Full ML vs Restricted ML

Maximum likelihood estimation is used as an estimation technique in most multilevel programs. Note, though, that this can be applied in two slightly different ways.

From Kreft & de Leeuw, pp 131-133:

"Multilevel models describe the dependent variable *y*, and apply the principle of maximum likelihood to this model. The distribution of *y* is assumed to be normal, with a mean depending on the regression coefficients and a dispersion depending on the variance components. These are the parameters that are estimated by the corresponding technique, which is simply called maximum likelihood, but sometimes also *full maximum likelihood*.

Alternatively, we can apply the principle of maximum likelihood to the least-squares residuals. This is known as *restricted* or *residual maximum likelihood*, or REML. It means we first remove the effect of the fixed variables: remember that the residuals are uncorrelated with all the fixed variables in the model. The distribution of the residuals is also normal, because computing residuals from *y* just involves taking weighted sums. But the distribution of the residuals no longer depends on the estimates of the fixed effects, it only depends on the variance components."

When ML is used, a deviance statistic (- 2 ln *L*) is printed to the HLM output. The deviance is a measure of lack of fit between model and data, but not directly interpretable. Comparison of the deviances of two nested models is typically used to evaluate whether the one model contributes significantly to the explanation of variation in the outcome compared to the second model (see faq on deviances). Note that deviances obtained under REML can be used only if the two models compared have the *same* fixed parts and *differ only* in their random parts. If this is not the case, the deviance obtained using Full ML should be used instead.

Note the following (Bryk and Raudenbush, 1992, p 53):

"Posterior variances will be larger - and more realistic- under REML than under MLF. This will be especially true when the number of level-2 units is small. Under MLF, the fixed effects are assumed known so that the posterior variance of β_{qj} does not reflect uncertainty about them. This assumption is realistic only when the number of level-2 units, *J*, is large. We strongly caution the reader that regardless of whether we use MLF or REML, these tests will be too liberal, with actual significance values substantially exceeding the nominal values, unless *J* is large. "

For more on this topic, please refer to

- Kreft & de Leeuw (1998), pp. 131-133
- Snijders & Boskers (2000), pp. 88-90
- Bryk and Raudenbush (1992), pp. 45-46.