ÖREBRO UNIVERSITY

WORKING PAPER 1/2023 (ECONOMICS AND STATISTICS)

Does eligibility requirements matter for academic achievements?

A quasi-experimental retrospective study of students studying intermediate statistics

Nicklas Pettersson, Niklas Karlsson and Daniela Andrén

ISSN 1403-0586

Örebro University School of Business
SE-701 82 Örebro, Sweden

Does eligibility requirements matter for academic achievements?

A quasi-experimental retrospective study of students studying intermediate statistics

Nicklas Pettersson, Niklas Karlsson and Daniela Andrén*
Örebro University School of Business, Sweden

January 17, 2023

Abstract

Student achievements are expected to be affected by both educational activities and learning during a course and previous teaching and learning, and earlier eligibility requirements. Using data from a quasi-experimental retrospective study, we estimated the effect on exam scores in an intermediate course in statistical theory from both earlier eligibility requirements, the realignment of a prerequisite course in introductory statistics and students' characteristics. We found that success in intermediate statistics was explained by the realignment and eligibility requirements in Mathematics, and also by the intersection between gender and foreign background.

Keywords: statistics education, aligned course, eligibility requirements, academic evaluation, quasi-experimental retrospective design.

*The authors are grateful to seminars and conferences participants for helpful comments and suggestions and the Faculty of Business, Science and Engineering (ENT) for financial support of our pedagogical project. Any remaining errors of fact, or omission, are ours. In addition, we would like to thank Ann-Marie Flygare for continuous support in developing and implementing the changes in the courses and Anna Hultgren for continuous support of our longitudinal pedagogical project.
email: Nicklas.Pettersson@oru.se; Niklas.Karlsson@oru.se; Daniela.Andren@oru.se

1 Introduction

Each student's ability to assimilate the content of higher education depends on their unique learning trajectory, influenced by their exposure to various instructional methods and learning objectives, with the progression dictated by the eligibility requirements. The level of the requirements and how well students passed them are likely to affect subsequent outcomes. In this paper, we aim to answer to what extent the academic achievements at the intermediate level were conditional on just passing earlier eligibility requirements and to what extent the design of an introductory course mattered to later achievements. We answer these questions by focusing on students studying statistics at a Swedish university.

The fulfillment of the eligibility requirements is intended to confirm that a student had acquired at least the minimal sufficient knowledge and skills needed for entrance. In favorably designed learning environments, as novel information is encountered and processed in working memory, the student will construct and accumulate relevant knowledge in long-term memory, later to be recalled at higher levels, in such a way that the requirements at each new stage are met and the student may progress further. Therefore, eligibility requirements are relevant to an educational programme curriculum, and courses that are necessary for knowledge acquisition at the current level of education have to be sequenced effectively and mapped to the programme's goals. However, as described by (Shepard et al. 2018), the design of the educational programme, instructional activities, and assessment strategies commonly used in higher education do not always lead to the desired student learning outcomes.

At the course level, a design that aligns the learning objectives and instructional design with the assessment may better direct the student to the intended learning activities and assist in their progression. Therefore, when evaluating effects of changes in an educational programme, one should also consider the adequacy of the intertwined eligibility requirements for gaining access to the subsequent courses.

At the core of statistics is the collection, organization and analysis of data and the interpretation and presentation of the analysis' results. Statistics is a branch of mathematics that deals with the study of data and is often used to make predictions or draw conclusions about a particular population or phenomenon. Courses at the introductory level are focused on surface learning through activities designed for consolidation of the basic concepts and skills that are necessary for ease of acquisition of knowledge and problem-solving. However, while most academic activities have positive influences on learning, their composition matters a lot for student achievements ((Hattie 2015); (Bendikson et al. 2011)). But students' behavior and learning depends on their view of success in terms passing examination requirements. Therefore, students are more likely to be guided to an intended successful academic achievement if course communication and activities are aligned with the examination requirements. In this context, adapting the course accordingly to provide sufficient challenges in the systematic manner of Scholarship of the Teaching and Learning (SoTL), leads to deeper learning outcome. For introductory statistics, this means that students need to repeat and 'wrangle' with basic concepts such that they become solidly consolidated in long-term memory.

During the last decades, the teaching of statistics has been impacted by the changing societal demands due to technological advances and the (sometimes intertwined) educational development (presumably hastened by the recent coronavirus pandemic). The focus has shifted from procedures to statistical concepts, and towards developing statistical literacy, reasoning and thinking rather than mere technical and computational skills. Specifically, the Guidelines for Assessment and Instruction in Statistics Education (Carver et al. 2016) advise college courses in introductory statistics to teach statistical thinking (as an investigative process of problem-solving and decision-making, and to give students experience with multivariable thinking); focus on conceptual understanding; integrate real data with a context and purpose; foster active learning; use technology to explore concepts and analyze data; and use assessments to improve and evaluate student learning. They suggested focusing less on probability theory; basic statistics already covered in upper secondary school; looking up standard distributional tables and constructing charts that are more easily produced with statistical software; and advanced software training.

The goals of this study are twofold. Firstly, we aim to evaluate whether the realignment of a course in introductory statistics (IS), intended to enhance learning and retention, had a medium-term achievement effect. Secondly, we aim to estimate the impact of the programme and course eligibility requirements on the students' success. The two aims are intertwined, and we use the same achievement outcome measure for both purposes, the examination results in the intermediate statistical theory (ST) course, for which IS is a prerequisite course.

Since experiments with random allocation to intervention groups seldom are feasible, it is more common to use quasi-experimental (or observational studies) to control for potential confounders. The strategy of evaluating learning by performance in a subsequent course, controlling for student characteristics, avoids the potential risk such as teaching-to-the-test associated with value-added-models comparing course grades or student evaluations pre-post intervention creating a false relation to contemporaneous course outcomes (Johnson 2006); (Weinberg et al. 2009); (Yunker & Yunker 2003). There is an inherent risk faced in higher education, to 'reward higher grades in the introductory course but punish professors who increase deep learning' (Carrell & West 2010).

2 Institutional settings and data

2.1 Swedish higher education and eligibility requirements

The Swedish higher education system is mainly funded by governmental grants. All university courses are classified into one or more fields of study. The government decides the fields of studies and the level of the associated financial compensation, and only the subject-related content should govern the classification of courses. Course classifications are decided by the educational institutions by specifying the percentage distribution by field of study. The total compensation is based on the distribution by field of study, the number of registered students and the amount of their completed credits (as converted to full-year performance) up to a ceiling determined by the government.

The basic entry requirements for university studies are an upper secondary school degree (or equivalent), in form of credits from at least one course in English and Mathematics each and two courses in Swedish. In addition, there may be special eligibility requirements considered to be absolutely necessary for a student to be able to assimilate new knowledge and skills. These are usually specified in form of field qualifications, a set of upper secondary school courses that a student must have passed to be eligible. Each field of study is also linked to at least one field qualification.

Courses may be combined into educational programmes for which there must be an education plan specifying the requirements for special eligibility to the programme. All students studying a programme, meets the requirements for a degree. The Bachelor degree programmes in subjects such as Statistics, Economics and Business Administration all belong to the 'Behavioral Sciences, Economics and Social Sciences' field of study and typically used the field qualification from 2014 (up until 2013) denoted by A4(4) for upper secondary school eligibility courses, while the subject Mathematics belongs to the 'Physics, Mathematics and Technology' field of study and typically used field qualification A8(8). Starting from June 2022, the field qualifications were abandoned and the new requirements were to directly specify the special eligibilities, which had always been an option, although rarely used in practice.

2.2 Education in Statistics at Örebro University

The education plan for the Bachelor's degree programme in Statistics at Örebro University, Sweden, analyzed in this study, was established in 2000. Before that, education in Statistics was conducted in form of courses. The programme underwent some changes over time, but it always started with an introductory course in statistics (IS), followed up during the intermediate studies by a course (among others) in statistical theory (ST).

A Bachelor in Statistics required field qualification A4(4), i.e., one course passed in Social studies, two courses passed in English, and three courses passed in Mathematics. A Bachelor in Mathematics required only part of field qualification A8(8), i.e., four courses in Mathematics.

ST was a ten weeks full-time (15 credits) intermediate course in probability and inference theory, offered at the beginning of the autumn semester. The course was held by the same teacher over the years and had on average about 30 students. The special eligibility requirements included IS but changed in 2014 from 30 credits in statistics to either 22.5 credits in statistics or 22.5 credits in Mathematics. The syllabus was unaltered between 2014 and 2017[1]. Teaching consisted of lectures composed of two or three 45 minute sessions each (in total 78 sessions), 7 computer practicals and a seminar. The examination was summative and based on an individual written exam (75% of the course grade), lab reports (15%) and a seminar report (10%).

Since 2004, the IS course was a ten weeks full-time (15 ECTS credits) course, covering basic probability, inference and survey sampling. It always

[1] In 2018, the requirements were enhanced after a major revision

had the same special eligibility requirements as the Bachelor programme. The course was given in the same format at the start of each semester. The course had 29 lectures (with approximately 120 students), 13 exercises and 4 computer practicals (with approximately 30 students in each group). Every occasion consisted of two 45 minutes sessions. The examination consisted of an individual written exam, including a mid-term test.[2] There were also four non-graded mandatory labs.

Starting in the Autumn 2015, the teaching activities and examination were updated and realigned with the recent development. The course contents, learning objectives and tuition time were however unaltered. Previously, the teaching followed a continuous series of lectures spanning the entire course, where the teaching period for the various parts was irregularly stretched over several weeks with varying composition of teaching forms between the weeks. The reviewed course was instead formed into weekly units with a consistent structure and recurring schedule each week. Except for the last week, which only had two summarizing lectures, each week started with a lecture followed by an exercise, then a second lecture followed by an exercise or a computer practical, and ended with a third lecture.

The two initial lectures condensed the material previously covered in up to three lectures, mainly by subduing repetitive technicalities (such as computational skills), and were then directly mapped to the forthcoming exercise. More visualizations for data exploration and analysis were also introduced in the revised course.

In line with the 'muddiest point in the lecture' (Mosteller 1989) technique and as input to the third and final lecture, students were asked to provide feedback (anonymously on paper, orally or via e-mail) on what had been hardest or least clear during the unit. As to not be overused (Keeler & Koretsky 2016), the feedback was primarily asked for at the end of the second exercise. In consideration of the feedback, the third lecture rehearsed and summarized the unit, while giving additional types of problems, integrated examples or in-depth applications based on those already presented. As a means of interleaving and spacing the rehearsal, more of the examples and exercises were integrated with material from previous units, especially at the end of the units. This integration was also reflected in the written examination, with an increased variation in the type of questions.

[2] The mid-term was replaced in Autumn 2016 by weekly digital tests.

2.3 Data

In total, 143 students participated in the examination of the ST-course following the 2014 syllabus. Seven male students never attempted the written exam. Except one of them, they were much older than the average student. We excluded these students since they were not our primary concern and imputing the dependent variable under a Missing-at-Random (MAR) model would not likely improve the precision of the regression estimates (Little 1992); (Von Hippel 2007). Our final sample thus included 136 students.

To estimate the effects of the changes of the IS-course, we created a variable with four categories with study groups, see Table 1. Our treatment study group, those who took ST within one year after the new IS, was denoted as NEW (n=25). The main control study group, denoted as OLD, contains students who took TS within one year after the old IS, and was the largest group (n=51).[3] The main bulk of students whose main subject was Statistics (93%), [4] was hereby retained within these two groups. All lecturers in the IS and ST courses for the OLD and NEW study groups had Swedish background. Over time, the IS-course was taught by a male lecturer (in OLD) and was split equally between another male and a female lecturer (in NEW). A third male lecturer held the ST course during the whole period.

Students taking ST more than one year after IS were denoted as LATE (n=33). Most remaining students never took the IS course (MATH(never), n=21) but a few took IS after ST (MATH(after), n=6). The latter group included one student who took IS and ST simultaneously. Since these two groups were small and admitted to ST via Mathematics requisites, we collapsed them and denoted the study group as MATH (n=27).

Most students (80%) where thus admitted to the ST-course via Statistics prerequisites (including the IS-course). Within this group, the latter the IS-course cohort a student belonged to, the fewer opportunities they had to complete the ST-course. The proportion of students taking the ST-course within a year after passing the IS-course also increased over time, from 76%(in

[3]It was therefore used as reference when variance inflation factors (VIF)s were computed.

[4]In Statistics, the correspondence between admission status and actual program completion was weak since students could easily change degree's programme. Switching between Bachelor degree programmes or taking double degrees is quite common. For example in 2017 few students switched to but many left Statistics early in the programme. We therefore used both previous (recent) studies and programme admission, if any, to classify 59 students as having had Statistics as their main subject.

2014), 81%(2015), 90%(2016) to 100%(2017).

Table 1: Categorization into study groups according to students IS- and ST-course semesters

ST-	IS-semester$_{year}$; Autumn(A), Spring(S) or Never(N).										
year	S_{13}	A_{13}	S_{14}	A_{14}	S_{15}	A_{15}	S_{16}	A_{16}	S_{17}	A_{17}	N
2014	9	21	4	1	0	2	0	0	0	0	6
2015	6	2	4	23	3	0	0	0	0	0	1
2016	1	1	1	1	4	18	1	0	0	3	5
2017	1	0	0	0	1	0	2	5	1	0	9

NEW	OLD	LATE	MATH(after)	MATH(never)

Our outcome variable was the exam score (SCORE; 0-60 points) on the first written ST exam.[5] The last resit for the ST-course was in August 2018.[6] In case of a blank submission the attempt was not registered, but such behaviour was rare. None of the five students who scored zero on their first exam later passed the written exam. There was little evidence to suggest any strategic exam attendance behaviour. We, therefore, argue that the score on the first written examination attempt for ST, a course at intermediate level, was a well standardized, simple and objective measure for outcome to be used when estimating the importance of the course in introductory statistics for students' medium term achievements.[7]

The distribution of SCORE deviated somewhat from a symmetrical distribution due to an excess of low values (Figure 1).

Figure 2 and Table 2 suggest that variation in SCORE was smallest in the study group NEW. Except for the study group OLD, which was lowest with least symmetric distribution, the average SCORE was fairly equal across the four study groups.

[5]There were no restrictions on the number of attempts to resit an exam, unless the course was dissolved in which there were three additional resits.

[6]Once a student passed ST by acquiring at least 50 points in total at the reports and the written exam, a resit was not allowed. Thus, later attempts were conditional on not passing before.

[7]It would have been possible to create a measure taking any subsequent attempts into account, although that would potentially run into complicated modelling where later attempts would be conditional on failing previous attempted written exam(s), where the failing itself depends on the score from the lab and seminar reports, and the potential influencing intermediate factors.

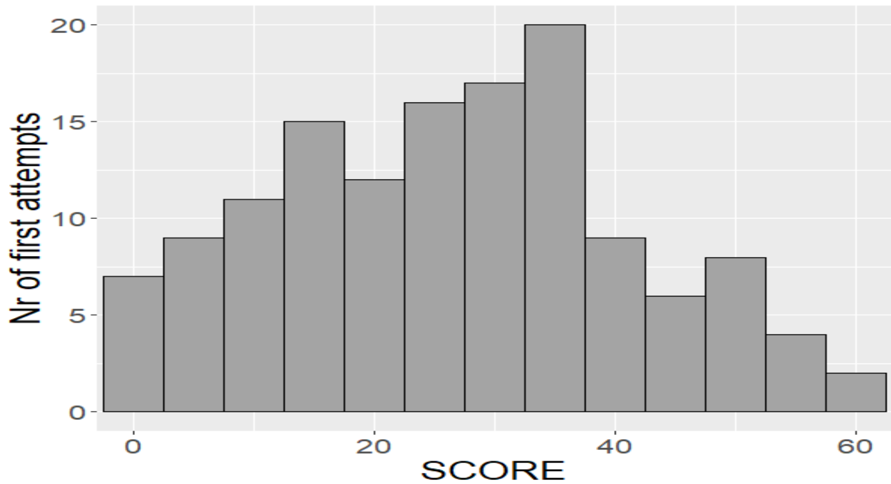

Figure 1: SCORE (first attempt at written exam in ST)

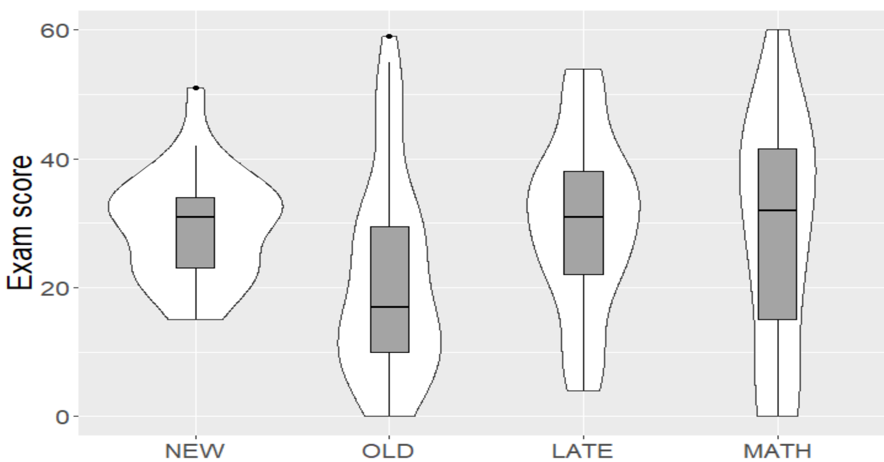

Figure 2: SCORE (first attempt at written exam in ST) by study group

Table 2: Mean and standard deviation of SCORE by study group

	mean	sd
NEW	29.64	8.43
OLD	21.35	15.25
LATE	30.06	13.53
MATH	28.41	17.95

Most students were in their twenties when they started studying the ST-course(Figure 3). The right skewness, also seen within each study group in Figure 4, diminished after a log-transformation, LAGE = log[AGE-min(AGE)+1].

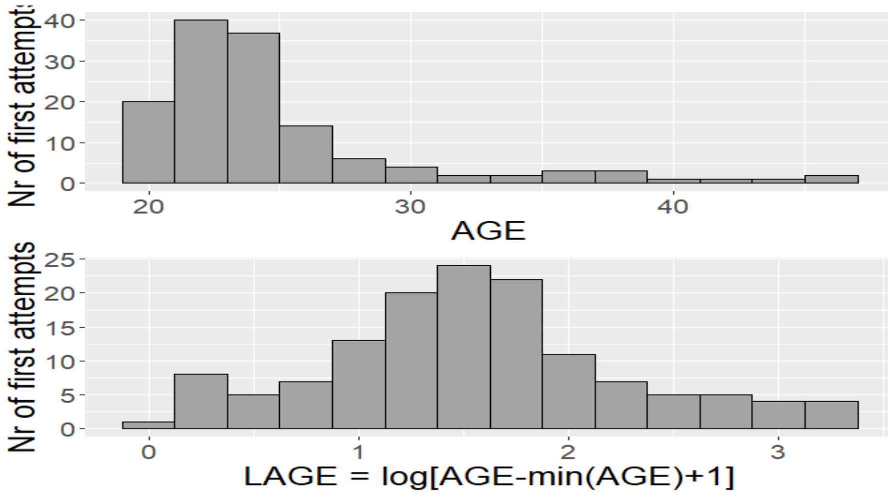

Figure 3: AGE and LAGE

Figure 4: AGE by study group

OLD was the youngest study group (with smallest sd) and LATE was the oldest (with largest sd), see Table 3.

Table 3: Mean and standard deviation of AGE by group

	mean	sd
NEW	25.28	4.63
OLD	22.70	2.53
LATE	28.06	7.05
MATH	25.13	5.53

We assumed that SCORE would best be approximated by a quadratic polynomial of LAGE, where LAGE2 = LAGE^2. There was also a peak at 23-24 years of age(in LAGE, just above 1.5), but Figure 5 suggest that SCORE might be better approximated by a sine function of LAGE, with the highest SCORE among the youngest, in the middle where the density was highest, and the oldest, although there was a risk of overfitting to the data since the uncertainty was greater at the boundaries due to few very young and very old students.

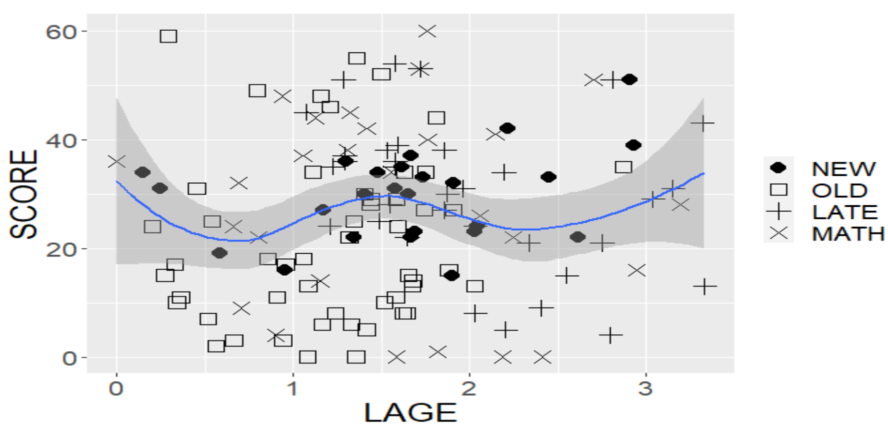

Figure 5: Locally weighted smoothing estimates (LOESS) of SCORE as a function of LAGE with 95% point-wise confidence intervals.

However, the sine function approximation of LAGE seemed to hold by group of study, see figure 6. Study group LATE seemed to be most truncated with the youngest in the group being 21.7 years.

When analyzing interaction between gender and foreign background (as classified by first name and surname), we used Swedish males as reference

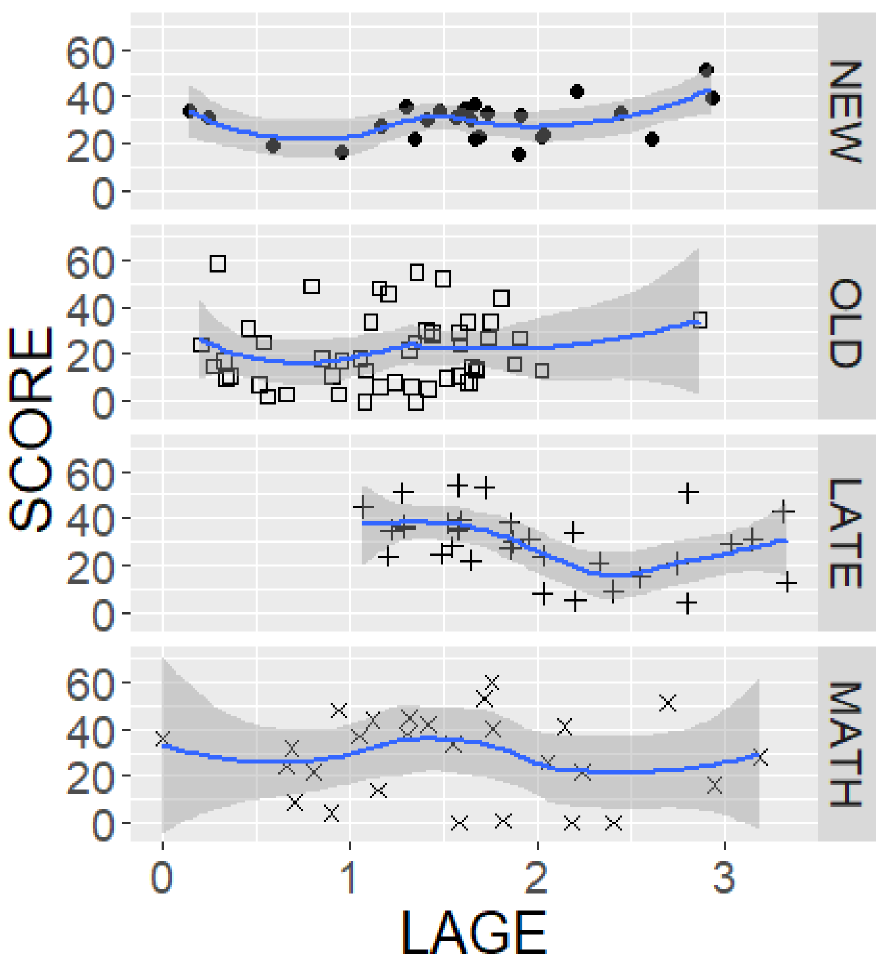

Figure 6: Locally weighted smoothing estimates (LOESS) of SCORE as a function of LAGE with 95% point-wise confidence intervals by group of study.

group since they made up the majority (n=76), followed by Swedish women (n=25), foreign males (n=20) and foreign women (n=15). See Table 4.

Table 4: Cross-tabulation of groups of study and interaction between gender and origin

	Swedish		Foreign	
	Male	Female	Male	Female
NEW	16	5	3	1
OLD	23	12	9	7
LATE	21	7	3	2
MATH	16	1	5	5

Figure 7 suggests that Swedish students, especially males, had higher exam scores than foreign students. Swedish males also had the most symmetric distribution of SCORE.

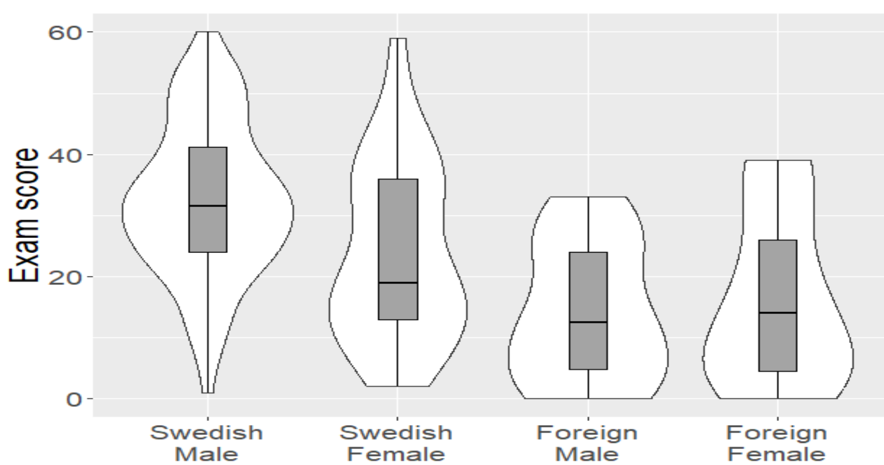

Figure 7: SCORE on first attempt at written exam by interaction between gender and origin

A common rating value scale (0-20) from the official conversion Table (UHR 2020) was used to equate the Bachelor degree's special eligibility requirement course grades (one course in Social studies, two courses in English and three courses in Mathematics at upper secondary school level) from three succeeding grading systems, see Table 5. The oldest system, (1962-1994), was a relative grading system without failing grades. The two recent grading systems were both goal-related and the lowest grade indicated no passing/failing the course. We used the rating values, divided by 10 as our grade measure.[8] From a goal-related perspective, the difference between passing(1) and failing(0) or between passing with excellence(2) and passing(1) are thus both interpreted as an increase of 1 in terms of our grade measure.

Table 5: Conversion of grading scales into a common rating value scale.

Rating value	Grading scale (time period in use) A-F (2011-)	IG-MVG (1994-2011)	1-5 (1962-1994)
0	F	IG	-
1.0	-	-	1
7.3	-	-	2
10.0	E	G	-
12.5	D	-	-
12.7	-	-	3
15.0	C	VG	-
17.0	-	-	4
17.5	B	-	-
20.0	A	MVG	5

Four foreign students and three older Swedish students had no grades registered in Mathematics, and two of them also missed all grades in Social studies and English. As part of the complete data regression analysis model, the imputation model was set up, and the missing values were multiply imputed (m=1 000 times) via chained equations.[9] In addition to the regression model variables, the rating value variables were imputed by up to five ratings in other subjects, ENGLISH and MATHEMATICS also by SAT scores, and MATHEMATICS also by upper secondary school major and age at the SAT test.

The average values for SOCIAL and ENGLISH were lowest in study groups OLD and MATH, while the average value for MATHEMATICS was highest in study groups LATE and MATH (Table 6). The correlations be-

[8]In subjects where more then one course were observed, the average rating value was used.

[9]Many of the additional variables eligible to the imputation model also had some missing values and were imputed as well.

tween grades were moderate: $\hat{\rho}_{SOCIAL,ENGLISH} = 0.31$, $\hat{\rho}_{SOCIAL,MATHEMATICS} = 0.39$ and $\hat{\rho}_{ENGLISH,MATHEMATICS} = 0.40$.

Table 6: Mean and standard deviation of SOCIAL, ENGLISH and MATHEMATICS by study group

	SOCIAL(n=134)		ENGLISH(n=134)		MATHEMATICS(n=129)	
	mean	sd	mean	sd	mean	sd
NEW	1.54	0.31	1.48	0.29	1.43	0.29
OLD	1.46	0.35	1.38	0.31	1.36	0.32
LATE	1.55	0.33	1.53	0.30	1.51	0.36
MATH	1.42	0.35	1.40	0.47	1.55	0.38

A positive relationship to SCORE was observed within all subjects; strongest for MATHEMATICS (with a plateau around 1.5) but weak for SOCIAL, see figure 8-10. One student (in study group LATE) had a grade value below 1 in ENGLISH.

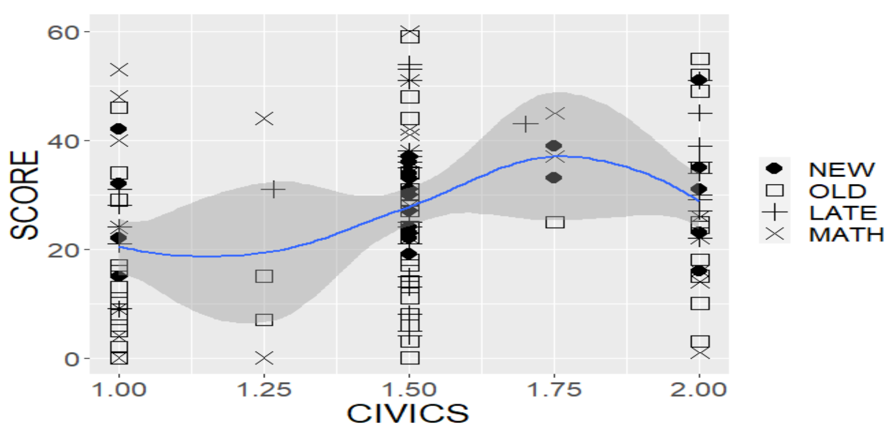

Figure 8: Locally weighted smoothing estimates (LOESS) of SCORE as a function of CIVICS (SOCIAL SCIENCES) with 95% point-wise confidence intervals.

All students' first attempts were within their year of study. Ten students attempted their first exam at one of five resits. In total there were thus nine exams for which the exam occasion random effect where estimated. In figure 11, it was seen that most students first attempted the main exam.

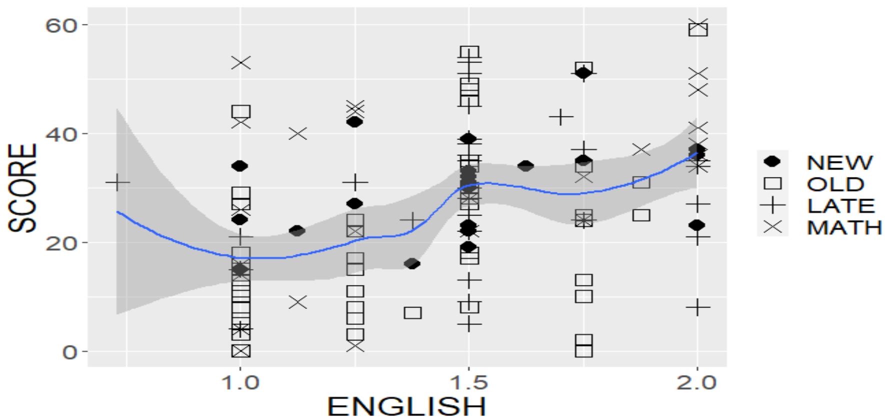

Figure 9: Locally weighted smoothing estimates (LOESS) of SCORE as a function of ENGLISH with 95% point-wise confidence intervals.

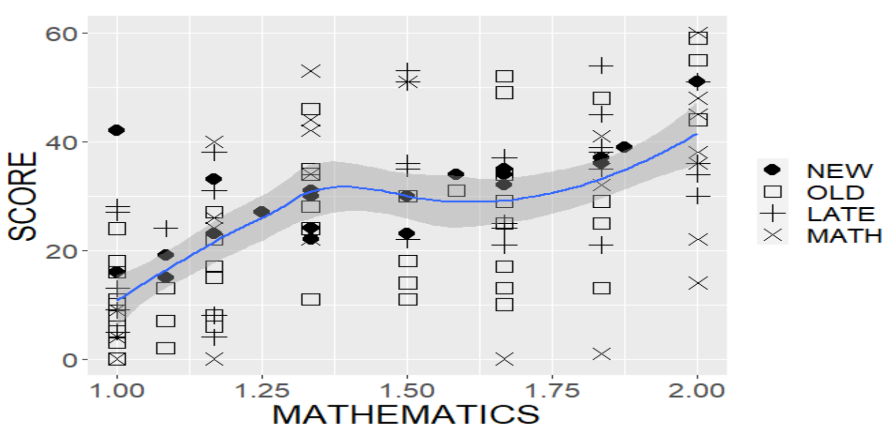

Figure 10: Locally weighted smoothing estimates (LOESS) of SCORE as a function of MATHEMATICS with 95% point-wise confidence intervals.

Exams and re-exams had on average fairly similar average SCORE (27.3 vs 24.8), but the standard deviation (5.3 vs 11.1) were higher among the less frequented re-exams.

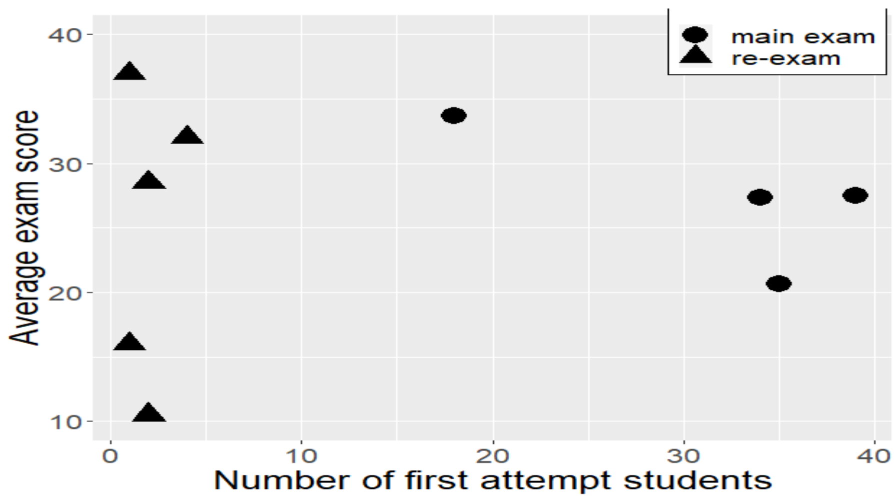

Figure 11: Number of students first exam attempts and average SCORE by main exam or re-exam

3 Empirical framework

In this study, our focus was medium-term achievement of students. As outcome measure, we used the score on the first attempt at the written exam for ST, a course given at the intermediate level. The written examination covered 75% of the course final grade and was an individual examination, which could not be directly influenced by the performance of other students as the other examinations (25%) in ST.

In order to estimate the medium-term achievement effect of realigning the IS course, a course at the introductory level, we choose to form quasi-experimental "treatment" and "control" groups among the ST students based on whether and when they had passed the IS course. We then controlled for historical performance (upper secondary school grades in prerequisites courses). Through this design, we were also able to achieve our second goal, to estimate the effect of the varying university eligibility requirements and the effect of upper secondary school special eligibility requirements on the course performance. Thus, our two goals were linked and therefore estimated simultaneously.

We also take in consideration the potential impact of demographic characteristics (age, sex and foreign background). Large differences related to demographics would suggest that students could have been differently treated. One such indication was found in a study by the Swedish School Inspectorate,[10] where a representative sample (n=11 650) of standardized test scores taken during Spring 2010 in English, Mathematics and Swedish at upper secondary school level was re-corrected anonymously. In contrast to universities, the tests that are graded in upper secondary schools are usually not anonymous. Therefore, it was suggested that boys were judged relatively harshly, and regardless of their sex, and students with a foreign background received relatively higher grades than other students (Tyrefors Hinnerich & Vlachos 2012).

With assumptions no stronger than ordinarily invoked by regression analyses, the effect parameters can be estimated by a linear model. We estimated an extended linear mixed-effects model, allowing flexible modelling of the covariance structure (Pinheiro & Bates 2000) to allow for heteroscedasticity within groups.

Consider first the case in which the outcome is linear in the covariates, unconfoundedness holds, and treatment effects are constant across all units. Then, the parameters are estimable from the regression presented in equation (1).

$$\begin{aligned} SCORE_{ij} =& \alpha + \beta \mathbf{X}_{ij} + b_j + e_{ij} \\ b_j, \sim& N(0, \sigma_b^2) \\ e_{ij} \sim& N(0, \sigma_e^2 \mathbf{\Lambda}_{ij}) \end{aligned} \tag{1}$$

where $SCORE_{ij}$, the educational outcome, and $\mathbf{X}_{ij}$, the covariates, are defined in Section 2.3; i denotes individuals and j denotes exam occasions; α is the intercept, β denotes fixed effects, b_j are the exam random effects with variance σ_b^2, e_{ij} are the residuals with variance defined by σ_e^2. $\mathbf{\Lambda}_{ij}$ structures the residual covariances as to control heteroscedasticity (Pinheiro & Bates 2000) through estimating a vector of parameters ($\boldsymbol{\delta}$) based on a power function of MATHEMATICS (with a parameter δ_0) and functions of the covariate

[10] The Swedish School Inspectorate is a governmental agency with the objective to ensure that all children and school students are provided with equal education of good quality in a safe environment.

values (with a parameter δ for each covariate value) as presented in equation (2).

$$\begin{aligned} &Var(e_{ij}) = \sigma_e^2 \mathbf{\Lambda}_{ij} = \sigma_e^2 MATHEMATICS_{ij}{}^{2\delta_0} \delta^2_{1,s_{ij}} \delta^2_{2,t_{ij}} \delta^2_{3,u_{ij}} \text{ where} \\ &\delta_{1,s_{ij}}, s \in \{ST_{2014}, ST_{2015}, ST_{2016}, ST_{2017}\} \text{ with the restriction } \delta_{1,ST_{2014}} = 1 \\ &\delta_{2,t_{ij}}, t \in \{NEW, OLD, LATE, MATH\} \text{ with the restriction } \delta_{2,NEW} = 1 \\ &\delta_{3,u_{ij}}, u \in \{SWE\_MALE, SWE\_FEMALE, FOR\_MALE, \\ &\qquad FOR\_FEMALE\} \text{ with the restriction } \delta_{3,SWE\_MALE} = 1. \end{aligned} \tag{2}$$

Models were compared using conditional AIC (cAIC) based on parametric conditional bootstrap (Greven & Kneib 2010), and marginal (fixed effects) and conditional (fixed and random effects) R-squared (Nakagawa et al. 2017). Multiple imputation via chained equations (Van Buuren 2018) was used to handle missing values. The imputation model was allowed to contain additional demographic, upper secondary school and university-related variables.

3.1 Realignment and university prerequisites

The ST course was compulsory both for a Bachelor's degree in Statistics and in Mathematics and students typically registered in the middle semesters. There was more variation among non-programme students or those from other programmes in when they took ST as part of their studies.

When studying the realignment effect, to reduce the risk of post-intervention effects between the IS and ST courses, such as relevant training in form of coursework or jobs, we focused on students taking ST within one year after IS. This was also consistent with the schedule in the curriculum for a bachelor's degree in Statistics, so the bulk of programme students was likely to be included.

Our treatment group took IS at the earliest in Autumn 2015 and ST within a year after. The main control group took IS before Autumn 2015 and ST within a year after. Students taking ST more than one year after IS, irrespective of when, were treated as an extraneous group. Students who had not taken IS but attained eligibility to ST via Mathematics (22.5 credits) prerequisites formed an additional control group.

The realignment was an in-subject work that was reasonably not widely known to the prospective student group. Although there is unlikely much self-selection into a specific academic year, differences might be driven by the variation of the IS-course and ST-course across years. To account for the variation across written exams, we primarily assumed a random effect with groups conforming to the written exam opportunities. Given that all students attempted their first exam before the next years upcoming study-exams, there were no crossed effects between ST-course years and exam attempts and therefore the between year variation was nested within and encompassed by the within study-exam years. Controlling for the variation between IS-course semesters would have demanded a more elaborate model since the extraneous group came from several semesters and a few of them credited the IS-course from another university.

3.2 Course achievement and secondary school special eligibility requirements

As presented above, in juridical terms, the special eligibility requirements are to be interpreted as the minimum requirements needed to be able to absorb the educational content. However, we argue that it is reasonable to assume that course grades rather are measures on an underlying latent factor, where not passing is truncated, such that the degree to which the eligibility requirement for courses was passed may be used as a proxy to estimate the counterfactual situation. We hypothesize that the difference in average effect on an outcome in the ST-course between students who passed or passed secondary school Mathematics with distinction was a reasonable proxy for the difference in average effect on an outcome in the ST-course between students who did not pass and who passed secondary school Mathematics. A weaker assumption would be that the underlying function was only monotonic, such that we would at least be able to estimate the direction of the effect of the difference between passing or not passing secondary school Mathematics and the outcome in ST.

The special eligibility requirements (as mentioned in the institutional section) linked to the contents of the Bachelor's degree in Statistics, [11] consisting of one course passed in Social Sciences, two courses passed in English, and three courses passed in Mathematics were controlled for in our model.

The special eligibility requirements for a Bachelor in Mathematics were

[11] The same as for a degree in Economics or Business Administration.

four courses in Mathematics. Except for the (few) students with a Mathematics background, not many other students had more than three upper secondary school courses in Mathematics. Instead of modelling the missing fourth course or adding additional controls, we chose to use the average of the available ratings for the first three courses in Mathematics only.

Since no student was admitted without fulfilling the eligibility requirements, the counterfactual situation where an outcome was measured in the ST-course on a student not meeting the requirements was in practice non-existent.

4 Results

Table 7 presents the estimated fixed effects for $\boldsymbol{\beta}$ from regression (1). All other study groups had lower estimated effects on SCORE compared to NEW, but the estimated parameter was smallest and insignificant for LATE. The direction of the age polynomial effects were opposite to the expected but also insignificant, both as parts or as a whole. Compared to SWE_MALE the other groups had significant negative effects on SCORE. There was a negative but insignificant effect from SOCIAL, a positive from ENGLISH and a huge positive and significant from MATHEMATICS.

None of the random exam effects stood out (Figure 12). However, there was little support to remove them (cAIC decreased by 8.6). The intraclass correlation (0.169) also supported keeping the random effects. The increase from marginal(0.607) to conditional (0.673) R-square were also considerable.

No outliers were detected. One initially influential observation (standardized residual -2.90) was normalized (standardized residual -2.40) after the covariances were structured to handle the heteroscedasticity, see equation (2). Multicollinearity was of no particular concern (condition number 10.6 and VIFs ranging between 1.1 and 1.8).[12] The residuals resembled a normal distribution.

When LAGE2 was excluded from the model specification, the estimated effect of LAGE became positive, as expected, but the estimated parameter was still insignificant (Table 8). Most other estimated fixed effects were not much affected, but the random effects where considerably reduced and almost became redundant. Also, marginal R-squared increased to 0.639 and

[12]VIFs were calculated with the largest study group OLD as reference and with LAGE and LAGE2 orthogonalized.

Table 7: Estimated fixed effects from regression on SCORE (0-60), n=136 observations and m=1000 multiply imputed datasets

	estimate	se	df	p
Intercept	0.28	7.09	108.8	0.968
NEW(ref)	0			
OLD	-5.65	3.01	118.1	0.063
LATE	-3.10	2.46	118.2	0.211
MATH	-7.17	2.93	111.9	0.016
LAGE	-5.63	4.16	98.1	0.179
LAGE2	1.87	1.20	97.4	0.124
SWE_MALE(ref)	0			
SWE_FEMALE	-6.73	2.19	120.0	0.003
FOR_MALE	-7.66	2.45	104.0	0.002
FOR_FEMALE	-8.02	1.85	118.0	0.000
SOCIAL(0-2)	-2.26	2.30	117.9	0.328
ENGLISH(0-2)	5.83	3.03	113.3	0.057
MATHEMATICS(0-2)	22.30	2.88	115.1	0.000

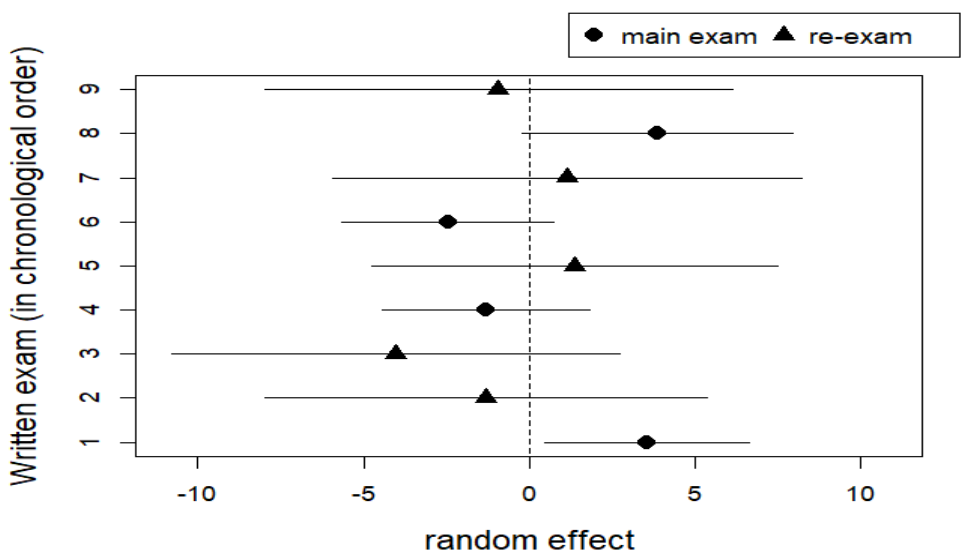

Figure 12: Estimated random effects of exams (in chronological order) with 95% confidence intervals

Table 8: Estimated fixed effects from regression on SCORE (0-60), n=136 observations and (m=1000) multiply imputed datasets. LAGE2 excluded.

	estimate	se	df	p
Intercept	-5.34	6.11	111.2	0.384
NEW(ref)	0.00			
OLD	-5.90	3.00	120.0	0.052
LATE	-3.07	2.54	118.8	0.228
MATH	-6.47	3.12	110.3	0.040
LAGE	0.77	1.42	84.9	0.591
SWE_MALE(ref)	0.00			
SWE_FEMALE	-6.77	2.19	119.8	0.002
FOR_MALE	-7.37	2.46	111.9	0.003
FOR_FEMALE	-8.19	1.87	116.4	0.000
CIVICS(0-2)	-1.78	2.24	121.0	0.429
ENGLISH(0-2)	5.98	3.09	115.2	0.055
MATHEMATICS(0-2)	22.42	2.94	116.1	0.000

Table 9: Estimated fixed effects from regression on SCORE (0-60), n=136 observations and (m=1000) multiply imputed datasets. SINE-function replace polynomial of LAGE.

	estimate	se	df	p
Intercept	-4.05	5.54	118.4	0.467
NEW(ref)	0.00			
OLD	-6.82	2.92	120.9	0.021
LATE	-3.56	2.56	119.5	0.166
MATH	-7.31	3.11	118.0	0.020
SINE(LAGE* π)	-0.85	1.16	118.1	0.464
SWE_MALE(ref)	0.00			
SWE_FEMALE	-7.12	2.17	120.1	0.001
FOR_MALE	-6.52	2.59	119.1	0.013
FOR_FEMALE	-8.65	1.96	116.2	0.000
SOCIAL(0-2)	-1.24	2.39	121.1	0.606
ENGLISH(0-2)	5.83	3.09	113.8	0.062
MATHEMATICS(0-2)	22.19	3.02	115.8	0.000

conditional decreased to 0.640, while cAIC decreased by 2.0 and the standardized residual was -2.33).Replacing the age polynomial by a sine function worsened the fit (cAIC decreased by 4.8 while marginal and conditional R-squared were 0.594 and 0.653 and the standardized residual was -2.28), see Table 9. The estimate of study group OLD strengthened and the estimate of SOCIAL weakened.

There were convergence problems in at least one of the imputed datasets when excluding the (pre-covariance structured) influential observation; including a polynomial of LAGE by group of study; replacing with a sine function of LAGE by group of study; including ST year fixed effects by group of study; or including ST year fixed effects. The ST year fixed effects represented the main exams to a large extent (93% of the observations), and in estimable situations, the random effects became redundant when these were added but the fit worsened (cAIC decreased by 4.2).

5 Discussion and conclusions

Our results supported that the realignment of the introductory IS-course had a positive medium-term effect on the achievement in the intermediate ST-course, indicating a better consolidation of basic knowledge without changing the syllabus or increasing the tuition time. We did not have access to data on students admitted to the IS-course, and therefore we could not directly compare the selection into the ST-course. However, it is not unlikely that the realignment of the introductory IS-course also influenced the choice to study the intermediate ST-course, and contributed to the observed baseline differences between the study groups NEW and OLD.

To fulfill the independence assumption, there are clear advantages of using outcome measures in subsequent courses when evaluating a course realignment, but there are also drawbacks. For example, if there are too many students who passed the IS course and then decided not to study the ST-course, the estimates might be biased due to self-selection. Given that at our university, the IS-course was compulsory for students in Economics and Business Administration and the ST-course was not, a considerable amount of students did not continue from IS to ST and the dataset was therefore smaller, leading to reduced precision and likely contributed to an inflated R-square.

A clear advantage was that the ST course was well-established with the same teacher and likely had a very small yearly variation with well calibrated examinations. A similar average score on exams and re-exam did not indicate any strong strategic behaviour and supported using the first written exam attempt as the outcome. Some students might otherwise have attended the ordinary exam, but if they believed they would fail, skip handing it in and later return at a re-exam in hope of a perceived more suitable exam.

Compared to all other groups, the study group NEW seemed to be more homogeneous with about half the standard deviation in SCORE. Compared to the study group OLD, unconditionally, the results for the study group NEW were 39% higher. Most of the difference remained in the estimated regression.

Comparing the study groups NEW to MATH, the unconditional difference is small (4% higher), but in the estimated regression, the difference was about the same as for the study group OLD. This result suggests that students who were eligible via Mathematics requirements seemed to have a disadvantage of not taking the realigned IS, but compensated (at least to some extent) through their stronger inclination towards mathematics, channeled through their higher grades in secondary school mathematics. From this point of view, the university eligibility requirements for the study group MATH seemed reasonably balanced. However, the outcome for these students was the most variable, indicating an increased challenge to the teaching and learning activities for this group in the ST-course. This result might reflect the self-regulatory competence of students, which is central to academic success at all levels of schooling (Duckworth & Carlson 2013).

The study group LATE was a deviant group, which surely had very different experiences during the on average four, but at most 18.5, years they had between the IS- and the ST-course. Unconditionally, the average SCORE was about the same in study group LATE as in NEW, and although not significantly, the regression estimate was negative. However, these results might by affected by selective drop out. We excluded the students who never attempted the written exam. Since all where males and four belonged to LATE, this suggest that these estimates might be upwardly biased since they likely would have acheived poorly had they attempted the exam.

The present study strongly supported that upper secondary school performance are ”the best predictor for university success” (Danilowicz-Gösele et al. 2017), although there was a huge difference across the special eligibility courses. The huge estimate of MATHEMATICS indicates the utter relevance of the associated (underlying) abilities and the impact of the special eligibility requirements. Although the teaching activities were in Swedish, the course literature was in English, and this might have had an impact on the positive estimate of ENGLISH. However, the estimated parameter for ENGLISH was likely also a measure of a more general (underlying) verbal ability.

Given that the ST course focused on probability theory and statistical inference, it seems reasonable that the Social Science eligibility requirement had a non-significant (negative) estimate. However, this eligibility requirement might be more relevant in other courses within a Bachelor's programme in Statisitcs. Further investigation would be needed to evaluate the necessity of such a requirement. Although entry exceptions were allowed at universities, the smooth path was likely to use the standardized field qualifications to simplify the communication of course and programme requirements to students and university staff.

Regarding demographics, the small sample might be a reason for not finding statistical support for age having an impact on the exam score. However, we found statistical support for the relationship between exam score and the interaction between gender and foreign background. Although a name-coded origin might be considered a crude measure for foreign background, the exam were anonymous, so the relatively high value of the estimated correlation outcome raises concerns of whether Swedish males had an unfair advantage or whether some measures of ability or other essential variables were omitted or if there were other errors involved.

We only had some of the relevant figures for the IS course, so it was difficult to draw any strong conclusions on how gender and background of students and teachers might have influenced the decision of the IS-students to continue to study the ST-course. The proportion with Swedish background was about the same (85%) in study group NEW compared to students registered to the Bachelor programme in Statistics and the IS course during 2015-2016 (82%), but much higher compared to the proportion in the OLD study group (69%). The proportion of women was lower in the study group NEW (24%) in comparison to students registered to the Bachelor programme in Statistics and the IS course during 2015-2016 (47%) and to the study group OLD (37%). This might ndicate that the realigned course and/or having teachers of both genders, gave, in relative terms, a disproportionately high incentive for males to continue to the ST course, and that a non-realigned course and/or having only a male teacher gave a disproportionate high incentive for students with foreign background to continue to the ST course. Such a selective drop out might also explain part of the differences between

Swedish males and other students. An alternative explanation is the indication of grade discrimination in upper secondary school as suggested by Tyrefors Hinnerich & Vlachos (2012). The sample of this study corresponds to the centre of the age distribution of the students in our sample when they attended upper secondary school. If males and/or students with Swedish background in our sample had received higher grades in English and Mathematics, the estimates would have been more similar.

One might also hypothesize that the largest groups (Swedish males) had a systematic advantage to the smaller groups in the learning activities within the ST course, leading to differences in exam scores. However, this hypothesis is not supported by our results. Given the very small size of our sample, such a phenomenon calls for further investigation.

Nonetheless, our results suggest the need of analyzing the development of the relationship between curricula in higher education and those in secondary education over time. As suggested by Carver et al. (2016), in Statistics, the higher education curriculum builds upon what has been taught in secondary education in a more-or-less cumulative manner. It is thus important to acknowledge that various phenomena might well be manifested earlier in the life span, which calls for bettering supporting the consolidation of basic concepts that were introduced in secondary education.

References

Bendikson, L., Hattie, J. & Robinson, V. (2011), 'Identifying the comparative academic performance of secondary schools', *Journal of educational administration* **49**(4), 433–449.

Carrell, S. E. & West, J. E. (2010), 'Does professor quality matter? evidence from random assignment of students to professors', *Journal of Political Economy* **118**(3), 409–432.

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Holmes Rowell, G., Velleman, P., Witmer, J. & Wood, B. (2016), 'Guidelines for assessment and instruction in statistics education college report', GAISE College Report ASA Revision Committee. http://www.amstat.org/education/gaise.

Danilowicz-Gösele, K., Lerche, K., Meya, J. & Schwager, R. (2017), 'Determinants of students' success at university', *Education economics* **25**(5), 513–532.

Duckworth, A. L. & Carlson, S. M. (2013), Self-regulation and school success, *in* 'Self-Regulation and Autonomy', Cambridge University Press, pp. 208–230.

Greven, S. & Kneib, T. (2010), 'On the behaviour of marginal and conditional aic in linear mixed models', *Biometrika* **97**(4), 773–789.

Hattie, J. (2015), 'The applicability of visible learning to higher education', *Scholarship of teaching and learning in psychology* **1**(1), 79–91.

Johnson, V. E. (2006), *Grade inflation: A crisis in college education*, Springer Science & Business Media.

Keeler, J. M. & Koretsky, M. D. (2016), 'Surprises in the muddy waters of high-enrollment courses', *Journal of Chemical Education* **93**(11), 1830–1838.

Little, R. J. (1992), 'Regression with missing x's: a review', *Journal of the American statistical association* **87**(420), 1227–1237.

Mosteller, F. (1989), 'The 'muddiest point in the lecture'as a feedback device', *On Teaching and Learning: The Journal of the Harvard-Danforth Center* **3**, 10–21.

Nakagawa, S., Johnson, P. C. & Schielzeth, H. (2017), 'The coefficient of determination r2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded', *Journal of the Royal Society Interface* **14**(134), 20170213.

Pinheiro, J. C. & Bates, D. M. (2000), *Mixed-Effects Models in S and S-Plus*, Springer. ISBN 0-387-98957-0.

Shepard, L. A., Penuel, W. R. & Pellegrino, J. W. (2018), 'Classroom assessment principles to support learning and avoid the harms of testing', *Educational measurement, issues and practice* **37**(1), 52–57.

Tyrefors Hinnerich, B. & Vlachos, J. (2012), 'Systematiska skillnader mellan interna och externa bedömningar av nationella prov. i: Resultatbilaga till lika för alla', *Omrättning av nationella prov i grundskolan och gymnasieskolan under tre år* pp. 63–122.

UHR (2020), 'Räkna ut ditt meritvärde', https://antagning.se/sv/betyg-och-behorighet/gymnasieskolan/avgangsbetyg-fore-1997/rakna-ut-ditt-meritvarde/. Universitets- och högskolerådet.

Van Buuren, S. (2018), *Flexible imputation of missing data*, CRC press.

Von Hippel, P. T. (2007), 'Regression with missing ys: an improved strategy for analyzing multiply imputed data', *Sociological Methodology* **37**(1), 83–117.

Weinberg, B. A., Hashimoto, M. & Fleisher, B. M. (2009), 'Evaluating teaching in higher education', *The Journal of Economic Education* **40**(3), 227–261.

Yunker, P. J. & Yunker, J. A. (2003), 'Are student evaluations of teaching valid? evidence from an analytical business core course', *Journal of education for business* **78**(6), 313–317.