When comparing models, the main points you want to look out for are:

* **Are the models nested?**
  + If yes, proceed with the other questions; if not, you can’t compare deviances, and are stuck using something like AIC to determine the better model, using ML estimation.
* **Are the fixed effects the same? Or do just the random effects differ?**
  + If fixed effects are different (but nested), use ML estimation (i.e., specify lmer(…, REML=FALSE))! If *just* the random effects are different, you can use REML estimation, the default in lmer()
* **Are you comparing a non-random effects model (e.g., lm) to a random effects model?**
  + Make sure to fit the random effects model with ML estimation, since that’s what lm() does. Either compare the deviances manually, or use this specific model ordering w/anova(mixed model, fixed model)

**Example Scenario:** Subjects study a list of word pairs and then are tested with *both* a free recall test (“Recall as many word pairs as you can”) and a cued recall test (present with one word, generate the pair). Further, all subjects see negative, positive, and neutral word pairs.

**Example 1**

Model 1: lm(Recall ~ Task)

Model 2: lmer(Recall ~ Task + (1|Subject))

Estimation for models:

Comparison:

**Example 2**

Model 1: lmer(Recall ~ Task + (1|Subject))

Model 2: lmer(Recall ~ Task + Valence + (1|Subject))

Estimation for models:

Comparison:

**Example 3**

Model 1: lmer(Recall ~ Task + (1|Subject))

Model 2: lmer(Recall ~ Valence + (1|Subject))

Estimation for models:

Comparison:

**Example 4**

Model 1: lmer(Recall ~ Task + (1|Subject))

Model 2: lmer(Recall ~ Task + (1 + Task |Subject))

Estimation for models:

Comparison:

**Example 5**

Model 1: lmer(Recall ~ Task \* Valence + (1 + Task|Subject))

Model 2: lmer(Recall ~ Task \* Valence + (1 + Task\*Valence|Subject))

Estimation for models:

Comparison: