SYMPOSIUM: INNOVATION IN ASSESSING DOMAIN-SPECIFIC CRITICAL ONLINE REASONING AND ITS GROWTH IN HIGHER EDUCATION

Sample Selection Bias in Longitudinal Assessments of Domain-Specific Critical Online Reasoning

Lukas Trierweiler, Susanne Schmidt, Olga Zlatkin-Troitschanskaia, Lisa Martin de los Santos, William Walstad

FOR CORE 5404 – Funded by the German Research Foundation (DFG)

EARLI 2026

Agenda

Introduction & Theoretical Background

Introduction and Theoretical Background

- Longitudinal studies face challenges with panel survivorship and test motivation (DaLomba et al., 2021; Malschinger et al., 2023)
- Non-random dropout and assessment participation are well-documented across academic disciplines (Porter & Whitcomb, 2005; Standish & Umbach, 2019)

Regression estimates without selection-bias correction show substantial deviations in the results **(Biele et al., 2019)**

Heterogeneities in test motivation drive potential panel dropouts **(Nielsen, 2012b)**

Random sampling in (higher) educational settings is often unrealizable **(Adelson, 2013)**

→ Selection bias may affect the representativeness and validity of COR assessment scores

COR Data & Analytical Strategy

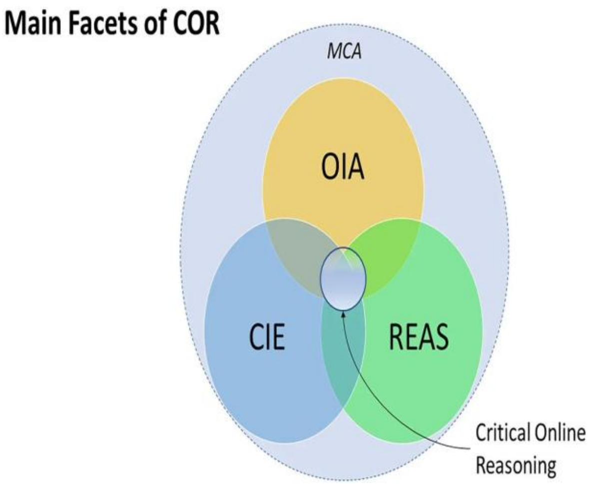
10.06.2026

REAS — Reasoning & Evidence Assessment Score

Definition (Molerov et al., 2020)
Synthesis of information accessed, including accounting for common errors and biases, and considering and weighting contradictory arguments and convert perspectives of partly conflicting information sources

Scoring
Standardized between 0 and 1 to reflect heterogeneity across COR task scenarios and reasoning contexts

Role in COR Framework
REAS scores serve as a proxy for Critical Online Reasoning (COR)



<u>Included Factors in the Score:</u>

- Coherence
- Arguments
- Balanced Argumentation
- Sources used to generate the statement & conclusion

Sample Description (domain-specific COR: Economics)

Population Sample				
Variable	**Economics t=0**	**Economics t=1**	**Social Science t=0**	**Social Science t=1**
N	996	1,045	541	549
Female, %	40.0%	40.7%	71.9%	72.5%
Semester, M (SD)	1.07 (0.43)	1.90 (1.46)	1.38 (1.03)	2.12 (1.46)
UEQ grade (1=best), M (SD)	1.97 (0.60)	2.18 (0.63)	2.11 (0.62)	2.25 (0.62)
▸ **Secondary Studies, %**	6.4%	5.5%	75.8%	68.5%
▸ **Migration Background., %**	29.4%	32.4%	25.9%	20.8%

REAS Participants Only (N = 632)				
Variable	**Economics t=0**	**Economics t=1**	**Social Science t=0**	**Social Science t=1**
N	196	206	113	117
Female, %	43.4%	45.6%	73.5%	70.9%
Semester, M (SD)	1.02 (0.12)	2.73 (1.35)	1.08 (0.30)	2.71 (1.43)
HZB grade (1=best), M (SD)	1.96 (0.58)	2.01 (0.62)	2.02 (0.57)	2.05 (0.63)
▸ **Secondary Studies, %**	8.7%	4.9%	72.6%	45.3%
▸ **Migration Background., %**	24.0%	17.0%	27.4%	13.7%
REAS score, M (SD)	0.51 (0.17)	0.54 (0.15)	0.50 (0.15)	0.52 (0.16)

Selection Bias: Theory

Selection Bias: Pre- and Post-Assessment Sources

Pre-Assessment

- Non-random university selection
- Non-random course/module selection within universities
- Non-random domain selection (focus on Economics, Social Science, (Medicine), (Physics))
- General population bias: Germany only (through country-specific heterogeneities)

→ Bias in the results and/or implications in their interpretations

Post-Assessment

- Non-randomness in initial survey participation
- Test-motivation heterogeneity → selective completion
- "Better students have more free time" → higher participation
- Panel survivorship: do assessment pools remain stable?

→ Research Question: To what extent can post-assessment bias be addressed with sample selection models?

Selection Bias (Pre-Assessment) — Possible Solutions

1.1

Non-Random Universities

Selection into universities is non-random, students differ systematically by institution. Assumption: (between-)institutional heterogeneities remain constant over time.

→ University-level (Two-Way) Fixed Effects (or random effects)

1.2

Non-Random Course / Module Selection

For Bachelor / “new” students this is less relevant: introductory courses were mostly mandatory, leaving little room for selective enrollment.

1.3

Non-Random Domain Selection

Students self-select into study domains (Economics, Social Science, ...). Only within-domain variation is comparable across individuals.

→ Domain-level Fixed / Random Effects -> can be combined with university-level Fixed Effects

Selection Bias: Estimation

Heckman Correction (1976) : Two-Stage Approach

Stage 1 — Participation Equation (Probit)

- **Binary DV**: participated in REAS **(1)** vs. not **(0)**

$Pr(Y=1|X) = \Phi(\beta \cdot Domain + \ldots + \gamma \cdot \chi'_{ik} + u_{ik})$

- **Exclusion instruments χ': predict participation but not REAS outcome itself**

- Output: individual IMRs $\lambda_i = \varphi(\hat{Z}_\gamma) / \Phi(\hat{Z}_\gamma)$

→

Stage 2 — Outcome Model (OLS)

- **DV:** REAS mean score (REAS participants only)
- REAS-Score$_i$ = λ_i (IMR) + $\chi'_{ik} + u_{ik}$
- λ captures direction & magnitude of selection bias
- Instruments excluded from Stage 2 (exclusion restriction)

Interpretation of λ (Inverse Mill's Ratio):

$\lambda > 0$ → positive selection (high scorers over-represented)

$\lambda < 0$ → negative selection (low scorers over-represented)

$\lambda \approx 0$, n.s. → no significant selection bias

Analytical Strategy — Approaches × Instruments

Three Approaches × Three Instrument Specifications

A

Pooled t0+t1 (Full Sample)

All observations; wave as covariate

B

Standalone t=0

T=0 cross-section only

C

Standalone t=1

t=1 cross-section only

Instrument Specifications

Model 1 Has minor subject (Secondary Studies)

Academic engagement proxy → higher participation

Model 2 Migration background

Cultural-institutional distance → lower participation

Model 3 Secondary Studies + Migration

Combined: Joint testing for both restrictions

| Used Significance Levels : * p<.10 ** p<.05 *** p<.01

Results:
Full Sample

Both Domains· Pooled t0+t1 Selection Results

Predictor	Model 1 — Secondary Studies	Model 2 — Migration	Model 3 — Both
Stage 1 — Probit			
Intercept	**-0.821***** (0.186) p< .001	**-0.635***** (0.180) p< .001	**-0.729***** (0.188) p< .001
Female	+0.074 (0.057) p.193	+0.078 (0.057) p.171	+0.086 (0.057) p.133
Semester	**+0.132***** (0.022) p< .001	**+0.133***** (0.022) p< .001	**+0.130***** (0.022) p< .001
Wave	**-0.117*** (0.061) p.055	**-0.126**** (0.061) p.039	**-0.129**** (0.061) p.035
Study domain	**+0.131*** (0.077) p.090	+0.013 (0.059) p.822	+0.099 (0.078) p.206
Parental educ.	-0.004 (0.019) p.830	-0.013 (0.020) p.496	-0.014 (0.020) p.490
UEQ grade	**-0.179***** (0.045) p< .001	**-0.173***** (0.045) p< .001	**-0.162***** (0.045) p< .001
ChatGPT daily use	+0.020 (0.020) p.318	+0.028 (0.020) p.169	+0.027 (0.020) p.174
▶ **Instruments (χ′)**			
▸ *Secondary Studies [Excl.]*	**-0.163**** (0.081) p.045	—	**-0.138*** (0.081) p.089
▸ *Migration [Excl.]*	—	**-0.232***** (0.063) p< .001	**-0.222***** (0.063) p< .001
Stage 2 — OLS (REAS)			
λ Inverse Mills Ratio (IMR)	**-0.341**** (0.132) p.010	**-0.074** (0.084) p.376	**-0.129*** (0.073) p.077
N probit / N sel / N out	2,971 / 607 / 626	2,971 / 607 / 626	2,971 / 607 / 626
Pseudo-R^2 (S1) / R^2 (S2)	0.037 / 0.066	0.042 / 0.057	0.044 / 0.061

Both Domains · t=0 Selection Results

Predictor	Model 1 — Secondary Studies	Model 2 — Migration	Model 3 — Both
Stage 1 — Probit			
Intercept	-0.230 (0.282) p.416	-0.232 (0.268) p.388	-0.194 (0.283) p.494
Female	+0.088 (0.080) p.271	+0.101 (0.081) p.212	+0.099 (0.081) p.221
Semester	**-0.473***** (0.112) p< .001	**-0.468***** (0.112) p< .001	**-0.469***** (0.112) p< .001
Study domain	+0.095 (0.116) p.413	+0.114 (0.084) p.174	+0.082 (0.116) p.480
Parental educ.	-0.041 (0.027) p.127	**-0.046*** (0.027) p.086	**-0.046*** (0.027) p.089
UEQ grade	-0.103 (0.065) p.110	-0.089 (0.064) p.162	-0.094 (0.065) p.148
ChatGPT daily use	**+0.068**** (0.032) p.035	**+0.071**** (0.032) p.027	**+0.071**** (0.032) p.027
▶ **Instruments (χ′)**			
▸ *Secondary Studies [Excl.]*	+0.041 (0.119) p.731	—	+0.048 (0.119) p.684
▸ *Migration [Excl.]*	—	-0.112 (0.086) p.192	-0.114 (0.086) p.186
Stage 2 — OLS (REAS)			
λ Inverse Mills Ratio (IMR)	**+0.246** (0.796) p.758	**-0.218** (0.248) p.379	**-0.161** (0.236) p.495
N probit / N sel	1,526 / 309	1,526 / 309	1,526 / 309
Pseudo-R^2 (S1) / R^2 (S2)	0.037 / 0.099	0.039 / 0.101	0.039 / 0.100

Both Domains · t=1 Selection Results

Predictor	Model 1 — Secondary Studies	Model 2 — Migration	Model 3 — Both
Stage 1 — Probit			
Intercept	**-1.294***** (0.295) p< .001	**-1.006***** (0.290) p< .001	**-1.175***** (0.299) p< .001
Female	+0.077 (0.084) p.363	+0.060 (0.084) p.476	+0.081 (0.085) p.338
Semester	**+0.217***** (0.026) p< .001	**+0.223***** (0.026) p< .001	**+0.213***** (0.027) p< .001
Study domain	**+0.251**** (0.109) p.022	+0.033 (0.088) p.706	**+0.195*** (0.111) p.078
Parental educ.	+0.049 (0.030) p.101	+0.036 (0.030) p.230	+0.038 (0.030) p.217
UEQ grade	**-0.250***** (0.066) p< .001	**-0.238***** (0.066) p< .001	**-0.226***** (0.067) p< .001
ChatGPT daily use	-0.002 (0.027) p.949	+0.011 (0.027) p.683	+0.010 (0.027) p.706
▶ **Instruments (χ′)**			
‣ *Secondary Studies [Excl.]*	**-0.336***** (0.120) p.005	—	**-0.291**** (0.121) p.016
‣ *Migration [Excl.]*	—	**-0.366***** (0.099) p< .001	**-0.339***** (0.099) p< .001
Stage 2 — OLS (REAS)			
λ Inverse Mills Ratio (IMR)	**-0.220**** (0.100) p.028	**+0.041** (0.086) p.630	**-0.065** (0.066) p.321
N probit / N sel	1,416 / 292	1,416 / 292	1,416 / 292
Pseudo-R^2 (S1) / R^2 (S2)	0.129 / 0.065	0.135 / 0.050	0.141 / 0.052

Results:
Domain-Specific

Economics (Domain 1) — Stage 1 Instruments & Stage 2 λ IMR

	Model 1 Secondary Studies	Model 2 Migration	Model 3 Both
A — Pooled t0+t1 N 1,926 / sel 382			
‣ Secondary Studies [Excl.]	+0.087 (0.138) p.530	—	+0.101 (0.138) p.467
‣ Migration [Excl.]	—	**-0.323***** (0.078) p< .001	**-0.324***** (0.078) p< .001
λ IMR (Stage 2)	+0.236 (0.338) p.485	+0.002 (0.078) p.984	+0.010 (0.078) p.895
C — Standalone t=0 N 987 / sel 196			
‣ Secondary Studies [Excl.]	+0.286 (0.184) p.119	—	+0.275 (0.184) p.134
‣ Migration [Excl.]	—	**-0.234**** (0.107) p.029	**-0.230**** (0.107) p.032
λ IMR (Stage 2)	+0.025 (0.195) p.898	-0.215 (0.153) p.163	-0.130 (0.126) p.302
D1 — Standalone t=1 N 927 / sel 182			
‣ Secondary Studies [Excl.]	-0.126 (0.229) p.581	—	-0.064 (0.231) p.783
‣ Migration [Excl.]	—	**-0.421***** (0.120) p< .001	**-0.418***** (0.120) p< .001
λ IMR (Stage 2)	+0.199 (0.379) p.600	**+0.173*** (0.092) p.061	**+0.155*** (0.090) p.087

Social Science (Domain 2) — Stage 1 Instruments & Stage 2 λ IMR

	Model 1 Secondary Studies	Model 2 Migration	Model 3 Both
A — Pooled t0+t1 N 1,045 / sel 225			
▸ Secondary Studies [Excl.]	**-0.370***** (0.100) p< .001	—	**-0.376***** (0.102) p< .001
▸ Migration [Excl.]	—	-0.040 (0.109) p.716	+0.033 (0.111) p.765
λ IMR (Stage 2)	**-0.237**** (0.078) p.003	-0.734 (0.687) p.287	**-0.235***** (0.078) p.003
C — Standalone t=0 N 539 / sel 113			
▸ Secondary Studies [Excl.]	-0.070 (0.150) p.638	—	-0.091 (0.151) p.548
▸ Migration [Excl.]	—	+0.149 (0.149) p.320	+0.160 (0.151) p.288
λ IMR (Stage 2)	-0.549 (0.626) p.383	-0.145 (0.291) p.619	-0.234 (0.252) p.355
D1 — Standalone t=1 N 489 / sel 110			
▸ Secondary Studies [Excl.]	**-0.537***** (0.149) p< .001	—	**-0.514***** (0.151) p< .001
▸ Migration [Excl.]	—	-0.253 (0.176) p.150	-0.129 (0.179) p.471
λ IMR (Stage 2)	**-0.195**** (0.080) p.017	**-0.375*** (0.222) p.094	**-0.195**** (0.077) p.013

Discussion, Limitations & Implications

Discussion of Findings

01 Bias in REAS scores without correction

Negative λ (Social Science) means that the true population mean is lower than the observed participant mean; OLS without Heckman correction yields downward-biased REAS estimates.

Positive λ (Economics) means that the true population mean is higher than the observed participant mean

02 Domain heterogeneity: Economics ≠ Social Science

Economics: Almost bias-free across all specifications

Social Science: consistent, significant negative bias. Selection effects differ in direction and magnitude between study domains

03 Bias on Observables

First stage estimates show a bias on observable factors:

Students with a better UEQ are more likely to participate

Students in higher semesters are more likely to participate

→ Whenever REAS Scores are used, a set of control variables is essential to avoid bias through observables

04 Actionable individual-level IMR weights

Individual **λ_i** values can be incorporated into any subsequent COR analysis (growth models, predictors of REAS change) to correct for participation selection bias.

This holds true if the exclusion restriction is stable

Limitations and Implications for Future Research

1. **Application towards subsequent measurement occasions and full participation:** Extend selection models across all COR panel waves; track whether bias grows or attenuates

2. **Full underlying sample (incl. Medicine and Physics):** Test whether negative selection in Economics and Social Sciences is replicated with Medicine and Physics populations

3. **Varying degrees of participation:** Differentiate full completion, partial completion (≥1 task), and non-participation

4. **Propensity Score Matching as the second method:** Validate Heckman findings and provide additional robustness checks

5. **Prediction models:** Assess the probability of achieving target sample sizes and panel survivorship from sociodemographics

Thanks for your attention!

References

Adelson, J. L. (2013). Educational Research with Real-World Data: Reducing Selection Bias with Propensity Scores. Practical Assessment, Research & Evaluation, 18(15).

Biele, G., et al. (2019). Bias from self selection and loss to follow-up. European Journal of Epidemiology, 34(10), 927–938.

Heckman, J. J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables. NBER. http://www.nber.org/chapters/c10491

Molerov, D., Zlatkin-Troitschanskaia, O., Nagel, M.-T., Brückner, S., Schmidt, S., & Shavelson, R. J. (2020). Assessing University Students' Critical Online Reasoning Ability: A Conceptual and Assessment Framework With Preliminary Evidence. Frontiers in Education, 5, 577843. https://doi.org/10.3389/feduc.2020.577843

Nielsen, T. (2012b). Drop-out from longitudinal educational studies and motivational design. ELSIN Conference, Cardiff, pp. 224–235.

Malschinger, P., Vogl, S. & Schels, B. Drop in, drop out, or stay on: Patterns and predictors of panel attrition among young people. Österreich Z Soziol 48, 427–450 (2023). https://doi.org/10.1007/s11614-023-00545-z

DaLomba E, Gramstad A, Johnson SG, Carstensen T, Stigen L, Mørk G, et al. (2021) Predictors of students' participation in a learning environment survey with annual follow-ups. PLoS ONE 16(6): e0253773. https://doi.org/10.1371/journal.pone.0253773

Porter, S.R., Whitcomb, M.E. Non-response in student surveys: The Role of Demographics, Engagement and Personality. Res High Educ 46, 127–152 (2005). https://doi.org/10.1007/s11162-004-1597-2

Standish, T., Umbach, P.D. Should We Be Concerned About Nonresponse Bias in College Student Surveys? Evidence of Bias from a Validation Study. Res High Educ 60, 338–357 (2019). https://doi.org/10.1007/s11162-018-9530-2