Estimation of transfer effects for the standardized Austrian matriculation exam

Gewessler, P., Steinfeld, J., & Themessl-Huber, M.
Federal Ministry of Education, Science and Research, Austria

INTRODUCTION

The goal is to estimate the shift in expected population ability between low-stakes item pilot tests and high-stakes matriculation exam (Mathematics) in Austria.

WHY

The size of the transfer effect is assumend to be mainly due to learning efforts in the student population and therefore of practical relevance.

Predicted item solution probabilities in the high-stakes test based on their difficulty estimate in the low-stakes test can be used to guide item selection by domain experts.

HOW

The principled approach (joint estimation) is infeasible due to increasingly large numbers of test takers.

Approximate joint estimation by a mulitstage procedure in three steps with propagation of uncertainty from one step to the next.

METHODS & PROCEDURE[1]

1. ESTIMATION OF PILOT STUDY PARAMETERS

Joint estimation of pilot studies using a 1PL IRT model (N = 12,047):

$$P(Y_{ij} = 1 \mid \theta_i, \beta_j) = \text{logit}^{-1}(\theta_i - \beta_j)$$
$$\theta_i \sim \text{Normal}(\mu_{c(i)}, \sigma_{c(i)})$$
$$\beta_j \sim \text{Normal}(0, 1)$$

2. ESTIMATION OF EXAM PARAMETERS

Separate estimation of the high-stakes exam in each cohort (each $N \approx 17{,}000$) using a 1PL IRT model.
An assumption of invariance of item difficulties up to some measurement error propagates uncertainty from the previous step:

$$\hat{\beta}_j \sim \text{Normal}(\beta_j, \hat{\sigma}_j)$$

where $\hat{\beta}_j$ and $\hat{\sigma}_j$ are the estimates derived in step 1.

3. REGRESSION OF POPULATION PARAMETERS

Regressing population means in the low-stakes pilot studies (μ_L) on population means in the high-stakes tests (μ_H) for each cohort (N = 3).

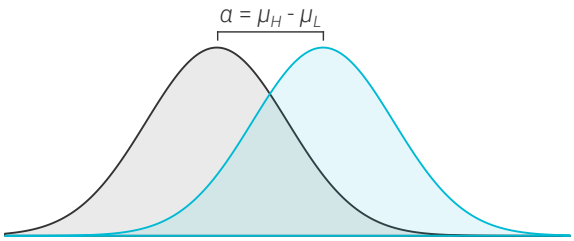
$$\mu_H \sim \text{Normal}(\alpha + \mu_L, \tau)$$

Measurement error is incorporated for the predictors:

$$\hat{\mu}_L \sim \text{Normal}(\mu_L, \hat{\sigma}_L)$$

RESULTS

EXPECTED TRANSFER EFFECT

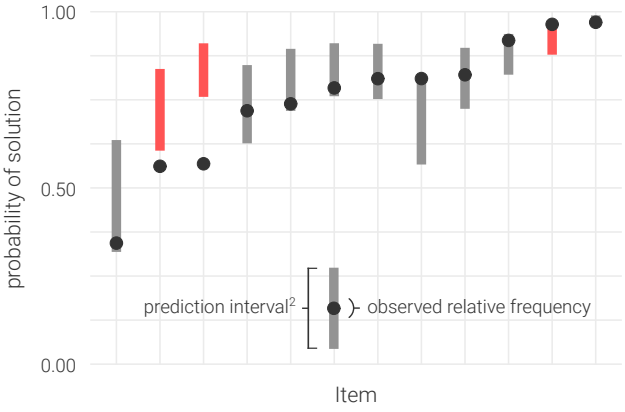


Large estimated transfer effect, α = 1.51 [1.24, 1.75][2].

PREDICTION AND VALIDATION

The predicted item solution probabilities for the latest high-stakes exam show appropriate coverage when compared to their relative frequencies of solution.

observed coverage: 75%
expected coverage: 80%



DISCUSSION

1. Very large shift in population means between low-stakes and high-stakes tests

2. Resulting credibility intervals show appropriate coverage for the prediction task, but Intervals are large due to small number of available observations in step 3.

3. Assumption of the 1PL model as the data generating process might prove to be too restrictive.

4. Assumption of a fixed shift parameter α might be inappropriate, but varying effects by cohort are difficult to estimate due to low number of observations.

1) Full model specifications are available in the supplementary material.
2) All reported intervals are 80% credibility/prediction intervals.

Federal Ministry
Republic of Austria
Education, Science
and Research

www.github.com/p-gw/imps-2019
philipp.gewessler@bmbwf.gv.at
www.bmbwf.gv.at

Find supplementary material here: