Example 4: Cross-Classified Models for Students Nested within Primary and Secondary Schools (as estimated using restricted maximum likelihood in SAS MIXED and STATA MIXED, although note that the results in STATA do not match exactly)

Crossed random effects models (also known as cross-classified models) are useful in situations in which people belong to more than one cluster (but the kinds of clusters are not nested). A simulated data example is shown below from Hox (2012) chapter 7, in which kids are nested within primary schools AND within secondary schools, but primary and secondary schools are crossed with each other at level 2 (1000 kids, 30 secondary schools, 50 primary schools).

SAS Syntax and Output for Data Import, Manipulation, and Description:

```
* Define global variable for file location to be replaced in code below;
* \\Client\ precedes actual path when using UIowa Virtual Desktop;
%LET filesave=C:\Dropbox\19_PSQF7375_Clustered\PSQF7375_Clustered_Example4;
LIBNAME example "&filesave.";

* Open output directory to save results to;
ODS RTF FILE="&filesave.\SAS_Example4_Output.rtf" STYLE=HTMLBlue STARTPAGE=NO;

* Import example 4 SAS dataset;
DATA work.pupcross; SET example.pupcross;
     LABEL pupSES="Student SES"
           achiev="Student Achievement in Secondary"; RUN;

* Getting means per primary school and secondary school of kid variables;
PROC SORT DATA=work.pupcross; BY pschool pdenom; RUN;
PROC MEANS NOPRINT DATA=work.pupcross; BY pschool pdenom; VAR pupSES achiev;
     OUTPUT OUT=work.Primary MEAN(pupSES achiev)=pmSES pmAchiev; RUN;
PROC SORT DATA=work.pupcross; BY sschool sdenom; RUN;
PROC MEANS NOPRINT DATA=work.pupcross; BY sschool sdenom; VAR pupSES achiev;
     OUTPUT OUT=work.Secondary MEAN(pupSES achiev)=smSES smAchiev; RUN;

* Label new variables;
DATA work.Primary; SET work.Primary;
       LABEL pmSES=    "Primary School Mean Student SES"
             pmAchiev= "Primary School Mean Student Achievement";
     DROP _TYPE_ _FREQ_; RUN;
DATA work.Secondary; SET work.Secondary;
       LABEL smSES=    "Secondary School Mean Student SES"
            smAchiev= "Secondary School Mean Student Achievement";
     DROP _TYPE_ _FREQ_; RUN;

* Merge back into individual data;
PROC SORT DATA=work.pupcross; BY pschool; RUN;
DATA work.pupcross; MERGE work.pupcross work.Primary; BY pschool; RUN;
PROC SORT DATA=work.pupcross; BY sschool; RUN;
DATA work.pupcross; MERGE work.pupcross work.Secondary; BY sschool; RUN;

* Center predictors;
DATA work.pupcross; SET work.pupcross;
       pupSES4 = pupSES - 4; LABEL pupSES4= "Student SES (0=4)";
       pmSES4  = pmSES - 4;  LABEL pmSES4=  "Primary Mean Student SES (0=4)";
       smSES4  = smSES - 4;  LABEL smSES4=  "Secondary Mean Student SES (0=4)"; RUN;

TITLE "SAS Primary School Descriptives";
PROC MEANS NDEC=2 DATA=work.Primary;
     VAR pdenom pmSES pmAchiev; RUN;

TITLE "SAS Secondary School Descriptives";
PROC MEANS NDEC=2 DATA=work.Secondary;
     VAR sdenom smSES smAchiev; RUN;

TITLE "SAS Student Descriptives";
PROC MEANS NDEC=2 DATA=work.pupcross;
     VAR pupSES achiev; RUN;
```

SAS Primary School Descriptives

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PDENOM	primary school denominational?	50	0.60	0.49	0.00	1.00
pmSES	Primary Mean Student SES	50	4.10	0.28	3.47	4.73
pmAchiev	Primary Mean Student Achievement	50	6.36	0.45	5.28	7.55

SAS Secondary School Descriptives

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
SDENOM	secondary school denominational?	33	0.67	0.48	0.00	1.00
smSES	Secondary Mean Student SES	33	4.14	0.34	3.47	5.00
smAchiev	Secondary Mean Student Achievement	33	6.32	0.32	5.54	6.91

SAS Student Descriptives

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PUPSES	Student SES	1000	4.10	1.40	1.00	6.00
ACHIEV	Student Achievement in Secondary	1000	6.34	0.87	3.90	9.90

STATA Syntax for Data Import, Manipulation, and Description:

```
// Define global variable for file location to be replaced in code below
// \\Client\ precedes actual path when using UIowa Virtual Desktop
global filesave "C:\Dropbox\19_PSQF7375_Clustered\PSQF7375_Clustered_Example4"
// Import example stata data file
use "$filesave\pupcross.dta", clear

// Save results to separate file
log using $filesave\PSQF7375_Clustered_Example4_STATA_Output.log, replace

// Get means per primary school and secondary school of kid predictor
egen pmSES    = mean(pupses), by (pschool)
egen pmAchiev = mean(achiev), by (pschool)
label variable pmSES    "Primary Mean Student SES"
label variable pmAchiev "Primary Mean Student Achievement"
egen smSES    = mean(pupses), by (sschool)
egen smAchiev = mean(achiev), by (sschool)
label variable smSES    "Secondary Mean Student SES"
label variable smAchiev "Secondary Mean Student Achievement"

// Center and label predictors
gen  pupSES4 = pupses - 4
gen  pmSES4  = pmSES - 4
gen  smSES4  = smSES - 4
label variable pupSES4 "Student SES (0=4)"
label variable pmSES4  "Primary Mean Student SES (0=4)"
label variable smSES4  "Secondary Mean Student SES (0=4)"

display as result "STATA Primary School Descriptives"
preserve  // Save for later use, then compute school-level dataset
collapse  pdenom pmSES pmAchiev, by(pschool)
format    pdenom pmSES pmAchiev %4.2f
summarize pdenom pmSES pmAchiev, format
restore   // Go back to student-level dataset

display as result "STATA Secondary School Descriptives"
preserve  // Save for later use, then compute school-level dataset
collapse  sdenom smSES smAchiev, by(sschool)
format    sdenom smSES smAchiev %4.2f
summarize sdenom smSES smAchiev, format
restore   // Go back to student-level dataset

display as result "STATA Student Descriptives"
format    pupSES achiev %4.2f
summarize pupSES achiev, format
```

Syntax and SAS Output for Empty Means Models 1a and 1b

We can start with a basic model in which we assume that academic achievement for child k who went to primary school p and secondary school s can be modeled by this equation: $\text{Achiev}_{kps} = \gamma_{000} + U_{00s} + e_{kps}$

in which achievement in 9th grade (i.e., in secondary school) is expected to be correlated among kids from the same secondary school (i.e., a random intercept at level 2 for secondary school).

```
TITLE "SAS Empty Means Model 1a: Secondary Random Intercept Only";
PROC MIXED DATA=work.pupcross COVTEST NOCLPRINT IC NAMELEN=100 METHOD=REML;
       CLASS pupil pschool sschool;
       MODEL achiev =    / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT / SUBJECT=sschool TYPE=UN; * Level 2 variance for secondary school;
       ODS OUTPUT InfoCrit=FitNested; * Save output for LRT;
 RUN; TITLE;

display as result "STATA Empty Means Model 1a: Secondary Random Intercept Only"
mixed achiev ,  ///
        || sschool: ,   variance reml covariance(un) ///
           dfmethod(satterthwaite) dftable(pvalue)
       estimates store FitNested  // Save for LRT
```

```
            Dimensions
Covariance Parameters             2
Columns in X                      1
Columns in Z Per Subject          1
Subjects                         30 # of secondary schools
Max Obs Per Subject              48 # kids per secondary school

                  Covariance Parameter Estimates
                                    Standard         Z
Cov Parm     Subject    Estimate       Error     Value        Pr Z
UN(1,1)      SSCHOOL     0.07206     0.02449      2.94      0.0016  Secondary Random Intercept Variance
Residual                  0.6833     0.03102     22.03      <.0001  Residual Pupil (Student) Variance

  Null Model Likelihood Ratio Test
    DF    Chi-Square      Pr > ChiSq
     1         53.70          <.0001
```

This is the LRT for the random intercept variance across secondary schools.

```
                            Information Criteria
Neg2LogLike     Parms       AIC       AICC       HQIC        BIC       CAIC
     2504.7         2    2508.7     2508.7     2509.6     2511.5     2513.5

                   Solution for Fixed Effects
                             Standard
Effect        Estimate          Error       DF    t Value    Pr > |t|
Intercept       6.3386        0.05583       29     113.53      <.0001  Grand Mean of Secondary Mean Achievement
```

The two-level ICC = 0.07206 / (0.07206 + 0.6883) = .0948, which is the correlation of students from the same secondary school, assuming that students are otherwise independent. However, because primary schools may have lasting effects, it might make sense to also allow a random intercept for primary school that is crossed at level 2 with the random intercept for secondary school: $\text{Achiev}_{kps} = \gamma_{000} + U_{0p0} + U_{00s} + e_{kps}$

```
TITLE "SAS Empty Means Model 1b: Primary by Secondary School Random Intercepts Crossed";
PROC MIXED DATA=work.pupcross COVTEST NOCLPRINT IC NAMELEN=100 METHOD=REML;
       CLASS pupil pschool sschool;
       MODEL achiev =    / SOLUTION DDFM=Satterthwaite OUTPM=PredEmpty;
       RANDOM INTERCEPT / SUBJECT=sschool TYPE=UN; * Level 2 variance for secondary;
       RANDOM INTERCEPT / SUBJECT=pschool TYPE=UN; * Level 2 variance for primary;
       ODS OUTPUT InfoCrit=FitCrossed CovParms=CovEmpty; * Save output for LRT and pseudo-R2;
 RUN; TITLE;
```

```
display as result "STATA Empty Means Model 1b: Primary by Secondary Crossed Random Intercepts"
mixed achiev ,  ///
        || _all: R.sschool ,     ///
        || _all: R.pschool ,  variance reml  ///
           dfmethod(satterthwaite) dftable(pvalue)
      estimates store FitCrossed  // Save for LRT
      lrtest FitCrossed FitNested  // Request LRT
```

STATA assumes that random effect levels are nested, so the _all R. is needed to override that default. The downside to this specification is that only a random intercept can be estimated per dimension because covariance_(unstructured) is not allowed with R.

```
           Dimensions
Covariance Parameters             3
Columns in X                      1
Columns in Z Per Subject         80
Subjects                          1
Max Obs Per Subject            1000
```

Notice that SAS thinks we have 1 subject with 1000 observations—that's ok. What it refers to is how many cases share the exact same **V** matrix, which never occurs here.

```
                 Covariance Parameter Estimates
                                   Standard         Z
Cov Parm     Subject    Estimate      Error     Value        Pr Z
UN(1,1)      SSCHOOL     0.06666    0.02190      3.04      0.0012  Secondary Random Intercept Variance
UN(1,1)      PSCHOOL      0.1719    0.04018      4.28      <.0001  Primary Random Intercept Variance
Residual                  0.5131    0.02390     21.47      <.0001  Residual Pupil Variance

                           Information Criteria
Neg2LogLike    Parms        AIC       AICC       HQIC        BIC       CAIC
     2321.1        3     2327.1     2327.1     2321.1     2321.1     2324.1

                 Solution for Fixed Effects
                             Standard
Effect        Estimate        Error       DF     t Value     Pr > |t|
Intercept       6.3486      0.07890       66       80.46       <.0001
```

```
* Calculate difference in model fit relative to nested model 1a;
  %FitTest(FitFewer=FitNested, FitMore=FitCrossed);
```

Do we need both random intercept variances?

```
Likelihood Ratio Test for FitNested vs. FitCrossed
              Neg2Log
   Name         Like      Parms        AIC        BIC     DevDiff     DFdiff     Pvalue
FitNested     2504.7          2     2508.7     2511.5           .          .          .
FitCrossed    2321.1          3     2327.1     2321.1     183.570          1          0
```

Rather than reporting traditional ICCs as in two-level nested models, it can be more intuitive to simply give the proportions of total variance attributable to each source:

Of the total variation of 0.75166 (from summing all three orthogonal variances):

0.06666 / 0.75166 = .089 reflects mean achievement differences between secondary schools

0.1719 / 0.75166 = .229 reflects mean achievement differences between primary schools

0.5131 / 0.75166 = .683 reflects achievement differences between kids with same schooling

95% random effect confidence interval for the intercept across each type of school:
Fixed effect ± 1.96*SQRT(random variance)

Secondary: 6.3486 ± 1.96*SQRT(0.06666) = 5.84 to 6.85

→ 95% of secondary schools are predicted to have school mean achievement from 5.84 to 6.85

Primary: 6.3486 ± 1.96*SQRT(0.1719) = 5.54 to 7.16

→ 95% of primary schools are predicted to have school mean achievement from 5.54 to 7.16

Syntax and SAS Output for Model 2: Adding primary school and secondary school denomination

$$\text{Achiev}_{kps} = \gamma_{000} + \gamma_{010}(\text{PrimDenom}_p) + \gamma_{001}(\text{SecDenom}_s) + U_{0p0} + U_{00s} + e_{kps}$$

```
TITLE "SAS Model 2: Add School Denomination Variables";
PROC MIXED DATA=work.pupcross COVTEST NOCLPRINT IC NAMELEN=100 METHOD=REML;
      CLASS pupil pschool sschool;
      MODEL achiev = pdenom sdenom / SOLUTION DDFM=Satterthwaite OUTPM=PredDenom;
      RANDOM INTERCEPT / SUBJECT=sschool TYPE=UN; * Level 2 variance for secondary;
      RANDOM INTERCEPT / SUBJECT=pschool TYPE=UN; * Level 2 variance for primary;
      ESTIMATE "Joint Test of Denomination" pdenom 1 sdenom 1;
      ODS OUTPUT CovParms=CovDenom; * Save output for pseudo-R2;
 RUN; TITLE;

display as result "STATA Model 2: Add School Denomination Variables"
mixed achiev pdenom sdenom,        ///
        || _all: R.sschool ,       ///
        || _all: R.pschool ,    variance reml   ///
           dfmethod(satterthwaite) dftable(pvalue)
      predict preddenom,          // save fixed-effect predicted outcomes
      corr achiev preddenom
      display as result r(rho)^2 // total R2
```

```
                Covariance Parameter Estimates
                                Standard         Z
Cov Parm     Subject    Estimate       Error     Value      Pr > Z
UN(1,1)      SSCHOOL     0.06017     0.02044      2.94      0.0016 leftover secondary intercept variance
UN(1,1)      PSCHOOL      0.1679     0.03976      4.22      <.0001 leftover primary intercept variance
Residual                  0.5129     0.02388     21.47      <.0001 residual pupil variance

                          Information Criteria
Neg2LogLike     Parms       AIC        AICC       HQIC        BIC       CAIC
     2320.3         3    2326.3      2326.4     2320.3     2320.3     2323.3

                 Solution for Fixed Effects
                            Standard
Effect       Estimate        Error       DF     t Value    Pr > |t|
Intercept      6.1115       0.1259     79.1       48.55     <.0001
PDENOM         0.1868       0.1276     48.2        1.46     0.1495 Effect of Denomination of Primary
SDENOM         0.1899      0.09853     46.5        1.93     0.0601 Effect of Denomination of Secondary
```

```
* Calculate PseudoR2 relative to empty means model 1b;
  %PseudoR2(NCov=3, CovFewer=CovEmpty, CovMore=CovDenom);
```

Which predictor caused the reduction in each random intercept variance?

PseudoR2 (% Reduction) for CovEmpty vs. CovDenom

Name	CovParm	Subject	Estimate	StdErr	ZValue	ProbZ	PseudoR2
CovEmpty	UN(1,1)	SSCHOOL	0.06666	0.02190	3.04	0.0012	.
CovEmpty	UN(1,1)	PSCHOOL	0.1719	0.04018	4.28	<.0001	.
CovEmpty	Residual		0.5131	0.02390	21.47	<.0001	.
CovDenom	UN(1,1)	SSCHOOL	0.06017	0.02044	2.94	0.0016	0.097296
CovDenom	UN(1,1)	PSCHOOL	0.1679	0.03976	4.22	<.0001	0.023550
CovDenom	Residual		0.5129	0.02388	21.47	<.0001	0.000525

```
* Calculate TotalR2 relative to empty means model 1b;
  %TotalR2(DV=achiev, PredFewer=PredEmpty, PredMore=PredDenom);
```

Total R2 (% Reduction) for PredEmpty vs. PredDenom

Name	Pred Corr	TotalR2	Total R2Diff
PredEmpty	0.00000	0.000000	.
PredDenom	0.14388	0.020701	0.020701

The total-R^2 is almost significant, as given by the multivariate Wald test for the two fixed effects.

```
                          Contrasts
                              Num      Den
Label                          DF       DF     F Value     Pr > F
Joint Test of Denomination      2     47.4        2.97     0.0609
```

Syntax and SAS Output for Model 3a: Adding fixed effect of student SES

$$\text{Achiev}_{kps} = \gamma_{000} + \gamma_{010}(\text{PrimDenom}_p) + \gamma_{001}(\text{SecDenom}_s) + \gamma_{100}(\text{pupSES}_{kps} - 4) + U_{0p0} + U_{00s} + e_{kps}$$

```
TITLE "SAS Model 3a: Add Student SES";
PROC MIXED DATA=work.pupcross COVTEST NOCLPRINT IC NAMELEN=100 METHOD=REML;
       CLASS pupil pschool sschool;
       MODEL achiev = pdenom sdenom pupSES4 / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT / SUBJECT=sschool TYPE=UN; * Level 2 variance for secondary;
       RANDOM INTERCEPT / SUBJECT=pschool TYPE=UN; * Level 2 variance for primary;
       ODS OUTPUT CovParms=CovPup1; * Save output for pseudo-R2;
 RUN; TITLE;

display as result "STATA Model 3a: Add Student SES"
mixed achiev c.pdenom c.sdenom c.pupSES4,  ///
        || _all: R.sschool ,     ///
        || _all: R.pschool ,   variance reml  ///
            dfmethod(satterthwaite) dftable(pvalue)
```

```
                 Covariance Parameter Estimates
                                  Standard         Z
Cov Parm     Subject    Estimate     Error     Value      Pr > Z
UN(1,1)      SSCHOOL     0.05710   0.01951      2.93      0.0017 leftover secondary intercept variance
UN(1,1)      PSCHOOL      0.1686   0.03966      4.25      <.0001 leftover primary intercept variance
Residual                  0.4915   0.02290     21.46      <.0001 leftover residual pupil variance

                          Information Criteria
Neg2LogLike    Parms        AIC       AICC       HQIC        BIC       CAIC
     2285.1        3     2291.1     2291.1     2285.1     2285.1     2288.1

                  Solution for Fixed Effects
                             Standard
Effect        Estimate        Error      DF    t Value    Pr > |t|
Intercept       6.1095       0.1246    78.9      49.02      <.0001
PDENOM          0.1890       0.1274    48.2       1.48      0.1446
SDENOM          0.1745      0.09620      46       1.81      0.0763
pupSES4         0.1066      0.01634     943       6.52      <.0001
```

What are we assuming in fitting this student-level SES effect by itself?

```
* Calculate PseudoR2 relative to denom model 2;
  %PseudoR2(NCov=3, CovFewer=CovDenom, CovMore=CovPup1);
```

PseudoR2 (% Reduction) for CovDenom vs. CovPup1

Name	CovParm	Subject	Estimate	StdErr	ZValue	ProbZ	PseudoR2
CovDenom	UN(1,1)	SSCHOOL	0.06017	0.02044	2.94	0.0016	.
CovDenom	UN(1,1)	PSCHOOL	0.1679	0.03976	4.22	<.0001	.
CovDenom	Residual		0.5129	0.02388	21.47	<.0001	.
CovPup1	UN(1,1)	SSCHOOL	0.05710	0.01951	2.93	0.0017	0.051054
CovPup1	UN(1,1)	PSCHOOL	0.1686	0.03966	4.25	<.0001	-0.004480
CovPup1	Residual		0.4915	0.02290	21.46	<.0001	0.041643

Syntax and SAS Output for Model 3b: Adding student SES contextual fixed effects

$$\text{Achiev}_{kps} = \gamma_{000} + \gamma_{010}(\text{PrimDenom}_p) + \gamma_{020}(\text{pmSES}_p - 4) + \gamma_{001}(\text{SecDenom}_s) + \gamma_{002}(\text{smSES}_s - 4) + \gamma_{100}(\text{pupSES}_{kps} - 4) + U_{0p0} + U_{00s} + e_{kps}$$

```
TITLE "SAS Model 3b: Add Pupil SES Contextual Effects";
PROC MIXED DATA=work.pupcross COVTEST NOCLPRINT IC NAMELEN=100 METHOD=REML;
      CLASS pupil pschool sschool;
      MODEL achiev = pdenom sdenom pupSES4 pmSES4 smSES4
                   / SOLUTION DDFM=Satterthwaite OUTPM=PredSES;
      RANDOM INTERCEPT / SUBJECT=sschool TYPE=UN; * Level 2 variance for secondary;
      RANDOM INTERCEPT / SUBJECT=pschool TYPE=UN; * Level 2 variance for primary;
      CONTRAST "Joint Test of SES" pupSES4 1, pmSES4 1, smSES4 1;
      ODS OUTPUT InfoCrit=FitFixSES CovParms=CovPup2; * Save output for LRT and pseudo-R2;
RUN; TITLE;
```

```
display as result "STATA Model 3b: Add Student SES Contextual Effects"
mixed achiev c.pdenom c.sdenom c.pupSES4 c.pmSES4 c.smSES4,  ///
        || _all: R.sschool ,     ///
        || _all: R.pschool ,   variance reml  ///
           dfmethod(satterthwaite) dftable(pvalue)
      test (c.pupSES4=0)(c.pmSES4=0)(c.smSES4=0), small // Multiv Wald test for SES
      predict predSES,           // save fixed-effect predicted outcomes
      corr achiev predSES
      display as result r(rho)^2 // total R2
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
UN(1,1)	SSCHOOL	0.05864	0.02072	2.83	0.0023
UN(1,1)	PSCHOOL	0.1737	0.04114	4.22	<.0001
Residual		0.4897	0.02284	21.44	<.0001

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
2284.5	3	2290.5	2290.5	2284.5	2284.5	2287.5

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	6.1216	0.1286	74.5	47.60	<.0001
PDENOM	0.1848	0.1292	47.2	1.43	0.1590
SDENOM	0.1214	0.1009	33.8	1.20	0.2372
pupSES4	0.1034	0.01646	920	6.28	<.0001
pmSES4	-0.03780	0.2283	48.2	-0.17	0.8692
smSES4	0.2942	0.1555	68.2	1.89	0.0626

What do the new SES effects represent?

```
* Calculate PseudoR2 relative to smushed model 3a;
  %PseudoR2(NCov=3, CovFewer=CovPup1, CovMore=CovPup2);
```

PseudoR2 (% Reduction) for CovPup1 vs. CovPup2

Name	CovParm	Subject	Estimate	StdErr	ZValue	ProbZ	PseudoR2
CovPup1	UN(1,1)	SSCHOOL	0.05710	0.01951	2.93	0.0017	.
CovPup1	UN(1,1)	PSCHOOL	0.1686	0.03966	4.25	<.0001	.
CovPup1	Residual		0.4915	0.02290	21.46	<.0001	.
CovPup2	UN(1,1)	SSCHOOL	0.05864	0.02072	2.83	0.0023	-0.027008
CovPup2	UN(1,1)	PSCHOOL	0.1737	0.04114	4.22	<.0001	-0.030163
CovPup2	Residual		0.4897	0.02284	21.44	<.0001	0.003648

```
* Calculate PseudoR2 relative to denom model 2;
  %PseudoR2(NCov=3, CovFewer=CovDenom, CovMore=CovPup2);
```

PseudoR2 (% Reduction) for CovDenom vs. CovPup2

Name	CovParm	Subject	Estimate	StdErr	ZValue	ProbZ	PseudoR2
CovDenom	UN(1,1)	SSCHOOL	0.06017	0.02044	2.94	0.0016	.
CovDenom	UN(1,1)	PSCHOOL	0.1679	0.03976	4.22	<.0001	.
CovDenom	Residual		0.5129	0.02388	21.47	<.0001	.
CovPup2	UN(1,1)	SSCHOOL	0.05864	0.02072	2.83	0.0023	0.025425
CovPup2	UN(1,1)	PSCHOOL	0.1737	0.04114	4.22	<.0001	-0.034777
CovPup2	Residual		0.4897	0.02284	21.44	<.0001	0.045139

```
* Calculate TotalR2 relative to denom model 2;
  %TotalR2(DV=achiev, PredFewer=PredDenom, PredMore=PredSES);
```

Total R2 (% Reduction) for PredDenom vs. PredSES

Name	Pred Corr	TotalR2	Total R2Diff
PredDenom	0.14388	0.020701	.
PredSES	0.23526	0.055346	0.034645

The change in total-R^2 from the three new SES fixed effects is significant, as given by the multivariate Wald test.

Contrasts

Label	Num DF	Den DF	F Value	Pr > F
Joint Test of SES	3	81.4	15.43	<.0001

Syntax and SAS Output for Model 3c: Adding random effect of student SES across secondary schools:

$$\text{Achiev}_{kps} = \gamma_{000} + \gamma_{010}(\text{PrimDenom}_p) + \gamma_{020}(\text{pmSES}_p - 4) + \gamma_{001}(\text{SecDenom}_s) + \gamma_{002}(\text{smSES}_s - 4)$$
$$+ \gamma_{100}(\text{pupSES}_{kps} - 4) + U_{0p0} + U_{00s} + U_{10s}(\text{pupSES}_{kps} - 4) + e_{kps}$$

```
TITLE "SAS Model 3c: Add Random Pupil SES across Secondary Schools";
PROC MIXED DATA=work.pupcross COVTEST NOCLPRINT IC NAMELEN=100 METHOD=REML;
       CLASS pupil pschool sschool;
       MODEL achiev = pdenom sdenom pupSES4 pmSES4 smSES4 / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT pupSES4 / SUBJECT=sschool TYPE=UN; * Level 2 variance for secondary;
       RANDOM INTERCEPT         / SUBJECT=pschool TYPE=UN; * Level 2 variance for primary;
       ODS OUTPUT InfoCrit=FitRandSESsec; * Save output for LRT;
 RUN; TITLE;
// STATA models with random slopes are not estimable correctly
// because covariance(un) is not allowed
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	Your turn to label these!
UN(1,1)	SSCHOOL	0.05583	0.01951	2.86	0.0021	Leftover sec int var
UN(2,1)	SSCHOOL	0.009256	0.005898	1.57	0.1166	cov sec int and sec slope
UN(2,2)	SSCHOOL	0.004633	0.003324	1.39	0.0817	sec random slope of pupSES
UN(1,1)	PSCHOOL	0.1710	0.04058	4.22	<.0001	
Residual		0.4833	0.02283	21.16	<.0001	

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
2278.6	5	2288.6	2288.6	2278.6	2278.6	2283.6

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	6.1049	0.1269	74.8	48.10	<.0001
PDENOM	0.1960	0.1283	47.3	1.53	0.1333
SDENOM	0.1597	0.09182	42	1.74	0.0893
pupSES4	0.1055	0.02073	27.7	5.09	<.0001
pmSES4	-0.03258	0.2267	48.2	-0.14	0.8863
smSES4	0.1213	0.1916	28.4	0.63	0.5317

```
* Calculate difference in model fit relative to fixed SES model 3b;
  %FitTest(FitFewer=FitFixSES, FitMore=FitRandSESsec);
```

The %FitTest macro provides the original *p*-value—see the excel sheet for the mixture *p*-value instead.

```
Likelihood Ratio Test for FitFixSES vs. FitRandSESsec
                  Neg2Log
    Name           Like      Parms       AIC        BIC     DevDiff     DFdiff      Pvalue
FitFixSES         2284.5       3       2290.5     2284.5       .           .          .
FitRandSESsec     2278.6       5       2288.6     2278.6    5.89779        2        0.052398
```

Do we need the random pupil SES slope over secondary schools?
What kind of effects would explain that variance?

Syntax and SAS Output for Model 3d: Adding random effect of student SES across primary schools:

$$\text{Achiev}_{kps} = \gamma_{000} + \gamma_{010}(\text{PrimDenom}_p) + \gamma_{020}(\text{pmSES}_p - 4) + \gamma_{001}(\text{SecDenom}_s) + \gamma_{002}(\text{smSES}_s - 4)$$
$$+ \gamma_{100}(\text{pupSES}_{kps} - 4) + U_{0p0} + U_{1p0}(\text{pupSES}_{kps} - 4) + U_{00s} + U_{10s}(\text{pupSES}_{kps} - 4) + e_{kps}$$

```
TITLE "SAS Model 3d: Add Random Pupil SES across Primary Schools";
PROC MIXED DATA=work.pupcross COVTEST NOCLPRINT IC NAMELEN=100 METHOD=REML;
       CLASS pupil pschool sschool;
       MODEL achiev = pdenom sdenom pupSES4 pmSES4 smSES4 / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT pupSES4 / SUBJECT=sschool TYPE=UN; * Level 2 variance for secondary;
       RANDOM INTERCEPT pupSES4 / SUBJECT=pschool TYPE=UN; * Level 2 variance for primary;
       ODS OUTPUT InfoCrit=FitRandSESprim; * Save output for LRT;
 RUN; TITLE;
```

```
                    Covariance Parameter Estimates
                                          Standard          Z
Cov Parm      Subject     Estimate         Error        Value         Pr Z   Your turn to label these!
UN(1,1)       SSCHOOL      0.05355       0.01884         2.84       0.0022
UN(2,1)       SSCHOOL     0.008731      0.005637         1.55       0.1214
UN(2,2)       SSCHOOL     0.004225      0.003229         1.31       0.0954
UN(1,1)       PSCHOOL       0.1615       0.03846         4.20       <.0001
UN(2,1)       PSCHOOL      0.01810      0.009498         1.91       0.0567
UN(2,2)       PSCHOOL     0.009637      0.004384         2.20       0.0140
Residual                    0.4656       0.02247        20.72       <.0001

                             Information Criteria
Neg2LogLike     Parms        AIC        AICC       HQIC        BIC       CAIC
     2263.8         7     2277.8      2277.9     2263.8     2263.8     2270.8

                     Solution for Fixed Effects
                             Standard
Effect         Estimate         Error       DF     t Value     Pr > |t|
Intercept        6.1126        0.1219     74.2       50.15       <.0001
PDENOM           0.1973        0.1209     46.5        1.63       0.1094
SDENOM           0.1414       0.09045       42        1.56       0.1255
pupSES4          0.1069       0.02469     35.7        4.33       0.0001
pmSES4         -0.01314        0.2148     48.3       -0.06       0.9515
smSES4           0.1562        0.1889     28.6        0.83       0.4153
```

```
* Calculate difference in model fit relative to random SES model 3c;
  %FitTest(FitFewer=FitRandSESsec, FitMore=FitRandSESprim);
```

```
Likelihood Ratio Test for FitRandSESsec vs. FitRandSESprim
                  Neg2Log
     Name          Like      Parms       AIC        BIC     DevDiff     DFdiff          Pvalue
FitRandSESsec     2278.6       5       2288.6     2278.6       .           .          .
FitRandSESprim    2263.8       7       2277.8     2263.8    14.7603        2        .000623499
```

Do we need the random pupil SES slope over primary schools?
What kind of effects would explain that variance?

95% random effect confidence interval for student SES slope across each type of school:
Fixed effect ± 1.96*SQRT(random variance)

Secondary Student SES Slope: 0.1069 ± 1.96*SQRT(0.004225) = −0.021 to 0.234
→ 95% of secondary schools are predicted to have student SES slopes from −0.021 to 0.234
Primary Student SES Slope: 0.1069 ± 1.96*SQRT(0.009637) = −0.086 to 0.299
→ 95% of primary schools are predicted to have student SES slopes from −0.086 to 0.299

Sample Results Section using SAS output (without "smushed" model) [indicates notes about what to change]

The extent to which student socioeconomic status (SES) could predict a student achievement outcomes in secondary school was examined in a series of multilevel models with crossed random effects (i.e., for student cross-classification). Specifically, the 1,000 students at level 1 were modeled as nested within 30 secondary schools at level 2, as well as nested within 50 primary schools, in which secondary and primary schools were crossed sampling dimensions at level 2. Residual maximum likelihood (REML) within SAS [or STATA] MIXED was used in estimating and reporting all model parameters. The significance of fixed effects was evaluated with Wald tests using Satterthwaite denominator degrees of freedom, whereas random effects were evaluated via likelihood ratio tests (i.e., −2ΔLL with degrees of freedom equal to the number of new random effects variances and covariances). Alpha was chosen as .05. Model-implied fixed effects were requested via ESTIMATE [or LINCOM] statements. Effect size was evaluated via pseduo-R^2 values for the proportion reduction in each variance component, as well as with total R^2, the squared correlation between the actual math outcomes and the math outcomes predicted by the fixed effects.

We first examined the extent of dependency due to mean differences for each type of school by including a random intercept variance for each. Relative to a model assuming independent students (i.e., with only a single model residual), adding a random intercept variance across secondary schools significantly improved model fit, −2ΔLL(1) = 53.70, $p < .001$. Adding a second random intercept variance across primary schools also significantly improved model fit, −2ΔLL(1) = 183.57, $p < .001$, providing empirical support for the need to model the cross-classification of students within primary and secondary schools. Of the total variation in student achievement, 0.089% reflected mean differences between secondary schools, 22.9% reflected mean differences between primary schools, and 68.3% reflected reamining between-student differences after controlling for schooling effects. A 95% random effects confidence interval was calculated for each source of intercept variation as the fixed intercept ± 1.96* SQRT(random intercept variance), which revealed that 95% of the secondary schools were predicted to have intercepts for school mean achievement between 5.84 and 6.85, whereas 95% of the primary schools were predicted to have intercepts for school mean achievement between 5.54 and 7.16. We then added the effects for the denomination status (0 = no, 1 = yes) for the primary school and for the secondary school. Both indicated nonsignificantly greater achievement outcomes for denominational schools. Primary school denomination captured 2.36% of the primary school random intercept variance and secondary school denomination captured 9.73% of the secondary school random intercept variance, resulting in a total-R^2 = 2.07%, which was not significant, $F(2,47) = 2.97$, $p = .061$. However, both denomination predictors were retained in the model as control variables.

We then considered the effects of student SES. In order to ensure proper interpretation of the student-level SES fixed effect as the within-school effect, we also included two level-2 contextual SES effects: primary school mean SES, and secondary school mean SES. All three predictors were centered at 4 (near the middle of the obtained range). These three effects together significantly improved the total-R^2 by 3.46%, new total-R^2 = .055, $F(3,81) = 15.43$, $p < .001$. The within-school level-1 SES effect was significant and accounted for 4.51% of the level-1 residual variance. It indicated that for every one-unit higher student SES relative to others in their primary and secondary school, students are expected to have higher achievement by 0.103. The level-2 contextual SES effect for primary schools indicated that primary school achievement was nonsignificantly lower by 0.038 for every unit higher primary mean SES after controlling for student SES (which did not account for any primary school random intercept variance). Likewise, the level-2 contextual SES effect for secondary schools indicated that secondary school achievement was nonsignificantly higher by 0.294 for every unit higher secondary mean SES after controlling for student SES (which accounted for 2.54% of the secondary school random intercept variance). Finally, we considered the potential for random slopes for the student SES effect. Significant slope variation was found across secondary schools, −2ΔLL(mixture of df=1 and df=2) = 5.90, $p = .034$, as well as across primary shcools, −2ΔLL(2) = 14.76, $p < .001$, indicating that the size of the relative SES advantage differed significantly across each type of school. A 95% random effects confidence interval for the student SES effect, calculated as fixed slope ± 1.96*SQRT(random slope variance), revealed that 95% of the secondary schools were predicted to have student SES slopes ranging from −0.021 to 0.234, and that that 95% of the primary schools were predicted to have student SES slopes ranging from −0.086 to 0.299.