Example 7: Explanatory IRT Models as Crossed Random Effects Models
(complete syntax and output available for STATA, R, and SAS electronically)

This example shows variants of "explanatory" item response theory (IRT) models, which can be estimated as generalized multilevel models with a random person intercept and either fixed item effects (in which the model just has level-1 trials nested in 155 level-2 persons) or a random item intercept (in which level-2 items are crossed with level-2 persons). These example data are from my dissertation: 36 items assessing attentional search via change detection scored incorrect (correct=0) or correct (correct=1), in which response time was also collected for the correct items. The items varied by four features: continuous visual clutter (clutter), whether the change was relevant to driving (relevant), continuous brightness of the change (bright), and whether the change was made to a legible sign (sign). These analyses require a "stacked" (or "long") data format in which each item for each person is stored on a separate row. Syntax and output for 1PL or Rasch IRT models on wide-format data are also available in the online materials for each program.

STATA Syntax for Importing and Preparing Data for Analysis:

```
// Define global variable for file location to be replaced in code below
// \\Client\ precedes path in Virtual Desktop outside H drive
   global filesave "C:\Dropbox\23_PSQF6272\PSQF6272_Example7"

// Open long-format example excel data file from sheet "long" and clear away existing data
   clear // clear memory in case of open data
   import excel "$filesave\Example7_Data.xlsx", firstrow case(preserve) sheet("long") clear

// Add variable labels
   label variable PersonID  "PersonID: Person Identifier"
   label variable PictureID "PictureID: Picture Identifier"
   label variable correct   "correct: Correct Response (0=no, 1=yes)"
   label variable RTsec     "RTsec: Response Time in Seconds if Correct"
   label variable clutter   "clutter: Rated Clutter in Picture Z-Score"
   label variable relevant  "relevant: Change Relevant to Driving (0=no, 1=yes)"
   label variable bright    "bright: Rated Change Brightness Z-Score"
   label variable sign      "sign: Change to Legible Sign (0=no, 1=yes)"

display "STATA Descriptive Statistics for Example Variables"
summarize correct RTsec clutter relevant bright sign
```

Descriptive Statistics for Example Variables (from SAS)

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
correct	correct: Correct Response (0=no, 1=yes)	5426	**0.7823**	0.4127	0.0000	1.0000
RTsec	RTsec: Response Time in Seconds if Correct	4244	15.6277	11.0487	1.5910	44.9950
clutter	clutter: Rated Clutter in Picture Z-Score	5426	0.0031	0.6509	-1.1832	1.4540
relevant	relevant: Change Relevant to Driving (0=no, 1=yes)	5426	0.5013	0.5000	0.0000	1.0000
bright	bright: Rated Change Brightness Z-Score	5426	-0.0229	0.4290	-1.0766	1.0276
sign	sign: Change to Legible Sign (0=no, 1=yes)	5426	0.3360	0.4724	0.0000	1.0000

R Syntax for Importing and Preparing Data for Analysis (after loading packages *readxl*, *TeachingDemos*, *psych*, *lme4*, *lmerTest*, and *performance*, as well as *mirt* for the IRT version):

```
# Define variables for working directory and data name -- CHANGE THESE
filesave = "C:\\Dropbox/23_PSQF6272/PSQF6272_Example7/"
filename = "Example7_Data.xlsx"
setwd(dir=filesave)

#  Import long-format example excel data file from sheet "long"
Example7 = read_excel(paste0(filesave,filename), sheet="long")
# Convert to data frame to use in analysis
Example7 = as.data.frame(Example7)

print("R Descriptive Statistics for Example Variables")
describe(x=Example7[ , c("correct","RTsec","clutter","relevant","bright","sign")])
```

Model 1: Single-Level Empty Means for Binary Correct Response
(t = trial, p = person, i = item)

Composite: $Log\left[\frac{prob(correct_{tpi}=1)}{prob(correct_{tpi}=0)}\right] = Logit\big(correct_{tpi} = 1\big) = \gamma_{000}$

```
display "STATA Model 1: Single-Level Empty Means for Binary Correct Response"
melogit correct ,  intmethod(laplace) nolog
display "-2LL = " e(ll)*-2      // Print -2LL for model
nlcom 1/(1+exp(-1*(_b[_cons]))) // Fixed intercept in probability

print("R Model 1: Single-Level Empty Means for Binary Correct Response")
Model1 = glm(data=Example7, family=binomial(link="logit"), formula=correct~1)
summary(Model1) # residual deviance = -2LL already

Coefficients:
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) 1.279380   0.032898  38.889 < 2.2e-16 gamma000 in logits

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5685.63  on 5425  degrees of freedom
Residual deviance: 5685.63  on 5425  degrees of freedom → -2LL for model
AIC: 5687.63

print("Convert logits to probability via inverse link")
Model1Prob=1/(1+exp(-1*coefficients(Model1))); Model1Prob
0.78234427 → gamma000 in probability
```

All two-level models from here use Laplace estimation for comparability across programs…

Model 2: Random Persons Only, Empty Means

$Logit\big(correct_{tpi} = 1\big) = \gamma_{000} + U_{0p0}$

```
display "STATA Model 2: Random Persons Only, Empty Means"
melogit correct , || PersonID: , intmethod(laplace) nolog
display "-2LL = " e(ll)*-2      // Print -2LL for model
nlcom 1/(1+exp(-1*(_b[_cons]))) // Fixed intercept in probability
estimates store Fit2            // Save fit for LRT

print("R Model 2: Random Persons Only, Empty Means")
Model2 = glmer(data=Example7, family=binomial(link="logit"), nAGQ=1,
               correct~1+(1|PersonID))
print("Show -2LL with more precision, results, and ICC using 3.29=residual variance")
-2*logLik(Model2); summary(Model2); icc(Model2)

'log Lik.' 5600.4498 (df=2) → -2LL for model

     AIC      BIC   logLik deviance df.resid
  5604.4   5617.6  -2800.2   5600.4     5424

Random effects:
 Groups   Name        Variance Std.Dev.
 PersonID (Intercept) 0.25229  0.50229       Var(U0p0)
Number of obs: 5426, groups:  PersonID, 155

Fixed effects:
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) 1.348536   0.053508  25.203 < 2.2e-16  gamma000

# Intraclass Correlation Coefficient
    Adjusted ICC: 0.071
  Unadjusted ICC: 0.071
```

```
print("LRT for person random intercept"); anova(Model2, Model1)
       npar     AIC     BIC   logLik deviance   Chisq Df Pr(>Chisq)
Model1    1 5687.63 5694.23 -2842.82  5685.63
Model2    2 5604.45 5617.65 -2800.22  5600.45 85.1826  1 < 2.22e-16

print("Convert logits to probability via inverse link")
Model2Prob=1/(1+exp(-1*fixef(Model2))); Model2Prob
0.79389026
```

Model 3: Random Persons Only, Rasch Fixed Items (via a Categorical Item ID Predictor):

$$Logit\left(correct_{tpi} = 1\right) = \gamma_{00,1}(Pic1_i) + \gamma_{00,2}(Pic2_i) + \cdots + \gamma_{00,36} + U_{0p0}$$

```
display "STATA Model 3: Random Persons Only, Rasch Fixed Items"
display "Rasch version of 1PL: Person Trait Variance Estimated, Single Discrimination=1"
melogit correct ibn.PictureID, noconstant  /// Removed intercept to get each item easiness
        || PersonID: , intmethod(laplace) nolog
display "-2LL = " e(ll)*-2      // Print -2LL for model
estimates store Fit3            // Save fit for LRT
lrtest Fit3 Fit2                // LRT for fixed item differences

print("R Model 3: Random Persons Only, Rasch Fixed Items")
print("Rasch version of 1PL: Person Trait Variance Estimated, Single Discrimination=1")
Model3 = glmer(data=Example7, family=binomial(link="logit"), nAGQ=1,
               correct~0+as.factor(PictureID)+(1|PersonID))
print("Show -2LL with more precision and results")
-2*logLik(Model3); summary(Model3)

'log Lik.' 4907.4892 (df=37) → -2LL for model
     AIC      BIC   logLik deviance df.resid
  4981.5   5225.7  -2453.7   4907.5     5389

Random effects:
 Groups   Name        Variance Std.Dev.
 PersonID (Intercept) 0.36752  0.60624  Var(U0p0)
```

This "fixed-items" model controls for differences in item easiness, but it does not allow any predictors of those differences—stay tuned for Model 5!

```
Fixed effects: → Item Easiness parameters
                       Estimate Std. Error z value  Pr(>|z|)
as.factor(PictureID)2    0.23102    0.17787  1.2988   0.19400  gamma00,1
as.factor(PictureID)6    3.36407    0.42029  8.0042 1.203e-15
as.factor(PictureID)7    1.14598    0.19799  5.7880 7.122e-09
as.factor(PictureID)10   2.22830    0.26554  8.3916 < 2.2e-16
as.factor(PictureID)11   0.96531    0.19123  5.0479 4.467e-07
as.factor(PictureID)13   0.11492    0.17618  0.6523   0.51422
as.factor(PictureID)22   1.27740    0.20496  6.2324 4.595e-10
as.factor(PictureID)23   2.92348    0.34900  8.3768 < 2.2e-16
as.factor(PictureID)26   1.03232    0.19392  5.3234 1.018e-07
as.factor(PictureID)33   1.63935    0.22301  7.3510 1.968e-13
as.factor(PictureID)35   1.31003    0.20682  6.3340 2.388e-10
as.factor(PictureID)42   0.94660    0.19220  4.9250 8.437e-07
as.factor(PictureID)52   2.71366    0.31956  8.4918 < 2.2e-16
as.factor(PictureID)59   2.25666    0.27282  8.2716 < 2.2e-16
as.factor(PictureID)61   0.88830    0.19203  4.6259 3.729e-06
as.factor(PictureID)62  -0.32835    0.18020 -1.8221   0.06844
as.factor(PictureID)66   2.53242    0.29729  8.5182 < 2.2e-16
as.factor(PictureID)97   1.37490    0.20797  6.6111 3.814e-11
as.factor(PictureID)117  3.06113    0.36795  8.3194 < 2.2e-16
as.factor(PictureID)123  2.36434    0.27974  8.4520 < 2.2e-16
as.factor(PictureID)128  1.14723    0.19798  5.7947 6.843e-09
as.factor(PictureID)135  3.36988    0.42022  8.0194 1.063e-15
as.factor(PictureID)136  1.27385    0.20829  6.1159 9.603e-10
as.factor(PictureID)137  1.49409    0.21497  6.9501 3.650e-12
as.factor(PictureID)140  0.38151    0.18049  2.1138   0.03453
as.factor(PictureID)146  1.79071    0.23249  7.7024 1.335e-14
as.factor(PictureID)152  0.73831    0.18679  3.9527 7.728e-05
as.factor(PictureID)155 -0.20935    0.17830 -1.1741   0.24035
as.factor(PictureID)161  2.90642    0.34946  8.3170 < 2.2e-16
as.factor(PictureID)162  3.03671    0.36838  8.2434 < 2.2e-16
```

```
as.factor(PictureID)171  1.57520    0.21978  7.1673 7.650e-13
as.factor(PictureID)172  0.86210    0.19035  4.5291 5.925e-06
as.factor(PictureID)173  3.21720    0.39120  8.2238 < 2.2e-16
as.factor(PictureID)174  1.69324    0.22596  7.4934 6.712e-14
as.factor(PictureID)177  0.89789    0.18900  4.7508 2.026e-06
as.factor(PictureID)179  1.78479    0.23266  7.6712 1.704e-14  gamma00,36

optimizer (Nelder_Mead) convergence code: 0 (OK)
Model failed to converge with max|grad| = 0.00431822 (tol = 0.002, component 1)

print("LRT for fixed item differences"); anova(Model3, Model2)
       npar     AIC     BIC   logLik deviance   Chisq Df Pr(>Chisq)
Model2    2 5604.45 5617.65 -2800.22  5600.45
Model3   37 4981.49 5225.65 -2453.74  4907.49 692.961 35 < 2.22e-16
```

Model 4: Random Persons, Empty Means Random Items

$$Logit(correct_{tpi} = 1) = \gamma_{000} + U_{0p0} + U_{00i}$$

```
display "STATA Model 4: Random Persons, Empty Means Random Items"
melogit correct , || _all: R.PictureID || PersonID: , intmethod(laplace) nolog
display "-2LL = " e(ll)*-2       // Print -2LL for model
estimates store Fit4             // Save fit for LRT
lrtest Fit4 Fit2                 // LRT for random item differences
```

```
print("R Model 4: Random Persons, Empty Means Random Items")
Model4 = glmer(data=Example7, family=binomial(link="logit"), nAGQ=1,
               correct~1+(1|PersonID)+(1|PictureID))
print("Show -2LL with more precision, results, and ICCs")
-2*logLik(Model4); summary(Model4); icc(Model4, by_group=TRUE)

'log Lik.' 5049.8399 (df=3) → -2LL for model

     AIC      BIC   logLik deviance df.resid
  5055.8   5075.6  -2524.9   5049.8     5423

Random effects:
 Groups    Name        Variance Std.Dev.
 PersonID  (Intercept) 0.36247  0.60206  Var(U0p0)
 PictureID (Intercept) 0.94895  0.97414  Var(U00i)
Number of obs: 5426, groups:  PersonID, 155; PictureID, 36

Fixed effects:
            Estimate Std. Error z value  Pr(>|z|)
(Intercept)  1.59260    0.17472  9.1151 < 2.2e-16  gamma000

Group     |   ICC
-----------------
PersonID  | 0.079
PictureID | 0.206

print("LRT for item random intercept"); anova(Model4, Model2)
       npar     AIC     BIC   logLik deviance  Chisq Df Pr(>Chisq)
Model2    2 5604.45 5617.65 -2800.22  5600.45
Model4    3 5055.84 5075.64 -2524.92  5049.84 550.61  1 < 2.22e-16
```

Model 5: Random Persons, LLTM-Predicted Random Items (Predicted Item Easiness)

$$Logit(correct_{tpi} = 1) = \gamma_{000} + \gamma_{001}(clutter_i) + \gamma_{002}(relevant_i) + \gamma_{003}(bright_i) + \gamma_{004}(sign_i) + U_{0p0} + U_{00i}$$

```
display "STATA Model 5: Random Persons, LLTM-Predicted Random Items"
melogit correct c.clutter c.relevant c.bright c.sign, ///
        || _all: R.PictureID || PersonID: , intmethod(laplace) nolog
display "-2LL = " e(ll)*-2       // Print -2LL for model
estimates store Fit5             // Save fit for LRT
```

```
print("R Model 5: Random Persons, LLTM-Predicted Random Items")
Model5 = glmer(data=Example7, family=binomial(link="logit"), nAGQ=1,
               correct~1+clutter+relevant+bright+sign+(1|PersonID)+(1|PictureID))
print("Show -2LL with more precision and results")
-2*logLik(Model5); summary(Model5)
```

```
'log Lik.' 5040.2043 (df=7) → -2LL for model
     AIC      BIC   logLik deviance df.resid
  5054.2   5100.4  -2520.1   5040.2     5419
Random effects:
 Groups    Name        Variance Std.Dev.
 PersonID  (Intercept) 0.36245  0.60203  Var(U0p0)
 PictureID (Intercept) 0.71353  0.84470  Var(U00i)

Fixed effects:
             Estimate Std. Error z value     Pr(>|z|)
(Intercept)  1.347412   0.259372  5.1949 0.0000002048  gamma000
clutter     -0.323771   0.241496 -1.3407      0.18002  gamma001
relevant     0.037214   0.425322  0.0875      0.93028  gamma002
bright       0.789280   0.498280  1.5840      0.11319  gamma003
sign         0.738577   0.335838  2.1992      0.02786  gamma004
```

Pseudo-R2 Relative to work.Cov4 in Model 4 (from SAS)

Name	CovParm	Subject	Estimate	PseudoR2
work.Cov4	UN(1,1)	PersonID	0.3625	.
work.Cov4	UN(1,1)	PictureID	0.9492	.
work.Cov5	UN(1,1)	PersonID	0.3625	0.00008
work.Cov5	UN(1,1)	PictureID	0.7138	0.24798

Do we need random item variance leftover? How does the LLTM exchangeable-item solution compare?

Model 6: Random Persons Only, LLTM-Predicted Fixed Items (Predicted Item Easiness)

$$Logit(correct_{tpi} = 1) = \gamma_{000} + \gamma_{001}(clutter_i) + \gamma_{002}(relevant_i) + \gamma_{003}(bright_i) + \gamma_{004}(sign_i) + U_{0p0}$$

```
display "STATA Model 6: Random Persons Only, LLTM-Predicted Fixed Items"
melogit correct c.clutter c.relevant c.bright c.sign, ///
        || PersonID: , intmethod(laplace) nolog
display "-2LL = " e(ll)*-2      // Print -2LL for model
estimates store Fit6            // Save fit for LRT
lrtest Fit5 Fit6                // LRT for random item variance leftover
```

```
print("R Model 6: Random Persons Only, LLTM-Predicted Fixed Items")
Model6 = glmer(data=Example7, family=binomial(link="logit"), nAGQ=1,
               correct~1+clutter+relevant+bright+sign+(1|PersonID))
print("Show -2LL with more precision and results")
-2*logLik(Model6); summary(Model6)
```

```
'log Lik.' 5439.799 (df=6) → -2LL for model

    AIC      BIC   logLik deviance df.resid
  5451.8   5491.4  -2719.9   5439.8     5420

Random effects:
 Groups   Name        Variance Std.Dev.
 PersonID (Intercept) 0.27332  0.5228   Var(U0p0)

Fixed effects:
             Estimate Std. Error z value  Pr(>|z|)
(Intercept)  1.082199   0.071967 15.0374 < 2.2e-16  gamma000
clutter     -0.267451   0.055230 -4.8425 1.282e-06  gamma001
relevant     0.220391   0.098888  2.2287   0.02583  gamma002
bright       0.474107   0.112466  4.2156 2.492e-05  gamma003
sign         0.661969   0.081797  8.0928 5.831e-16  gamma004
```

```
Fixed effects from Random Item Model 5:
             Estimate Std. Error   Pr(>|z|)
(Intercept)  1.347412   0.259372  0.0000002048
clutter     -0.323771   0.241496  0.18002
relevant     0.037214   0.425322  0.93028
bright       0.789280   0.498280  0.11319
sign         0.738577   0.335838  0.02786
```

```
print("LRT for random item variance leftover"); anova(Model5, Model6)
       npar    AIC     BIC  logLik deviance   Chisq Df Pr(>Chisq)
Model6    6 5451.8 5491.39 -2719.9   5439.8
Model5    7 5054.2 5100.40 -2520.1   5040.2 399.595  1 < 2.22e-16
```

Comparing Results: Fixed vs. Random Effects for Item Easiness Predictions (Compiled in SAS)

Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Source
fixed_b	36	1.61152	1.02673	58.01488	-0.32837	3.36982	Predicted from Model 6 (no random item)
random_b	36	1.54605	0.93089	55.65778	-0.25853	3.06788	Predicted from Model 5 (with random item)
fixed_pred_b	36	1.40093	0.44574	50.43335	0.60327	2.28516	Predicted from Model 3 (Rasch fixed items)
random_pred_b	36	1.59248	0.49180	57.32927	0.53565	2.58004	Predicted from Model 4 (empty, random items)

Pearson Correlation Coefficients, N = 36 Prob > |r| under H0: Rho=0				
	fixed_b	random_b	fixed_pred_b	random_pred_b
fixed_b	1.00000	0.99950 <.0001	0.47623 0.0033	0.48546 0.0027
random_b	**0.99950** <.0001	1.00000	0.47770 0.0032	0.48561 0.0027
fixed_pred_b	**0.47623** 0.0033	0.47770 0.0032	1.00000	0.98472 <.0001
random_pred_b	0.48546 0.0027	**0.48561** 0.0027	**0.98472** <.0001	1.00000

The R^2 for item easiness is very close to the pseudo-R^2 for the proportion reduction in random item variance after including the 4 item predictors.

The difference lies in the significance of the item predictor effects, whose SEs are based on the wrong error term in the random-subjects-only LLTM.

Explanatory Items Model with Random Subjects AND Random Items (Predicted Item Easiness)
Model 7a: Adding a Random Slope of Item Predictor "Bright" over Persons *(Models 7bcd were NPD)*

$$Logit\left(correct_{tpi}=1\right)=\gamma_{000}+\gamma_{001}(clutter_i)+\gamma_{002}(relevant_i)+\gamma_{003}(bright_i)+\gamma_{004}(sign_i)$$
$$+U_{0p0}+U_{0p3}(bright_i)+U_{00i}$$

```
display "STATA Model 7a: Add Random Brightness Slope (NPD)"
melogit correct c.clutter c.relevant c.bright c.sign, || _all: R.PictureID  ///
        || PersonID: bright , cov(un) intmethod(laplace) nolog difficult
display "-2LL = " e(ll)*-2      // Print -2LL for model
estimates store Fit7            // Save fit for LRT
lrtest Fit7 Fit5                // LRT for random brightness slope over subjects

print("R Model 7a: Add Random Brightness Slope over Persons (NS)")
Model7a = glmer(data=Example7, family=binomial(link="logit"), nAGQ=1,
                correct~1+clutter+relevant+bright+sign+(1+bright|PersonID)+(1|PictureID))
print("Show -2LL with more precision and results")
-2*logLik(Model7a); summary(Model7a)

'log Lik.' 5038.6144 (df=9) → -2LL for model

     AIC      BIC   logLik deviance df.resid
  5056.6   5116.0  -2519.3   5038.6     5417

Random effects:
 Groups    Name        Variance Std.Dev. Corr
 PersonID  (Intercept) 0.36249  0.60207          Var(U0p0)
           bright      0.14089  0.37536  -0.105  Var(U0p3) Cor(U0p0,U0p3)
 PictureID (Intercept) 0.71760  0.84711          Var(U00i)
```

```
Fixed effects:
             Estimate Std. Error z value     Pr(>|z|)
(Intercept)  1.354590   0.260173  5.2065 0.0000001924  gamma000
clutter     -0.324530   0.242126 -1.3403      0.18014  gamma001
relevant     0.032053   0.426599  0.0751      0.94011  gamma002
bright       0.791084   0.501149  1.5785      0.11444  gamma003
sign         0.739749   0.336698  2.1971      0.02802  gamma004

print("LRT for random brightness slope over persons"); anova(Model7a, Model5)
        npar     AIC     BIC   logLik deviance   Chisq Df Pr(>Chisq)
Model5     7 5054.20 5100.40 -2520.10  5040.20
Model7a    9 5056.61 5116.01 -2519.31  5038.61 1.58992  2     0.4516
```

Now let's predict response time instead using a general version of a subset of the same models…

Model 8: Random Persons Only, Empty Means Predicting Continuous Response Time

$$RTsec_{tpi} = \gamma_{000} + U_{0p0} + e_{tpi}$$

```
display "STATA Model 8: Random Persons, Empty Means for Response Time"
mixed RTsec , || PersonID: , ///
      reml dfmethod(satterthwaite) dftable(pvalue) nolog
display "-2LL = " e(ll)*-2      // Print -2LL for model
estimates store Fit8            // Save fit for LRT

print("R Model 8: Random Persons Only, Empty Means for Response Time")
Model8 = lmer(data=Example7, REML=TRUE, RTsec~1+(1|PersonID))
print("Show -2LL as deviance, results, and ICC")
llikAIC(Model8, chkREML=FALSE); summary(Model8); icc(Model8)

       AIC        BIC      logLik   deviance   df.resid
 32406.299  32425.358 -16200.149  32400.299   4241.000  deviance = -2LL for model

Random effects:
 Groups   Name        Variance Std.Dev.
 PersonID (Intercept)   3.8361  1.9586  Var(U0p0)
 Residual             118.3161 10.8773  Var(etpi)

Fixed effects:
             Estimate Std. Error        df t value  Pr(>|t|)
(Intercept)  15.72135    0.23005 147.64349   68.34 < 2.2e-16  gamma000

    Adjusted ICC: 0.031
  Unadjusted ICC: 0.031

print("LRT for person random intercept"); ranova(Model8)
               npar   logLik     AIC     LRT Df       Pr(>Chisq)
<none>            3 -16200.1 32406.3
(1 | PersonID)    2 -16217.7 32439.5 35.1556  1 0.0000000030438
```

Model 9: Random Persons, Empty Means Random Items

$$RTsec_{tpi} = \gamma_{000} + U_{0p0} + U_{00i} + e_{tpi}$$

```
display "STATA Model 9: Random Persons, Empty Means Random Items"
mixed RTsec , || _all: R.PictureID || PersonID: , ///
      reml dfmethod(satterthwaite) dftable(pvalue) nolog
display "-2LL = " e(ll)*-2      // Print -2LL for model
estimates store Fit9            // Save fit for LRT
lrtest Fit9 Fit8                // LRT for random items

print("R Model 9: Random Persons, Empty Means Random Items")
Model9 = lmer(data=Example7, REML=TRUE, RTsec~1+(1|PersonID)+(1|PictureID))
print("Show -2LL as deviance, results, and ICCs")
llikAIC(Model9, chkREML=FALSE); summary(Model9); icc(Model9, by_group=TRUE)
```

```
       AIC         BIC      logLik   deviance   df.resid 
 31944.657   31970.070 -15968.328  31936.657   4240.000   deviance = -2LL for model

Random effects:
 Groups    Name        Variance Std.Dev.
 PersonID  (Intercept)   5.1758  2.2750   Var(U0p0)
 PictureID (Intercept)  15.5655  3.9453   Var(U00i)
 Residual              102.6873 10.1335   Var(etpi)

Fixed effects:
            Estimate Std. Error     df t value  Pr(>|t|)
(Intercept)   16.271      0.701 39.929  23.211 < 2.2e-16  gamma000

Group     |   ICC
-----------------
PersonID  | 0.042
PictureID | 0.126

print("LRT for each random intercept"); ranova(Model9)
                npar   logLik     AIC     LRT Df Pr(>Chisq)
<none>             4 -15968.3 31944.7
(1 | PersonID)     3 -16002.6 32011.2  68.539  1 < 2.22e-16
(1 | PictureID)    3 -16200.1 32406.3 463.642  1 < 2.22e-16
```

Model 10: Random Persons, LLTM-Predicted Random Items

$$RTsec_{tpi} = \gamma_{000} + \gamma_{001}(clutter_i) + \gamma_{002}(relevant_i) + \gamma_{003}(bright_i) + \gamma_{004}(sign_i) + U_{0p0} + U_{00i} + e_{tpi}$$

```
display "STATA Model 10: Random Persons, LLTM-Predicted Random Items"
mixed RTsec c.clutter c.relevant c.bright c.sign, || _all: R.PictureID  ///
      || PersonID: , reml dfmethod(satterthwaite) dftable(pvalue) nolog
display "-2LL = " e(ll)*-2      // Print -2LL for model
estimates store Fit10           // Save fit for LRT

print("R Model 10: Random Persons, LLTM-Predicted Random Items")
Model10 = lmer(data=Example7, REML=TRUE,
               RTsec~1+clutter+relevant+bright+sign+(1|PersonID)+(1|PictureID))
print("Show -2LL as deviance and results")
llikAIC(Model10, chkREML=FALSE); summary(Model10);

       AIC         BIC      logLik   deviance   df.resid 
 31924.156   31974.982 -15954.078  31908.156   4236.000   deviance = -2LL for model

Random effects:
 Groups    Name        Variance Std.Dev.
 PersonID  (Intercept)   5.1925  2.2787   Var(U0p0)
 PictureID (Intercept)   9.6355  3.1041   Var(U00i)
 Residual              102.6762 10.1329   Var(etpi)

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept) 17.49470    0.96793 34.04328 18.0743  < 2e-16   gamma000
clutter      0.97399    0.88838 30.90588  1.0964  0.28139   gamma001
relevant    -0.46804    1.57692 31.32691 -0.2968  0.76857   gamma002
bright      -4.86674    1.86869 32.05753 -2.6044  0.01384   gamma003
sign        -3.30457    1.22717 30.73121 -2.6928  0.01137   gamma004

print("LRT for random item variance leftover"); ranova(Model10)
                npar   logLik     AIC     LRT Df Pr(>Chisq)
<none>             8 -15954.1 31924.2
(1 | PersonID)     7 -15988.5 31991.0  68.886  1 < 2.22e-16
(1 | PictureID)    7 -16081.5 32177.1 254.902  1 < 2.22e-16
```

Pseudo-R2 Relative to work.Cov9 (from SAS)

Name	CovParm	Subject	Estimate	PseudoR2
work.Cov9	UN(1,1)	PersonID	5.1757	.
work.Cov9	UN(1,1)	PictureID	15.5655	.
work.Cov9	Residual		102.69	.
work.Cov10	UN(1,1)	PersonID	5.1887	-0.00251
work.Cov10	UN(1,1)	PictureID	9.6356	0.38097
work.Cov10	Residual		102.68	0.00009

Model 11: Add Random Brightness Slope over Persons

$$RTsec_{tpi} = \gamma_{000} + \gamma_{001}(clutter_i) + \gamma_{002}(relevant_i) + \gamma_{003}(bright_i) + \gamma_{004}(sign_i) \\ + U_{0p0} + U_{0p3}(bright_i) + U_{00i} + e_{tpi}$$

```
display "STATA Model 11: Add Random Brightness Slope over Persons"
mixed RTsec c.clutter c.relevant c.bright c.sign, || _all: R.PictureID  ///
      || PersonID: bright, cov(un) reml dfmethod(satterthwaite) dftable(pvalue) nolog difficult
display "-2LL = " e(ll)*-2      // Print -2LL for model
estimates store Fit11           // Save fit for LRT
lrtest Fit11 Fit10              // LRT for random brightness slope over subjects

print("R Model 11: Add Random Brightness Slope over Persons (NS)")
Model11 = lmer(data=Example7, REML=TRUE,
               RTsec~1+clutter+relevant+bright+sign+(1+bright|PersonID)+(1|PictureID))
print("Show -2LL as deviance and results")
llikAIC(Model11, chkREML=FALSE); summary(Model11)

      AIC        BIC     logLik   deviance   df.resid
 31927.312  31990.844 -15953.656  31907.312   4234.000  deviance = -2LL for model

Random effects:
 Groups    Name        Variance Std.Dev. Corr
 PersonID  (Intercept)   5.1981  2.2799        Var(U0p0)
           bright        2.3294  1.5262   0.094 Var(U0p3) Cor(U0p0,U0p3)
 PictureID (Intercept)   9.6383  3.1046        Var(U00i)
 Residual              102.2688 10.1128        Var(etpi)

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept) 17.49573    0.96794 34.03945 18.0751  < 2e-16  gamma000
clutter      0.97204    0.88839 30.90145  1.0942  0.28234  gamma001
relevant    -0.46742    1.57694 31.32332 -0.2964  0.76887  gamma002
bright      -4.86027    1.87311 32.33634 -2.5948  0.01411  gamma003
sign        -3.30638    1.22721 30.72930 -2.6942  0.01133  gamma004

print("LRT for random brightness slope over persons"); ranova(Model11)
                                npar  logLik     AIC     LRT Df Pr(>Chisq)
<none>                            10 -15953.7 31927.3
bright in (1 + bright | PersonID)  8 -15954.1 31924.2   0.844  2    0.65573
(1 | PictureID)                    9 -16081.4 32180.8 255.508  1    < 2e-16
```

Sample Results Section using R Output

[indicates notes about what to customize or also include; note that SE and p-values are not needed if you provide tables for the model solutions]

The extent to which item features could predict binary accuracy and continuous response time (RT; in seconds) in a change detection task was examined in a series of multilevel models predicting trial-level responses to 36 items from 155 persons. Binary accuracy was predicted using a logit link function and Bernoulli conditional outcome distribution, whereas response time was predicted directly (i.e., using an identity link function) using a normal conditional distribution. The binary outcome models were estimated via full-information marginal maximum likelihood (MML) using the Laplace method (via the glmer function in R lme4), whereas the continuous RT models were estimated via residual maximum likelihood (using the lmer function in R lme4). In the binary outcome models, all fixed effects should be interpreted as unit-specific (i.e., as the fixed effect specifically for persons and items in which their corresponding random effect = 0). The significance of fixed effects was evaluated with Wald tests (i.e., the ratio of each estimate to its standard error using no denominator degrees of freedom for accuracy, but using Satterthwaite denominator degrees of freedom for continuous RT), whereas the significance of random effects was evaluated via likelihood ratio tests (i.e., −2ΔLL with degrees of freedom equal to the number of new random effects variances and covariances). Effect size was evaluated via pseduo-R^2 values for the proportion reduction in each variance component for level-2 random item variance when appropriate [as well as odds ratios for individual slopes]. Person predictors were not examined [although they easily could be].

Accuracy: An empty means (no predictor) model with only a person random intercept variance had an intraclass correlation of ICC = .071 (using 3.29 as the logit-scale level-1 residual variance), indicating that 7.1% of the variance in accuracy was due to person mean differences, which was significant, −2ΔLL(1) = 85.18, p < .0001. The extent of differences in item easiness was initially examined by treating items as fixed effects (i.e., in which item ID was a categorical predictor as a factor variable, otherwise known as a "Rasch" psychometric model). Significant differences in item easiness were observed, −2ΔLL(35) = 692.96, p < .0001. To be able to quantify (and then predict) the extent of those item differences, we removed the categorical item ID predictor and instead added an item random intercept variance. The revised empty means model (with persons and items as crossed random effects at level 2) indicated that 7.9% of the variance in accuracy was due to person mean differences in ability and 20.6% was due to item mean differences in easiness; the latter was also significant, −2ΔLL(1) = 550.61, p < .0001. We then attempted to predict those item easiness differences by using four item feaures: continuous visual clutter (z-score metric), whether the change was relevant to driving (binary), continuous brightness of the change (z-score metric), and whether the change was made to a legible sign (binary). Although the four item feaures accounted for 24.8% of the item random intercept variance, only one predictor was uniquely significant: Changes to legible signs had higher accuracy (logit estimate = 0.739, SE = 0.336, p = .028). Significant variance in item easiness remained, as indicated by a model comparison without the item random intercept variance, −2ΔLL(1) = 399.585, p < .0001. Thus, items should be examined for further unmeasured characteristics that translate into differences in easiness. We then examined the potential for person differences in the effects of the item features (i.e., a random slopes over level-2 persons). Nonconvergence occurred when trying to estimate a person random slope variance for three of the four feaures; for brightness, a preson random slope variance did not improve model fit, −2ΔLL(2) = 1.590, p = .452. Consequently, it appears that the effects of the item features were largely comparable across persons.

Response Time: An empty means (no predictor) model with only a person random intercept variance had an intraclass correlation of ICC = .031 (using an estimated level-1 residual variance), indicating that 3.1% of the variance in accuracy was due to person mean differences, which was significant, −2ΔLL(1) = 35.16, p < .0001. Adding an item random intercept variance significantly improved model fit, −2ΔLL(1) = 463.64, p < .0001. The revised empty means model indicated that 4.2% of the variance in response time was due to person mean differences in ability, whereas 12.6% was due to item mean differences in easiness. We then attempted to predict those item easiness differences from four item feaures: continuous visual clutter (z-score metric), whether the change was relevant to driving (binary), continuous brightness of the change (z-score metric), and whether the change was made to a legible sign (binary). Although the four feaures accounted for 38.1% of the item random intercept variance, only two predictors were uniquely significant: Changes to legible signs had faster response times (estimate = −3.305, SE = 1.227, p = .011), and changes that were visually brighter also had faster response times (estimate = −4.867, SE = 1.869, p = .014). Significant variance in item easiness remained, as indicated by a model comparison without the item random intercept variance, −2ΔLL(1) = 255.508, p < .0001. Thus, items should be examined for further unmeasured characteristics that translate into differences in easiness. We then examined the potential for person differences in the effects of the item characteristics (i.e., random slopes over level-2 persons). A random slope for brightness did not improve model fit, −2ΔLL(2) = 0.844, p = .656. Consequently, it appears that the effect of change brightness were comparable across persons. [Other random slope tests could be reported as well.]