Longitudinal Stability of Repeated Covariate Equating in High-Stakes Assessments

Wednesday 23 July 2025 11:45 (15 minutes)

Poster

Longitudinal Stability of Repeated Covariate Equating in High-Stakes Assessments

Author

Michaela Vařejková

Affiliation

Institute of Computer Science of the Czech Academy of Sciences and Faculty of Mathematics and Physics, Charles University

Abstract

In high-stakes educational assessments, ensuring score comparability across multiple test forms is essential for fairness and validity, particularly when test-taker groups differ in ability. Traditional test equating methods rely on anchor items, which allow direct score adjustments between test forms. However, when anchor items are unavailable, covariate equating provides an alternative by using external variables correlated with the test scores (such as grades, prior test scores, or school type) to adjust for ability differences between test-taker groups. One challenge in this approach arises when the covariates themselves are measured using multiple test forms, which may introduce bias into the equating process. A proposed solution is repeated covariate equating, where the covariates are equated before being incorporated into the primary equating process. By analyzing trends across multiple exam years, we aim to determine how repeated covariate equating evolves over time and how changes in the tested population influence test score comparability.

This study explores the longitudinal stability of repeated covariate equating by analyzing test score comparability over multiple years (2016 –2024) in the Czech Republic's national Matura exam, with a focus on the English Language test. Given that the test is administered in both spring and autumn terms, differences in student characteristics across sessions must be accounted for to ensure comparability. Additionally, as test-taking patterns shift over time —due to changes in student cohorts, evolving test-taking behaviors, and shifts in the composition of non-mandatory subject selections —the effectiveness of equating methods may vary across years. Our study examines whether score comparability is maintained consistently when applying repeated covariate equating over an extended period.

Using kernel equating for non-equivalent groups, we adjust for school type, gender, and Czech Language scores as covariates. We equate the autumn test administration to the corresponding spring administration within each year and examine year-to-year equating outcomes to assess whether and how the equating results vary over time. Specifically, we analyze differences in equated scores across years, evaluate the stability of the regression coefficients for covariates, and assess changes in the standard errors of equating. Furthermore, we investigate whether pre-equating the Czech Language test scores (which are themselves measured using multiple test forms) enhances the stability of equated English scores over time.

Keywords

test equating

Primary author: VAŘEJKOVÁ, Michaela (Institute of Computer Science of the Czech Academy of Sciences and Faculty of Mathematics and Physics, Charles University)

Co-authors: POTUŽNÍKOVÁ, Eva (Faculty of Education, Charles University); MARTINKOVÁ, Patrícia (Institute of Computer Science of the Czech Academy of Sciences and Faculty of Education, Charles University)

Presenter: VAŘEJKOVÁ, Michaela (Institute of Computer Science of the Czech Academy of Sciences and Faculty of Mathematics and Physics, Charles University)

Session Classification: Poster Session 1

Track Classification: Statistical analyses: Statistical analyses