

**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

Vår 2022

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

Abstract

Research Topic

Grade point averages (GPA) play monopolistic roles in Norway’s tertiary admission processes. Earlier studies from the UK (He et al., 2018) and the Netherlands (Korobko et al., 2008), however, raised methodological and fairness concerns over GPA as an appropriate measure for graduates’ academic competency. Violations of the unidimensionality assumption arose when different subjects contribute to the final GPA scores at different weights, causing invalid statistical inferences under the item response theory (IRT) framework. Additionally, misaligned subject difficulties distort candidates’ incentives, leading to material misallocation of youth’s time and effort at a critical point in their studies. This paper aims to examine whether Norway’s GPA subjects exhibit comparable difficulty levels, both across candidate cohorts (e.g., medical school vs general tertiary applicants) and across time. It further investigates covariates that associated strongly with any discrepancies in subject difficulties for policy considerations.

Theoretical Framework

IRT is particularly suitable in the educational measurement literature for extracting item difficulty parameters. This study considers each GPA subject as an IRT item and each candidate as an IRT person. It primarily focuses on the item parameters β while treating person competencies θ as “nuisance parameters” by integrating them out using marginal maximum likelihood estimates. Additionally, since students had self-selected into GPA subjects with highest expected payoffs, the observed GPA datasets are reasonably expected to involve the missing-not-at-random (MNAR) mechanism. Leaving untreated, such non-ignorable missingness would cause over- and under-estimates of person and item parameters, respectively (Rose, 2013). This study addresses MNAR using a multiple imputation procedure prior to IRT analyses.

Methodology

Expected Results

Based on their similarities in design and administration, this study expects comparable patterns of inter-subject difficulties in Norway's GPA and in the UK's GCSEs and A Levels. More specifically, we expect GPA grades to misalign by one grade between the easiest and the hardest subject following He et al.'s (2018) results. Missing data correction is expected to produce material differences in estimations based on the Dutch study using Central Examinations in Secondary Education data by Korobko et al. (2008).

Relevance to Nordic Educational Research

References

- 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>
- Rose, N. (2013). *Item nonresponses in educational and psychological measurement* [PhD Thesis, Friedrich-Schiller-Universität Jena]. Open Access Thesis and Dissertations. https://www.db-thueringen.de/servlets/MCRFileNodeServlet/dbt_derivate_00027809/Diss/NormanRose.pdf