**Tentative Title**

Identifying inter-subject difficulties in Norwegian GPA data

**Summary**

Upholding fairness is one major objective in every educational assessment process. High-stake decisions such as school graduation and university admission in Norway rely heavily on the point grade average (GPA) system as a sum-score measure that maps students’ competency into a numeric value. Concerns arise however over the construct-irrelevant variance such that not only candidates’ personal competence, but also characters relating to subject difficulties, can affect GPA scores. This paper investigates inter-subject difficulties using 2019 Year 10 academic achievement data from Norwegian national register using a Rasch model. Results are consistent with prior reports from the UK and the Netherlands that Norway’s Year 10 subjects did differ in difficulties, particularly so towards the low-performer end among teacher-assign grades and higher end among exam grades. This conclusion calls a rethinking about GPA’s construct validity and its potential fairness implication among specific student cohorts.

**Research Questions**

RQ1: Do Norwegian Year 10 subjects differ in their difficulty levels?

RQ2: Do subject difficulties differ by source such as between teachers and external examiners?

RQ3: Do subject difficulties differ across achievement levels?

RQ4: Do subject difficulties differ across demographic attributes such as socioeconomic status, gender, and immigration background?

**Main Data**

Norwegian register 2019 grunnskole academic record, TSD file #157: W21\_4952\_TAB\_KAR\_GRS.csv

**Main Analytical Approaches**

IRT, Rasch model, partial credit model

**Tentative Title**

IRT parameter estimations with non-ignorable missing data: A case study using Norwegian GPA data

**Summary**

Conventional item response theory (IRT) relies on the "ignorability" assumption behind any missing data. Such foundation appears no longer solid in Norwegian high school graduation tests where observed grades result from students choosing subjects with highest payoffs. Such self-selection yields under-estimations of subject difficulties and over-estimations of person competencies. Consensus remains scarce over the appropriate IRT procedures in the presence of missing-not-at-random (MNAR) data. This paper compares joint modelling (JM) against multiple imputation (MI) approaches in producing unbiased IRT parameters with non-ignorable missing data. Using 2019 Norwegian senior high school grades, we show that MI is superior to JM in terms of bias reduction and efficiency. A two-stage procedure for estimating subject difficulties and graduates' competencies are proposed for the purpose of enhancing fairness in educational assessment.

**Research Questions**

RQ1: Does the presence of missing-not-at-random (MNAR) among Norwegian VG3 GPA statistically alter conventional IRT estimates? If so, in which direction and by how much?

RQ2: Does joint modelling (JM) correct IRT estimation biases resultant from MNAR process?

RQ3: Does multiple imputation (MI) correct IRT estimation biases resultant from MNAR process?

RQ4: Between JM and MI, which procedure is more efficient for the purpose of restoring IRT estimation of item- and person-parameters?

**Main Data**

Norwegian register 2019 VG3 academic record, TSD file #158: W21\_4952\_TAB\_KAR\_VG.csv

**Main Analytical Approaches**

IRT (both “conventional” and in the presence of “non-ignorable” missings)

Missing data treatment: JM (Schafer, 1997) and MI (Little & Rubin, 2019; van Buuren, 2019)

**Tentative Title**

A difference-in-differences analysis on the impact of COVID-19 school closures on Norwegian students’ national test performance

**Summary**

The sudden suspension of in-person learning in early 2020 brought about profound disruptions in youth’s academic and social development. Numeral studies have been devoted into investigating the magnitudes of learning loss resultant from school closures, but their results are hardly converging due to data and/or research design issues. To overcome sampling problems, this study uses the entire Norwegian student population as its main data source, and it employs a difference-in-differences research design to reach a causal conclusion. Results suggest that Norway’s school closures did slow down national test grade growth. Among various demographic attributes, socioeconomic status stood out as the most influential factor in explaining the magnitude of growth slow-down. This paper lends weight to post-COVID remedial measures by identifying the most vulnerable learner groups.

**Research Questions**

RQ1: Did national test grades grow at a comparable rate in 2020 than in previous years?

RQ2: If the answer to RQ1 is “no”, how much of the slow-down can be attributed to school closures?

RQ3: Which learner groups were most disadvantaged by school closures?

**Main Data**

Norwegian register 2016--2022 national test data, TSD file #100: W21\_4952\_NATIONALE\_PROVER.csv

**Main Analytical Approaches**

Difference-in-differences regression design

Cross-classification multi-level design

**Tentative Title**

From grades to learning: A Bayesian approach to competency estimation

**Summary**

Grades and learnings are not interchangeable concepts. While a good test inventory provides close approximation to the given amount of students’ learning, this process does not answer students’ amount of learning given the grades signal. The Bayes formular links “learning given grades” and “grades given learning” systematically and is able to accommodate large-scale missing data such as the complete absence of exam grades due to COVID school closures using partial information updates.

**Research Questions**

RQ1: Does a uniform prior provide comparable frequentist properties to non-Bayesian approaches?

RQ2: Does teacher-assigned grades as the prior enhance Bayesian updates for the purpose of estimating learner competency?

RQ3: Is Bayesian update robust against systemic missing data?

**Main Data**

Norwegian register 2012--2022 Year 10 data, TSD file #157: W21\_4952\_TAB\_KAR\_GRS.csv

**Main Analytical Approaches**

IRT, Rasch model

Bayesian approach turning grades given learning to learning given grades

Multiple imputation for missing data treatment