Example 5a: Crossed Random Effects Models for Trials nested within Subjects and within Items
(complete data, syntax, and output available for SAS, SPSS, and STATA electronically as SLPH 861 Ex. 6)

Source: Locker Jr., L., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling in the analysis of psycholinguistic data. *Behavior Research Methods, 39*(4), 723-730.

Response time data for a lexical decision task (decide as quickly as you can whether this is a word or a non-word) were collected for 39 items from 38 subjects (total possible observations = 1482; total actual observations = 1392 after removing inaccurate responses). Items are words that varied systematically in two characteristics: Semantic Frequency (low/high) and Neighborhood Size (small/large).

SAS Data Manipulation:

```
* SAS: Bringing in data from folder to work library;
* Adding another version of predictors to be coded 0/1 for low/high;
DATA work.example5a; SET filesave.Example5a;
       IF freq=-.5 THEN freq01=0; IF freq=.5 THEN freq01=1;
       IF size=-.5 THEN size01=0; IF size=.5 THEN size01=1;
RUN;
```

Model 1: Empty means baseline model with only residual variance → $RT_{tis} = \gamma_{000} + e_{tis}$
(default REPEATED statement if not included is TYPE=VC)

```
TITLE1 "1. Empty Means Model: No Random Intercepts (E only)";
PROC MIXED DATA=work.example5a COVTEST NOCLPRINT NAMELEN=100 IC METHOD=REML;
       CLASS Subject Item;
       MODEL rt = / SOLUTION DDFM=Satterthwaite;
       ODS OUTPUT InfoCrit=FitEonly; * Save fit for comparison;
RUN; TITLE1;
```

```
           Covaria nce Parameter Estimates
                      Standard         Z
Cov Parm     Estimate    Error      Value        Pr Z
Residual       21340     809.19     26.37      <.0001  All the variance in RT in one pile of e (TYPE=VC)

                         Information Criteria
Neg2LogLike    Parms       AIC        AICC       HQIC        BIC       CAIC
    17820.7        1    17822.7    17822.7    17824.7    17827.9    17828.9

                  Solution for Fixed Effects
                      Standard
Effect       Estimate    Error      DF     t Value    Pr > |t|
Intercept      632.38    3.9154    1391     161.51      <.0001  grand mean RT across all obs
```

Model 2: Is there significant mean RT variation across subjects? → $RT_{tis} = \gamma_{000} + U_{00s} + e_{tis}$

```
TITLE1 "2. Random Intercept for Subjects Model";
PROC MIXED DATA=work.example5a COVTEST NOCLPRINT NAMELEN=100 IC METHOD=REML;
      CLASS Subject Item;
      MODEL rt = / SOLUTION DDFM=Satterthwaite;
      RANDOM INTERCEPT /SUBJECT=Subject TYPE=UN;     * Level 2 variance for subjects;
      ODS OUTPUT InfoCrit=FitRandSubjects;           * Save fit for comparison;
RUN; TITLE1;
```

```
                  Covariance Parameter Estimates
                                    Standard         Z
Cov Parm     Subject    Estimate       Error     Value        Pr Z
UN(1,1)      subject     5167.07     1305.09      3.96      <.0001 Variance across SUBJECTS in mean RT
Residual                   16307      626.74     26.02      <.0001 Leftover trial-to-trial variance

  Null Model Likelihood Ratio Test
    DF    Chi-Square      Pr > ChiSq
     1        280.44          <.0001

                            Information Criteria
Neg2LogLike     Parms         AIC        AICC        HQIC         BIC        CAIC
    17540.3         2     17544.3     17544.3     17545.4     17547.5     17549.5

                   Solution for Fixed Effects
                         Standard
Effect        Estimate      Error      DF     t Value     Pr > |t|
Intercept       631.42    12.1540      37       51.95       <.0001 grand mean RT across all obs
```

If total RT variance = 21,474, then

5,167 / 21,474 = 24% is between subjects

16,307 / 21,474 = 76% is within subjects

Is there significant variation in mean RT across subjects—is that new 24% > 0%?

```
* Calculate difference in model fit relative to e-only model;
%FitTest(FitFewer=FitEonly, FitMore=FitRandSubjects);
```

```
Likelihood Ratio Test for FitEonly vs. FitRandSubjects
                   Neg2Log
Name                  Like      Parms        AIC        BIC     DevDiff     DFdiff     Pvalue
FitEonly           17820.7          1    17822.7    17827.9          .           .          .
FitRandSubjects    17540.3          2    17544.3    17547.5     280.439          1          0
```

Note that in this case, this LRT for the improvement in model fit appears elsewhere on the page!

Model 3: Is there significant mean RT variation across items? → $RT_{tis} = \gamma_{000} + U_{00s} + U_{0i0} + e_{tis}$

```
TITLE1 "3. Random Intercepts for Subjects and Items: Crossed Model";
PROC MIXED DATA=work.example5a COVTEST NOCLPRINT NAMELEN=100 IC METHOD=REML;
      CLASS Subject Item;
      MODEL rt =  / SOLUTION DDFM=Satterthwaite;
      RANDOM INTERCEPT / SUBJECT=Item TYPE=UN;            * Level 2 variance for items;
      RANDOM INTERCEPT / SUBJECT=Subject TYPE=UN;         * Level 2 variance for subjects;
      ODS OUTPUT InfoCrit=FitRandItems CovParms=CovEmpty; * Save fit, variances to compare;
RUN; TITLE1;
```

```
                  Covariance Parameter Estimates
                                    Standard         Z
Cov Parm     Subject    Estimate       Error     Value      Pr Z
UN(1,1)      item        2409.36      678.04      3.55    0.0002 Intercept Variance across ITEMS in mean RT
UN(1,1)      subject     5166.81     1292.78      4.00    <.0001 Intercept Variance across SUBJECTS in mean RT
Residual                   14344      559.99     25.61    <.0001 Leftover trial-to-trial residual variance

  Null Model Likelihood Ratio Test
    DF    Chi-Square      Pr > ChiSq
     2        380.84          <.0001

                            Information Criteria
Neg2LogLike     Parms        AIC       AICC       HQIC        BIC       CAIC
    17439.9         3    17445.9    17445.9    17439.9    17439.9    17442.9

                   Solution for Fixed Effects
                         Standard
Effect        Estimate      Error       DF    t Value    Pr > |t|
Intercept       635.33    14.4301     59.4      44.03      <.0001
```

If total variance now = 21,920, then…

5,167 / 21,920 = 24% is between subjects
2,409 / 21,920 = 11% is between items
14,344 / 21,920 = 65% is within subjects and items (subject x item interaction)

Is there significant variation in mean RT across items—is that new 11% > 0%?

```
* Calculate difference in model fit relative to random subjects model;
%FitTest(FitFewer=FitRandSubjects, FitMore=FitRandItems);
```

```
Likelihood Ratio Test for FitRandSubjects vs. FitRandItems
                   Neg2Log
     Name            Like      Parms        AIC        BIC    DevDiff    DFdiff    Pvalue
FitRandSubjects   17540.3        2      17544.3    17547.5       .          .         .
FitRandItems      17439.9        3      17445.9    17439.9    100.399       1         0
```

Calculate 95% random effects confidence intervals for each random intercept:
95% CI = fixed effect ± 1.96*SQRT(variance)

Subject Intercept CI = 635 ± 1.96*SQRT(5167) = 494 to 776
95% of the individual subject mean RTs are expected to fall between 494 and 776 ms

Item Intercept CI = 635 ± 1.96*SQRT(2409) = 539 to 732
95% of the individual item mean RTs are expected to fall between 539 and 732 ms

Model 4a: Are there significant fixed effects of the item predictors (continuous Frequency and Size)?

$\rightarrow RT_{tis} = \gamma_{000} + \gamma_{010}(Freq_i) + \gamma_{020}(Size_i) + \gamma_{030}(Freq_i)(Size_i) + U_{00s} + U_{0i0} + e_{tis}$

```
TITLE1 "4a. Random Subjects by Random Items Crossed Predictive Model";
TITLE2 "Freq01 and Size01 are not on CLASS statement, so are continuous";
PROC MIXED DATA=work.example5a COVTEST NOCLPRINT NAMELEN=100 IC METHOD=REML;
      CLASS Subject Item;
      * | operator estimates all possible main effects and interactions up to @ order;
      MODEL rt = freq01|size01@2 / SOLUTION DDFM=Satterthwaite OUTPM=ItemPred;
      RANDOM INTERCEPT / SUBJECT=Item TYPE=UN;      * Level 2 variance for items;
      RANDOM INTERCEPT / SUBJECT=Subject TYPE=UN;   * Level 2 variance for subjects;
      ODS OUTPUT InfoCrit=FitItem CovParms=CovItem; * Save fit, variances to compare;
* Getting cell means (traditional for Regression);
      ESTIMATE "RT for Low Freq, Small Size"  intercept 1 freq01 0 size01 0 freq01*size01 0;
      ESTIMATE "RT for Low Freq, Large Size"  intercept 1 freq01 0 size01 1 freq01*size01 0;
      ESTIMATE "RT for High Freq, Small Size" intercept 1 freq01 1 size01 0 freq01*size01 0;
      ESTIMATE "RT for High Freq, Large Size" intercept 1 freq01 1 size01 1 freq01*size01 1;
* Getting marginal means (traditional for ANOVA);
      ESTIMATE "RT for Low Freq"        intercept 1 freq01  0 size01 .5 freq01*size01  0;
      ESTIMATE "RT for High Freq"       intercept 1 freq01  1 size01 .5 freq01*size01 .5;
      ESTIMATE "RT for Small Size"      intercept 1 freq01 .5 size01  0 freq01*size01  0;
      ESTIMATE "RT for Large Size"      intercept 1 freq01 .5 size01  1 freq01*size01 .5;
      ESTIMATE "Grand Mean for All"     intercept 1 freq01 .5 size01 .5 freq01*size01 .25;
* Getting all possible simple effects (more useful);
      ESTIMATE "Simple Freq Effect for Small Size"   freq01 1 freq01*size01 0;
      ESTIMATE "Simple Freq Effect for Large Size"   freq01 1 freq01*size01 1;
      ESTIMATE "Simple Size Effect for Low Freq"     size01 1 freq01*size01 0;
      ESTIMATE "Simple Size Effect for High Freq"    size01 1 freq01*size01 1;
* Getting all possible marginal effects (traditional for ANOVA, less useful);
      ESTIMATE "Marginal Freq Effect"   freq01 1 freq01*size01 .5;
      ESTIMATE "Marginal Size Effect"   size01 1 freq01*size01 .5;
RUN; TITLE2;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z	
UN(1,1)	item	1692.07	526.60	3.21	0.0007	Intercept Variance across ITEMS in mean RT
UN(1,1)	subject	5168.48	1293.11	4.00	<.0001	Intercept Variance across SUBJECTS in mean RT
Residual		14341	559.79	25.62	<.0001	Leftover trial-to-trial residual variance

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
2	356.19	<.0001

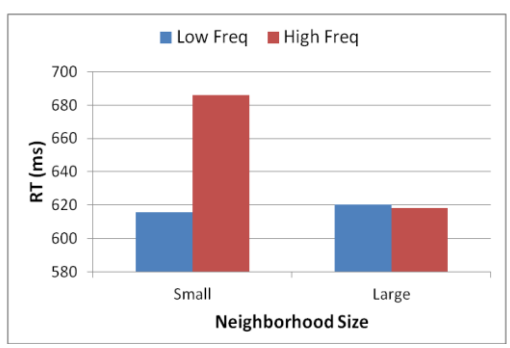
This is the test of whether we need anything in the **G** matrix. **G** still has 2 random intercept variances (subjects, items).

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
17402.4	3	17408.4	17408.5	17402.4	17402.4	17405.4

Solution for Fixed Effects → are SIMPLE MAIN EFFECTS (0=0)

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	615.78	18.5739	60.7	33.15	<.0001
freq01	70.0204	20.5952	32.4	3.40	0.0018
size01	4.4350	20.4202	31.4	0.22	0.8295
freq01*size01	-72.0301	29.3756	31.8	-2.45	0.0199



Type 3 Tests of Fixed Effects → are SIMPLE MAIN EFFECTS (0=0) STILL

Effect	Num DF	Den DF	F Value	Pr > F
freq01	1	32.4	11.56	0.0018
size01	1	31.4	0.05	0.8295
freq01*size01	1	31.8	6.01	0.0199

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	
RT for Low Freq, Small Size	615.78	18.5739	60.7	33.15	<.0001	**CELL MEANS**
RT for Low Freq, Large Size	620.22	18.5482	60.3	33.44	<.0001	
RT for High Freq, Small Size	685.80	18.7416	62.7	36.59	<.0001	
RT for High Freq, Large Size	618.21	19.1504	58.8	32.28	<.0001	
RT for Low Freq	618.00	15.5006	62.5	39.87	<.0001	**MARGINAL MEANS**
RT for High Freq	652.01	15.7322	63.5	41.44	<.0001	
RT for Small Size	650.79	15.5588	63.4	41.83	<.0001	
RT for Large Size	619.21	15.6749	62.7	39.50	<.0001	
Grand Mean for All	635.00	13.7824	53.9	46.07	<.0001	
Simple Freq Effect for Small Size	70.0204	20.5952	32.4	3.40	0.0018	**SIMPLE EFFECTS**
Simple Freq Effect for Large Size	-2.0097	20.9460	31.2	-0.10	0.9242	
Simple Size Effect for Low Freq	4.4350	20.4202	31.4	0.22	0.8295	
Simple Size Effect for High Freq	-67.5951	21.1176	32.2	-3.20	0.0031	
Marginal Freq Effect	34.0053	14.6873	31.8	2.32	0.0272	**MARGINAL EFFECTS**
Marginal Size Effect	-31.5801	14.6880	31.8	-2.15	0.0393	

```
* Get total R2;
PROC CORR DATA=ItemPred; VAR pred rt; RUN;
```

	Pred	rt	
rt	**0.17421**	1.00000	→ **.17421^2 = Overall R^2 = .03**
Response Time in Milliseconds	<.0001		

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

PsuedoR2 (% Reduction) for CovEmpty vs. CovItemPred

Name	CovParm	Subject	Estimate	StdErr	ZValue	ProbZ	PseudoR2
CovEmpty	UN(1,1)	item	2409.36	678.04	3.55	0.0002	.
CovEmpty	UN(1,1)	subject	5166.81	1292.78	4.00	<.0001	.
CovEmpty	Residual		14344	559.99	25.61	<.0001	.
CovItemPred	UN(1,1)	item	1692.07	526.60	3.21	0.0007	0.29771
CovItemPred	UN(1,1)	subject	5168.48	1293.11	4.00	<.0001	-0.00032
CovItemPred	Residual		14341	559.79	25.62	<.0001	0.00018

Why didn't we explain any subject or residual variance?

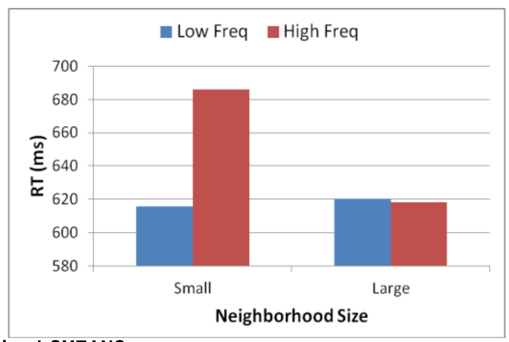
Model 4b: Are there significant fixed effects of the item predictors Frequency and Size? SAME MODEL, JUST USING CLASS STATEMENT INSTEAD FOR DEMONSTRATION

→ $RT_{tis} = \gamma_{000} + \gamma_{010}(Freq_i) + \gamma_{020}(Size_i) + \gamma_{030}(Freq_i)(Size_i) + U_{00s} + U_{0i0} + e_{tis}$

```
TITLE1 "4b. Random Subjects by Random Items Crossed Predictive Model";
TITLE2 "Using CLASS statement to get cell means and comparisons VIA LSMEANS";
PROC MIXED DATA=work.example5a COVTEST NOCLPRINT NAMELEN=100 IC METHOD=REML;
       * Add freq and size to CLASS statement to use LSMEANS;
       CLASS Subject Item freq01 size01;
       * | operator estimates all possible main effects and interactions up to @ order;
       MODEL rt = freq01|size01@2 / SOLUTION DDFM=Satterthwaite OUTPM=ItemPred;
       RANDOM INTERCEPT / SUBJECT=Item TYPE=UN;       * Level 2 variance for items;
       RANDOM INTERCEPT / SUBJECT=Subject TYPE=UN;    * Level 2 variance for subjects;
       ODS OUTPUT InfoCrit=FitRandItem CovParms=CovItemPred; * Save fit, variances to compare;
       * Requesting marginal means per condition (what Type 3 tests are for);
       LSMEANS freq01 size01;
       * Requesting F-tests for simple main effects (more useful than marginal);
       LSMEANS freq01*size01 / SLICE=freq01 SLICE=size01;
RUN; TITLE1; TITLE2;
```

Solution for Fixed Effects → are SIMPLE MAIN EFFECTS (BUT highest=0 given CLASS statement)

Effect	freq01	size01	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			618.21	19.1504	58.8	32.28	<.0001
freq01	0		2.0097	20.9460	31.2	0.10	0.9242
freq01	1		0	.	.	.	.
size01		0	67.5951	21.1176	32.2	3.20	0.0031
size01		1	0	.	.	.	.
freq01*size01	0	0	-72.0301	29.3756	31.8	-2.45	0.0199
freq01*size01	0	1	0	.	.		
freq01*size01	1	0	0	.	.		
freq01*size01	1	1	0	.	.		



Type 3 Tests of Fixed Effects → THESE MAIN EFFECTS ARE NOW MARGINAL

Effect	Num DF	Den DF	F Value	Pr > F
freq01	1	31.8	5.36	0.0272
size01	1	31.8	4.62	0.0393
freq01*size01	1	31.8	6.01	0.0199

Least Squares Means → Means per condition and/or cell as requested via LSMEANS

Effect	freq01	size01	Estimate	Standard Error	DF	t Value	Pr > |t|	
freq01	0		618.00	15.5006	62.5	39.87	<.0001	MARGINAL MEANS
freq01	1		652.01	15.7322	63.5	41.44	<.0001	
size01		0	650.79	15.5588	63.4	41.83	<.0001	
size01		1	619.21	15.6749	62.7	39.50	<.0001	
freq01*size01	0	0	615.78	18.5739	60.7	33.15	<.0001	CELL MEANS
freq01*size01	0	1	620.22	18.5482	60.3	33.44	<.0001	
freq01*size01	1	0	685.80	18.7416	62.7	36.59	<.0001	
freq01*size01	1	1	618.21	19.1504	58.8	32.28	<.0001	

Tests of Effect Slices → TESTS OF SIMPLE MAIN EFFECTS via LSMEANS SLICE

Effect	freq01	size01	Num DF	Den DF	F Value	Pr > F	
freq01*size01	0		1	31.4	0.05	0.8295	size effect for low freq
freq01*size01	1		1	32.2	10.25	0.0031	size effect for high freq
freq01*size01		0	1	32.4	11.56	0.0018	freq effect for small size
freq01*size01		1	1	31.2	0.01	0.9242	freq effect for large size

<u>Model 5: Should items still be treated as a random effect?</u>

Is there still significant variance in mean RT across items after controlling for frequency and size?

→ $RT_{tis} = \gamma_{000} + \gamma_{010}(Freq_i) + \gamma_{020}(Size_i) + \gamma_{030}(Freq_i)(Size_i) + U_{00s}$ ______ $+ e_{tis}$

```
TITLE1 "5. Drop Random Item Intercept?";
PROC MIXED DATA=work.example5a COVTEST NOCLPRINT NAMELEN=100 IC METHOD=REML;
       CLASS Subject Item;
       MODEL rt = freq01|size01@2 / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT / SUBJECT=Subject TYPE=UN; * Level 2 variance for subjects ONLY;
       ODS OUTPUT InfoCrit=FitNoRandItem;          * Save fit to compare;
RUN; TITLE1;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
UN(1,1)	subject	5171.97	1302.28	3.97	<.0001
Residual		15688	603.61	25.99	<.0001

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	292.19	<.0001

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
17466.4	2	17470.4	17470.4	17471.6	17473.7	17475.7

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	614.64	13.3976	54.8	45.88	<.0001
freq01	62.5713	9.5910	1352	6.52	<.0001
size01	5.4273	9.2634	1351	0.59	0.5580
freq01*size01	-64.6343	13.4592	1351	-4.80	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
freq01	1	1351	20.22	<.0001
size01	1	1351	15.97	<.0001
freq*size	1	1351	23.06	<.0001

Is there still significant item variance remaining?

```
* Calculate difference in model fit relative to random subjects and items predictive model;
%FitTest(FitFewer=FitNoRandItem, FitMore=FitRandItem);
```

Likelihood Ratio Test for FitNoRandItem vs. FitRandItem

Name	Neg2Log Like	Parms	AIC	BIC	DevDiff	DFdiff	Pvalue
FitNoRandItem	17466.4	2	17470.4	17473.7	.	.	.
FitRandItem	17402.4	3	17408.4	17402.4	63.9914	1	1.2212E-15

Model 6: Is there a significant random subject slope for the item predictor of frequency?

```
TITLE1 "6. Random Slope for Effect of Freq over Subjects";
PROC MIXED DATA=work.example5a COVTEST NOCLPRINT NAMELEN=100 IC METHOD=REML;
       CLASS Subject Item;
       MODEL rt = freq01|size01@2 / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT        / SUBJECT=Item TYPE=UN;       * Level 2 variance for items;
       RANDOM INTERCEPT freq01 / SUBJECT=Subject TYPE=UN;    * Level 2 variances for subjects;
       ODS OUTPUT InfoCrit=FitRandFreq;                      * Save fit to compare;
RUN; TITLE1; TITLE2;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	
UN(1,1)	item	1700.03	527.91	3.22	0.0006	Residual item variance after predictors
UN(1,1)	subject	5231.22	1307.42	4.00	<.0001	Variance over SUBJECTS in mean RT @ freq=0
UN(2,1)	subject	1058.11	571.78	1.85	0.0642	Intercept, freq slope covariance
UN(2,2)	subject	371.65	447.45	0.83	0.2031	Random freq slope variance over subjects
Residual		14244	563.58	25.28	<.0001	Leftover trial-to-trial residual variance

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
17397.6	5	17407.6	17407.6	17397.6	17397.6	17402.6

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	615.85	17.9378	55.8	34.33	<.0001
freq01	69.8447	20.8577	33.5	3.35	0.0020
size01	4.4434	20.4461	31.4	0.22	0.8294
freq01*size01	-72.0683	29.4136	31.8	-2.45	0.0200

Does the effect of frequency vary over subjects?

```
* Calculate difference in model fit relative to random subjects and items predictive model;
%FitTest(FitFewer=FitRandItem, FitMore=FitRandFreq);
```

Likelihood Ratio Test for FitRandItem vs. FitRandSize

Name	Neg2Log Like	Parms	AIC	BIC	DevDiff	DFdiff	Pvalue
FitRandItem	17402.4	3	17408.4	17402.4	.	.	.
FitRandSize	17402.4	5	17412.4	17402.4	0.081671	2	0.95999

Calculate 95% random effects confidence intervals for the frequency effect across subjects:

95% CI = fixed effect ± 1.96*SQRT(variance)

Subject Frequency Slope CI = 69.84 ± 1.96*SQRT(371.65) = 32 to 107

95% of the individual subject simple frequency slopes are expected to fall between 32 and 107 ms

Writing out a single-level combined equation for this last model to illustrate the random slopes:

$RT_{tis} = \gamma_{000} + \gamma_{010}(Freq_i) + \gamma_{020}(Size_i) + \gamma_{030}(Freq_i)(Size_i) + U_{00s} + U_{01s}(Freq_i) + U_{0i0} + e_{tis}$

$RT_{tis} = 615.85 + (69.84*Freq_i) + (4.44*Size_i) + (-72.07*Freq_i*Size_i)$

$+ U_{00s}$ → increment to *mean RT* depending on which subject after controlling for NOTHING

$+ U_{0i0}$ → increment to *mean RT* depending on which item after controlling for freq and size

$+ U_{01s}(Freq_i)$ → increment to *slope of frequency* depending on which subject

$+ e_{tis}$ → increment to *trial RT* depending on which trial after controlling for everything

Sample Results Section:

The extent to which semantic frequency (coded low = 0, high = 1) and phonological neighborhood size (coded small = 0, large = 1) could predict response time (RT) in milliseconds in a lexical decision task was examined for 39 items administered to 38 subjects. Because RTs for incorrect responses were not included, the data were unbalanced, such that each subject had a different number of trials included for each condition. Accordingly, rather than aggregating the individual trial RTs into potentially biased item condition means (that would assume items are fixed) and conducting an analysis of variance, all possible RTs were examined instead in a multilevel model with crossed random effects, in which individual trials (the combination of each subject with each item) were nested within subjects and within items, which were crossed random effects. Restricted maximum likelihood within SAS PROC MIXED was used to estimate all models; denominator degrees of freedom were estimated with the Satterthwaite method.

The extent to which systematic variability in mean RT existed for each dimension of sampling was first examined in a series of empty means models (i.e., only a fixed intercept and no predictors). Relative to a model with only a residual variance, the addition of a random intercept variance for subjects significantly improved model fit, $-2\Delta LL(\sim 1) = 280.4$, $p < .001$ (and the smaller AIC and BIC concur), indicating significant differences between subjects in mean RT, and that trials from the same subject were positively correlated. The addition of a random intercept for items also significantly improved model fit, $-2\Delta LL(\sim 1) = 100.4$, $p < .001$ (and the smaller AIC and BIC concur), indicating significant differences between items in mean RT as well, and that trials for the same item were also positively correlated. Of the total estimated RT variance, 24% was due to between-subject differences in mean RT (given by the subject random intercept), 11% was due to between-item differences in mean RT (given by the item random intercept), and the remaining 65% was due to the subject by item interaction (i.e., residual variance). Construction of 95% random effects confidence intervals as described in Snijders and Bosker (1999) revealed that 95% of subject mean RTs are expected to fall between 494 and 776 ms, whereas 95% of the item mean RTs are expected to fall between 539 and 732 ms. Thus, there was relatively more variability across subjects than across items. The extent to which the main and interaction effects of semantic frequency and neighborhood size could account for between-item differences in mean RT was then examined in a conditional model; results are provided in Table 1.

ANOVA-like description of the results: There was a significant semantic frequency by neighborhood size interaction, $F(1,31.8) = 6.01$, $p = .0199$; the pattern of the interaction is shown in Figure 1 and was decomposed by examining simple main effects of each predictor. First, with respect to the effect of neighborhood size, for low frequency words, there was no significant difference between words with small or large neighborhood size (M = 615.78, M = 620.22), $F(1,31.4) = 0.05$, $p = .8295$, whereas for high frequency words, responses were significantly slower to words with smaller than larger neighborhoods (M = 685.80, M = 618.21), $F(1,32.2) = 10.25$, $p = .0031$. With respect to the effect of frequency, for small neighborhood words, responses were significantly faster to words of low than high frequency (M = 615.78, M = 685.80), $F(1,32.4) = 11.56$, $p = .0018$, whereas for large neighborhood words, there was no significant difference between words of low or high frequency (M = 620.22, M = 618.21), $F(1,31.2) = 0.01$, $p = .9242$.

Regression-like description of the same results: The fixed intercept for the predicted RT for a word of low frequency and small size was $\gamma_{000} = 615.78$. There was a significant simple main effect for the mean difference between low and high frequency words of small size of $\gamma_{010} = 70.02$ ($p = .002$). There was a nonsignificant simple main effect for the mean difference between small and large size words of low frequency of $\gamma_{020} = 4.44$ ($p = .830$). However, there was a significant frequency by size interaction of $\gamma_{030} = -72.03$ ($p = .020$), such that relative to the frequency effect for small words of $\gamma_{010} = 70.02$, the frequency effect for large words was significantly less positive by −72.03 (yielding a nonsignificant simple effect of frequency for large words of $\gamma_{010} + \gamma_{030} = -2.01$, $p = .924$). Similarly, relative to the size effect for low frequency words of $\gamma_{020} = 4.44$, the size effect for high frequency words was significantly more negative by −72.03 (yielding a significant simple effect of size for high frequency words of $\gamma_{020} + \gamma_{030} = -67.56$, $p = .003$). Thus, as shown in Figure 1, a positive frequency effect was found only for words of small size, and a negative size effect was found only for high frequency words.

The effects of frequency and size explained approximately 30% of the item intercept variance. Given that 11% of the total RT variance was due to mean differences between items, this translates into a total reduction in all RT variance of 3.28%. The extent to which these effects were sufficient to describe all between-item differences in mean RT was then examined by removing the item random intercept variance from the conditional model. The resulting significant decrease in model fit, $-2\Delta LL(\sim 1) = 64.4$, $p < .001$ (and the larger AIC and BIC) suggest that significant differences remain between items after controlling for their primary design features, or that items should not be treated as fixed effects. Finally, the potential for individual subject differences in the frequency effect was examined by adding a random subject frequency slope (and its covariance with the subject random intercept) to the model. Model fit did not significantly improve, $-2\Delta LL(\sim 2) = 4.8$, $p = .091$ (although the AIC and BIC were smaller), indicating that each subject does not need his or her own random deviation from the fixed effect of frequency.