Identifying Inter-subject Difficulties in Norwegian GPA Data Using Item Response Theory

Tony C. A. Tan

Centre for Educational Measurement, University of Oslo

Continuous Draft

Prof Rolf V. Olsen & Dr Astrid M. J. Sandsør

Autumn 2021

Identifying Inter-subject Difficulties in Norwegian GPA Data Using Item Response Theory

Ever since men walked on this Earth, we have always been wondering about one thing: What's for dinner? (Coe, 2008; He et al., 2018; Korobko et al., 2008)

Theoretical Framework

The Norwegian GPA System

Missing Data Treatment

IRT item parameter estimation demands strong assumptions on the data missing mechanism. Joint, conditional and marginal maximum likelihood procedures are only valid under "ignorable non-response" conditions where missing responses are related to neither item nor person parameters (Molenaar, 1995). When test takers skip items after seeing their content, for example, the ignorablity condition is unlikely to hold (Mislevy & Stocking, 1987), neither are items not reached due to time limitation (Lord, 1974, 1983).

Methods

Sample

For this study, students' GPA records will be extracted from the Norwegian registry covering the period between 2009 (first year post-2006 reform data became available) and 2019 (last "normal" year before COVID). GDPR registration is lodged through the NSD Portal and the UiO ethics approval is also obtained. All data import, storage, and analyses are to be conducted within the secured infrastructure TSD provided by the UiO Central IT Division. TSD logs all activities and no data or results can be copied out of the restricted system without prior approval from project leaders.

Under the advisory of He et al. (2018), subjects with fewer than 1,000 candidates and students taking fewer than two GPA subjects will be excluded from subsequent analyses. Each year's record (score matrix) will contain N rows representing the number of valid candidates and L columns reflecting the usable number of GPA subjects in that year. Since no student took all the GPA subjects, a large proportion of the score matrices will remain missing by design. The existence of missing data does not pose any problems for using the Rasch model as the model functions at the individual subject and as long as there is sufficient overlap across subjects in the score matrix. The ability to deal with incomplete data is one major advantage of using the Rasch model for studying inter-subject comparability.

Missing Value Treatment

Missing patterns are not missing at random. If a candidate chose to do physics, he was also highly likely to have chosen advanced maths. So the presence and absence of data tend to group in clusters.

Subject Choice

This study explicitly models candidates' subject choice decisions by introducing an indicator variable d_{ni} such that

$$d_{ni} = \begin{cases} 1 \text{ if Candidate } n \text{ chose Subject } i \\ 0 \text{ if Candidate } n \text{ did not choose Subject } i, \end{cases} \tag{1}$$

for Candidate $n = 1, \ldots, N$ and GPA Subject $i = 1, \ldots, L$.

Generalised Partial Credit Model (GPCM)

A unidemensional generalised partial credit model (Muraki, 1992) with the probability that Candidate n's score in Subject i (x_{ni}) being Grade j ($j = 0, \ldots, m$) is given by

$$p\left(x_{ni} = j | d_{ni} = 1; \theta_n\right) = \frac{\exp\left\{j\alpha_i\theta_n - \sum_{h=1}^{j} \beta_{ih}\right\}}{1 + \sum_{h=1}^{m} \exp\left\{h\alpha_i\theta_n - \sum_{k=1}^{h} \beta_{ik}\right\}}, \tag{2}$$

where θ_n is the unidemensional proficiency parameter that represents the overall proficiency of Candidate n.

Log-likelihood

In MML, a likelihood function (ℓ) is maximised where the candidates' proficiency parameters (θ) are integrated out of the likelihood. The marginal log-likelihood for a unidemensional GPCM is given by

$$\ell_{\text{unidimensional}} = \sum_{p} \sum_{n|p} \log \int \prod_{i} p(x_{ni}|d_{ni}; \theta) g(\theta; \mu_p, \sigma^2) \mathrm{d}\theta, \tag{3}$$

where x_{ni} is the observed grade, $p(\cdot)$ is equal to Equation (2) evaluated at x_{ni} if $d_{ni} = 1$, and $p(\cdot) = 1$ if $d_{ni} = 0$. In addition, $g(\theta; \mu_p, \sigma^2)$ is the normal pdf with mean μ_p and variance σ^2. The model can be identified by choosing a standard normal $\mathcal{N}(0, 1)$ (Korobko et al., 2008).

Multidimensionality

There exists strong believes among educational scientists that learners' proficiency is multidimensional, such as one proficiency factor for STEM subjects, for example, and another one for languages. If F proficiency dimensions are required to model the grades, the proficiency can be represented by a vector of proficiency parameters $\boldsymbol{\theta}_n = (\theta_{n1}, \cdots, \theta_{nF})^\intercal$ with the corresponding GPCM:

$$p\left(x_{ni} = j | d_{ni} = 1; \boldsymbol{\theta}_n\right) = \frac{\exp\left\{j\left(\sum_{f=1}^{F} \alpha_{if}\theta_{nf}\right) - \sum_{h=1}^{j} \beta_{ih}\right\}}{1 + \sum_{h=1}^{m} \exp\left\{h\left(\sum_{f=1}^{F} \alpha_{if}\theta_{nf}\right) - \sum_{k=1}^{h} \beta_{ik}\right\}}. \tag{4}$$

with $\boldsymbol{\theta}_n$ following a multivariate normal distribution with mean $\boldsymbol{\mu}_p$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Similar to the unidemensional case, Equation (4) is identified by setting $\boldsymbol{\mu}_p = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}$ the identity matrix. The log-likelihood of a multidimensional GPCM then becomes:

$$\ell_{\text{multidimensional}} = \sum_{p} \sum_{n|p} \log \int \cdots \int \prod_{i} p(x_{ni} | d_{ni}; \boldsymbol{\theta}) g(\boldsymbol{\theta}; \boldsymbol{\mu}_p, \boldsymbol{\Sigma}) \mathrm{d}\boldsymbol{\theta}, \tag{5}$$

with each component sharing similar interpretations to the unidemensional counterpart in Equation (3).

Interaction between Subject Choice and Proficiency

Under the advisory of Korobko et al. (2008), a latent variable θ^+ is introduced to reflect student's propensity of choosing a particular subject. Augmenting θ^+ to $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_F)^\intercal$ yields $\boldsymbol{\theta}^+ = (\theta_1, \cdots, \theta_F, \theta^+)^\intercal$, with a corresponding marginal likelihood:

$$\ell_{\text{interaction}} = \sum_{p} \sum_{n|p} \log \int \cdots \int \prod_{i} \left[p\left(x_{ni} | d_{ni}; \boldsymbol{\theta}\right) p\left(d_{ni}; \theta^+\right)\right] g(\boldsymbol{\theta}^+; \boldsymbol{\mu}_p, \boldsymbol{\Sigma}) \mathrm{d}\boldsymbol{\theta}^+. \tag{6}$$

Results

Model 1

Model 2

Model 3

Lots of tables here.

Discussions

What does all this mean? Well, let me make you a cup of tea first.

References

Coe, R. (2008). Comparability of GCSE examinations in different subjects: An application of the Rasch model. *Oxford Review of Education*, *34*(5), 609–636. https://doi.org/10.1080/03054980801970312

He, Q., Stockford, I., & Meadows, M. (2018). Inter-subject comparability of examination standards in GCSE and GCE in England. *Oxford Review of Education*, *44*(4), 494–513. https://doi.org/10.1080/03054985.2018.1430562

Korobko, O. B., Glas, C. A. W., Bosker, R. J., & Luyten, J. W. (2008). Comparing the difficulty of examination subjects with item response theory. *Journal of Educational Measurement*, *45*(2), 139–157. https://doi.org/10.1111/j.1745-3984.2007.00057.x

Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, *39*(2), 247–264. https://doi.org/10.1007/bf02291471

Lord, F. M. (1983). Maximum likelihood estimation of item response parameters when some responses are omitted. *psychometrika*, *48*(3), 477–482. https://doi.org/10.1007/bf02293689

Mislevy, R. J., & Stocking, M. L. (1987). A consumer's guide to LOGIST and BILOG. *ETS Research Report*, *1987*(2), 1–73. https://doi.org/10.1002/j.2330-8516.1987.tb00247.x

Molenaar, I. W. (1995). Estimation of item parameters. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 39–51). Springer-Verlag. https://doi.org/10.1007/978-1-4612-4230-7_3

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *ETS Research Report Series*, *1992*(1), 1–30. https://doi.org/10.1002/j.2333-8504.1992.tb01436.x