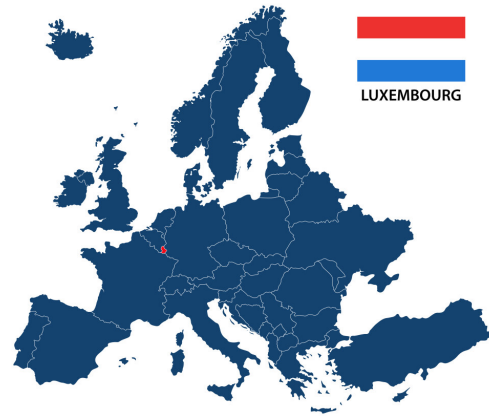


# Grade Retention in Primary Education in Luxembourg

Effectiveness regarding  
academic achievement + noncognitive outcomes

Martha Ottenbacher, Rachel Wollschläger,  
Ulrich Keller, Sonja Ugen

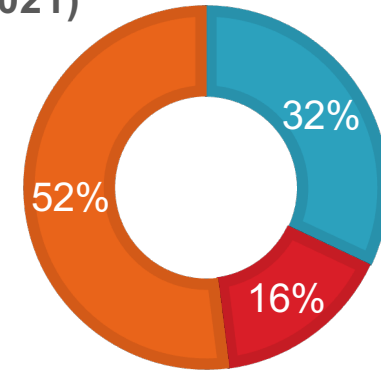
# The setting Luxembourg



VectorStock VectorStock.com/18695372

## HOME LANGUAGES in grade 1 (2021)

- Mono-Linguistic
- Multi-Linguistic
- No Instruction Language



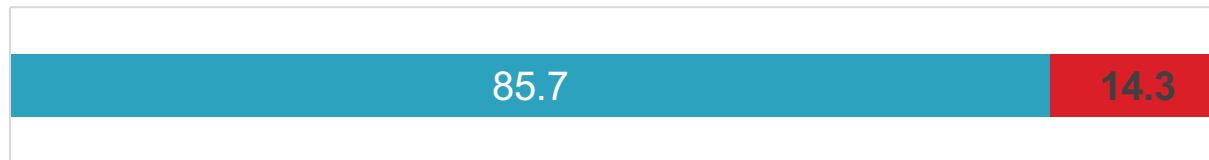
## RETENTION by age 15 (PISA, 2018)

Belgium	31%
Germany	20%
France	17%
Netherlands	17%
Scandinavia	< 5%

**Luxemb.** **32%**

Ottenbacher et al.

## RETENTION up to grade 3 (EpStan 2017/2018)



■ Promoted ■ Retained

## Pro Retention

Time to learn (AA) ✓

Social comparison theory + reference group (PS)

Attitudes towards self (PS)

## Contra Retention

Social comparison theory + reference group (AA) ✓

Unpleasant emotions (PS) ✓

Age-appropriateness of curriculum (PS)

School identification (PS)

Labled Stigma/Labeling theory (PS)

Inconsistent / non-significant effects

Hwang/Cappella (2019), Wu et al. (2010)

Meta studies: average zero effects / favoring promotion

Goos et al. (2021), Jimerson (2001)

# Large Scale Assessment in Luxembourg: EpStan

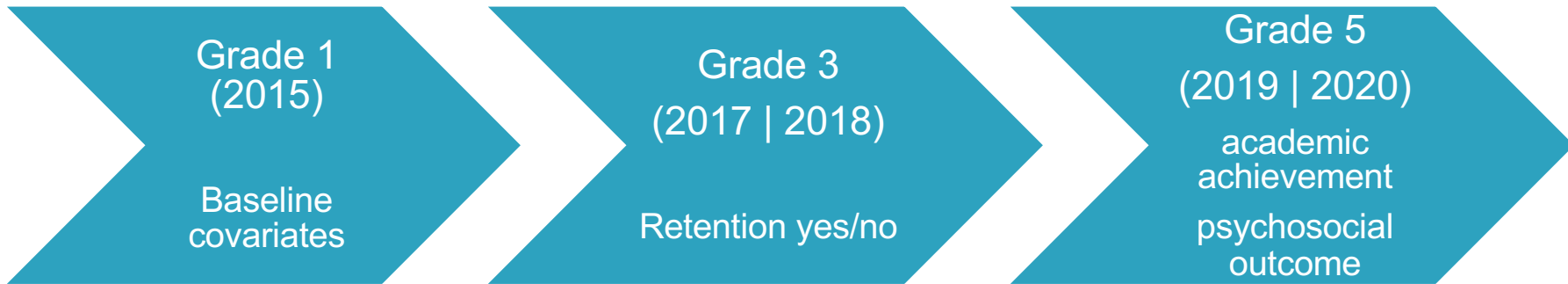


- Years: 2010 – 2024
- Panel: grades 1, 3, 5, 7, 9
- Participation is mandatory → nearly full sample
- Size in 2024: 26,600 students
- Assessment of core competences: based on national curriculum, but independent of teacher
- Academic achievement + psychosocial outcomes
- Student perspective + parent perspective
- Data matching for research is possible



VectorStock VectorStock.com/18695372

# EpStan Data – Present Study



Demographics,  
Academic achievement,  
Psychosocial outcomes

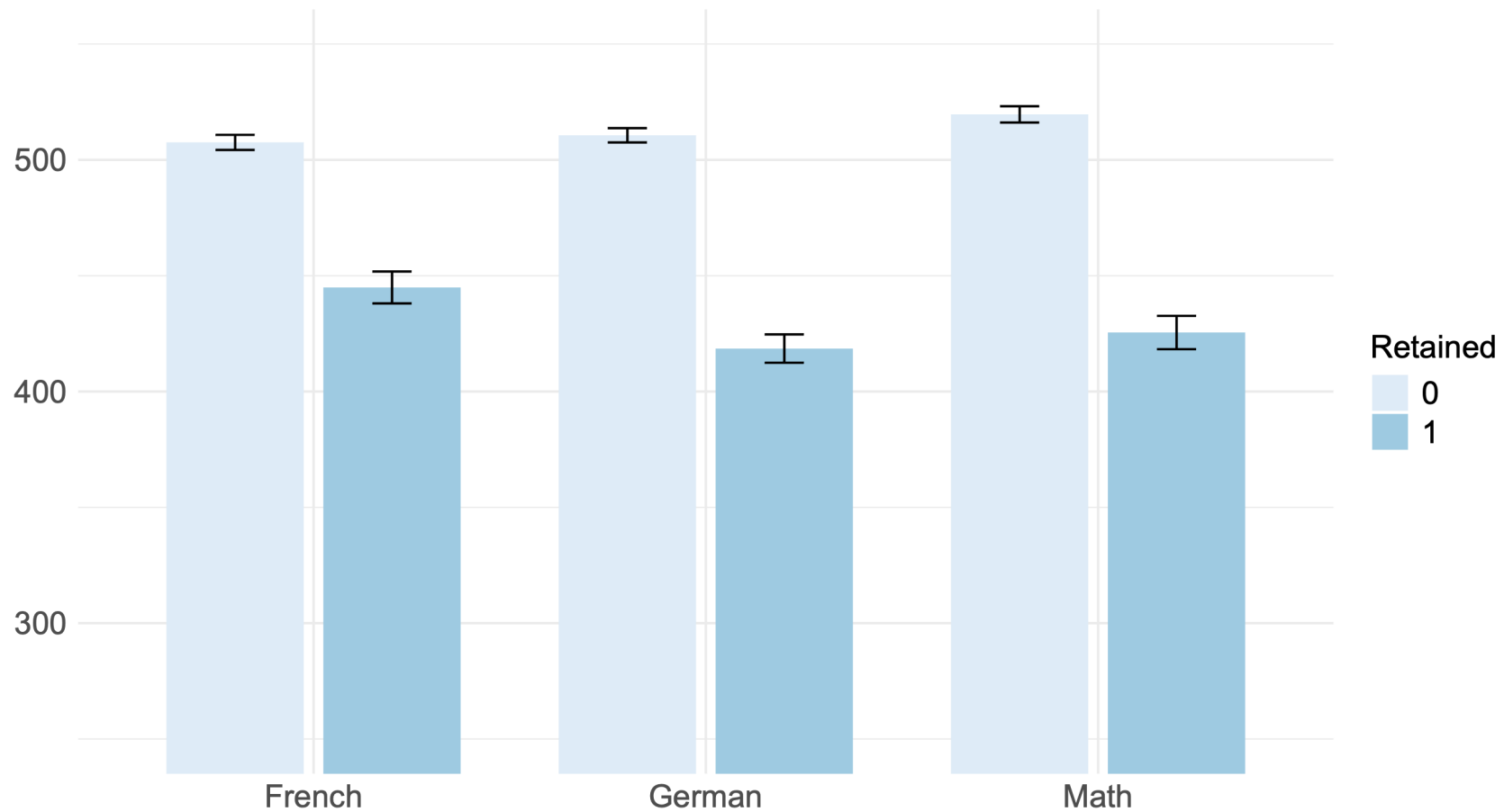
Academic achievement:  
Math, German reading, French  
reading

Psychosocial outcome:  
anxiety, interest, self-concept

**Final sample after restrictions: 3,801 students (545 were retained)**

- Descriptives on achievement (2019|2020)

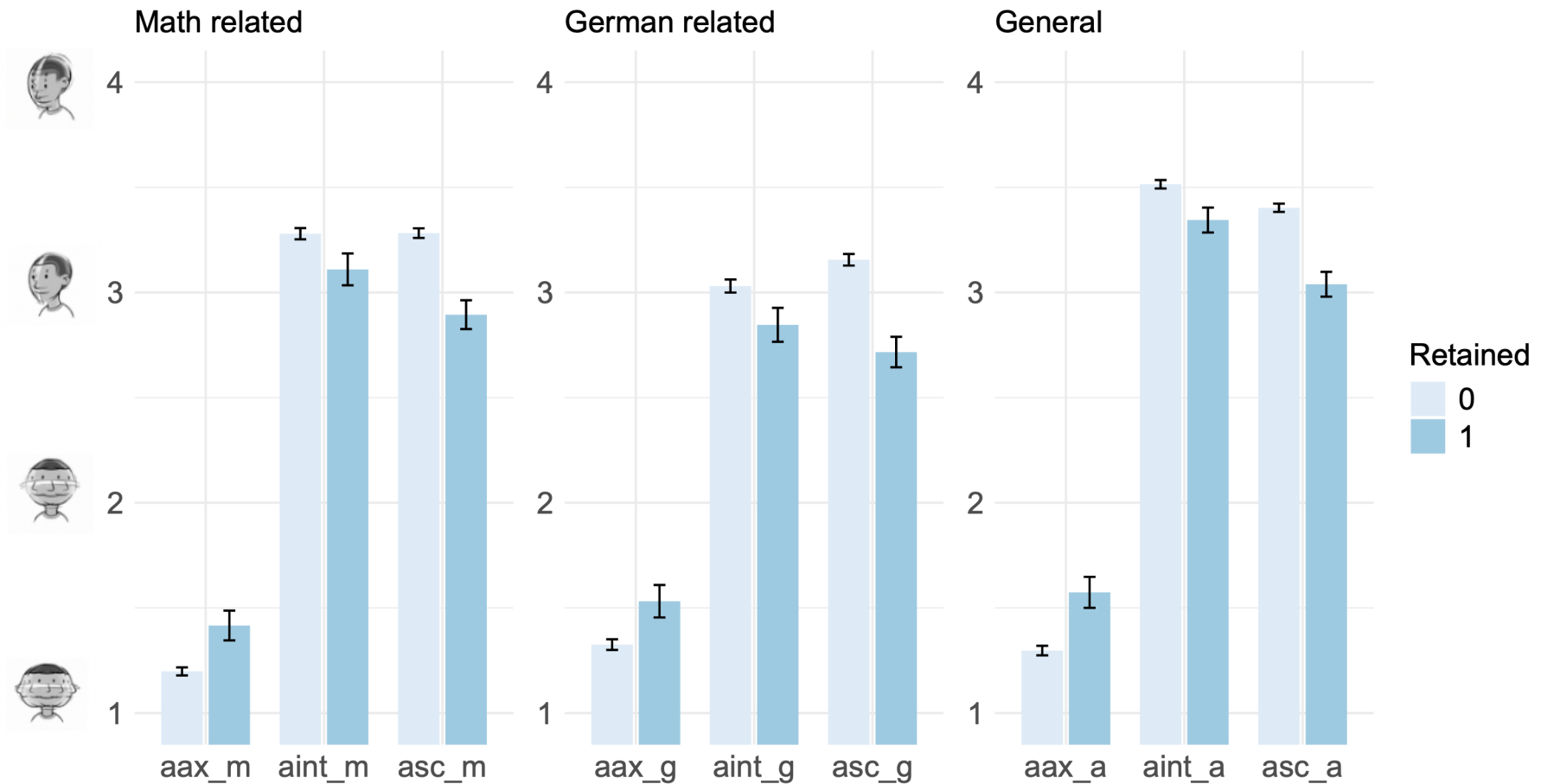
Achievement in Grade 5 by Retention



# Descriptives

- Descriptives on noncognitive outcomes (2019|2020)

Psychosocial outcomes in Grade 5 by Retention



## ■ Steps to prepare analysis

### Missings

- Imputation: nonparametric missing value imputation using Random Forest

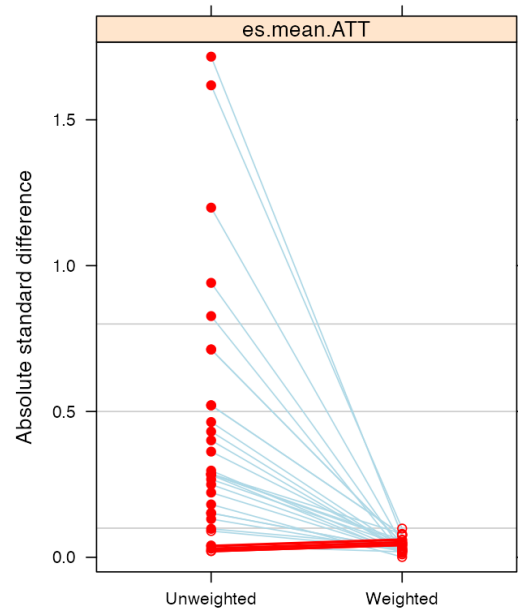
### Establish Comparability

- Student grade 1 baseline characteristics
- Multiple gradient boosted logistic regressions and diagnoses  
→ best balance between retained and promoted → propensity scores
- Weight the whole sample

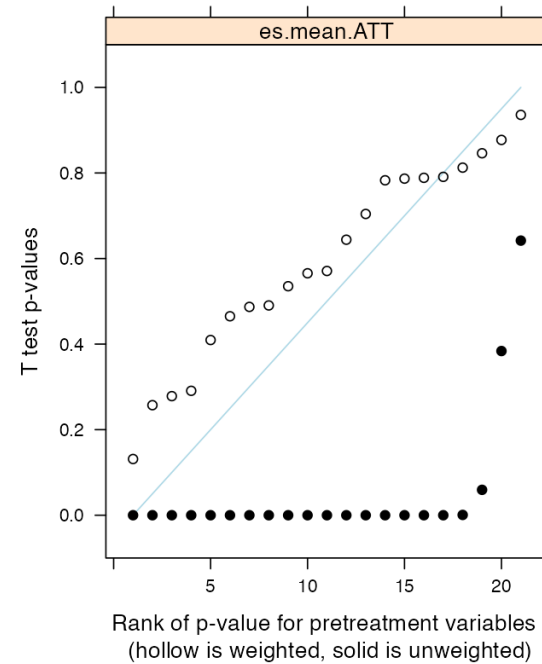
- Establishing Comparability

twang: control group matches characteristics of unweighted treatment group

Standardized effect size of pretreatment variables



t-test p-values for weighted pretreatment variables



## ■ Steps to prepare analysis

### Missings

- Imputation: nonparametric missing value imputation using Random Forest

### Establish Compara- bility

- Student grade 1 baseline characteristics
- Multiple gradient boosted logistic regressions and diagnoses  
→ best balance between retained and promoted → propensity scores
- Weight the whole sample

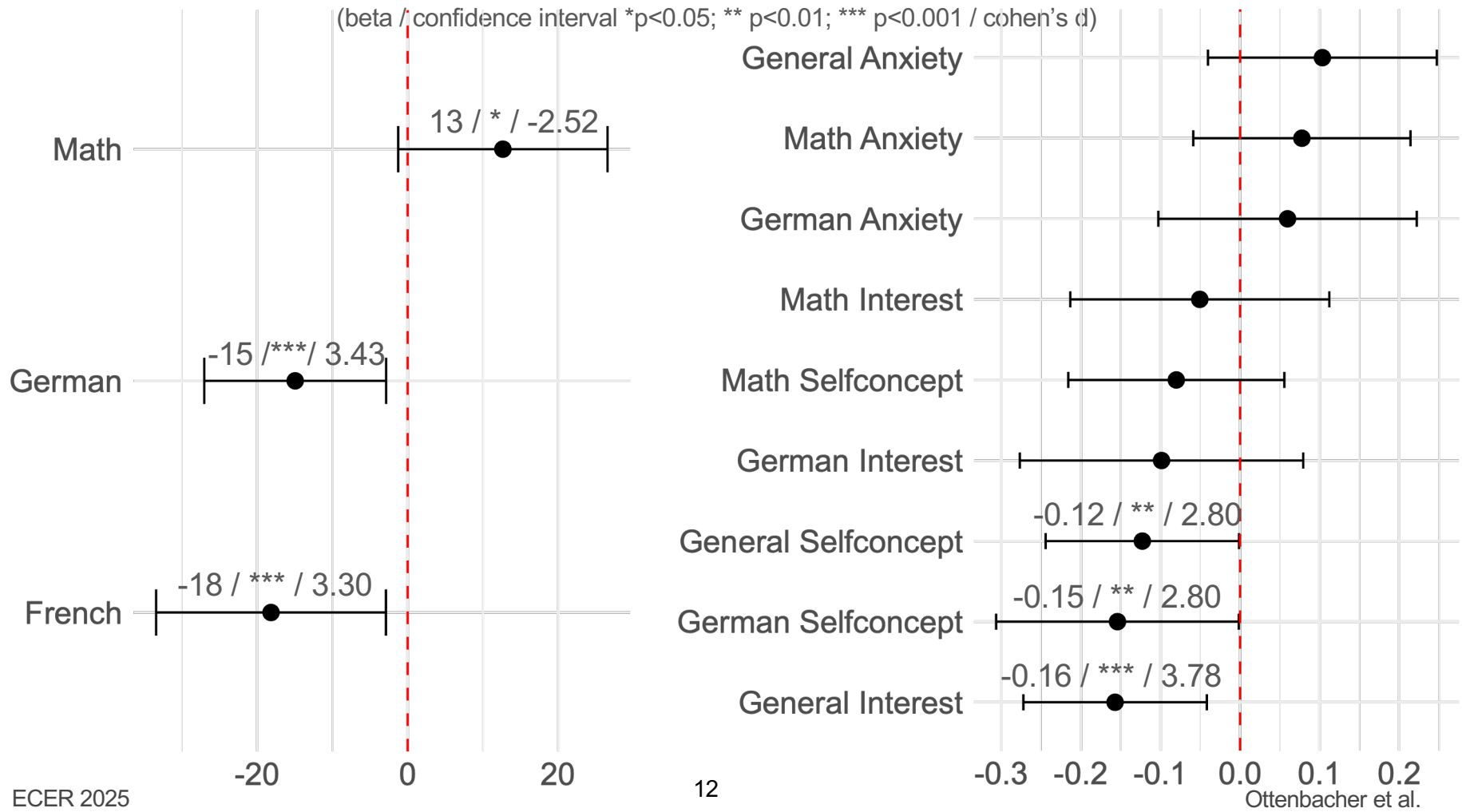
### Causal effect

- ATT via survey-weighted generalised linear models

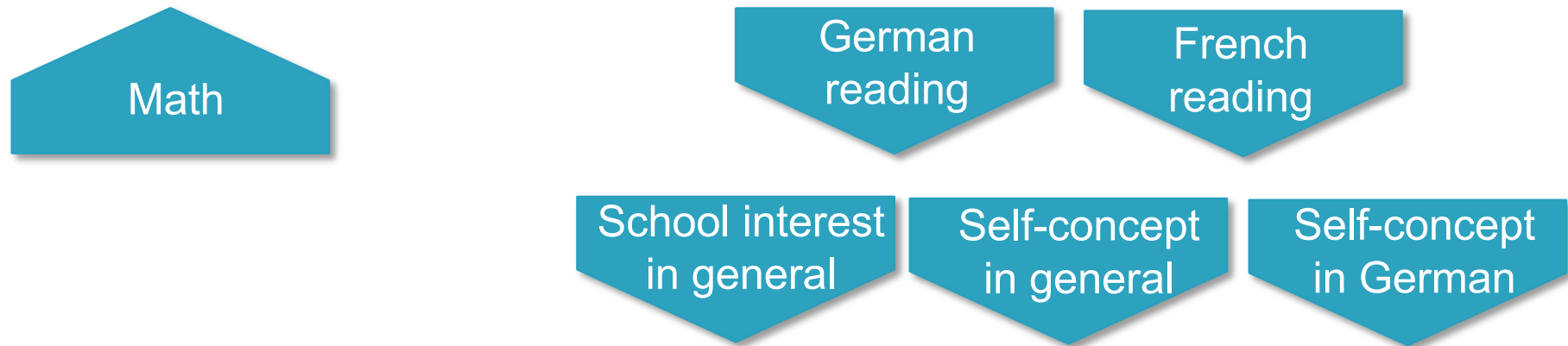
# Results



- Differences in predicted outcomes: Retained – Promoted students



- What is the causal effect?



- Limitations
  - Missings
  - Grade retention dummy (no information if retention in grade 1, grade 2 or grade 3)
  - Year of birth (not month of birth)

- EpStan data:
  - Assess core competences in primary
  - Causal effect can be estimated
- Retention effect in our study:
  - Predominantly non-significant or significant negative
- Conclusions for Luxembourg:
  - No general recommendation for retention
  - Increasing diversity among students in terms of migration/home language
  - But hints that tailored support helps
- Broader implications for countries with increasing immigration

# Questions or data request?



[data@epstan.lu](mailto:data@epstan.lu)



# Appendix

- **Cost-benefit perspective**
  - economically relevant
  - **negative spillover on classmates of receiving class:** first time course takers have lower achievements (Hill 2014)
  - **increased dropout rates from school:** school content doesn't match interests of older retained students? (Jimerson/Ferguson 2007, Hughes et al. 2017, Jimerson et al. 2002, Andrew 2014)
  - **labour market outcomes decreased** : repetition records in CV decrease likelihood for positive call-back (Baert/Picchio 2021)
  
- **Causal effects remain unclear**
  - **educational setting:** pedagogic measures accompanying retention or mere repetition
  - **grade:** positive in Kindergarten (Vandecandelaere et al. 2016a), detrimental in primary (Ikeda/García 2021), positive in high school + beyond (Kinder/Knecht 2011, Tafreschi/Thiemann 2016)
  - **time perspective:** short term effects are higher or positive; long term effects are lower/negative (e.g. Wu et al. 2008, Allen et al. 2009, Wu et al. 2010)
  - **quality of the study:** correlational studies present larger (negative) effects

## ■ Potential effects of grade retention

	Pro	Contra
Academic achievement	<p><u>Time to learn</u> (Caroll 1963, Bloom 1976): Consolidate knowledge before stepping on.</p>	<p><u>Social comparison theory</u> (Festinger 1954): We compare upwards. Lower performing students achieve better results in group with slightly higher performing mates. Via increased motivation, goal settings, information on task completion.</p>
Psychosocial wellbeing	<p><u>Social comparison theory</u> (Festinger 1954): We compare upwards. Lower performing students have higher self-concept if retained and surrounded by equally/lower performing mates.</p> <p><u>Attitudes towards self</u> (Festinger 1954): Retainees have advantage over classmates (age, experience from previous curriculum).</p>	<p><u>Triggers dismissal by children</u>: curriculum for retainees is no longer age-appropriate which lowers interest.</p> <p><u>Anxiety</u>: Being retained and need to integrate into new group is associated with unpleasant emotions like fear, threat, anxiety.</p> <p><u>Labled Stigma/Labeling theory</u> (Becker 1963): Retention is visible. Association with failure. Others perceive retainees negatively. Retainees may internalize the belief of weak performer. Self-fulfilling prophecy.</p>

## Pro Retention

**Math** Klapproth et al. (2016)

## Contra Retention

**Languages** Gómez Soler et al. (2021),  
Hwang/Cappella (2019) , Klapproth et  
al. (2016), Vandecandelare et al. (2016)

**Self-concept (general + in German)**  
Klapproth et al. (2016), Martin (2011)

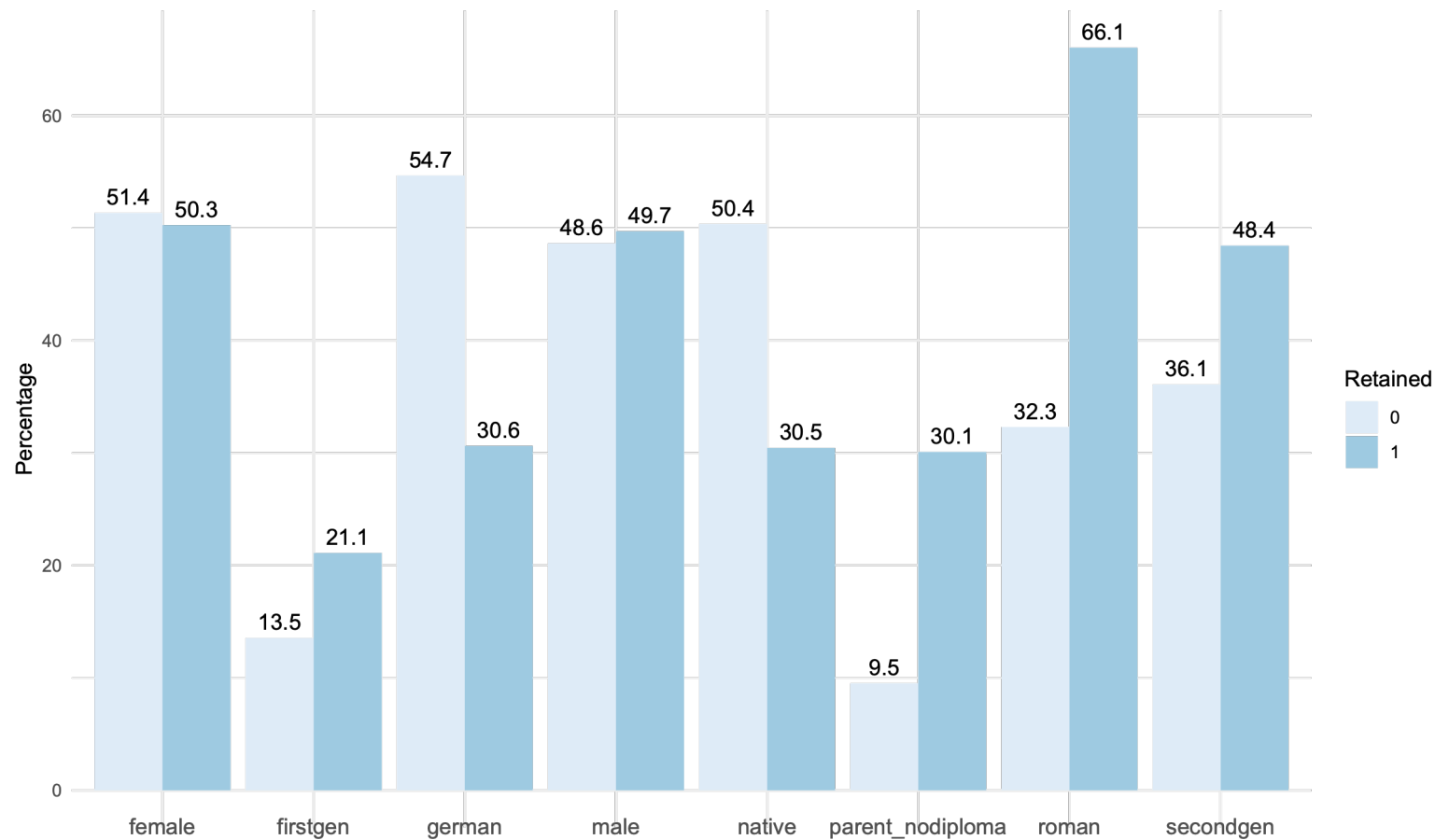
## Inconsistent/non-significant effects

Hwang/Cappella (2019), Wu et al. (2010)

## Meta studies: average zero effects or favoring promotion

Goos et al. (2021), Jimerson (2001)

- Descriptives on some baseline characteristics (2015)



## ■ Establishing Comparability

Balance table before and after weighting

Covariates	Unweighted		Weighted	
	unw_retainees	unw_promoters	w_retainees	w_promoters
ysex:M	0.497	0.486	0.497	0.518
ybirthyr	2008.648	2008.628	2008.648	2008.638
ahisei	40.030	50.666	40.030	39.900
ahedu:craftsman's diploma	0.013	0.010	0.013	0.007
ahedu:no diploma	0.312	0.097	0.312	0.298
ahedu:non-university degree	0.050	0.078	0.050	0.054
ahedu:secondary education	0.261	0.244	0.261	0.287
ahedu:technical secondary education	0.288	0.244	0.288	0.270
ahedu:university degree	0.077	0.328	0.077	0.084
aimmig:1st generation	0.209	0.135	0.209	0.210
aimmig:2nd generation	0.486	0.361	0.486	0.496
aimmig:native	0.305	0.503	0.305	0.294
alh_ger:TRUE	0.306	0.547	0.306	0.342
alh_rom:TRUE	0.661	0.323	0.661	0.638

... table continued

Covariates	Unweighted		Weighted	
	unw_retainees	unw_promoters	w_retainees	w_promoters
pcreche:yes	0.594	0.669	0.594	0.581
pinlu10	0.967	0.951	0.967	0.972
alh_e	435.949	535.977	435.949	439.813
all_e	421.127	527.668	421.127	427.620
am_e	431.433	534.957	431.433	435.944
asc_m	3.095	3.559	3.095	3.120
aint_m	3.288	3.602	3.288	3.304
aax_m	1.908	1.533	1.908	1.985
asc_a	3.347	3.654	3.347	3.324
aint_a	3.546	3.748	3.546	3.528
aax_a	1.842	1.495	1.842	1.896
asc_g	3.061	3.502	3.061	3.102
aint_g	3.278	3.589	3.278	3.334
aax_g	1.932	1.548	1.932	2.040

# Results (Detailed)

Characteristic	am_e		ad_e		af_e	
	Beta <sup>1</sup>	SE	Beta <sup>1</sup>	SE	Beta <sup>1</sup>	SE
(Intercept)	-31	42.0	338***	33.4	328***	46.4
retained						
0	—	—	—	—	—	—
1	13*	5.12	-15***	4.31	-18***	5.12
female	-27***	4.76	12**	4.10	27***	4.86
cybirthyr	9.9	6.72	8.2	5.42	-8.5	6.65
cybirthyr2	-13	10.4	-13*	6.15	-21**	6.97
ahisei	0.21	0.201	0.00	0.177	0.41	0.239
parent_nodiploma	-11	8.92	-30***	8.36	-49***	10.8
parent_craftsman	-12	22.7	3.7	13.5	-29	18.3
parent_nonuniversity	3.3	10.6	-24*	10.1	-21	12.1
parent_secondary	-12	7.82	-28***	7.83	-48***	10.5
parent_technicalsecondary	-17*	8.05	-28***	8.14	-33**	11.0
secondgen	-1.1	6.13	-13*	6.02	9.8	6.59
firstgen	0.79	8.36	-8.1	7.21	5.6	7.93
german	-4.5	6.17	7.3	5.85	-18**	6.05
roman	-0.93	5.15	-26***	4.94	26***	5.92
creche	-2.5	5.18	-3.6	4.40	15**	5.30
pinlu10	0.44	10.6	-17	10.4	-21*	10.8
alh_e	-0.04	0.035	0.26***	0.033	0.12***	0.036
all_e	0.19***	0.042	0.06	0.040	0.13**	0.042
am_e	0.57***	0.045	0.05	0.041	0.08	0.045
asc_m	2.5	2.45	-2.1	2.32	3.2	2.40
aint_m	6.0*	2.92	-2.6	2.28	-0.04	2.66
aax_m	1.7	2.08	0.17	1.87	1.8	1.79
asc_a	1.8	3.93	-2.9	2.64	-3.8	2.73
aint_a	-1.3	3.38	2.8	2.84	4.2	3.65
aax_a	-1.4	2.15	1.7	1.90	-2.6	1.96
asc_g	-7.4**	2.81	3.3	2.15	-1.2	2.30
aint_g	-1.9	3.21	2.8	2.36	-2.8	2.55
aax_g	-3.3	2.11	-4.6*	1.93	0.49	1.92
ad_e_future	0.39***	0.039				

<sup>1</sup> \*p<0.05; \*\*p<0.01; \*\*\*p<0.001

Abbreviations: CI = Confidence Interval, SE = Standard Error

# Results (Detailed)

Characteristic	asc_m		aint_m		aax_m	
	Beta <sup>†</sup>	SE	Beta <sup>†</sup>	SE	Beta <sup>†</sup>	SE
(Intercept)	2.1***	0.351	2.1***	0.297	2.1***	0.396
retained						
0	—	—	—	—	—	—
1	-0.08	0.050	-0.05	0.058	0.08	0.053
female	-0.37***	0.048	-0.21***	0.057	0.14**	0.052
cybirthyr	0.03	0.067	0.14	0.080	-0.08	0.071
cybirthyr2	0.03	0.078	0.16	0.129	-0.05	0.127
ahisei	0.00	0.002	0.00	0.002	0.00	0.002
parent_nodiploma	0.04	0.106	0.03	0.112	0.27**	0.095
parent_craftsman	-0.21	0.212	-0.17	0.301	0.10	0.200
parent_nonuniversity	0.03	0.119	-0.01	0.117	0.06	0.091
parent_secondary	0.05	0.101	0.01	0.097	0.16*	0.078
parent_technicalsecondary	0.10	0.099	0.12	0.096	0.08	0.078
secondgen	0.02	0.066	0.18*	0.078	0.03	0.070
firstgen	0.12	0.088	0.10	0.104	0.08	0.087
german	0.07	0.069	-0.04	0.097	0.08	0.086
roman	-0.14*	0.054	-0.08	0.067	0.06	0.061
creche	-0.04	0.051	-0.10	0.058	0.03	0.053
pinlu10	-0.26*	0.111	-0.24	0.129	-0.11	0.153
alh_e	0.00	0.000	0.00	0.000	0.00	0.000
all_e	0.00	0.000	0.00	0.001	0.00	0.000
am_e	0.00***	0.000	0.00***	0.001	0.00	0.000
asc_m	0.03	0.027	-0.01	0.031	0.04	0.029
aint_m	0.06*	0.027	0.12***	0.035	-0.11***	0.033
aax_m	0.01	0.020	0.01	0.025	0.00	0.023
asc_a	0.04	0.036	0.19**	0.058	-0.05	0.032
aint_a	-0.03	0.036	-0.10*	0.043	-0.01	0.039
aax_a	0.00	0.023	0.03	0.028	-0.01	0.026
asc_g	-0.06*	0.026	-0.08**	0.030	0.00	0.025
aint_g	0.05	0.028	0.05	0.031	-0.01	0.030
aax_g	-0.02	0.021	-0.04	0.028	0.00	0.027

<sup>†</sup> \*p<0.05; \*\*p<0.01; \*\*\*p<0.001  
Abbreviations: CI = Confidence Interval, SE = Standard Error

Characteristic	asc_g		aint_g		aax_g	
	Beta <sup>†</sup>	SE	Beta <sup>†</sup>	SE	Beta <sup>†</sup>	SE
(Intercept)	3.4***	0.455	3.4***	0.375	1.2**	0.394
retained						
0	—	—	—	—	—	—
1	-0.15**	0.058	-0.10	0.063	0.06	0.060
female	0.02	0.055	0.23***	0.061	0.05	0.058
cybirthyr	0.13	0.072	0.19*	0.084	-0.12	0.080
cybirthyr2	0.01	0.079	0.19*	0.080	0.04	0.116
ahisei	0.00	0.002	-0.01*	0.003	0.00	0.003
parent_nodiploma	-0.10	0.107	-0.05	0.126	0.08	0.131
parent_craftsman	-0.02	0.255	-0.33	0.255	-0.11	0.205
parent_nonuniversity	-0.04	0.146	-0.17	0.159	-0.05	0.151
parent_secondary	-0.01	0.097	-0.07	0.117	0.01	0.118
parent_technicalsecondary	-0.05	0.100	-0.09	0.118	0.01	0.115
secondgen	-0.10	0.077	-0.01	0.089	0.22**	0.082
firstgen	-0.09	0.093	-0.09	0.130	0.25*	0.122
german	0.18	0.093	0.10	0.087	0.15	0.097
roman	-0.36***	0.066	-0.22***	0.067	0.16*	0.070
creche	-0.13*	0.058	-0.16**	0.061	-0.01	0.064
pinlu10	-0.13	0.143	-0.22	0.165	0.02	0.152
alh_e	0.00**	0.000	0.00	0.000	0.00	0.000
all_e	0.00**	0.001	0.00	0.001	0.00	0.001
am_e	0.00***	0.001	0.00**	0.001	0.00	0.001
asc_m	0.00	0.030	-0.03	0.036	0.04	0.045
aint_m	-0.02	0.033	0.04	0.034	-0.02	0.034
aax_m	0.01	0.024	-0.04	0.026	0.03	0.027
asc_a	0.00	0.033	0.07	0.063	0.04	0.038
aint_a	0.02	0.040	-0.01	0.046	-0.04	0.037
aax_a	0.05	0.027	0.06	0.029	-0.09***	0.027
asc_g	0.02	0.027	-0.03	0.036	-0.04	0.036
aint_g	0.04	0.029	0.05	0.032	0.02	0.033
aax_g	-0.04	0.028	0.00	0.027	0.05	0.028

<sup>†</sup> \*p<0.05; \*\*p<0.01; \*\*\*p<0.001  
Abbreviations: CI = Confidence Interval, SE = Standard Error

Characteristic	asc_a		aint_a		aax_a	
	Beta <sup>†</sup>	SE	Beta <sup>†</sup>	SE	Beta <sup>†</sup>	SE
(Intercept)	2.5***	0.245	3.0***	0.254	1.6***	0.352
retained						
0	—	—	—	—	—	—
1	-0.12**	0.044	-0.16***	0.038	0.10	0.056
female	-0.09*	0.043	0.07	0.039	0.10	0.057
cybirthyr	0.01	0.061	0.01	0.053	-0.03	0.071
cybirthyr2	0.01	0.115	0.01	0.058	0.02	0.112
ahisei	0.00	0.002	0.00	0.002	0.00	0.002
parent_nodiploma	-0.12	0.089	0.05	0.079	0.29**	0.098
parent_craftsman	-0.21	0.140	-0.52**	0.174	-0.19	0.140
parent_nonuniversity	0.07	0.103	0.16	0.094	0.07	0.107
parent_secondary	-0.04	0.078	0.08	0.074	0.20*	0.091
parent_technicalsecondary	-0.06	0.079	0.08	0.078	0.09	0.087
secondgen	0.04	0.057	0.19***	0.053	0.03	0.087
firstgen	0.07	0.074	0.21**	0.073	-0.05	0.098
german	0.02	0.072	0.09	0.048	0.03	0.090
roman	-0.15**	0.055	-0.09*	0.042	0.01	0.068
creche	-0.02	0.044	0.01	0.040	0.01	0.057
pinlu10	-0.12	0.110	-0.14	0.094	-0.07	0.143
alh_e	0.00	0.000	0.00	0.000	0.00	0.000
all_e	0.00	0.000	0.00	0.000	0.00	0.001
am_e	0.00	0.000	0.00*	0.000	0.00	0.001
asc_m	0.05*	0.024	0.01	0.022	-0.01	0.035
aint_m	0.01	0.027	0.07**	0.023	-0.06	0.034
aax_m	-0.01	0.019	-0.04*	0.018	0.02	0.023
asc_a	0.02	0.036	0.05	0.027	0.02	0.031
aint_a	0.02	0.032	0.03	0.029	0.01	0.039
aax_a	0.05*	0.021	0.03	0.020	-0.01	0.024
asc_g	-0.01	0.025	-0.02	0.021	-0.03	0.030
aint_g	0.03	0.026	0.03	0.023	0.00	0.032
aax_g	-0.01	0.022	0.02	0.018	0.00	0.023

<sup>†</sup> \*p<0.05; \*\*p<0.01; \*\*\*p<0.001  
Abbreviations: CI = Confidence Interval, SE = Standard Error

## ■ Scales construction noncognitive outcomes

Scale	Math	Deutsch	Alle Schulfächer
asc_m	1. Ich bin gut im Fach Rechnen.	1. Ich bin gut im Fach Deutsch.	1. Ich bin gut in den meisten Schulfächern.
aint_m	2. Ich interessiere mich für das Fach Rechnen.	2. Ich interessiere mich für das Fach Deutsch.	2. Ich interessiere mich für die meisten Schulfächer.
aax_m	3. Das Fach Rechnen macht mir Spaß.	3. Das Fach Deutsch macht mir Spaß.	3. Die meisten Schulfächer machen mir Spaß.
aax_m	4. Ich habe Angst vor dem Fach Rechnen.	4. Ich habe Angst vor dem Fach Deutsch.	4. Ich habe Angst vor den meisten Schulfächern.
aax_m	5. Im Fach Rechnen lerne ich schnell.	5. Im Fach Deutsch lerne ich schnell.	5. In den meisten Schulfächern lerne ich schnell.

## Estimating ATE: controls in other studies

	1	2-3	4-6	7-9	10-12
Dep Var grade 5	<b>Math</b>	<b>Language</b>	<b>Math related psychosocial</b>	<b>Language related psychosocial</b>	<b>General psychosocial</b>
Age	Wu et al.2008, Ikede/Garcia 2013, Hu/Hannum 2019, Schwerdt et al. 2015	Wu et al.2008, Ikede/Garcia 2013, Hu/Hannum 2019, Schwerdt et al. 2015			Ikede/Garcia 2013, Hu/Hannum 2019, Wu et al. 2010
Age, Age <sup>2</sup>					Nagin et al. 2003
Age cohort	Peixoto et al. 2025 (subsamples)				
Month of birth	Vandecandelaere 2016c, Vandecandelaere 2016b, Goos et al. 2013, Hwang/Capella 2019	Hwang/Capella 2019	Hwang/Capella 2019	Hwang/Capella 2019	Vandecandelaere 2016a, Hwang/Capella 2019
Age no	Winter/Green2012, Nuñez et al. 2015, Klapproth/Schaltz 2015, McCombs et al. 2009, Lorence/Dwerkin 2006	Nuñez et al. 2015, Lorence 2014, Roderick/Nagaoka 2005, Moser et al 2012, McCombs et al. 2009, Lorence/Dwerkin 2006			Martorelli/Mariano 2018
AA grade 5 yes					
AA grade 5 no			Goos et al. 2013	Goos et al. 2013	Vandecandelaere 2016a, Goos et al. 2013