + \epsilon_{ijk}$, $\rho_3 = \text{var}(\zeta_{jk})/[\text{var}(\zeta_k) + \text{var}(\phi_{jk}) + \text{var}(\epsilon_{ijk})]$
Reading $\rho_3$ (three-level)

$\rho_3$ for reading

Grade 3

Grade 4

Grade 5

Grade 6

Grade 7

Grade 8

State mean = 0.049

State mean = 0.053

State mean = 0.055

State mean = 0.045

State mean = 0.046

State mean = 0.040

3rd grade not available for FL

model is $Y_{ijk} = \mu_{ijk} + \zeta_k + \phi_{jk} + \epsilon_{ijk}$, $\rho_3 = \frac{\text{var}(\zeta_{jk})}{\text{var}(\zeta_k) + \text{var}(\phi_{jk}) + \text{var}(\epsilon_{ijk})}$
Implications

• Plausible: the ICCs based on national surveys may be larger than actual field conditions.
• Specific designs that model the district effects might be able to assume smaller ICC values and employ fewer schools
Thank you

Questions?