Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2001/05

Specification issues in stratified variance component ordinal response models

L. Grilli - C. Rampichini

Università degli Studi
di Firenze

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

Statistics

Specification issues in stratified variance component ordinal response models

Leonardo Grilli
Research assistant
Dept. of Statistics "G. Parenti", University of Florence
Viale Morgagni, 59, 50134 Florence, Italy
Tel: +39 0554237271
Fax: +39 0554223560
e-mail: grilli@ds.unifi.it

Carla Rampichini
Professor of Statistics
Dept. of Statistics "G. Parenti", University of Florence
Viale Morgagni, 59, 50134 Florence, Italy
Tel: +39 0554237246
Fax: +39 0554223560
e-mail: carla@ds.unifi.it

Abstract: The paper presents some criteria for the specification of ordinal variance component models when the second level units are grouped in few strata. The base model is specified using a latent variable approach, allowing the first level variance, the second level variance and the thresholds to vary according to the strata. However this model is not identifiable. The paper discusses some alternative assumptions that overcome the identification problem and illustrates a possible general strategy for the model selection. The proposed methodology is applied to the analysis of course program evaluations based on student ratings, referring to three different schools of the University of Florence. The adopted model takes into account both the ordinal scale of the ratings and the hierarchical nature of the phenomenon. In this framework, the identification of the latent variable distributions is crucial, since a different first level variance among the schools would change substantially the interpretation of model parameters. This is not the case in our application. Results show that both the latent average evaluation of the courses and the measurement scale vary with the school, suggesting to be careful in the interpretation of raw ratings based on an ordinal scale.
Keywords: ordinal response models, variance component models, varying thresholds, course program evaluation.

1 The standard model

Suppose that an observed ordinal response variable Y, with $k = 1, 2, \ldots, K$ levels, derives, through a set of thresholds, from a latent continuous variable $\tilde{Y}$ following a variance component model (Hedeker and Gibbons, 1994):

$$\tilde{y}_{ij} = \alpha + \beta x_{ij} + \tau u_j + \varepsilon_{ij}, \tag{1}$$

with $i = 1, 2, \ldots, n_j$ respondents for the j-th group ($j = 1, 2, \ldots, J$). In (1) α is the intercept; x_{ij} is a covariate and β the corresponding slope; the random variables ε_{ij} and u_j are the disturbances, respectively at the first (individual) and second (group) level; and σ^2 and τ^2 are the variance components, respectively at the first and second level.

For the disturbances of model (1) the usual hypotheses are:

(i) $E(\varepsilon_{ij}) = 0$ and $Var(\varepsilon_{ij}) = \sigma^2$;

(ii) $u_j \overset{iid}{\sim} N(0, 1)$;

(iii) the ε_{ij}'s and u_j's are mutually independent.

The observed ordinal variable Y is linked to the latent one $\tilde{Y}$ through the following relationship:

$$\{y_{ij} = k\} \quad \Leftrightarrow \quad \{\gamma_{k-1} < \tilde{y}_{ij} \leq \gamma_k\},$$

where the thresholds satisfy $-\infty = \gamma_0 \leq \gamma_1 \leq \ldots \leq \gamma_{K-1} \leq \gamma_K = +\infty$. Therefore, conditional to u_j, the model probabilities are:

$$\begin{aligned}
& P(y_{ij} = k) \\
= \; & P(\gamma_{k-1} < \tilde{y}_{ij} \leq \gamma_k) \\
= \; & P(\tilde{y}_{ij} \leq \gamma_k) - P(\tilde{y}_{ij} \leq \gamma_{k-1}) \\
= \; & P(\varepsilon_{ij} \leq \gamma_k - [\alpha + \beta x_{ij} + \tau u_j]) - P(\varepsilon_{ij} \leq \gamma_{k-1} - [\alpha + \beta x_{ij} + \tau u_j]) \\
= \; & P\left(\frac{\varepsilon_{ij}}{\frac{\sigma}{c}} \leq \frac{\gamma_k}{\frac{\sigma}{c}} - \left[\frac{\alpha}{\frac{\sigma}{c}} + \frac{\beta}{\frac{\sigma}{c}} x_{ij} + \frac{\tau}{\frac{\sigma}{c}} u_j\right]\right) + \\
& -P\left(\frac{\varepsilon_{ij}}{\frac{\sigma}{c}} \leq \frac{\gamma_{k-1}}{\frac{\sigma}{c}} - \left[\frac{\alpha}{\frac{\sigma}{c}} + \frac{\beta}{\frac{\sigma}{c}} x_{ij} + \frac{\tau}{\frac{\sigma}{c}} u_j\right]\right) \\
= \; & F\left(\frac{\gamma_k}{\frac{\sigma}{c}} - \left[\frac{\alpha}{\frac{\sigma}{c}} + \frac{\beta}{\frac{\sigma}{c}} x_{ij} + \frac{\tau}{\frac{\sigma}{c}} u_j\right]\right) - F\left(\frac{\gamma_{k-1}}{\frac{\sigma}{c}} - \left[\frac{\alpha}{\frac{\sigma}{c}} + \frac{\beta}{\frac{\sigma}{c}} x_{ij} + \frac{\tau}{\frac{\sigma}{c}} u_j\right]\right) \\
= \; & F(\gamma_{\sigma,k} - [\alpha_\sigma + \beta_\sigma x_{ij} + \tau_\sigma u_j]) - F(\gamma_{\sigma,k-1} - [\alpha_\sigma + \beta_\sigma x_{ij} + \tau_\sigma u_j]),
\end{aligned}$$

where $F(\cdot)$ is the distribution function of the "standardized" first level error term $\varepsilon_{ij}\frac{c}{\sigma}$, which has variance c^2. The value of the constant c is arbitrarily chosen, usually on the basis of $F(\cdot)$: typical choices are $c = 1$ for the normal distribution, $c = \sqrt{\frac{\pi^2}{3}}$ for the logistic distribution and $c = \sqrt{\frac{\pi^2}{6}}$ for the complementary log-log distribution.

Note that all the model parameters are defined in terms of $\frac{\sigma}{c}$, the standard deviation of the "standardized" first level error term, which depends on the unknown σ (this fact is denoted by the presence of the symbol σ in the subscript of the parameters). Thus only the ratios of the model parameters to the standard deviation of the first level error term are identifiable; one popular way to overcome this identifiability problem is to fix σ, usually to 1.

2 Model specification in presence of strata

Suppose now that the second level units can be grouped in few strata $h = 1, \ldots, H$. If the number of strata is small (say, below ten), this new third level is most appropriately modelled by allowing the parameters to vary among the strata. Denoting with the superscript (h) the stratum to which the quantities are referred, the model for stratum h is:

$$\tilde{y}_{ij}^{(h)} = \alpha^{(h)} + \beta^{(h)} x_{ij} + \tau^{(h)} u_j + \varepsilon_{ij}^{(h)}, \tag{2}$$

with the following hypotheses:

(i) $E(\varepsilon_{ij}^{(h)}) = 0$ and $Var(\varepsilon_{ij}^{(h)}) = \sigma^2 (1 + \theta^{(h)})^2$;

(ii) $u_j \overset{iid}{\sim} N(0, 1)$;

(iii) the $\varepsilon_{ij}^{(h)}$'s and u_j's are mutually independent.

Therefore the parameters of model (2) are

$$\begin{array}{l} \alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(H)}, \\ \tau^{(1)}, \tau^{(2)}, \ldots, \tau^{(H)}, \\ \sigma, \theta^{(2)}, \ldots, \theta^{(H)}. \end{array}$$

Note that the $\theta^{(h)}$'s are intended to measure the difference in the first level variance from the reference stratum $h = 1$, for which it is $\theta^{(1)} \equiv 0$.

Consequently, conditional to u_j, the probabilities of model (2) for stratum h are:

$$\begin{aligned} &= P\left(\tilde{y}_{ij}^{(h)} \leq \gamma_k^{(h)}\right) = P\left(\frac{\varepsilon_{ij}^{(h)}}{\frac{\sigma}{c}(1+\theta^{(h)})} \leq \right. \\ &\left. \leq \frac{\gamma_k^{(h)}}{\frac{\sigma}{c}(1+\theta^{(h)})} - \left[\frac{\alpha^{(h)}}{\frac{\sigma}{c}(1+\theta^{(h)})} + \frac{\beta^{(h)}}{\frac{\sigma}{c}(1+\theta^{(h)})} x_{ij} + \frac{\tau^{(h)}}{\frac{\sigma}{c}(1+\theta^{(h)})} u_j\right]\right) \\ &= F\left(\frac{\gamma_{\sigma,k}^{(h)}}{(1+\theta^{(h)})} - \left[\frac{\alpha_\sigma^{(h)}}{(1+\theta^{(h)})} + \frac{\beta_\sigma^{(h)}}{(1+\theta^{(h)})} x_{ij} + \frac{\tau_\sigma^{(h)}}{(1+\theta^{(h)})} u_j\right]\right) \\ &= F\left(\gamma_{\sigma,k}^{(h)^*} - \left[\alpha_\sigma^{(h)^*} + \beta_\sigma^{(h)^*} x_{ij} + \tau_\sigma^{(h)^*} u_j\right]\right), \end{aligned}$$

where the superscript $(h)^*$ indicates that the parameter is in relative terms, i.e. divided by $(1 + \theta^{(h)})$.

The h^*-parameters can be easily estimated by allowing the intercept, the second level variance and the thresholds to vary among the H strata.

However some additional assumptions are to be made in order to estimate the original h-parameters, which are the ones of interest. For example, if a certain stratum has a different h^*-intercept, it can be that the true h-intercept is different, but it can also be that its level one variance is different, or a mixture of the two cases.

Three possible assumptions that overcome this identifiability problem are the following:

1. $\theta^{(2)} = \theta^{(3)} = \ldots = \theta^{(H)} = 0$ (common first level variance): in this case the h-parameters are the same as the h^*-parameters.

2. $\tau_\sigma^{(1)} = \tau_\sigma^{(2)} = \ldots = \tau_\sigma^{(H)}$ (common second level variance): in this case the parameters $\theta^{(h)}$ can be estimated from the following identity:

$$\frac{\tau_\sigma^{(1)^*}}{\tau_\sigma^{(h)^*}} = \frac{\tau_\sigma^{(1)}}{\frac{\tau_\sigma^{(h)}}{(1+\theta^{(h)})}} = \frac{\tau_\sigma^{(1)}}{\tau_\sigma^{(h)}}(1+\theta^{(h)}) = 1+\theta^{(h)}.$$

   The original intercepts and thresholds are then easily calculated by multiplying the h^*-parameters by $(1+\theta^{(h)})$. For example,

$$\alpha_\sigma^{(h)} = \alpha_\sigma^{(h)^*}(1+\theta^{(h)}). \tag{3}$$

3. $\beta_\sigma^{(1)} = \beta_\sigma^{(2)} = \ldots = \beta_\sigma^{(H)}$ (common regression coefficient): we can proceed like in case 2, using the following identity:

$$\frac{\beta_\sigma^{(1)^*}}{\beta_\sigma^{(h)^*}} = \frac{\beta_\sigma^{(1)}}{\frac{\beta_\sigma^{(h)}}{(1+\theta^{(h)})}} = \frac{\beta_\sigma^{(1)}}{\beta_\sigma^{(h)}}(1+\theta^{(h)}) = 1+\theta^{(h)}.$$

Often the researcher acts as in case 1, tacitly assuming identical first level variances. It should be stressed that such an assumption, that is crucial for the interpretation of the results, is not testable and its validity in the data at hand is difficult to assess. However, there are other ways to proceed. The second choice (identical second level variances) is similar to the first one, since it simply shifts the assumption from the first to the second level variance. But the third choice (identical regression slopes) is somewhat different, since such an assumption concerns not a variance parameter, but an association parameter and is consequently more easy to justify. In fact, it is more common to have some a priori knowledge on the regression coefficients than on the variances. Moreover, the validity of such an assumption in the data at hand can be investigated through some technique which can help to explore the association among the latent variable and the covariate of interest. For example, one might assign a set of scores to the levels of the ordinal variable, estimate a separate regression slope for each of the H strata and then compare the slopes; when also the covariate is ordinal, a closely related technique is to compare the H Spearman correlations between the response variable and the covariate. This strategy has a theoretical justification in the well-known fact that association parameter estimators are usually more robust to model misspecifications than variance parameter estimators. In particular, Fielding (1999) found that, for the fixed regression coefficients, the ordinal variance component model leads essentially to the same conclusions as the linear variance component model on the scores of the ordinal variable (using various scoring systems); on the other hand, the conclusions on the variance components are significantly different.

A possible general strategy is the following:

1. Choose a covariate which has a slope sufficiently stable among the H strata (using a priori information, separate regression lines, Spearman correlations or the like).

2. Fit the most general model in which all the parameters (the h^*- parameters in our notation) are allowed to vary among the H strata.

3. Use the estimated $\beta_\sigma^{(h)^*}$'s to obtain an estimate of the $\theta^{(h)}$'s.

4. Test the hypothesis that the $\theta^{(h)}$'s are jointly null; if such an hypothesis is rejected, perform a sequence of tests to identify the subset of $\theta^{(h)}$'s which are significantly different from zero.

5. If the hypothesis that all the $\theta^{(h)}$'s are null is not rejected, then go on with model selection in the usual manner (the interpretation of the results is straightforward, since in this case the h-parameters equal the h^*-parameters).

6. Otherwise, for the strata whose corresponding $\theta^{(h)}$ is significantly different from zero, it is necessary to correct the h^*-estimates with an estimate of the factor $(1 + \theta^{(h)})$, like in (3). In this case the model selection should be modified to take into account the restrictions on

the parameters. For example, if $\theta^{(2)}$ is different from zero, testing the hypothesis $\alpha_\sigma^{(1)} = \alpha_\sigma^{(2)}$ amounts to testing the hypothesis

$$\frac{\alpha_\sigma^{(1)*}}{\alpha_\sigma^{(2)*}} = \frac{\beta_\sigma^{(1)*}}{\beta_\sigma^{(2)*}} \tag{4}$$

One way of testing (4) is to carry out a Wald test with the aid of the delta method, thought this technique is not always adequate (Godfrey, 1991). Alternatively, one should fit a restricted model which satisfies the non linear constraint (4) and carry out a deviance test.

3 The data

We apply the proposed method to the data gathered in the survey on course evaluation carried out by the University of Florence, in all the schools of the University, for classes in the second semester of the 1999-2000 academic year. Specifically, we will refer to the results from the schools of Engineering, Science and Letters.

The data have a hierarchical structure: respondents are nested in courses that are nested in schools. The total number of groups, represented by the courses, is 370, while the number of strata, represented by the schools, is three. Table 1 reports, for each school, the number of respondents, the number of courses evaluated and the minimum, median and maximum number of respondents per course.

Table 1: Number of respondents, number of courses evaluated and minimum, median and maximum number of respondents per course. The University of Florence, academic year 1999-2000, second semester.

School	*N. respondents*	*N. courses*	*Respondents per course*		
			min	*median*	*max*
Engineering	3165	150	4	16	71
Science	1633	103	4	13	52
Letters	1932	117	3	10	118
TOT	6730	370	3	13	118

In the present application we focus on the item relative to the overall satisfaction, which required a response on a 4-level ordinal scale: 1) decidedly no; 2) more no than yes; 3) more yes than no; 4) decidedly yes. The aim of the analysis it to establish if the different evaluations expressed by the students in the three schools might, to some extent, be attributed to a different "measurement scale", i.e. to a different way of interpreting the levels of the ordinal scale (obviously, the evaluations expressed by each student are influenced also by their characteristics and expectations).

4 Empirical results

The first step of the analysis is the estimation of the most general model presented in section 2, that is the model allowing the intercept, the second and first level variance and the thresholds to vary among the schools, with the first threshold for all the schools fixed to 0 and the first level variance for Engineering fixed to 1. In the present application we always assume that the first level disturbances have a Gaussian distribution, leading to a probit model specification.

In order to identify the parameters, we introduce into the model a covariate with an assumed common slope. In this case the covariate is the answer (on a 4-level ordinal scale) to the question whether the student will take the exam at the first examination session (covariate *exam*). The choice of this covariate is motivated by our knowledge of the phenomenon, since we have no reason to suppose a differential effect on the latent evaluation among the schools. Moreover, this covariate shows very similar values of the Spearman correlation coefficient with the overall satisfaction among the three schools (0.32, 0.34 and 0.36 respectively, for Engineering, Science and Letters).

The estimation is carried out by the NLMIXED procedure of the SAS software (SAS Institute, 1999), which performs a dual quasi-Newton optimization with adaptive Gaussian quadrature. Table 2 presents the results relative to various models.

Comparing models 1 and 2 in terms of deviance, the hypothesis of equal first level variances among the schools is not to be rejected. Model selection can now proceed in the usual way. First of all, from model 2 it seems that Engineering and Science have the same thresholds and second level variance

Table 2: Results of model selection

	Model						
	with covariate					without cov.	
	1	2	3	4	5	6	7
Intercept							
α^E	0.3160	0.2539	0.2549	0.2479	0.1636	1.4704	1.4745
$\alpha^S - \alpha^E$	*0.1305*	0.2301	0.2274	0.2244	0.2141	0.2578	0.2425
$\alpha^L - \alpha^E$	*0.1149*	0.2450	0.2444	0.2909	0.6143	0.2410	0.2370
Exam							
β^E	0.4260	0.4436	0.4434	0.4428	0.4459		
β^S	0.4692	”	”	”	”		
β^L	0.4572	”	”	”	”		
Rand. par.							
θ^S	*-0.0801*						
θ^L	*-0.1035*						
τ^E	0.7505	0.7525	0.7208	0.6692	0.7096	0.7620	0.7596
τ^S	0.6902	0.6907	”	”	0.6587	0.7554	”
τ^L	0.5000	0.5008	0.5006	”	0.5788	0.5600	0.5600
ρ^E		0.3615	0.3464	0.3093	0.3349	0.3674	0.3659
ρ^S		0.3220	”	”	0.3026	0.3633	”
ρ^L		0.2005	0.2004	”	0.2509	0.2387	0.2387
Thresholds							
γ_2^E	0.9984	1.0022	1.0131	1.0081	0.9603	0.9360	0.9488
γ_2^S	1.0408	1.0.388	”	”	”	0.9787	”
γ_2^L	0.8032	0.7985	0.7984	0.8070	”	0.7226	0.7226
γ_3^E	2.5334	2.5438	2.5368	2.5236	2.3856	2.3751	2.3736
γ_3^S	2.5333	2.5278	”	”	”	2.3756	”
γ_3^L	2.0132	2.0022	2.0021	2.0279	”	1.8299	1.8299
N. par.	15	13	10	9	9	12	9
-2 logL	14649	14651	14653	14663	14717	15445	15446

The superscript denotes the school: E=Engineering, S=Science, L=Letters; the estimates in italics are not significant at the 95% level; the symbol ” indicates that the value is, by definition, equal to the value in the above cell.

and this impression is confirmed by the very small increase of the deviance when fitting model 3. Further simplifications of the model are not supported by the data: for example, Table 2 reports the estimates for the model with constant second level variance and varying thresholds for Letters (model 4) and for the model with varying second level variance but fixed thresholds (model 5).

Note that imposing fixed thresholds causes an important loss of fit, so the data strongly support the hypothesis of different "measurement scales" among the students of the three schools. The consequences of this fact can be appreciated by comparing models 3 and 5: in model 5 the higher ratings obtained by the courses of Letters with respect to those of Engineering are totally attributed to higher latent evaluations ($\alpha_L - \alpha_E = 0.6143$), while in model 3 are attributed partly to higher latent evaluations ($\alpha_L - \alpha_E = 0.2444$) and partly to a more favorable "measurement scale" (the lower values of the thresholds for Letters imply that, for the same latent evaluation, the expressed rating on the ordinal scale is greater or equal).

Another interesting feature of model 3 is that, although the first level variance is constant among the schools, the second level one is not, with Letters having a significantly lower value: 0.5006 versus 0.7208. The intraclass correlation coefficient, $\rho = \tau/(1 + \tau)$, is 0.2004 for Letters and 0.3464 for the other schools: this means that in the school of Letters the proportion of variance attributable to the courses is substantially lower and, consequently, the student ratings have a lesser discriminant power.

Note that the conclusions just outlined heavily rely on the hypothesis of constant first level variance among the schools, highlighting the practical importance of devising a procedure to assess the validity of such an hypothesis.

Since in the present application the inclusion of the covariate *exam* is instrumental and not of direct interest, after testing for equal first level variances we also performed the model selection without the covariate: Table 2 reports the fitting of models 6 and 7, which are the no-covariate counterparts of models 2 and 3, respectively. As was to be expected, the omission of the covariate *exam* leads to the same substantive conclusions, since its effect is approximately constant among the strata. Figure 1 represents, for each school, the marginal distribution and the mean of the latent variable and the corresponding thresholds estimated from Model 7 of Table 2. It is worth noting that, with respect to the distribution of Engineering, the distribution of Science is simply shifted on the right, while the distribution of Letters is shifted on the right and has a lower variance. Moreover, the thresholds of

Letters are shifted on the left, with the last threshold being almost equal to the mean, so that the area under the density function on the right of the last threshold (i.e. the model estimated proportion of very satisfied students) is about 0.5.

5 Concluding remarks

The paper has discussed the issues that arise in the specification of ordinal variance component models in presence of strata, with an application to the analysis of student ratings.

In the paper we suggest to overcome the identification problem due to the stratum-dependent first level variance by introducing a covariate with a common slope among the strata. In a specific application, this choice can be justified on the grounds of prior knowledge and, to a lesser extent, on the basis of techniques which attempt to describe the behavior of the latent variable. The reliability of these techniques in various practical situations need to be assessed through a careful simulation study.

It should be noted that the same identification problems affect the standard one level ordinal model; the role of the second level variance is to lead to a more realistic model for the phenomenon under study, also allowing a broader and more interesting discussion.

Since in our application the use of the covariate *exam* was merely instrumental, Table 2 also presents the results from the model with no covariate. Indeed, the exclusion of the covariate used to test for equal first level variances is not expected to modify the conclusions on such an hypothesis; on the contrary, the inclusion of further covariates is potentially harmful, especially if the new covariates have varying effects among the strata. To avoid such problems the hypothesis of equal first level variances should be tested in the more general model (i.e. the model including all the relevant covariates), imposing the appropriate restrictions on the slopes of the covariates which are assumed to have a constant effect among the strata.

As for the distributional form of the disturbances, the trials we made suggest that this issue is not crucial in the present application. However, it would be very useful to carry out a sensitivity analysis and to develop a formal procedure for the selection of the distributional form; such a selection is particularly relevant for the first level disturbances, whose distribution determines the link function of the model.

Acknowledgements

Financial support from MURST research projects “Evaluation of quality and effectiveness in human services, with particular attention to health and education” and “Statistical education and evaluation: instruments, methods and new technologies” is gratefully acknowledged.

References

[1] Fielding A. (1999) Why use arbitrary point scores?: ordered categories in models of educational progress, *JRSS A* **162**, 303-328.

[2] Godfrey L.G. (1991) *Misspecification test in econometrics: the Lagrange multiplier principle and other approaches*. New York: Cambridge University Press.

[3] Hedeker D., Gibbons R.D. (1994) A random-effects ordinal regression model for multilevel analysis. *Biometrics* **50**, 933-944.

[4] SAS Institute (1999) *SAS/STAT User's Guide*, Version 8., Cary: SAS Institute Inc.

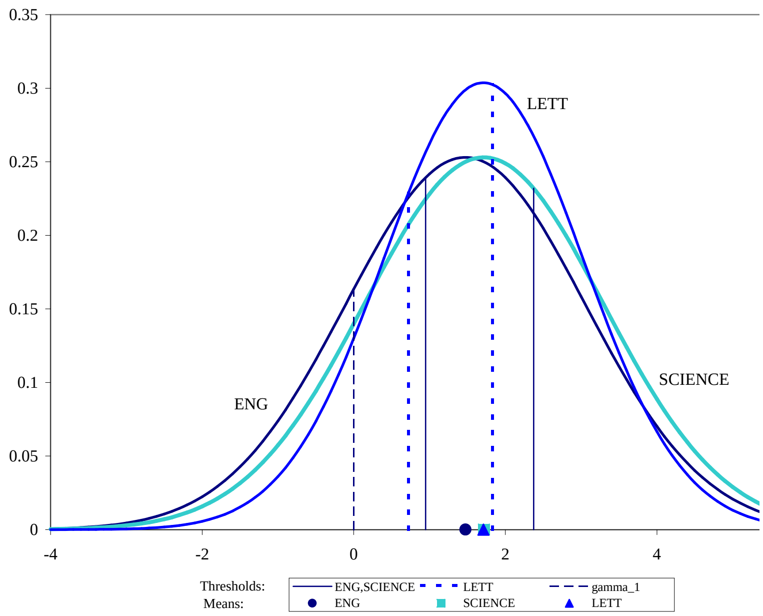

Figure 1: Latent variable distributions, thresholds and means by school from Model 7 of Table 2.