Intraclass Correlations and Covariate Outcome Correlations for Planning 2 and 3 Level Cluster Randomized Experiments in Education

Larry V. Hedges[1]

E. C. Hedberg[2]

1: Board of Trustees Professor at The Institute for Policy Research at Northwestern University

847 491 8899

L-Hedges@Northwestern.edu

2: Senior Research Scientist at NORC at the University of Chicago

3016 E Wescott Drive

Phoenix, AZ 85050

312 485 5376

Hedberg-Eric@NORC.org

Acknowledgement: 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.

Abstract

Background

Cluster randomized experiments that assign intact groups such as schools or school districts to treatment conditions are increasingly common in educational research. Such experiments are inherently multilevel designs whose sensitivity (statistical power and precision of estimates) depends on the variance decomposition across levels. This variance decomposition is usually summarized by the intraclass correlation structure and, if covariates are used, the effectiveness of the covariates in explaining variation at each level of the design.

Objectives

This paper provides a compilation of school and district level intraclass correlation values of academic achievement and related covariate effectiveness based on state longitudinal data systems. These values are designed to be used for planning group-randomized experiments in education. The use of these values to compute statistical power and plan 2 and 3 level group randomized experiments is illustrated.

Research Design

We fit several hierarchical linear models to state data by grade and subject to estimate intraclass correlations and covariate effectiveness. We then compare our average of state estimates with the national work by Hedges and Hedberg (2007).

A fundamental problem in education and other applied social sciences is determining the causal effects of interventions designed to improve educational or social conditions. Randomized experiments are widely appreciated because they offer the strongest designs for making causal inferences about treatment effects (Mosteller and Boruch, 2002). For this reason, randomized experiments have become much more frequently used in the last decade to evaluate educational interventions, products, and services. The most common experimental designs in education have been designs that assign intact groups (such as classrooms, schools, or school districts) to treatment conditions. These designs are called group or cluster randomized because the intact groups (e.g.,, schools or districts) can be considered statistical clusters.

The sampling plans for cluster randomized experiments typically involve multistage cluster samples in which clusters (such as schools) are sampled, and then individuals are sampled within clusters. In some cases there are three or four stages of sampling where school districts are sampled first, then schools within districts, then classrooms within schools, then individuals within classrooms. Because cluster randomized experiments involve multistage sampling, it is often natural to think of them in terms of multilevel statistical models—as multilevel experiments.

One aspect of planning experimental designs is assuring that the design has adequate sensitivity to detect the treatment effects of interest. We use the word *sensitivity* to include the precision (standard error) of estimates of treatment effects, the statistical power to detect effects, and the minimum effect size that is detectable with a given level of certainty (the minimum detectable effect size). The sensitivity of multilevel designs depends on the variance decomposition between and within schools

(Raudenbush 1997; Bloom, Bos et al. 1999; Bloom 2005; Konstantopoulos, 2009; Hedges and Rhoads 2011). This variance decomposition is typically summarized by a system of intraclass correlation coefficients, which are the proportion of the total variance that occurs between units at various levels of the design. For example, suppose there are three levels in the design (school districts as level 3, schools as level 2, and individuals within schools as level 1 and that districts will be assigned to treatments. Then the variance decomposition determining statistical power could be summarized by a level 3 (district level) intraclass correlation ρ_3 that expresses the fraction of the total variation in the outcome that is between district means and a level 2 (school level) intraclass correlation ρ_2 that expresses the fraction of the total variation in the outcome that is between school means but within districts. The fraction of the total variation that is between individuals but within schools is the complement of ρ_2 and ρ_3, namely $\bar{\rho} = 1 - \rho_2 - \rho_3$.

If the design uses covariates, the effectiveness of covariates in explaining variation (variance components) at different levels of the design also has an impact on the sensitivity of the design. The effectiveness of covariates in explaining variance at each level of the design is often characterized by a measure of variance accounted for (an R^2) at each level of the design. For example if the three level design mentioned above used a pretest as a covariate at each level, the effectiveness of the pretest as a covariate would be characterized by three R^2 values: R_3^2, the variance in district (level 3) means that is explained by district mean pretest scores; R_2^2, the variance in school (level 2) means within districts that is explained by school mean pretest scores; and R_1^2, the variance in individual (level 1) scores within schools explained by individual pretest scores.

Because cluster randomized designs are inherently multilevel experiments, rational planning of sample sizes cluster randomized designs is more complex than planning sample sizes in single level experiments that randomize individuals within simple random samples. For example, there are multiple components of the total sample size: one for each level of the design. While decisions about sample size in one level designs involve determining a single number (total sample size) that yields a design with the required sensitivity, decisions about sample size in multilevel designs involve determining appropriate sample sizes at each of several levels. In single level designs, a larger total sample size always leads to greater statistical power (all other things equal). However, the relationship between sample size and design sensitivity is not straightforward in multilevel designs. Given the same *total* sample size, different allocations of sample sizes across levels can lead to very different statistical power and designs with smaller total sample size can have greater statistical power than other designs that have larger total sample size. For this reason, so called optimal design or optimal allocation methods (which maximize precision or statistical power for a given cost function) are often used to assist in planning multilevel designs (see, e.g., Raudenbush, 1997; Konstantopoulos, 2009). Optimal allocation depends on cost data, but also on the intraclass correlation structure and the effectiveness of covariates in explaining variation in the outcome variable at different levels of the design.

Because intraclass correlation structure and covariate effectiveness is crucial in planning cluster randomized experiments, we refer to these values as *design parameters*. The purpose of this paper is to provide empirical evidence about design parameters that

can be used in the planning of two and three level cluster randomized experiments that use academic achievement as the outcome variable. This paper is in the same spirit as Hedges and Hedberg, 2007) but instead of using data from national surveys as they did, this paper uses data from state longitudinal data systems to estimate design parameters and includes information on school districts as a level of analysis.

The Present Study and Key Findings

Data from state longitudinal data systems in seven states was used to estimate parameters useful for designing two and three level cluster randomized trials. We considered two cases. In one case, school district is ignored in the design, there is one intraclass correlation that reflects total variation across schools, and between-district variation is pooled into between-school variation. This might happen when the design calls for schools from several districts, there are few schools per district, and district is not used as a blocking factor. In the second case, school district is explicitly included in the design as level of sampling. This might occur if randomization to treatments occurred at the school district level or schools were randomized to treatments but districts were used as blocking factor assumed not to interact with treatments.

In the first case (where district variation is pooled into between school variation) intraclass correlation estimates in grades 3 to 8 averaged about $\rho = 0.20$ in mathematics achievement and $\rho = 0.17$ in reading achievement, but there was considerable variation across states. There was a slight trend for intraclass correlations to be larger at the higher grades (from $\rho = 0.18$ in grade 3 to $\rho = 0.22$ in grade 8 in reading and from $\rho = 0.16$ in grade 3 to $\rho = 0.19$ in grade 8 in mathematics achievement). A pretest on academic

achievement was a substantially more effective covariate than demographic variables, explaining an average of R_2^2 = 80% of the variation in mathematics achievement at level 2 (the school level) and an average of R_2^2 = 87% of the variance in reading achievement at level 2, while explaining an average of R_1^2 = 64% of the variance in mathematics achievement at level 1 (the individual level) and an average of R_1^2 = 57% of the variance in reading achievement at level 1. As in the case of the intraclass correlations, there was considerable variation across states in the effectiveness of the covariates in explaining variation in reading and mathematics achievement. As in the case of intraclass correlations, there is a tendency for the effectiveness of covariates to increase with grade level. For example, the effectiveness of pretest in explaining level 2 variation in mathematics achievement increased from an average of R_2^2 = 75% in grade 3 to an average of R_2^2 = 88% in grade 8. Similarly, the effectiveness of pretest in explaining level 1 variation in mathematics achievement increased from an average of R_1^2 = 58% in grade 3 to an average of R_1^2 = 68% in grade 8. Variation across grades in the effectiveness of covariates in explaining reading achievement was similar to that in mathematics achievement.

In the second case (where school districts are explicitly included in the design) there are two intraclasss correlations to be estimated, ρ_3 at the school district level and ρ_2 at the school within-district level. District level intraclass correlation estimates in grades 3 to 8 averaged about ρ_3 = 0.05 in both reading and mathematics achievement and there was little variation across either states or grade levels. School within-district intraclass correlation estimates in grades 3 to 8 averaged about ρ_2 = 0.13 in mathematics achievement and about ρ_2 = 0.10 in reading achievement, but there was some variation

across states. There was a slight trend for intraclass correlations to be larger at the higher grades (from $\rho_2 = 0.09$ in grade 3 to $\rho_2 = 0.13$ in grade 8 in reading and from $\rho_2 = 0.11$ in grade 3 to $\rho_2 = 0.16$ in grade 8 in mathematics achievement). A pretest on academic achievement was a substantially more effective covariate than demographic variables, explaining an average of $R_3^2 = 84\%$ of the variation in mathematics achievement at level 2 and an average of $R_3^2 = 89\%$ of the variance in reading at level 3, explaining an average of $R_2^2 = 72\%$ of the variation in mathematics achievement at level 2 and an average of $R_2^2 = 81\%$ of the variance in reading at level 2, while explaining an average of $R_1^2 = 64\%$ of the variance in mathematics achievement at level 1 and $R_1^2 = 58\%$ of the variance in reading at level 1. The substantially smaller school level intraclass correlations within districts demonstrates the greater sensitivity of designs that might be carried out within a single school district or within a few districts in which it might be reasonable to expect no district by treatment interactions.

Methods

Data Sources

The evidence reported in this paper was derived from the state longitudinal data systems in seven U.S. States: Arkansas, Arizona, Florida, Kentucky, Massachusetts, North Carolina, and Wisconsin. Students were assessed in the spring using a different test in each state, namely the Augmented Benchmark Examination (Arkansas), Arizona's Instrument to Measure Standards, the Florida Comprehensive Assessment Test, the Commonwealth Accountability Testing System (Kentucky), Massachusetts Comprehensive Assessment System, the North Carolina End of Grade Tests, and the Wisconsin Knowledge and Concepts Examination. These states are a subset of the states

we approached to be part of our study. All data was from the 2009-2010 school year with the exception of Florida, which supplied data from the 2006-2007 school year.

Data quality was assessed by comparing the number of students in our cleaned dataset to the number of students listed in the CCD. Since our data was comprised of the number tested, and since No Child Left Behind (Bush 2001) requires only 95 percent participation, we allowed no more than a 5 percent deficit between our number of students and the CCD. For this analysis, we removed students who were cognitively disabled at the time of assessment and those attending charter schools. In many states, charter schools are their own district and so an estimation of the (between school within district) intraclass correlation is not possible. Since many such districts would be removed a priori, we maintained comparability by removing all charter schools.

Choices of Covariates

It is often advantageous to use one or more covariates to improve design sensitivity. Covariates chosen can be any variables that are correlated with the outcomes that cannot be influenced by the treatment, but the covariates usually used are pretests on the same construct as the outcome variable or demographic variables such as gender, race/ethnicity, indicators of socio-economic status (SES), and indicators of potential difficulty in school such as English language learner status. We evaluated the effectiveness of three covariate models: one involving only a pretest on the outcome variable of interest, one involving only demographic variables, and one involving both a pretest and demographic variables. The details of the analytic models are given in the appendix and the details of the covariate sets are given below.

The pretest covariate model. If pretest scores on achievement are available, they can be a powerful covariate and considerably increase the sensitivity of an experimental design. The two-level pretest covariate model involves using the cluster-centered (school mean-centered) pretest score at the individual level and the school mean pretest score at the school level. The three-level pretest covariate model involves using the subcluster-centered (school mean-centered) pretest score at the individual level, the cluster-centered (school district mean-centered) school mean pretest score at the school level, and the school district mean at the school district level.

The demographic covariates model. Sometimes pretest scores are not available but other background information about individuals is available to serve as covariates. The demographic covariates model includes five covariates at each level. At the individual-level, the covariates are dummy variables for male gender, for Black or Hispanic status, for eligibility for free or reduced price lunch as a proxy for socioeconomic status, and an indicator that the student is classified as an English language learner The two-level demographics covariate model involves using the cluster-centered (school mean-centered) covariate score at the individual level and the school mean covariate score at the school level. The three-level demographic covariate model involves using the subcluster-centered (school mean-centered) covariate score at the individual level, the cluster-centered (school district mean-centered) school mean covariate score at the school level, and the school district mean at the school district level. We experimented with different centering techniques and confirmed that when group means are included the variance components did not change.

The pretest and demographic covariates model. The pretest and demographic covariates model combines the use of an achievement pretest and demographic covariates at each level.

Analysis Models

The data analysis was carried out using STATA version 12.1's "XTMIXED" routine for mixed linear model analysis, with residual variance components estimated by restricted maximum likelihood. For each sample and achievement domain, analyses were carried out based on four different models. The first model, (the unconditional model) involved no covariates at any level. The second model (the pretest covariate model) used a test on the same individuals in the same achievement domain one year earlier as a covariate. The third model (the demographic covariates model) used dummy variables for male gender, Black or Hispanic race or ethnicity, free or reduced price lunch (as an indicator of socio-economic status, and limited English proficiency status. The fourth model (the pretest and demographic covariates model) use both the covariates in the second and third models together. We describe these explicitly in the Appendix using hierarchical linear model notation. The standard errors of the intraclass correlations were computed using large sample results given in Hedges, Hedberg, and Kuyper (2012). The standard errors of R^2 values were computed from the large sample estimates of the variance of the squared multiple correlation,

$$\mathrm{Var}\left\{R^2\right\} = \frac{4R^2\left(1-R^2\right)^2}{n},$$

(see, e.g., Fisher, 1925/1990).

Results

We present results by first giving a summary of sample sizes. Then we present estimates of design parameters when the between school district variance is pooled into the between school variance. To produce tables of reasonable size we present results for grades 1 to 6 and grades 7 to 11 in separate tables. Finally we present design parameters for three level analyses where between district and between school-within-district variance are considered separately.

A summary of the sample sizes used in the analyses is given in Table 1. The body of the table is organized into horizontal panels by grade where each row represents sample sizes for each state within the grade defined by the horizontal panel. The table has three vertical panels for district, school, and student sample sizes. The first, second, and third vertical panel shows that the data in each state are based on from 73 to 416 school districts, 294 to 1,841 schools, and 29,882 to 160,821 students in each state.

Insert Table 1 About Here

Design Parameters Pooling Between-District into Between-School Variance

In this section we present design parameters that are appropriate when research designs will include schools from several districts, but there will be no attempt to use districts as blocking variables (essentially district dummy variables as covariates). In such cases, the between-district variation is pooled into the between-school variation. The design parameters for mathematics achievement in grades 1 to 6 are given in Table 2

and the design parameters for reading achievement in grades 1 to 6 are given in Table 3. We do not report values of design parameters for the models including both pretest and demographic covariates because the demographic covariates generally differ very little from those based on pretest alone. That is, the demographic covariates have little explanatory power beyond that of the pretest. The structure of these tables is similar to that of Table 1 in that the tables are organized into horizontal panels by grade and each row represents information for each state within the grade defined by the horizontal panel. The tables are organized into three vertical panels. The left hand panel gives the intraclass correlation (unadjusted for any covariates) and its standard error, the middle panel gives the R^2 values reflecting the effectiveness of the pretest as a covariate at level 2 and level 1 (along with their standard errors), and the right hand panel gives the gives the R^2 values reflecting the effectiveness of the demographic variables as covariates at level 2 and level 1 (along with their standard errors). Design parameters are reported for all grades included in the state longitudinal data. In some cases (e.g., Grade 3 in Arizona or Grade 1 in Arkansas), design parameters corresponding to pretest as a covariate are missing because there was no assessment at an earlier grade to use as a pretest.

Insert Tables 2 and 3 About Here

The design parameters for mathematics achievement in grades 7 to 11 are given in Table 4 and the design parameters for reading achievement in grades 7 to 10 are given in Table 5. These tables have the same format as Tables 2 and 3 but reflect different grade levels. Tables 4 and 5 are somewhat sparser, which reflects state practices of less

frequent assessments at higher grade levels. One striking feature of Table 4 is that some of the intraclass correlations in mathematics (e.g., in Florida) are large than at the lower grades. Similarly, some of the intraclass correlations in reading (e.g., in Florida and Massachusetts) in Table 5 are also larger than at the lower grades. These findings could have important implications for research design.

Insert Tables 4 and 5 About Here

Design Parameters Pooling Between District into Between School Variance

In this section we present design parameters that are appropriate for designing two-level cluster randomized designs that will include only schools from a single school district. The results in this section are also useful for evaluating the proportion of total variance across schools that is accounted for when district is used as a fixed blocking variable in a two level design that involves several districts, since the relevant R^2 for the set of district dummy variables is $R_2^2 = \rho_3/(\rho_2 + \rho_3)$. Finally, the results in this section are appropriate for three level cluster randomized designs assigning school districts to treatments. In such cases, the between district variation is pooled into the between school variation. As in the previous section, we do not report values of design parameters for the models including both pretest and demographic covariates because the demographic covariates generally differ very little from those based on pretest alone. That is, the demographic covariates have little explanatory power beyond that of the pretest.

The design parameters for mathematics achievement in grades 1 to 6 are given in Table 6 and the design parameters for reading achievement in grades 1 to 6 are given in

Table 7. The structure of these tables is similar to that of Tables 2 and 3 in that the tables are organized into horizontal panels by grade and each row represents information for each state within the grade defined by the horizontal panel and the tables are organized into three vertical panels. The left hand panel gives the intraclass correlations at level 3 (district level) and 2 (school level), unadjusted for any covariates and their standard errors. The middle panel gives the R^2 values reflecting the effectiveness of the pretest as a covariate at levels 3, 2 and 1 (district, school, and students, respectively), along with their standard errors, and the right hand panel gives the gives the R^2 values reflecting the effectiveness of the demographic variables as covariates at levels 3, 2, and (along with their standard errors).

Insert Tables 6 and 7 About Here

The design parameters for mathematics achievement in grades 7 to 11 are given in Table 8 and the design parameters for reading achievement in grades 7 to 10 are given in Table 9. These tables have the same format as Tables 6 and 7 but reflect different grade levels. Tables 8 and 9 are somewhat sparser, which reflects state practices of less frequent assessments at higher grade levels. One striking feature of Table 8 is that the level 3 (district level) intraclass correlations are all relatively small but some of the level 2 (school level) intraclass correlations in mathematics (e.g., in Florida) are larger than at the lower grades. Table 9 reflects the same pattern of intraclass correlations in reading achievement. These findings could have important implications for research design.

Insert Tables 8 and 9 About Here

Using the Results of this Paper

The results of this paper can be used alone or with software designed to help plan the details of the design of cluster randomized experiments. One application is for planning the sample size required to achieve a specific design sensitivity. The design sensitivity might be specified as a particular statistical power to detect a treatment effect of a given size, a standard error of an estimate of a treatment effect given a particular total standard deviation, or a specific minimum detectable effect size. The design parameter values given in this paper can provide inputs for comparing the sensitivity different alternative designs and for planning optimal allocations once a design type has been chosen. We illustrate several applications of these design parameters in planning designs using a commercial program that computes statistical power for cluster randomized studies, *CRT-Power* (Borenstein, 2012), but most of the same calculations could be done using other software, including freeware such as *Optimal Design* (Spybrook, et al., 2012).

Suppose that we are planning a cluster randomized experiment to evaluate an intervention that involves teacher professional development in grade 5 in Kentucky (or a state we believe is very similar to Kentucky). We might start by assuming that we will have a sample of schools dispersed across many school districts, so the design parameters in Table 2 would be appropriate. Entering Table 2 in the row for grade 5 in Kentucky, we see that the intraclass correlation for mathematics at the school level ignoring districts is $\rho = 0.151$. We might envision a sample size of $n = 25$ students per school and an effect

size (standardized by the total standard deviation) of $\delta = 0.25$. We might enter this intraclass correlation, effect size $\delta = 0.25$, and level 1 sample size $n = 25$ into *CRT-Power* and note that a total of $m = 48$ schools per treatment group would be necessary to obtain power of 80%, with no covariates. However if a pretest was available, we might enter Table 2 on the row for grade 5 in Kentucky and go to the second vertical panel of the table to obtain the values $R_2^2 = 0.551$ and $R_1^2 = 0.584$ for the effectiveness of the covariate at level 2 and level 1. Entering these values into *CRT-Power*, we see that only $m = 22$ schools per treatment group would be necessary to obtain statistical power of 80%. If a pretest was not available, but instead demographic data were available, Table 2 shows that the effectiveness of the pretest as a covariate would be characterized by $R_2^2 = 0.322$ and $R_1^2 = 0.085$, where each R^2 values involves 4 covariates. Entering these values into *CRT-Power* (along with $\rho = 0.151$, $n = 25$, and noting that there are 4 covariates) we see that $m = 35$ schools per treatment group would be necessary to obtain statistical power of 80%. This difference ($m = 35$ versus $m = 22$ schools required per treatment group) reflects how much more effective pretest is as a covariate relative to the demographic variables.

Another design option might be to obtain all of the schools for the study from a single large district. In that case the design parameters in Table 6 would be appropriate because the level 2 design parameters reflect the variation of schools within districts. Entering Table 6 in the row for grade 5 in Kentucky, we see that the intraclass correlation for mathematics at the school level ignoring districts is $\rho_3 = 0.020$ and $\rho_2 = 0.117$. We might envision a sample size of $n = 25$ students per school and an effect size (standardized by the total standard deviation) of $\delta = 0.25$. We might enter this intraclass

correlation $\rho_2 = 0.117$, effect size $\delta = 0.25$, and level 1 sample size $n = 25$ into *CRT-Power* and note that a total of $m = 40$ schools per treatment group would be necessary to obtain power of 80%, with no covariates. Note that this is less than the $m = 48$ schools per treatment group required in the design with many unblocked districts. Now consider the sensitivity of the design if a pretest was available, Table 6 shows that the effectiveness of the pretest as a covariate would be characterized by $R_2^2 = 0.500$ and $R_1^2 = 0.584$. Entering these values into *CRT-Power*, we see that only $m = 20$ schools per treatment group would be necessary to obtain statistical power of 80%. If a pretest was not available, but instead demographic data were available, Table 6 shows that the effectiveness of the demographic variables as covariates would be characterized by $R_2^2 = 0.315$ and $R_1^2 = 0.085$, where each R^2 values involves 4 covariates. Entering these values into *CRT-Power* (along with $\rho = 0.117$, $n = 30$, and noting that there are 4 covariates) we see that $m = 30$ schools per treatment group would be necessary to obtain statistical power of 80%. This difference ($m = 30$ versus $m = 20$ schools required per treatment group) reflects how much more effective pretest is as a covariate relative to the demographic variables.

Alternatively, one might consider a design that sampled schools from say 3 districts and used district dummy variables with the assumption that there is no district by treatment interaction. In that case Table 2 presenting design parameters from two level analyses provides the intraclass correlation data because the intraclass correlations in this table reflect the total variation across schools (and districts). However some of the variation is accounted for by the district dummy variables being used as covariates. Entering Table 6 on the row for Grade 5 in Kentucky, we see that the intraclass

correlation for mathematics at the district level is $\rho_3 = 0.020$ and the intraclass correlation at the school within district level is $\rho_2 = 0.117$. The three district level covariates (the district dummy variables) account for variance corresponding to an R_2^2 of $\rho_3/(\rho_2 + \rho_3) = 0.020/ (0.117 + 0.020) = 0.146$. Continue to consider a sample size of $n = 25$ students per school and an effect size of $\delta = 0.25$. However now we have three district level covariates corresponding to the district dummy variables, and these dummy variables correspond to an R_2^2 of $R_2^2 = \rho_3/(\rho_2 + \rho_3) = 0.02/ (0.117 + 0.020) = 0.146$. Entering the intraclass correlation of $\rho_2 = 0.151$, three covariates at level 2 with a combined R^2 value of $R_2^2 = 0.146$, an effect size $\delta = 0.25$, and level 1 sample size $n = 25$ into *CRT-Power* and note that a total of $m = 42$ schools per treatment group would be necessary to obtain power of 80%, with no other covariates.

One more option that might be considered is planning a three level cluster randomized trial that assigned school districts to treatments at grade 5 in Kentucky. Continue to assume a sample size of $n = 25$ students per school and an effect size (standardized by the total standard deviation) of $\delta = 0.25$, except now we consider a three level design in which districts are assigned to treatments and there are $p = 4$ schools per district. Entering Table 6 in the row for grade 5 in Kentucky, to obtain the intraclass correlation for mathematics at the district level of $\rho_3 = 0.020$ the intraclass correlation at the school within district level of $\rho_2 = 0.117$. Entering the intraclass correlations $\rho_2 = 0.117$ and $\rho_3 = 0.020$, effect size $\delta = 0.25$, the level 1 sample size $n = 25$, and the level 2 sample size of $p = 4$ into *CRT-Power* and note that a total of $m = 16$ districts (and 64 schools) per treatment group would be necessary to obtain power of 80%, with no covariates. However if a pretest was available, Table 6 shows that the effectiveness of

the pretest as a covariate would be characterized by $R_3^2 = 0.617$, $R_2^2 = 0.500$, and $R_1^2 = 0.584$. Entering these values into *CRT-Power*, we see that only $m = 8$ districts (and 32 schools) per treatment group would be necessary to obtain statistical power of 80%.

Of the options considered here, the option of using schools from a single district yields a design that requires the smallest sample size to achieve statistical power of 80%. However this design has the disadvantage that it may limit generalizability, since it involves only a single school district. Moreover, it might be infeasible because it requires so many (40) schools from a single district. The design involving many districts and pooling between district variation into the between school variation (not blocking by school district) is nearly as sensitive (requiring a total of 42 schools) and probably has advantages in external validity and feasibility. The results in this example are driven by the fact that, at grade 5 in Kentucky, the level 3 (between district) intraclass correlation is so small relative to the level 2 (between schools within districts) intraclass correlation ($\rho_3 = 0.020$ and $\rho_2 = 0.117$). The corresponding results in Arizona (where $\rho_3 = 0.122$, $\rho_2 = 0.099$, and the intraclass correlation ignoring districts is $\rho = 0.201$) would have been substantially different.

The design parameters provided here can be used to determine optimal allocations of sample between levels of a design to achieve 80% power for the smallest relative cost. To do so we must also specify a cost structure in terms of the relative cost of units at each level of the design. Consider a three level design assigning school districts at grade 5 in Kentucky, and assume that the cost of adding a new district to the study is 5 times the cost of adding a school in an existing district, which is 10 times the cost of adding an individual within an existing school. That is, the relative cost of level 3 units (districts) is

$c_3 = 50$, the relative cost of level 2 units (schools) is $c_2 = 10$, and the relative cost of level 1 units students) is $c_1 = 1$. Entering these cost parameters along with $\rho_2 = 0.117$ and $\rho_3 = 0.020$ into CRT-Power, and using the optimal design wizard, we obtain an optimal allocation of $m = 55$ districts per treatment, each with $p = 5$ schools, each with $n = 9$ students in the study. With the same cost structure, the optimal allocation for the same design at grade 5 in Arizona (where $\rho_3 = 0.122$, $\rho_2 = 0.099$) would be $m = 17$ districts per treatment, each with $p = 2$ schools, each with $n = 9$ students in the study.

Finally, the design parameters provided here could be used to explore the minimum detectable effect size, such as the minimum effect size detectable with 80% power. Consider a two level cluster randomized design assigning schools to treatments in grade 5 in Kentucky, where the intraclass correlation ignoring districts is $\rho = 0.151$ and the covariate effects can be summarized as $R_2^2 = 0.551$ and $R_1^2 = 0.584$. Suppose that we anticipate using $n = 25$ students per school. Entering these design parameters into *CRT-Power*, we can explore the minimum detectable effect size for various possible designs. If we can only afford $m = 10$ schools per treatment, the minimum detectable effect size is $\delta = 0.382$. If we could afford $m = 15$ schools per treatment group, the minimum detectable effect size drops to $\delta = 0.305$, and if we could afford $m = 20$ schools per treatment group, the minimum detectable effect size drops to $\delta = 0.261$. If we could afford $m = 22$ schools per treatment group, the minimum detectable effect size drops to $\delta = 0.249$, which corresponds to the finding above that 22 schools per treatment group were necessary to obtain 80% in this design.

Discussion

In this section we discuss the general patterns of our findings and contrast our results with Hedges and Hedberg's (2007) work. We focus our discussion on grades 3 through 8 because we have the most states represented in these grades. Obviously, we have estimated a large number of parameters. To keep our findings tractable, we offer comparisons only with the averages of state parameters.

The central finding in this study is that states vary in their variance decomposition patterns and may not be adequately summarized by the national estimates from Hedges and Hedberg's earlier work. This is not a criticism of the earlier estimates, but it is instead an added nuance and warning that national estimates may not fit local contexts. Yet, this variation seems to occur due to district structure, since many within-district intraclass correlations are consistent.

While many states produced similar within-district ICCs, the estimates from Florida in grades 6, 7, and 8 for both math and reading are far larger than the other states. One conjecture is that this is related to district size, since Florida has fewer but much larger districts than the other states. Small auxiliary analyses show that the natural log of the ratio of students to districts is a powerful predictor of our within-district measures, explaining from two thirds to four fifths of the variance among the state estimates in math and a quarter to four fifths in reading in grades 4 through 8. However, this is simply conjecture with such a small sample of estimates, and we further explore this in another manuscript.

Comparing our two-level estimates to Hedges and Hedberg (2007), we find that the local context differs substantively from the national context, with the assumption that these parameters are relatively stable over time. Comparing the average of the state

estimates with the earlier published national tables, we find that our state estimates are smaller for the elementary grades. For example our the average grade 3 math result from our states is 0.180, compared to 0.241 nationally, and 0.175 vs. 0.232 for grade 4, 0.187 vs. 0.216 in grade 5, and 0.200 vs. 0.264 for grade 6. We find a similar pattern in reading: 0.156 here vs. 0.271 nationally for grade 3, 0.170 vs. 0.242 for grade 4, 0.164 vs. 0.263 for grade 5, and 0.170 vs. 0.260 for grade 6. This pattern is reversed for the secondary grades 7 and 8, where our average state results are larger than those published based on national sources for both reading and math. However, the differences are not as large.

We also find differences with our estimates of R^2 parameters associated with using a pretest. In particular, our two-level models produced higher R^2 statistics for math at both level 1 and level 2. At level 1, these differences are 0.58 here vs. 0.49 nationally for grade 4, 0.61 vs. 0.51 for grade 5, and 0.62 vs. 0.50 in grade 6. At level 2, we also find greater values: 0.74 vs. 0.67 in grade 4, 0.76 vs. 0.63 in grade 5. Similar to the values of ICCs, however, we found similar values in grade 8 at level 1: 0.68 here vs. 0.65 nationally (grade 7 was not available from a national survey) and grades 6 and 7 for level 2: 0.73 vs. 0.74 and 0.88 vs. 0.82, respectively.

R^2 values in reading, in contrast, were similar to those found in national surveys for most grades, with the largest difference coming from the grade 6 level 1 estimate of 0.57 vs. 0.51. This points to the importance of the local environment in parameters for mathematics intervention evaluation design, whereas the local context appears to be less important in reading studies. Also of interest is that in our local estimates, reading R^2

values were lower, on average, than math R^2 values at level 1, whereas the opposite is true at a national level.

This study is one of the first to estimate parameters in a systematic way using three level models. Since districts are key stakeholders in the process of designing evaluations, it is key to understand how the school-level parameters differ from the three level models compared to the two level models. Overall, when we examine the average of the state estimates for both reading and math, we find that the school level estimates from the three level models are generally 40 percent smaller than the estimates from three level models in the elementary grades. This difference in reduced to about a third reduction for the early secondary grades.

When we examine the value of the pre-test in our three level models, the R^2 values at level 1 in the three level models were generally consistent, but this was not true of the R^2 values at higher levels. Overall, R^2 values at level 1 increased with grades for both reading and math, averaging 0.58 in grade 4 to 0.68 in grade 8 for math, and 0.56 to 0.59 for reading. We also found that level-2 R^2s increased with grade level, but the level 2 R^2 values were not consistent across states as Kentucky had much lower values. Finally, values for the level 3 R^2s were high at the district level, but also less consistent with Arkansas and Kentucky showing lower values.

Conclusions

We have presented empirical evidence about design parameters useful in planning two, three, and four level cluster randomized experiments using academic achievement as outcomes. All estimates are presented along with standard errors that provide some sense

of the sampling error inherent in these estimates, which tends to be rather small. We have illustrated the use of the design parameters in planning cluster randomized studies.

The intraclass correlation values reported in this tabulation differ somewhat from the national values reported by Hedges and Hedberg (2007). The values in the lower grades are generally consistent with those reported in Tables 2 and 3, albeit with variation from state to state. However, for higher grades, many of the intraclass correlation values reported in Tables 4 and 5 are larger than those in Hedges and Hedberg (2007). More over the decreasing trend in intraclass correlations with grade level that they found is not evident in Tables 4 and 5. There are many potential explanations of these differences. For example, they might reflect the combination of states in the national representative samples used by Hedges and Hedberg (2007) or they might reflect the fact that state assessments used here are better aligned with instruction, but these are just speculations

While the evidence reported in this paper is based on data from state longitudinal data systems that essentially correspond to censuses in seven states, it has some limitations. It is data from only seven states and in grades 3 to 11, but covers grades 3 to 8 in most states. Obviously not all grades are covered in every state. While states are smaller than the nation, and therefore state estimates should be more relevant than national estimates like those of Hedges and Hedberg (2007), there are heterogeneities even within single states. There is, however, a conundrum in seeking estimates from smaller areas that better represent the region in which the sample will be drawn. While such estimates may have less bias, they will also have greater variance, and at some point reduction in bias is more than compensated for by the increase in variance. The data is based on state assessments, and while these may be very relevant to many studies

because the state assessments are likely to be aligned with instruction, the evidence reported here would be less relevant to studies that will use achievement tests that are not aligned with instruction.

We are currently attempting to extend this database by adding evidence from additional states and additional years. As we do so, it is our intention to make these values available on a website SOMEWEBSITE so that the entire collection can be made available to researchers designing new studies. We also hope that states will consider making information about design parameters available routinely to assist researchers in planning evaluation studies in their states.

Appendix: Multilevel Models Defining Design Parameters

This appendix describes the specific multilevel models on which the computations of design parameters are based.

Two-Level Hierarchical Designs

No Covariates

Suppose that m clusters and there are n_i observations in the i^{th} cluster. Let Y_{ij} be the j^{th} observation in the i^{th} cluster. Then the level 1 (individual-level) model is

$$Y_{ij} = \beta_{0i} + \varepsilon_{ij} \quad , i = 1, \ldots, m; j = 1, \ldots, n_i,$$

where β_{0i} is the mean of the i^{th} cluster and the ε_{ij} are independently normally distributed with mean 0 and variance σ_1^2.

The level 2 (cluster-level) model is

$$\beta_{0i} = \gamma_0 + \eta_i, i = 1, ..., m,$$

where γ_0 is the grand mean, and the η_i are independently distributed with mean 0 and variance σ_2^2. The intraclass correlation coefficient ρ is defined in terms of the variances as $\rho = \sigma_2^2 / \left(\sigma_1^2 + \sigma_2^2\right)$.

With Covariates

If there are q covariates $X1, \ldots, Xq$ at level 1 and q covariate $W1, \ldots, Wq$ at level 2, the level 1 model becomes

$$Y_{ij} = \beta_{0i} + \beta_1 X_{ij} + \cdots + \beta_q Xq_{ij} + \varepsilon_{ij} \quad , i = 1, \ldots, m; j = 1, \ldots, n_i,$$

where β_{0i} is the covariate-adjusted mean of the i^{th} cluster, β_a is the (fixed) effect of the a^{th} individual-level covariate, Xa_{ij} is the value of the individual-level covariate Xa (centered

on cluster means), and the ε_{ij} are independently normally distributed with mean 0 and variance σ_{A1}^2.

The level 2 (cluster-level) model is

$$\beta_{0i} = \gamma_0 + \gamma_1 W_i + \cdots + \gamma_q W q_i + \eta_i,\ i = 1, ..., m,$$

where γ_0 is the covariate-adjusted grand mean, γ_1, is the effect of the a^{th} cluster-level covariate, Wk_i is the (grand mean centered) value of the cluster-level covariate Wa for cluster i, and the η_i are independently distributed with mean 0 and variance σ_{A2}^2.

In this model, the intraclass correlation ρ is still defined in terms of the unadjusted variances as given in the model with no covariates. In this model the covariate outcome correlations are defined in terms of the adjusted and unadjusted residual variances as $R_1^2 = 1 - \sigma_{A1}^2 / \sigma_1^2$ and $R_2^2 = 1 - \sigma_{A2}^2 / \sigma_2^2$.

Three-Level Hierarchical Designs

No Covariates

Suppose that there is a three stage cluster sampling design m clusters, so that there are p_i subclusters in the i^{th} cluster and the j^{th} subcluster has n_{ij} observations. Let Y_{ijk} be the k^{th} observation in j^{th} subcluster of the i^{th} cluster. Thus, the level 1 model is

$$Y_{ijk} = \beta_{0ij} + \varepsilon_{ijk},\ i = 1, \ldots, m; j = 1, \ldots, p_i; k = 1, \ldots, n_{ij},$$

where β_{0ij} is the mean of the j^{th} subcluster in the i^{th} cluster, and the ε_{ijk} are independently normally distributed with mean 0 and variance σ_1^2.

The level 2 (subcluster-level) model is

$$\beta_{0ij} = \gamma_{0i} + \eta_{ij},\ i = 1, ..., m, j = 1, ..., p_i,$$

where γ_0 is the mean of the i^{th} cluster and the η_{ij} are independently distributed with mean 0 and variance σ_2^2.

The level 3 (cluster-level) model is

$\gamma_{0i} = \pi_0 + \xi_i$, $i = 1, ..., m$,

where π_0 is the grand mean, and the ξ_i are independently normally distributed with mean 0 and variance σ_3^2. The level 2 intraclass correlation is defined in terms of the variances as $\rho_2 = \sigma_2^2 \Big/ \left(\sigma_1^2 + \sigma_2^2 + \sigma_3^2\right)$ and the level 3 intraclass correlation is $\rho_3 = \sigma_3^2 \Big/ \left(\sigma_1^2 + \sigma_2^2 + \sigma_3^2\right)$.

With Covariates

If there are q covariate $W1, ..., Wq$ at the cluster level, and q covariates $Z1, ..., Zq$ at the subcluster level, and q covariates $X1, ..., Xq$ at the individual level, the level 1 (individual-level) model is

$Y_{ijk} = \beta_{0ij} + \beta_1 X1_{ij} + \cdots + \beta_q Xq_{ij} + \varepsilon_{ijk}$, $i = 1, \ldots, m; j = 1, \ldots, p_i; k = 1, \ldots, n_{ij}$,

where β_{0ij} is the covariate-adjusted mean of the j^{th} subcluster in the i^{th} cluster, β_a is the effect of the a^{th} individual-level covariate (which is a fixed effect), X_{ijk} is the values of the individual-level covariate (centered on cluster means), and the ε_{ijk} are independently normally distributed with mean 0 and variance σ_{A1}^2.

The level 2 (subcluster-level) model is

$\beta_{0ij} = \gamma_{0i} + \gamma_1 Z1_{ij} + \cdots + \gamma_q Zq_{ij} + \eta_{ij}$, $i = 1, ..., m; j = 1, ..., p_i$,

where γ_{0i} is the covariate-adjusted mean of the i^{th} cluster, γ_a is the effect of the a^{th} level 2 covariate (which is a fixed effect), Za_{ij} is the values of the subcluster-level covariate Za (centered on subcluster means), and the η_{ij} are independently distributed with mean 0 and variance σ_{A2}^2.

The level 3 (cluster-level) model is

$$\gamma_{0i} = \pi_0 + \pi_1 W1_i + \cdots + \pi_q Wq_i + \xi_i,\ i = 1, ..., m,$$

where π_0 is the covariate-adjusted grand mean, π_1 is the effect of the level 3 covariate, Wa_i is the (grand mean centered) value of the a^{th} cluster-level covariate Wa, and the ξ_i are independently distributed with mean 0 and variance σ_{A3}^2.

In the model with covariates, the three intraclass correlations ρ_2 and ρ_3 (and their complement $\bar{\rho}$) are still defined in terms of the unadjusted variances as given in the model with no covariates. In this model the covariate outcome correlations are defined in terms of the adjusted and unadjusted variances as $R_1^2 = 1 - \sigma_{A1}^2 / \sigma_1^2$, $R_2^2 = 1 - \sigma_{A2}^2 / \sigma_2^2$, and $R_3^2 = 1 - \sigma_{A3}^2 / \sigma_3^2$.

References

Bloom, H. S. (1995). Minimum detectable effects: A simple way to report statistical power of experimental designs. *Evaluation Review*, *19*,547-556.

Bloom, H. S., L. Richburg-Hayes, et al. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis*, *29(1),* 30-59.

Bloom, H. S., Bos, J. M., & Lee, S. W. (1999). Using cluster random assignment to measure program impacts: statistical implications for the evaluation of educational programs. *Evaluation Review*, *23*, 445-469.

Borenstein, M., Hedges, L. V., & Rothstein, H. (2012). *CRT-Power*. Teaneck, NJ: Biostat, Inc,

Bush, G. W. (2001). No child left behind, US Department of Education Washington, DC.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences (2nd Edition)*. New York: Academic Press.

Fisher, R. A. (1925/1990). Statistical methods for research workers. Oxford: Oxford University Press.

Hedges, L. V. and E. C. Hedberg (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29(1),* 60-87.

Hedges, L. V., Hedberg, E.C., & Kuyper, A.. (2012). The variance of intraclass correlations in three-and four-level models. *Educational and Psychological Measurement 72(6)*, 893-909.

Konstantopoulos, S. (2008). The power of the test for treatment effects in three-level cluster randomized designs. *Journal of Research on Educational Effectiveness, 1,* 66-88.

Konstantopoulos, S. (2009). Incorporating cost in power analysis for three-level cluster randomized designs. *Evaluation Review, 33,* 335-357.

Konstantopoulos, S. (2010). Power analysis in two-level unbalanced designs. *Journal of Experimental Education, 78,* 291-317.

Konstantopoulos, S. (2011). Optimal sampling of units in three-level cluster randomized designs: An ANCOVA framework. *Educational and Psychological Measurement*, *71*, 798-813.

Mosteller, F. & Boruch, R. (Eds.) (2002). *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution Press.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized experiments. *Psychological Methods*, *2*, 173-185.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA: Sage Publications.

Schochet, P. Z. (2005). *Statistical power for random assignment evaluations of educational programs*. Princeton, NJ: Mathematica Policy Research.

Spybrook, J., Raudenbush, S. W., et al. (2012). Optimal design for longitudinal and multilevel research: Documentation for the *Optimal Design* software." Survey Research Center of the Institute of Social Research at University of Michigan.

Snijders, T. & Bosker, J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics*, *18*, 237-259.

Table 1
Sample Sizes by Subject, State, and Grade

	Estimation Sample					
	District Level		School Level		Student Level[a]	
	Math	*Reading*	*Math*	*Reading*	*Math*	*Reading*
Grade 1						
Arkansas	245	245	488	488	32,510	32,350
Grade 2						
Arkansas	245	245	489	489	32,381	32,347
Grade 3						
Arkansas	246	246	490	490	32,002	31,962
Arizona	193	193	919	919	64,768	64,769
Kentucky	174	174	721	721	43,439	43,439
Massachusetts	279	279	958	958	61,673	61,204
North Carolina	119	118	1,324	1,323	98,993	98,676
Wisconsin	414	414	1,046	1,046	51,340	51,165
Grade 4						
Arkansas	246	246	488	488	31,714	31,677
Arizona	195	195	920	920	64,581	64,569
Florida	73	73	1,842	1,841	144,368	144,375
Kentucky	174	174	723	723	44,171	44,171
Massachusetts	278	278	942	942	62,736	62,345
North Carolina	119	119	1,318	1,318	96,196	95,902
Wisconsin	415	415	1,043	1,043	52,016	51,920
Grade 5						
Arkansas	245	245	434	434	31,550	31,512
Arizona	192	192	913	913	64,508	64,509
Florida	73	73	1,841	1,841	147,967	147,967
Kentucky	174	174	716	716	44,279	44,279
Massachusetts	277	277	870	870	63,231	62,872
North Carolina	120	120	1,302	1,302	94,707	94,454
Wisconsin	415	415	999	999	51,628	51,535
Grade 6						
Arkansas	244	244	340	340	31,161	31,137
Arizona	191	191	744	744	63,364	63,361
Florida	74	74	1,109	1,106	145,611	145,668
Kentucky	174	174	409	409	44,475	44,475
Massachusetts	273	273	535	534	64,283	63,958
North Carolina	125	123	643	641	92,967	92,714
Wisconsin	416	416	632	632	51,778	51,712
Grade 7						
Arkansas	245	245	299	299	31,085	31,048
Arizona	189	189	538	537	64,349	64,351
Florida	74	74	1,002	1,008	151,181	151,335
Kentucky	174	174	326	326	43,743	43,743
Massachusetts	240	240	440	441	63,704	63,448
North Carolina	128	128	614	614	91,774	91,532
Wisconsin	416	416	563	563	52,523	52,467
Grade 8						
Arkansas	245	245	296	296	30,634	30,610
Arizona	190	190	539	539	65,044	65,059
Florida	74	74	1,024	1,030	148,099	148,284
Kentucky	174	174	323	323	43,926	43,926
Massachusetts	240	240	435	435	65,119	64,818
North Carolina	128	128	620	621	91,517	91,278
Wisconsin	416	416	565	565	53,263	53,208
Grade 9						
Arkansas	245	245	294	294	30,964	29,882
Florida	73	73	1,117	1,112	160,207	160,589
Grade 10						
Arizona	118	118	261	261	60,933	61,448
Florida	73	73	925	939	156,205	160,821
Kentucky		169		229		43,647
Massachusetts	258	258	341	341	65,648	65,747
Wisconsin	380	380	465	465	58,728	58,692
Grade 11						
Kentucky	169		230		40,770	

Source: Arkansas, Arizona, Florida, Kentucky, Massachusetts, North Carolina, and Wisconsin Departments of Education and the Common Core of Data (CCD). Notes: a: States varied in the detail provided for special needs students. When possible, we removed only students with cognitive disabilities. Typically, this removed 10 percent of the population. Higher rates reflect coarse disability data where details of the disability were not available. We also removed students who were members of a chater school

Table 2
Intraclass Correlations (ICCs) and R^2estimates for Mathematics Achievement by State: Two-level Models, Grades 1-6

	Unconditional		Pretest Covariate[a]				Demographic Covariates[b]			
	School Level		School Level		Student Level		School Level		Student Level	
	ρ_2	$SE(\rho_2)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$
Grade and State										
Grade 1										
Arkansas	0.175	0.010					0.527	0.031	0.077	0.003
Grade 2										
Arkansas	0.177	0.010	0.631	0.027	0.471	0.004	0.574	0.029	0.081	0.003
Grade 3										
Arkansas	0.164	0.010	0.509	0.032	0.451	0.004	0.467	0.033	0.076	0.003
Arizona	0.184	0.008					0.589	0.021	0.125	0.002
Kentucky	0.146	0.007					0.265	0.028	0.072	0.002
Massachusetts	0.243	0.009					0.647	0.018	0.089	0.002
North Carolina	0.162	0.006					0.643	0.016	0.130	0.002
Wisconsin	0.182	0.007					0.729	0.014	0.082	0.002
Average	**0.180**	**0.003**					**0.557**	**0.009**	**0.096**	**0.001**
Grade 4										
Arkansas	0.154	0.010	0.685	0.024	0.521	0.004	0.488	0.032	0.080	0.003
Arizona	0.188	0.008	0.770	0.013	0.578	0.003	0.644	0.019	0.110	0.002
Florida	0.165	0.005	0.832	0.007	0.568	0.002	0.670	0.013	0.079	0.001
Kentucky	0.153	0.008	0.472	0.027	0.540	0.003	0.326	0.029	0.083	0.003
Massachusetts	0.225	0.009	0.758	0.014	0.577	0.003	0.632	0.019	0.083	0.002
North Carolina	0.164	0.006	0.772	0.011	0.637	0.002	0.630	0.016	0.137	0.002
Wisconsin	0.174	0.007	0.873	0.007	0.608	0.003	0.752	0.013	0.084	0.002
Average	**0.175**	**0.003**	**0.737**	**0.006**	**0.576**	**0.001**	**0.592**	**0.008**	**0.094**	**0.001**
Grade 5										
Arkansas	0.159	0.010	0.701	0.024	0.577	0.004	0.502	0.034	0.085	0.003
Arizona	0.201	0.008	0.801	0.012	0.612	0.002	0.636	0.019	0.122	0.002
Florida	0.180	0.006	0.838	0.007	0.610	0.002	0.688	0.012	0.079	0.001
Kentucky	0.151	0.008	0.551	0.025	0.584	0.003	0.322	0.029	0.085	0.003
Massachusetts	0.242	0.009	0.796	0.012	0.649	0.002	0.699	0.017	0.093	0.002
North Carolina	0.178	0.006	0.795	0.010	0.656	0.002	0.611	0.017	0.132	0.002
Wisconsin	0.199	0.008	0.847	0.009	0.605	0.003	0.641	0.018	0.087	0.002
Average	**0.187**	**0.003**	**0.761**	**0.006**	**0.613**	**0.001**	**0.586**	**0.008**	**0.098**	**0.001**
Grade 6										
Arkansas	0.146	0.011	0.637	0.031	0.613	0.004	0.471	0.039	0.092	0.003
Arizona	0.202	0.009	0.736	0.017	0.639	0.002	0.573	0.024	0.116	0.002
Florida	0.295	0.012	0.897	0.006	0.634	0.002	0.807	0.010	0.094	0.001
Kentucky	0.124	0.009	0.394	0.038	0.582	0.003	0.277	0.038	0.098	0.003
Massachusetts	0.232	0.012	0.790	0.016	0.682	0.002	0.746	0.019	0.104	0.002
North Carolina	0.186	0.010	0.762	0.016	0.638	0.002	0.636	0.023	0.153	0.002
Wisconsin	0.212	0.011	0.880	0.009	0.648	0.003	0.744	0.018	0.093	0.002
Average	**0.200**	**0.004**	**0.728**	**0.008**	**0.634**	**0.001**	**0.608**	**0.010**	**0.107**	**0.001**

Source: Arkansas, Arizona, Florida, Kentucky, Massachusetts, North Carolina, and Wisconsin Departments of Education. a: Model includes the student pretest, the school mean of current students' pretest, and district mean of current students' pretest. b: Model includes gender, race, low SES indicator, English learner status, and school

Table 3
Intraclass Correlations (ICCs) and R^2estimates for Reading Achievement by State: Two-level Models, Grades 1-6

	Unconditional		Pretest Covariate[a]				Demographic Covariates[b]			
	School Level		School Level		Student Level		School Level		Student Level	
	ρ_2	$SE(\rho_2)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$
Grade and State										
Grade 1										
Arkansas	0.130	0.008					0.447	0.033	0.059	0.003
Grade 2										
Arkansas	0.138	0.009	0.714	0.022	0.467	0.004	0.666	0.025	0.080	0.003
Grade 3										
Arkansas	0.147	0.009	0.724	0.021	0.473	0.004	0.598	0.028	0.114	0.003
Arizona	0.183	0.008					0.741	0.015	0.170	0.003
Kentucky	0.102	0.006					0.416	0.028	0.074	0.002
Massachusetts	0.212	0.008					0.753	0.014	0.091	0.002
North Carolina	0.143	0.005					0.769	0.011	0.146	0.002
Wisconsin	0.147	0.006					0.774	0.012	0.092	0.002
Average	**0.156**	**0.003**					**0.675**	**0.008**	**0.115**	**0.001**
Grade 4										
Arkansas	0.143	0.009	0.772	0.018	0.565	0.004	0.579	0.029	0.117	0.003
Arizona	0.194	0.008	0.872	0.008	0.596	0.003	0.834	0.010	0.153	0.003
Florida	0.159	0.005	0.912	0.004	0.522	0.002	0.809	0.008	0.078	0.001
Kentucky	0.108	0.006	0.584	0.024	0.480	0.003	0.423	0.028	0.083	0.003
Massachusetts	0.279	0.010	0.819	0.011	0.512	0.003	0.707	0.016	0.106	0.002
North Carolina	0.146	0.005	0.920	0.004	0.610	0.002	0.794	0.010	0.161	0.002
Wisconsin	0.162	0.007	0.897	0.006	0.662	0.002	0.811	0.011	0.098	0.002
Average	**0.170**	**0.003**	**0.825**	**0.005**	**0.564**	**0.001**	**0.708**	**0.007**	**0.114**	**0.001**
Grade 5										
Arkansas	0.137	0.009	0.805	0.017	0.604	0.004	0.664	0.026	0.128	0.004
Arizona	0.187	0.008	0.893	0.007	0.592	0.003	0.788	0.012	0.174	0.003
Florida	0.147	0.005	0.941	0.003	0.559	0.002	0.811	0.008	0.074	0.001
Kentucky	0.110	0.006	0.605	0.023	0.482	0.003	0.404	0.028	0.092	0.003
Massachusetts	0.246	0.010	0.867	0.008	0.550	0.003	0.766	0.014	0.113	0.002
North Carolina	0.151	0.006	0.917	0.004	0.591	0.002	0.784	0.011	0.151	0.002
Wisconsin	0.167	0.007	0.899	0.006	0.673	0.002	0.793	0.012	0.106	0.003
Average	**0.164**	**0.003**	**0.847**	**0.004**	**0.579**	**0.001**	**0.716**	**0.007**	**0.120**	**0.001**
Grade 6										
Arkansas	0.121	0.010	0.779	0.021	0.578	0.004	0.570	0.035	0.139	0.004
Arizona	0.180	0.009	0.870	0.009	0.567	0.003	0.774	0.015	0.170	0.003
Florida	0.230	0.011	0.951	0.003	0.538	0.002	0.871	0.007	0.080	0.001
Kentucky	0.081	0.006	0.569	0.032	0.491	0.003	0.441	0.037	0.114	0.003
Massachusetts	0.245	0.012	0.875	0.010	0.565	0.003	0.821	0.014	0.127	0.002
North Carolina	0.147	0.008	0.889	0.008	0.608	0.002	0.790	0.015	0.174	0.002
Wisconsin	0.186	0.010	0.928	0.005	0.635	0.003	0.859	0.010	0.120	0.003
Average	**0.170**	**0.004**	**0.837**	**0.006**	**0.569**	**0.001**	**0.732**	**0.008**	**0.132**	**0.001**

Source: Arkansas, Arizona, Florida, Kentucky, Massachusetts, North Carolina, and Wisconsin Departments of Education. a: Model includes the student pretest, the school mean of current students' pretest, and district mean of current students' pretest. b: Model includes gender, race, low SES indicator, English learner status, and school

Table 4
Intraclass Correlations (ICCs) and R^2estimates for Mathematics Achievement by State: Two-level Models, Grades 7-11

	Unconditional		Pretest Covariate[a]				Demographic Covariates[b]			
	School Level		School Level		Student Level		School Level		Student Level	
	ρ_2	$SE(\rho_2)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$
Grade and State										
Grade 7										
Arkansas	0.160	0.012	0.216	0.042	0.591	0.004	0.554	0.038	0.092	0.003
Arizona	0.202	0.011	0.842	0.013	0.666	0.002	0.703	0.021	0.131	0.002
Florida	0.346	0.014	0.939	0.004	0.586	0.002	0.815	0.011	0.082	0.001
Kentucky	0.121	0.009	0.590	0.035	0.624	0.003	0.392	0.042	0.099	0.003
Massachusetts	0.266	0.014	0.913	0.008	0.722	0.002	0.833	0.015	0.101	0.002
North Carolina	0.209	0.011	0.879	0.009	0.675	0.002	0.661	0.022	0.132	0.002
Wisconsin	0.238	0.012	0.921	0.006	0.685	0.002	0.775	0.017	0.102	0.003
Average	**0.220**	**0.005**	**0.757**	**0.008**	**0.650**	**0.001**	**0.676**	**0.010**	**0.106**	**0.001**
Grade 8										
Arkansas	0.134	0.011	0.826	0.018	0.687	0.003	0.591	0.037	0.116	0.003
Arizona	0.202	0.011	0.865	0.011	0.657	0.002	0.607	0.026	0.118	0.002
Florida	0.384	0.013	0.968	0.002	0.658	0.001	0.828	0.010	0.101	0.001
Kentucky	0.119	0.009	0.752	0.024	0.665	0.003	0.381	0.043	0.093	0.003
Massachusetts	0.259	0.014	0.931	0.006	0.746	0.002	0.800	0.017	0.097	0.002
North Carolina	0.264	0.013	0.884	0.009	0.661	0.002	0.645	0.023	0.134	0.002
Wisconsin	0.199	0.011	0.901	0.008	0.704	0.002	0.741	0.019	0.099	0.002
Average	**0.223**	**0.004**	**0.875**	**0.005**	**0.683**	**0.001**	**0.656**	**0.010**	**0.108**	**0.001**
Grade 9										
Arkansas	0.120	0.010	0.869	0.014	0.610	0.004	0.649	0.033	0.096	0.003
Florida	0.405	0.013	0.912	0.005	0.630	0.001	0.710	0.015	0.100	0.001
Grade 10										
Arizona	0.299	0.020					0.666	0.034	0.100	0.002
Florida	0.424	0.014	0.876	0.008	0.586	0.002	0.670	0.018	0.102	0.001
Massachusetts	0.279	0.017					0.836	0.016	0.097	0.002
Wisconsin	0.194	0.012	0.871	0.011	0.637	0.003	0.820	0.015	0.115	0.002
Grade 11										
Kentucky	0.081	0.008					0.406	0.050	0.069	0.002

Source: Arkansas, Arizona, Florida, Kentucky, Massachusetts, North Carolina, and Wisconsin Departments of Education. a: Model includes the student pretest, the school mean of current students' pretest, and district mean of current students' pretest. b: Model includes gender, race, low SES indicator, English learner status, and school and district means of these covariates.

Table 5
Intraclass Correlations (ICCs) and R^2 estimates for Reading Achievement by State: Two-level Models, Grades 7-1

	Unconditional Model		Pretest Covariate[a]				Demographic Covariates[b]			
	School Level		School Level		Student Level		School Level		Student Level	
	ρ_2	$SE(\rho_2)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$
Grade and State										
Grade 7										
Arkansas	0.118	0.010	0.415	0.044	0.597	0.004	0.675	0.031	0.131	0.004
Arizona	0.177	0.010	0.911	0.007	0.589	0.003	0.805	0.015	0.178	0.003
Florida	0.238	0.011	0.976	0.001	0.549	0.002	0.887	0.007	0.079	0.001
Kentucky	0.087	0.007	0.709	0.027	0.521	0.003	0.466	0.040	0.113	0.003
Massachusetts	0.313	0.016	0.918	0.007	0.588	0.003	0.826	0.015	0.148	0.003
North Carolina	0.172	0.010	0.951	0.004	0.622	0.002	0.830	0.013	0.158	0.002
Wisconsin	0.191	0.011	0.949	0.004	0.634	0.003	0.895	0.008	0.111	0.003
Average	**0.185**	**0.004**	**0.833**	**0.008**	**0.586**	**0.001**	**0.769**	**0.008**	**0.131**	**0.001**
Grade 8										
Arkansas	0.114	0.010	0.846	0.016	0.586	0.004	0.618	0.035	0.142	0.004
Arizona	0.167	0.010	0.893	0.009	0.582	0.003	0.760	0.018	0.183	0.003
Florida	0.330	0.013	0.975	0.002	0.580	0.002	0.848	0.009	0.095	0.001
Kentucky	0.088	0.007	0.766	0.023	0.531	0.003	0.423	0.042	0.115	0.003
Massachusetts	0.269	0.014	0.940	0.006	0.594	0.003	0.854	0.013	0.140	0.003
North Carolina	0.193	0.010	0.960	0.003	0.620	0.002	0.825	0.013	0.176	0.002
Wisconsin	0.180	0.010	0.933	0.005	0.621	0.003	0.831	0.013	0.121	0.003
Average	**0.192**	**0.004**	**0.902**	**0.004**	**0.588**	**0.001**	**0.737**	**0.009**	**0.139**	**0.001**
Grade 9										
Arkansas	0.100	0.009	0.772	0.023	0.461	0.004	0.696	0.030	0.111	0.003
Florida	0.319	0.012	0.936	0.004	0.560	0.002	0.761	0.012	0.109	0.001
Grade 10										
Arizona	0.262	0.019					0.779	0.024	0.166	0.003
Florida	0.335	0.013	0.953	0.003	0.549	0.002	0.723	0.015	0.113	0.001
Kentucky	0.058	0.006					0.508	0.046	0.104	0.003
Massachusetts	0.328	0.018					0.853	0.015	0.153	0.003
Wisconsin	0.193	0.012	0.900	0.009	0.588	0.003	0.845	0.013	0.116	0.002

Source: Arkansas, Arizona, Florida, Kentucky, Massachusetts, North Carolina, and Wisconsin Departments of Education. a: Model includes the student pretest, the school mean of current students' pretest, and district mean of current students' pretest. b: Model includes gender, race, low SES indicator, English learner status, and school and district means of these covariates.

Table 6
Intraclass Correlations (ICCs) and R^2estimates for Mathematics Achievement by States: Three-level Models, Grades 1-6

	Unconditional Model				Pretest Covariate[a]						Demographic Covariates[b]					
	District Level		School Level		District Level		School Level		Student Level		District Level		School Level		Student Level	
	ρ_3	$SE(\rho_3)$	ρ_2	$SE(\rho_2)$	R^2_3	$SE(R^2_3)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$	R^2_3	$SE(R^2_3)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$
Grade and State																
Grade 1																
Arkansas	0.039	0.009	0.125	0.010							0.235	0.047	0.569	0.029	0.077	0.003
Grade 2																
Arkansas	0.053	0.011	0.114	0.010	0.536	0.043	0.595	0.028	0.471	0.004	0.395	0.049	0.597	0.028	0.081	0.003
Grade 3																
Arkansas	0.041	0.010	0.115	0.010	0.420	0.048	0.491	0.032	0.451	0.004	0.396	0.048	0.465	0.033	0.075	0.003
Arizona	0.109	0.016	0.091	0.005							0.711	0.035	0.472	0.024	0.125	0.002
Kentucky	0.021	0.005	0.109	0.007							c	d	0.281	0.028	0.072	0.002
Massachusetts	0.075	0.009	0.123	0.007							0.815	0.020	0.389	0.025	0.089	0.002
North Carolina	0.045	0.009	0.126	0.005							0.661	0.050	0.658	0.015	0.130	0.002
Wisconsin	0.039	0.005	0.082	0.005							0.714	0.024	0.490	0.022	0.083	0.002
Average	**0.055**	**0.004**	**0.108**	**0.003**							**0.659**	**0.017**	**0.459**	**0.010**	**0.096**	**0.001**
Grade 4																
Arkansas	0.033	0.008	0.114	0.009	0.802	0.023	0.633	0.026	0.521	0.004	0.276	0.048	0.524	0.031	0.080	0.003
Arizona	0.117	0.017	0.088	0.005	0.953	0.007	0.579	0.021	0.578	0.003	0.759	0.030	0.479	0.024	0.110	0.002
Florida	0.026	0.007	0.143	0.005	0.934	0.015	0.819	0.008	0.568	0.002	0.387	0.089	0.709	0.011	0.079	0.001
Kentucky	0.016	0.005	0.122	0.007	0.522	0.052	0.409	0.028	0.540	0.003	c	d	0.350	0.029	0.083	0.003
Massachusetts	0.083	0.009	0.098	0.006	0.920	0.009	0.497	0.023	0.577	0.003	0.776	0.024	0.318	0.025	0.083	0.002
North Carolina	0.049	0.009	0.122	0.005	0.905	0.017	0.722	0.013	0.637	0.002	0.623	0.055	0.630	0.016	0.137	0.002
Wisconsin	0.038	0.005	0.076	0.005	0.935	0.006	0.729	0.014	0.608	0.003	0.581	0.031	0.588	0.020	0.084	0.002
Average	**0.052**	**0.004**	**0.109**	**0.002**	**0.853**	**0.009**	**0.627**	**0.008**	**0.576**	**0.001**	**0.567**	**0.021**	**0.514**	**0.009**	**0.094**	**0.001**
Grade 5																
Arkansas	0.046	0.011	0.108	0.010	0.521	0.044	0.717	0.023	0.577	0.004	0.272	0.049	0.588	0.030	0.085	0.003
Arizona	0.122	0.017	0.099	0.006	0.968	0.005	0.646	0.019	0.612	0.002	0.753	0.031	0.479	0.024	0.122	0.002
Florida	0.022	0.007	0.159	0.006	0.986	0.003	0.819	0.008	0.610	0.002	0.297	0.090	0.713	0.011	0.079	0.001
Kentucky	0.020	0.005	0.117	0.007	0.617	0.046	0.500	0.026	0.584	0.003	0.091	0.042	0.315	0.029	0.085	0.003
Massachusetts	0.098	0.011	0.100	0.006	0.935	0.008	0.570	0.022	0.649	0.002	0.837	0.018	0.425	0.025	0.093	0.002
North Carolina	0.052	0.010	0.131	0.005	0.924	0.013	0.752	0.012	0.656	0.002	0.682	0.048	0.627	0.016	0.132	0.002
Wisconsin	0.056	0.007	0.092	0.006	0.869	0.012	0.732	0.015	0.605	0.003	0.459	0.036	0.520	0.022	0.087	0.002
Average	**0.059**	**0.004**	**0.115**	**0.003**	**0.831**	**0.010**	**0.677**	**0.007**	**0.613**	**0.001**	**0.484**	**0.019**	**0.524**	**0.009**	**0.098**	**0.001**
Grade 6																
Arkansas	0.035	0.012	0.107	0.014	0.727	0.030	0.609	0.033	0.613	0.004	0.395	0.049	0.516	0.038	0.092	0.003
Arizona	0.084	0.014	0.112	0.007	0.824	0.023	0.651	0.021	0.639	0.002	0.607	0.044	0.495	0.026	0.116	0.002
Florida	0.040	0.012	0.246	0.012	0.863	0.030	0.908	0.005	0.634	0.002	0.674	0.062	0.830	0.009	0.094	0.001
Kentucky	0.027	0.007	0.087	0.008	0.454	0.056	0.365	0.038	0.582	0.003	0.071	0.038	0.269	0.037	0.098	0.003
Massachusetts	0.087	0.011	0.101	0.008	0.894	0.012	0.599	0.027	0.682	0.002	0.898	0.012	0.538	0.029	0.104	0.002
North Carolina	0.035	0.009	0.140	0.009	0.828	0.028	0.743	0.018	0.638	0.002	0.727	0.042	0.665	0.022	0.153	0.002
Wisconsin	0.042	0.007	0.089	0.007	0.788	0.018	0.788	0.015	0.648	0.003	0.471	0.036	0.582	0.025	0.093	0.002
Average	**0.050**	**0.004**	**0.126**	**0.004**	**0.768**	**0.012**	**0.666**	**0.009**	**0.634**	**0.001**	**0.549**	**0.016**	**0.556**	**0.011**	**0.107**	**0.001**

Source: Arkansas, Arizona, Florida, Kentucky, Massachusetts, North Carolina, and Wisconsin Departments of Education. a: Model includes the student pretest, the school mean of current students' pretest, and district mean of current students' pretest. b: Model includes gender, race, low SES indicator, English learner status, and school and district means of these covariates. c: Model produced a negative R2 and was truncated to 0. d: Standard error not computed.

Table 7
Intraclass Correlations (ICCs) and R^2estimates for Reading Achievement by State: Three-level Models, Grades 1-6

	Unconditional Model				Pretest Covariate[a]						Demographic Covariates[b]					
	District Level		School Level		District Level		School Level		Student Level		District Level		School Level		Student Level	
	ρ_3	$SE(\rho_3)$	ρ_2	$SE(\rho_2)$	R^2_3	$SE(R^2_3)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$	R^2_3	$SE(R^2_3)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$
Grade and State																
Grade 1																
Arkansas	0.026	0.007	0.100	0.008							c	d	0.593	0.028	0.059	0.003
Grade 2																
Arkansas	0.034	0.008	0.098	0.008	0.466	0.047	0.753	0.019	0.467	0.004	0.493	0.045	0.720	0.021	0.080	0.003
Grade 3																
Arkansas	0.034	0.008	0.107	0.009	0.698	0.032	0.716	0.022	0.473	0.004	0.715	0.031	0.547	0.030	0.114	0.003
Arizona	0.109	0.015	0.083	0.005							0.769	0.029	0.671	0.018	0.170	0.003
Kentucky	0.016	0.004	0.076	0.005							c	d	0.466	0.027	0.074	0.002
Massachusetts	0.071	0.008	0.087	0.005							0.874	0.014	0.473	0.023	0.091	0.002
North Carolina	0.037	0.008	0.114	0.005							0.691	0.047	0.806	0.010	0.146	0.002
Wisconsin	0.029	0.004	0.069	0.004							0.715	0.024	0.597	0.019	0.092	0.002
Average	**0.049**	**0.004**	**0.089**	**0.002**							**0.753**	**0.014**	**0.593**	**0.009**	**0.115**	**0.001**
Grade 4																
Arkansas	0.026	0.007	0.112	0.009	0.763	0.026	0.763	0.019	0.565	0.004	0.263	0.048	0.651	0.025	0.117	0.003
Arizona	0.127	0.017	0.081	0.005	0.989	0.002	0.721	0.016	0.596	0.003	0.861	0.018	0.741	0.015	0.153	0.003
Florida	0.026	0.007	0.138	0.005	0.945	0.013	0.909	0.004	0.522	0.002	0.782	0.045	0.820	0.008	0.078	0.001
Kentucky	0.013	0.003	0.084	0.005	0.692	0.039	0.509	0.026	0.480	0.003	c	d	0.455	0.027	0.083	0.003
Massachusetts	0.107	0.011	0.113	0.006	0.930	0.008	0.596	0.020	0.512	0.003	0.848	0.017	0.370	0.025	0.106	0.002
North Carolina	0.042	0.008	0.111	0.005	0.965	0.006	0.906	0.005	0.610	0.002	0.712	0.045	0.822	0.009	0.161	0.002
Wisconsin	0.032	0.004	0.074	0.004	0.940	0.006	0.790	0.012	0.662	0.002	0.739	0.022	0.660	0.017	0.098	0.002
Average	**0.053**	**0.003**	**0.102**	**0.002**	**0.889**	**0.007**	**0.742**	**0.006**	**0.564**	**0.001**	**0.701**	**0.014**	**0.646**	**0.007**	**0.114**	**0.001**
Grade 5																
Arkansas	0.033	0.009	0.099	0.009	0.756	0.027	0.811	0.016	0.604	0.004	0.399	0.049	0.741	0.021	0.128	0.004
Arizona	0.132	0.017	0.079	0.005	0.985	0.002	0.774	0.013	0.592	0.003	0.841	0.021	0.675	0.018	0.174	0.003
Florida	0.026	0.007	0.125	0.004	0.979	0.005	0.934	0.003	0.559	0.002	0.703	0.058	0.823	0.007	0.074	0.001
Kentucky	0.021	0.005	0.081	0.005	0.614	0.046	0.551	0.025	0.482	0.003	0.144	0.049	0.419	0.028	0.092	0.003
Massachusetts	0.093	0.010	0.096	0.006	0.975	0.003	0.663	0.019	0.550	0.003	0.907	0.011	0.473	0.025	0.113	0.002
North Carolina	0.048	0.009	0.111	0.005	0.960	0.007	0.904	0.005	0.591	0.002	0.726	0.043	0.815	0.009	0.151	0.002
Wisconsin	0.032	0.005	0.080	0.005	0.898	0.010	0.821	0.010	0.673	0.002	0.718	0.023	0.665	0.017	0.106	0.003
Average	**0.055**	**0.004**	**0.096**	**0.002**	**0.881**	**0.008**	**0.780**	**0.006**	**0.579**	**0.001**	**0.634**	**0.015**	**0.659**	**0.007**	**0.120**	**0.001**
Grade 6																
Arkansas	0.027	0.009	0.091	0.011	0.776	0.025	0.779	0.021	0.578	0.004	0.163	0.043	0.765	0.022	0.139	0.004
Arizona	0.089	0.014	0.086	0.006	0.943	0.008	0.796	0.013	0.567	0.003	0.756	0.031	0.702	0.018	0.170	0.003
Florida	0.027	0.008	0.193	0.010	0.910	0.020	0.955	0.003	0.538	0.002	0.737	0.053	0.874	0.007	0.081	0.001
Kentucky	0.013	0.004	0.062	0.006	0.594	0.047	0.561	0.033	0.491	0.003	0.232	0.056	0.453	0.036	0.114	0.003
Massachusetts	0.090	0.011	0.096	0.008	0.932	0.008	0.741	0.019	0.565	0.003	0.941	0.007	0.652	0.024	0.128	0.002
North Carolina	0.041	0.010	0.101	0.007	0.913	0.015	0.888	0.008	0.608	0.002	0.662	0.050	0.846	0.011	0.174	0.002
Wisconsin	0.028	0.004	0.066	0.005	0.873	0.012	0.839	0.012	0.635	0.003	0.801	0.017	0.648	0.023	0.120	0.003
Average	**0.045**	**0.003**	**0.099**	**0.003**	**0.849**	**0.009**	**0.794**	**0.007**	**0.569**	**0.001**	**0.613**	**0.015**	**0.706**	**0.008**	**0.132**	**0.001**

Source: Arkansas, Arizona, Florida, Kentucky, Massachusetts, North Carolina, and Wisconsin Departments of Education. a: Model includes the student pretest, the school mean of current students' pretest, and district mean of current students' pretest. b: Model includes gender, race, low SES indicator, English learner status, and school and district means of these covariates. c: Model produced a negative R2 and was truncated to 0. d: Standard error not computed.

Table 8
Intraclass Correlations (ICCs) and R^2estimates for Mathematics Achievement by State: Three-level Models, Grades 7-11

	Unconditional Model				Pretest Covariate[a]						Demographic Covariates[b]					
	District Level		School Level		District Level		School Level		Student Level		District Level		School Level		Student Level	
	ρ_3	$SE(\rho_3)$	ρ_2	$SE(\rho_2)$	R^2_3	$SE(R^2_3)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$	R^2_3	$SE(R^2_3)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$
Grade and State																
Grade 7																
Arkansas	0.039	0.016	0.117	0.017	c	d	0.535	0.039	0.591	0.004	0.436	0.048	0.597	0.036	0.092	0.003
Arizona	0.078	0.015	0.116	0.009	0.976	0.003	0.734	0.020	0.666	0.002	0.806	0.025	0.654	0.024	0.131	0.002
Florida	0.004	0.005	0.341	0.014	0.751	0.050	0.948	0.003	0.586	0.002	c	d	0.844	0.009	0.082	0.001
Kentucky	0.020	0.007	0.093	0.009	0.416	0.057	0.598	0.034	0.624	0.003	0.148	0.050	0.397	0.042	0.099	0.003
Massachusetts	0.098	0.013	0.114	0.010	0.955	0.006	0.824	0.015	0.723	0.002	0.944	0.007	0.706	0.024	0.101	0.002
North Carolina	0.041	0.012	0.158	0.011	0.982	0.003	0.849	0.011	0.675	0.002	0.652	0.050	0.691	0.021	0.133	0.002
Wisconsin	0.043	0.007	0.098	0.009	0.882	0.011	0.837	0.013	0.685	0.002	0.322	0.038	0.683	0.022	0.103	0.003
Average	**0.046**	**0.004**	**0.148**	**0.004**	**0.827**	**0.013**	**0.761**	**0.009**	**0.650**	**0.001**	**0.551**	**0.016**	**0.653**	**0.010**	**0.106**	**0.001**
Grade 8																
Arkansas	0.033	0.014	0.098	0.015	0.802	0.023	0.824	0.019	0.687	0.003	0.154	0.042	0.741	0.026	0.116	0.003
Arizona	0.066	0.013	0.126	0.009	0.942	0.008	0.809	0.015	0.657	0.002	0.738	0.033	0.558	0.028	0.118	0.002
Florida	0.006	0.007	0.378	0.014	0.687	0.060	0.976	0.001	0.658	0.001	0.008	0.021	0.843	0.009	0.101	0.001
Kentucky	0.019	0.007	0.095	0.009	0.738	0.034	0.746	0.024	0.665	0.003	0.120	0.046	0.416	0.042	0.093	0.003
Massachusetts	0.105	0.013	0.104	0.010	0.970	0.004	0.843	0.014	0.746	0.002	0.899	0.012	0.633	0.028	0.097	0.002
North Carolina	0.055	0.016	0.199	0.013	0.973	0.005	0.860	0.010	0.661	0.002	0.617	0.053	0.698	0.020	0.134	0.002
Wisconsin	0.036	0.007	0.089	0.008	0.938	0.006	0.789	0.016	0.704	0.002	0.721	0.023	0.484	0.030	0.099	0.002
Average	**0.046**	**0.004**	**0.156**	**0.004**	**0.864**	**0.011**	**0.835**	**0.006**	**0.683**	**0.001**	**0.465**	**0.013**	**0.625**	**0.010**	**0.108**	**0.001**
Grade 9																
Arkansas	0.025	0.011	0.092	0.013	0.997	0.000	0.815	0.019	0.610	0.004	0.579	0.041	0.693	0.030	0.096	0.003
Florida	0.024	0.009	0.377	0.014	0.942	0.013	0.915	0.005	0.630	0.001	0.757	0.050	0.725	0.014	0.100	0.001
Grade 10																
Arizona	0.083	0.026	0.219	0.022							0.842	0.027	0.691	0.032	0.099	0.002
Florida	0.007	0.008	0.418	0.015	0.744	0.052	0.884	0.007	0.586	0.002	0.220	0.086	0.695	0.017	0.102	0.001
Massachusetts	0.097	0.016	0.142	0.016							0.985	0.002	0.684	0.028	0.097	0.002
Wisconsin	0.040	0.006	0.066	0.007	0.806	0.018	0.740	0.021	0.637	0.003	0.610	0.031	0.679	0.025	0.115	0.002
Grade 11																
Kentucky	0.001	0.004	0.080	0.009							c	d	0.524	0.045	0.069	0.002

Source: Arkansas, Arizona, Florida, Kentucky, Massachusetts, North Carolina, and Wisconsin Departments of Education. a: Model includes the student pretest, the school mean of current students' pretest, and district mean of current students' pretest. b: Model includes gender, race, low SES indicator, English learner status, and school and district means of these covariates. c: Model produced a negative R2 and was truncated to 0. d: Standard error not computed.

Table 9
Intraclass Correlations (ICCs) and R^2estimates for Reading Achievement by State: Three-level Models, Grades 7-10

	Unconditional Model				Pretest Covariate[a]						Demographic Covariates[b]					
	District Level		School Level		District Level		School Level		Student Level		District Level		School Level		Student Level	
	ρ_3	$SE(\rho_3)$	ρ_2	$SE(\rho_2)$	R^2_3	$SE(R^2_3)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$	R^2_3	$SE(R^2_3)$	R^2_2	$SE(R^2_2)$	R^2_1	$SE(R^2_1)$
Grade and State																
Grade 7																
Arkansas	0.029	0.011	0.085	0.012	c	d	0.806	0.02	0.597	0.004	0.43	0.048	0.776	0.023	0.131	0.004
Arizona	0.075	0.014	0.097	0.008	0.949	0.007	0.877	0.01	0.589	0.003	0.843	0.021	0.791	0.016	0.178	0.003
Florida	0.013	0.008	0.224	0.011	0.932	0.015	0.981	0.001	0.549	0.002	0.614	0.07	0.903	0.006	0.079	0.001
Kentucky	0.009	0.004	0.074	0.007	0.596	0.047	0.705	0.027	0.521	0.003	0.132	0.048	0.508	0.039	0.113	0.003
Massachusetts	0.116	0.015	0.13	0.012	0.977	0.003	0.811	0.016	0.588	0.003	0.968	0.004	0.648	0.027	0.149	0.003
North Carolina	0.049	0.012	0.114	0.008	0.947	0.009	0.952	0.004	0.622	0.002	0.762	0.037	0.859	0.011	0.158	0.002
Wisconsin	0.027	0.004	0.069	0.006	0.895	0.01	0.891	0.009	0.635	0.003	0.694	0.025	0.836	0.013	0.111	0.003
Average	**0.045**	**0.004**	**0.113**	**0.004**	**0.883**	**0.009**	**0.860**	**0.006**	**0.586**	**0.001**	**0.635**	**0.016**	**0.760**	**0.008**	**0.131**	**0.001**
Grade 8																
Arkansas	0.02	0.01	0.092	0.012	e	d	0.804	0.02	0.586	0.004	0.143	0.041	0.752	0.025	0.142	0.004
Arizona	0.061	0.012	0.099	0.008	0.935	0.009	0.857	0.011	0.582	0.003	0.836	0.022	0.72	0.02	0.183	0.003
Florida	0.013	0.008	0.316	0.013	0.955	0.01	0.976	0.001	0.58	0.002	0.548	0.078	0.869	0.008	0.094	0.001
Kentucky	0.013	0.005	0.072	0.007	0.635	0.044	0.778	0.022	0.531	0.003	0.06	0.035	0.469	0.04	0.115	0.003
Massachusetts	0.106	0.013	0.101	0.01	0.988	0.002	0.834	0.015	0.594	0.003	0.892	0.013	0.754	0.02	0.14	0.003
North Carolina	0.044	0.012	0.136	0.01	0.963	0.006	0.958	0.003	0.62	0.002	0.763	0.037	0.853	0.011	0.176	0.002
Wisconsin	0.026	0.005	0.078	0.007	0.916	0.008	0.869	0.01	0.621	0.003	0.763	0.02	0.679	0.022	0.122	0.003
Average	**0.040**	**0.004**	**0.128**	**0.004**	**0.899**	**######**	**0.868**	**0.005**	**0.588**	**0.001**	**0.572**	**0.015**	**0.728**	**0.009**	**0.139**	**0.001**
Grade 9																
Arkansas	0.02	0.011	0.079	0.012	e	d	0.682	0.031	0.461	0.004	0.216	0.047	0.842	0.017	0.111	0.003
Florida	0.02	0.009	0.3	0.013	0.979	0.005	0.936	0.004	0.56	0.002	0.731	0.054	0.775	0.012	0.109	0.001
Grade 10																
Arizona	0.067	0.024	0.197	0.022							0.949	0.009	0.799	0.022	0.166	0.003
Florida	0.001	0.004	0.333	0.013	0.812	0.04	0.966	0.002	0.549	0.002	c	d	0.714	0.016	0.113	0.001
Kentucky	0.003	0.003	0.055	0.006							c	d	0.533	0.045	0.104	0.003
Massachusetts	0.106	0.018	0.168	0.018							e	d	0.717	0.026	0.153	0.003
Wisconsin	0.034	0.005	0.065	0.006	0.887	0.011	0.755	0.02	0.588	0.003	0.797	0.019	0.648	0.026	0.116	0.002

Source: Arkansas, Arizona, Florida, Kentucky, Massachusetts, North Carolina, and Wisconsin Departments of Education. a: Model includes the student pretest, the school mean of current students' pretest, and district mean of current students' pretest. b: Model includes gender, race, low SES indicator, English learner status, and school and district means of these covariates. c: Model produced a negative R^2 and was truncated to 0. d: Standard error not computed. e: Model produced a greater than 1 R^2 and was truncated to 1.