CSEF

Centre for Studies in Economics and Finance

WORKING PAPER NO. 506

Rank Concerns, Peer Effects, and Ability Tracking in University. Evidence from a Randomized Experiment

Marco Bertoni and Roberto Nisticò

July 2018

University of Naples Federico II

University of Salerno

Bocconi University, Milan

CSEF - Centre for Studies in Economics and Finance
DEPARTMENT OF ECONOMICS – UNIVERSITY OF NAPLES
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unisa.it

ISSN 2240-9696

WORKING PAPER NO. 506

Rank Concerns, Peer Effects, and Ability Tracking in University. Evidence from a Randomized Experiment

Marco Bertoni* and Roberto Nisticò**

Abstract

If relative rank within classes enhances student achievement, tracking will help low-ability students and may harm high achievers. Using data from a randomized experiment generating a wide range of support of group ability composition, we show that students with higher ordinal ability rank within groups have better academic outcomes. We use our flexible education production function and the ample support of the data to predict the effects of alternative grouping polices. When we unpack the mechanisms behind ability tracking, we show that rank and peer effects work in opposite directions in generating outcomes for low- and high-ability students.

Keywords: ability tracking, rank concerns, peer effects.

JEL Classification: I21, I24, J24.

Acknowledgements: We thank Giorgio Brunello for helpful comments. Marco Bertoni acknowledges funding from a CARIPARO Foundation Starting Grant. The usual disclaimer applies.

* Università di Padova. Corresponding author: Department of Economics and Management "Marco Fanno", Università di Padova, Via del Santo 33, 35123 Padova, Italy. Phone: +39 049 8274002. Email: marco.bertoni@unipd.it.

** Università di Napoli Federico II and CSEF. E-mail: roberto.nistico@unina.it.

Table of contents

1. Introduction

A long-standing debate in the economics of education concerns the effectiveness of ability grouping to improve student performance.

On the one hand, learning spillovers from interaction with high-ability peers should make ability mixing more favourable to low ability students, while ability tracking should be more beneficial for high ability students. Although ability peer effects are context-dependent (see Carrel et al., 2013), there is by now a consensus on their heterogeneous and non-linear nature (see Sacerdote, 2014), that makes it hard to evaluate their implications from a general perspective. On the other hand, within homogeneous groups teaching should be easier, as it can be targeted to the ability level of the whole class, favouring ability tracking (see Duflo et al., 2011).

This paper highlights a mechanism behind the effect of ability grouping policies that has been so far overlooked, and that goes through ordinal ability rank within groups. A recent literature (see Murphy and Weinhardt, 2013, Elsner and Isphording, 2017 and 2018, Cicala et al., 2017, and Tincani, 2017) has shown that ability rank has a positive causal effect on achievement. By changing the whole ability distribution within groups (see Booij et al., 2017), different assignment policies will simultaneously affect also students' relative ability ranking within groups.

The contribution of this paper is twofold. First, we innovate on the existing literature on rank effects by demonstrating the presence of positive causal rank effects on educational achievement in a randomized experiment carried out at the University of Amsterdam by Booij et al., 2017, who randomly allocated first year students in economics to tutorial groups to achieve an unprecedentedly wide support of group ability composition, enhancing the external validity of our estimates. To do so, we estimate a very flexible education production function which permits both individual relative rank and the ability composition of peers to affect outcomes. This also allows us to assess how the omission of relative ability rank among the inputs in the education production function biases the coefficients related to own ability and to peer characteristics.

Second, using this flexible model and the large support of group ability configurations available in the data, we estimate the overall effect of a broad set of possible group assignment policies on student achievement, and unpack it into two components: a rank effect and a peer composition effect. Our results show that rank and peer effects contribute in opposite direction to generate outcomes for low and high-ability students. For instance, as we move from a system based on ability mixing to two- or three-way tracking students at the bottom of the ability distribution will lose out in terms of average ability of peers and will face a more homogeneous environment, but they will gain in terms of relative ability rank within their group.

By overlooking this mechanism, previous studies on ability tracking gave a misleading picture of the relevance of learning externalities due to peer effects for education production. Finally, as in Booij et al., 2017, our results from survey data show little evidence of teacher responses to changing group ability composition.

The paper unfolds as follows. Section 2 describes the experimental setup and the data. Section 3 illustrates the empirical methodology. We present the empirical results on our education production function with rank and peer effects in Section 4, while Section 5 discusses some robustness checks and extensions. Section 6 presents our results on ability tracking, while Section 7 discusses evidence on survey data on teacher responses and peer interactions. Conclusions follow thereafter.

2. Experimental setup and data

2.1. Experimental setting

Our analysis relies on data from a randomized experiment carried out by Booij et al., 2017, among about 2,000 students starting the three-year bachelor programme in economics and business at the University of Amsterdam in September 2009, 2010 and 2011. Close to 60% of the total teaching time of this program takes place into tutorial groups of roughly 40 students, whose composition is fixed throughout the first year. The experiment randomly allocated students to groups, with the aim of achieving a very wide support of ability composition.

In this context, ability is measured in terms of the grade point average at standardized nationwide secondary school final exams. Only a binned measure of GPA below 6.5, between 6.5 and 7, or above 7 was available at the time of assigning students to tutorial groups, before the beginning of the academic year. Hence, ability composition was manipulated by assigning to each group a different share of students from each GPA category (see Booij et al., 2017, for additional details).

Two additional features of the assignment mechanism are worth mentioning. First, students who took advanced math at high school were grouped together. Second, while in 2010 and 2011 the assignment was carried out in September, when the applications were closed, in 2009 students were assigned to a given tutorial group *at the moment* of application. As more able students applied earlier, high-ability groups were filled quickly. This may have given rise to a positive correlation between peer ability and peer motivation (proxied by application order). As a result, all regressions will include a saturated set of own GPA category, advanced math, and cohort-dummies, interacted with application order, that are necessary to grant conditional randomization of group ability composition.

Booij et al., 2017, provide evidence on the validity of the randomization by showing that background characteristics such as gender, age and previous attendance of professional college are uncorrelated with peer ability. In addition, they show that – as information about the randomized experiment was not provided to the staff involved in the assignment of teachers to tutorial groups - there is no systematic correlation between some available background characteristics of teachers assigned to a given tutorial group and group ability composition.

A final remark about the design concerns students who enrolled in the program, were assigned to a group, but never showed up. First, the share of no-shows is low and equal to less than 5% of students per group on average. Second, their decision to drop out cannot have been affected by ability composition of their assigned tutorial group, as no information about it was revealed to students before the courses started. As a result, all measures of group composition and ordinal rank within group used in the paper are constructed among the group of students actually attending the program. Still, in a robustness test we show that results are robust when we instrument actual rank and peer group features with their beginning-of-year counterparts.

2.2. Data

The data comes from the administration of the department of economics and business of the University of Amsterdam (see Booij et al., 2017).

Table 1 reports descriptive statistics for the main variables considered in this study. As in Booij et al., 2017, we measure student performance using three different outcome variables. First, our main outcome is the number of credits attained throughout the first year. The maximum number of credits attainable is 60, but only close to 20% of students reach this target and the average is close to 30, ruling out ceiling effects. Throughout the analysis, we proceed as Booij et al., 2017, and standardize credits to have zero mean and unit standard deviation within cohort. Second, we consider average grade at the exams completed during the first year, that we also standardize within cohort throughout the analysis. As on average students complete only 7 out of 13 exams that are scheduled for the first year, the validity of this otherwise commonly used performance measure is debatable in this context, because of self-selection issues. Finally, our third outcome is a "dropout" dummy, that in compliance with the University of Amsterdam's policies for enrolment in the second year of the program is equal to one if a student failed to complete at least 45 out of 60 credits during the first year, and zero otherwise. The choice of this variable is policy-motivated, as this core is a performance measure adopted by the University of Amsterdam. As shown in Table 1, only slightly more than half of the students in our sample pass the threshold for admission at the second year.

The main explanatory variables used in this paper concern student ordinal ability rank within tutorial group and tutorial group peer ability composition. As commonly done in the literature since Murphy and Weinhardt, 2013, we measure student rank as their percentile rank in the end-of-high-school GPA distribution within tutorial group. Since not all groups are of the same size, we normalize the raw ordinal rank by group size, according to the following formula:

$$RANK_{ig} = \frac{n_{ig} - 1}{N_g - 1}$$

where n_{ig} is the ordinal rank of individual i assigned to group g, and N_g is group size. As reported in Table 1, average individual rank in our sample is 0.488.

Figure 1 portrays the relationship between $RANK_{ig}$ and GPA_i. Panel a reports the raw data, and shows that – given the very wide support of ability composition generated by the randomization – there is large variability in $RANK_{ig}$ *throughout* the distribution of GPA_i. For instance, a student can be in the top 10% of the GPA_i distribution within his/her group even if he ranks only at the 35th percentile of the overall GPA_i distribution. On the other hand, there are students who rank at the 80th percentile of the overall GPA_i distribution and are in the bottom 10% of the distribution of GPA_i in their tutorial group. As the distribution of GPA_i has long tails, Panel b of Figure 1 underlines this feature of the data by reporting box-whisker plots of the distribution of $RANK_{ig}$ by decile of GPA_i (see Elsner and Isphording, 2017).

Second, as in Booij et al., 2017, we describe the ability composition of a student's tutorial group peers with the mean $\overline{GPA}_{-i}$ and the standard deviation $SD(GPA_{-i})$ of their end-of-secondary-school GPA_i. These are constructed after standardizing GPA_i to have zero mean and unit standard deviation within cohort, and leaving out individual i.

Finally, the data also contains information on the exact individual end-of-secondary school GPA_i - that we also standardize to have zero mean and unit standard deviation within cohort – on gender, age (categorized in tertiles within cohort), previous attendance of a professional college before university enrolment, as well as on the variables used to carry out the randomization (descriptive statistics not reported), that include cohort of enrolment, GPA_i category, advanced maths at high school and application order.

3. Empirical methodology

We build on Booij et al., 2017 and estimate the following education production function:

$$y_{ig} = \alpha_0 + \alpha_1 RANK_{ig} + \alpha_2 GPA_i + \alpha_3 \overline{GPA}_{-i} + \alpha_4 SD(GPA_{-i}) + \alpha_5 \overline{GPA}_{-i} \times SD(GPA_{-i}) + \quad (1)$$

$$+\alpha_6 GPA_i \times \overline{GPA}_{-i} + \alpha_7 GPA_i \times SD(GPA_{-i}) + \alpha_8 GPA_i \times \overline{GPA}_{-i} \times SD(GPA_{-i}) +$$

$$+x_i'\beta + \varepsilon_{ig}$$

where y_{ig} is the outcome of student i in tutorial group g (measured by credits, average grade or dropout); $RANK_{ig}$ is student i's percentile rank within the assigned tutorial group g; and GPA_i

measures his/her own prior ability. $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$ measure the mean and the standard deviation of peer GPA_i, respectively, while the interaction term $\overline{GPA}_{-i} \times SD(GPA_{-i})$ allows for the possibility that the mean and the SD of peer GPA_i are not perfect substitutes in shaping student performance. To allow for further flexibility, the peer variables are also interacted with own GPA_i. The vector of covariates x_i includes both randomization controls - a saturated set of own GPA category, advanced math, and cohort-dummies, interacted with application order - and background controls - male, being in the youngest third of the age distribution, being in the oldest third of the age distribution, and professional college.

In some specifications, we also include a set of tutorial group fixed-effects θ_g, which allow us to control for further unobserved group characteristics - such as group size, teacher effects and group atmosphere - in the estimation of rank effects. In those instances, the fixed effects θ_g absorb the peer variables $\overline{GPA}_{-i}$, $SD(GPA_{-i})$ and $\overline{GPA}_{-i} \times SD(GPA_{-i})$.

Finally, ε_{ig} is an error term. In all regressions, standard errors are clustered at the tutorial group level to allow for correlation among the outcomes of students assigned to the same tutorial group.

Conditional on the set of randomization controls included in the model, we can take group composition as being as good as randomly assigned. Therefore, we reach consistent and efficient estimates of the model parameters with Ordinary Least Squares (OLS). Still, in a robustness test, we will use Two-Stage Least Squares (TSLS) to instrument actual rank and peer group composition with their beginning-of-year counterparts, and deal with potentially endogenous attrition.

Finally, as in Booij et al., 2017, in Section 6 we use the flexibility granted by our education production function and the ample support of group configurations present in the data to estimate student outcomes under different grouping policies, and unpack the contribution of rank concerns and peer effects to generate ability tracking effects. We refer to Section 6 for details.

4. Empirical results

We present the estimates of Equation (1) on credits completed during the first year – our main outcome – in Table 2. In column (1) we estimate a "pure peer effects" model that replicates the specification in Table 4, column (5) of Booij et al., 2017. In column (2) we enrich this specification by including also $RANK_{ig}$ among the inputs of the education production function. The coefficient on $RANK_{ig}$ is positive and of an economically relevant magnitude: moving from the bottom to the top of the within-group ability distribution increases the number of credits achieved by close to half of a standard deviation. This effect is statistically significant at the 1% level. Consequently, the inclusion of $RANK_{ig}$ improves the fit of the model to the data, as confirmed by the lower values of both the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC).

This positive effect is in line with other studies on the effect of ordinal rank on educational attainment, such as Murphy and Weinhardt, 2013, and Elsner and Isphording, 2017, but is obtained in a context where ability group composition is randomized. Cicala et al., 2017, also estimate rank effects using experimental data (see Duflo et al., 2011), but their data have been purposively generated to estimate only the effects of two-way ability tracking relative to ability mixing. On the contrary, our setup spans a much wider set of ability group configurations, thereby allowing us to enhance external validity.

At this stage, it is also instructive to study how the coefficients related to peer ability composition change as we include rank among the explanatory variables. Conditional on GPA_i, we expect a negative correlation between $RANK_{ig}$ and both $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$. In fact, students with the same GPA_i shall have a lower $RANK_{ig}$ in groups with higher $\overline{GPA}_{-i}$ (or $SD(GPA_{-i})$), holding fixed $SD(GPA_{-i})$ (or $\overline{GPA}_{-i}$). Therefore, the omission of rank shall generate a negative omitted variable bias on the coefficients related to these variables. The coefficient related to GPA_i should also be affected, but in this case the bias is supposedly positive, given that students with higher GPA_i are ranked higher within groups.

All these predictions are confirmed when we compare columns (1) and (2) of Table 2 to assess the quantitative relevance of the bias. Although in no case the magnitude of the omitted variable bias is as severe as to reverse the signs of the coefficients of the peer variables, the point estimates change significantly. For instance, the coefficient of $\overline{GPA}_{-i}$ more than doubles, while the one on GPA_i more than halves. This evidence is further corroborated by the estimates of heterogeneous effects reported in Appendix Figure A1, that replicates Figure 3 in Booij et al., 2017. In particular, the top and bottom left panels show that, unlike in their case, once we control for $RANK_{ig}$ the effect of $\overline{GPA}_{-i}$ is positive and significant even in homogeneous groups.

Column (3) of Table 2 shows that our estimated rank effect is very stable even when we include tutorial group fixed effects, which control for any remaining group-specific unobservable characteristic that could be correlated with rank and the outcome. As noted above, the group fixed effects absorb the levels of the peer characteristics. In addition, their inclusion also improves the fit of the model, as verified by the lower AIC and BIC. In column (4), instead, we verify that omitting the set of individual background controls is not harmful for the estimation of rank effect. This is not surprising, given the conditionally random allocation of subjects to tutorial groups.

We report in Appendix Table A1 the estimation results for our two other outcomes, average grade and dropout. For each outcome, we report results for specifications as in columns (1) to (3) of Table 2. On the one hand, the rank effect on average grade is positive but imprecisely estimated. On the other hand, the effect on dropout is negative and statistically significant. In addition, in both cases the bias in the coefficients of the peer variables and of GPA_i due to the omission of $RANK_{ig}$ is qualitatively similar to the case of credits.

5. Extensions

5.1. Robustness tests

The estimated rank effects reported in Table 2 are robust to several sensitivities that we now describe. To save space, we report all robustness tests in the Appendix, and we consider only the most comprehensive specification of column (3), Table 2. Additional results are available from the authors.

First, all estimates discussed so far refer to students actually attending the program, thereby excluding a minority of students who enrolled in the program but never showed up. We verify that this endogenous choice does not invalidate our results by instrumenting actual rank and actual peer group $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$ with their beginning-of-year counterparts, including also the no-shows. Results for all three outcome variables are reported in Appendix Table A2. Given that there are very few no-shows, the first-stage F statistics are very high (close to 50). Additionally, the rank effects estimated with TSLS are statistically indistinguishable from the ones obtained with OLS, confirming that endogenous dropout is not a major concern.

Second, in Appendix Table A3 we relax the assumption of a linear relationship between $RANK_{ig}$ and the outcomes. We do so by using a quadratic instead of a linear functional form for $RANK_{ig}$ and by replacing the linear trend with bins for belonging to the middle or the upper third of the rank distribution. For credits, the data do not support the quadratic specification: the high collinearity between the linear and the quadratic term decreases the precision of the estimated rank effect dramatically, and the AIC and BIC are higher than for the linear model. Similarly, the specification with bins does not show strong evidence of non-linear effects: moving from the first to the second rank tertile causes a gain in credits that is roughly equal to half of the effect of moving from the first to the third tertile. Even in this case, the AIC and BIC are higher than for the linear model. Therefore, we conclude that the linear specification does a good job at approximating the data. Results for the other outcomes are comparable, although for average grade the estimates using bins suggest some evidence of a non-linear effect.

Next, as both panels of Figure 1 highlight a non-linear relationship between GPA_i and $RANK_{ig}$, in Appendix Table A4 we verify that the linear functional form that we have so far used to account for prior achievement is not overly restrictive. We do so by substituting the linear trend in GPA_i with a quadratic one or with dummies for belonging to the middle or the upper third of the GPA_i distribution.

In all cases and for all outcomes the estimated rank effect is comparable to the one obtained with the linear specification, albeit sometimes less precisely estimated. In addition, the AIC and BIC tend to support the linear specification.

We have also carried out the following additional robustness tests: (i) excluding the 2009 cohort, for which the assignment to groups was carried out at the moment of application and not after the collection of all applications, like in the following cohorts; (ii) excluding outlier students in the top and bottom 1% of the distribution of GPA_i. Results are in Appendix Table A5, and confirm our main findings.

5.2. Heterogeneous effects

We now investigate how the estimated rank effects varies with respect to two individual background characteristics – prior achievement and gender – and two features of the tutorial group – size and within-group heterogeneity in prior achievement.

Table 3 reports heterogeneous effects of rank on credits using the specification in column (3) of Table 2. Results for the other outcomes are presented in Appendix Table A6, and are in line with the ones for credits.

First, column (1) shows that the effect of rank is not significantly different when we distinguish between students with prior ability above and below the median, although the point estimate is larger for the former group than for the latter.

Second, in column (2) we show that while for males the rank effect is large and significant, it is small and not statistically significant for females. In this case, the difference in the effect across groups is statistically significant at the 5% level. This finding is in line with Murphy and Weinhardt, 2013, who also find heterogeneous rank effects by gender among English secondary school students, and is also consistent with a large literature on the heterogeneous gender attitudes towards competitiveness (see e.g. Gneezy et al., 2003).

In column (3) we investigate whether rank effects are heterogeneous by group size, distinguishing between groups of size above and below the median. In the data, group size varies between 32 and 45, and the median size is 40. Results highlight that rank matters more in smaller groups, the difference being statistically significant at 10% level. This finding confirms results from Elsner and Isphording, 2017, and supports the idea that students in smaller groups could be better informed about their rank, and thus be more responsive to this margin.[1]

[1] As actual group size is affected by the problem of no-shows, in a robustness test we have also instrumented actual group size and its interaction with $RANK_{ig}$ with the number of no-shows by group and its interaction with $RANK_{ig}$, as done by Booij et al., 2017. Results – available from the authors – are indistinguishable from the ones reported in Table 3.

Finally, column (4) reports heterogeneous effects by the SD of GPA, distinguishing between groups with SD above and below median. Like Elsner and Isphording, 2017, we do not find any evidence of heterogeneous effects along this margin.

6. Implications for Ability Tracking

We now discuss the educational effects at the student population level of different group assignment policies. As pointed by Booij et al., 2017, there are no obvious implications for group assignments when one departs from a linear-in-means peer effects model to employ a more flexible specification. In particular, if $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$ are allowed to be imperfect substitutes in affecting student achievement, then different grouping strategies will reflect different trade-offs between $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$. We therefore use our flexible education production function - Equation (1) - to compare the average performance under ability mixing with the average performance under five alternative grouping configurations, namely:

1) *Two-way tracking*: high-ability (GPA above median) and low-ability (GPA below median) students are grouped separately.
2) *Three-way tracking*: top-ability (GPA in top tertile), middle-ability (GPA in middle tertile) and bottom-ability (GPA in bottom tertile) students are grouped separately.
3) *Track low*: bottom-ability students are grouped together, while middle- and top-ability students are mixed.
4) *Track middle*: middle-ability students are grouped together, while bottom- and top-ability students are mixed.
5) *Track high*: top-ability students are grouped together, while bottom- and middle-ability students are mixed.

We use our estimates from the model in column (2) of Table 2 to compute the average treatment effects of the five different tracking policies relative to mixing. Following Booij et al., 2017, we proceed in two steps. First, we compute the mean values of both rank and the peer variables in the alternative grouping configurations. Second, we derive the mean predicted performance in our sample using the estimates (and the relative standard errors) reported in Table 2, column (2).[2] We do so for the whole population and by tertile of prior achievement.

[2] The (conditional) average treatment effects of tracking are computed as $(\overline{x}_{track} - \overline{x}_{mix})\hat{\beta}$ while the standard errors as $\sqrt{(\overline{x}_{track} - \overline{x}_{mix})'V(\hat{\beta})(\overline{x}_{track} - \overline{x}_{mix})}$ where $\overline{x}_{track}$ and $\overline{x}_{mix}$ are vectors of sample mean covariates that include the leave-out means of the rank and peer variables under alternative grouping strategies, and $\hat{\beta}$ the coefficients from the regression in Table 2, column (2).

Relative to Booij et al., 2017, our estimates are obtained from a specification that includes also rank among the covariates. This allows us to elaborate on the mechanisms behind ability tracking, and to unpack the total tracking effect into a rank effect and a peer effect. Total effects are obtained by changing both peer characteristics and rank as we move from ability mixing to tracking. Rank (Peer) effects are obtained by holding peer characteristics (rank) fixed and moving rank (peer characteristics) when switching from mixing to tracking.

Table 4 shows the estimated tracking effects on first-year credits. Results in columns (1a)-(1d) are for the whole population, while results in the following columns split students by tertile of ability (above-below median for two-way tracking). Total effects are reported in columns (1a), (2a), (3a), (4a), while rank and peer effects are shown in columns (1b), (2b), (3b), (4b) and (1c), (2c), (3c), (4c), respectively.

Results in columns (1a), (2a), (3a), (4a) are very similar, though larger in magnitude to those reported in columns (1) to (4) of Table 5 in Booij et al., 2017. This is expected, given the omitted variable bias discussed above, and confirms their two main findings:

(i) any grouping policy will enhance average student achievement compare to mixing,

(ii) the gains of switching from mixing to tracking are mostly concentrated at students in the lower two-thirds of the ability distribution.

The key contribution of this paper, however, is to qualify that these effects are at least in part due to rank effects, and cannot be entirely attributed to a direct effect of peer group composition, as often argued by the extant literature on this topic.

The estimates in columns (1b), (2b), (3b), (4b) and (1c), (2c), (3c), (4c) suggest in fact that rank and peer effects work in opposite directions in the production of student outcomes. This is particularly evident when looking at the effects differentiated by student ability category (Low, Middle, High).

For instance, reading across the estimates in the first row of Table 5, we find that, on average, students under two-way tracking experience an increase of 10% of a SD in the number of first-year credits compared to mixing. This effect is larger (16%) for low-ability students and smaller (5%) and insignificant for high-ability ones.

However, our separate rank and peer effects estimates indicate two new findings:

(i) for low achievers, the total effect is mainly driven by the rank effect. Hence, low-ability students are not advantaged by a tracked system because of interactions with peers of lower quality or higher peer group homogeneity. Instead, our results show that low-ability students gain because of the tracking-induced increase in relative ordinal rank within groups;

(ii) for high achievers, rank and peer effects have a similar magnitude but opposite sign, hence balancing out the total tracking effect. Therefore, high achievers do benefit from interacting with better peers or a higher homogeneity, but at the same time the presence of more able peers negatively affects their relative rank within groups, thereby harming their outcomes.

Our results also provide new evidence about the "track-middle" option that Carrel et al., 2013, viewed as optimal on the basis of their estimates on pre-treatment data. As found by both Carrel et al., 2013, and Booij et al., 2017, our estimates suggest that this grouping strategy has an insignificant and close to zero overall effect for low-ability students. However, we find that this zero effect is the sum of a positive and significant rank effect and a negative and significant peer composition effect of a similar magnitude. The former is due to the increase in average rank of low-ability students when switching from mixing to "track-middle" grouping, the latter is instead likely attributable to the increase in the heterogeneity of peer composition associated with track-middle grouping.

We gain additional insights also when we look at the effect of "track-high" grouping on middle- and high-ability students. In this case, we see that the positive tracking effect on middle-ability students is entirely attributable to rank concerns, while the overall zero effect for high-ability students hides a very negative rank effect and a positive effect of peer ability composition.

The estimated tracking effects on the other two outcomes, average grade and dropout, are qualitatively similar to those in Table 4 and are reported in Appendix Tables A7 and A8, respectively.

7. Evidence from survey data

To learn about the mechanisms behind peer effects, Booij et al., 2017, complemented their evidence from administrative data with a survey among the students involved in the experiment. The survey was carried out three months after the beginning of the academic year, and investigated aspects related to the teaching environment and interactions with peers. A total of 26 questions were asked throughout the 3 years, although the detailed content of the questionnaire changed slightly between years (see Booij et al., 2017, for further details). The response rate was fairly high, close to 70%. Importantly, Booij et al., 2017, show that survey response was unrelated to the ability composition of tutorial groups.

We follow Booij et al., 2017, and study the mechanisms behind both rank and peer effects on six index variables that summarize the informational content of the 26 survey items, standardized to have zero mean and unit standard deviation. The mapping between the indexes and the survey questions is as follows:

1. *Too fast:* tutorial group teachers are too fast, spend too little time on simple things, or give complicated answers;
2. *Too slow:* tutorial group teachers are too slow, spend too much time on simple things, or focus too much on weak students;
3. *Stimulating*: the student learns a lot from tutorial group teachers, group meetings are stimulating or teacher asks questions to test our understanding;
4. *Conducive:* there is a good atmosphere in tutorial group, the student learns from students in tutorial group, tutorial group influences performance positively;
5. *Interactive:* the student studies together with others, helps other students or is helped by other students
6. *Involved:* the student or others frequently ask questions; the level of other students demotivates the student (-), the student dislikes to ask questions (-); unquietness makes it difficult to concentrate (-).

The effects of rank and of the peer variables on these outcomes, estimated using the specifications in columns (2) and (3) of Table 2, are reported in Appendix Table A9.

On the one hand, irrespective of the inclusion of group fixed effects all our estimates of rank effects are too imprecise to be significant, but reassuringly they have the expected sign. In fact, we see that students with higher rank are seemingly more likely to state that teachers are too slow and less likely to say that they are too fast, less likely to find the peer environment stimulating and to benefit from learning from others in the tutorial group, more likely to interact and being involved with them.

On the other hand, given the insignificance of rank effects, estimates without group fixed effects reveal that the coefficients related to the peer variables are substantially in line with the ones estimated by Booij et al., 2017. These results suggest that teachers are not very responsive to group ability composition, while there is evidence that low-ability students are more likely to feel involved in the class when surrounded by peers of similar ability, and vice-versa.

8. Conclusions

This paper provides evidence that ability rank positively affects student achievement using data from a randomized experiment carried out at the University of Amsterdam by Booij et al., 2017, that spans a very broad support of group ability composition. Thanks to our very flexible education production function, which allows both a student relative rank within group and the ability composition of the others in the group to affect outcomes, we assess the extent to which omitting rank among the inputs in the production function biases the estimates of the coefficients associated to own ability and peer characteristics.

Moreover, we provide novel evidence that ordinal ability rank within groups is an important mechanism behind the effect of ability grouping policies. We do so by unpacking the overall effect of a battery of grouping scenarios on student achievement into two components: a rank effect and a peer composition effect. Our analysis indicates that rank and peer effects contribute in opposite direction in the production of student outcomes for low and high achievers. For instance, when switching from ability mixing to two- or three-way tracking students at the bottom of the ability distribution will lose out in terms of average ability of peers, but they will gain in terms of relative ability rank within groups.

Our results therefore highlight that the effects of ability tracking cannot be entirely attributed to peer composition effects. Instead, they are in large part due to rank concerns. For instance, as highlighted by Murphy and Weinhardt, 2013, this finding suggests that it could be possible to improve outcomes of high ability students, who face negative rank effects in a tracked system, by suggesting teachers to provide salient targeted information on their global instead of local ability ranking, and vice-versa for the low ability. In addition, our results also warn that the "big fish in a small pond" rank effect that could motivate students and parents to choose "low-tier" schools shall be compounded with the positive externalities that would instead motivate the choice of "top-tier" ones.

References

Booij, A. S., Leuven, E., & Oosterbeek, H. (2017). Ability peer effects in university: Evidence from a randomized experiment. *Review of Economic Studies*, *84*(2), 547-578.

Carrell, S. E., Sacerdote, B. I., & West, J. E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica*, *81*(3), 855-882.

Cicala, S., Fryer, R. G., & Spenkuch, J. L. (2017). Self-Selection and Comparative Advantage in Social Interactions. *Journal of the European Economic Association*.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101(5), 1739-74.

Elsner, B., & Isphording, I. E. (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics*, 35(3), 787-828.

Elsner, B., & Isphording, I. E. (2018). Rank, Sex, Drugs and Crime. Forthcoming, *Journal of Human Resources*, 0716-8080R.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *Quarterly Journal of Economics*, *118*(3), 1049-1074.

Murphy, R., & Weinhardt, F. (2013). *The Importance of Rank Position*. CEP Discussion Paper No. 1241. Centre for Economic Performance.

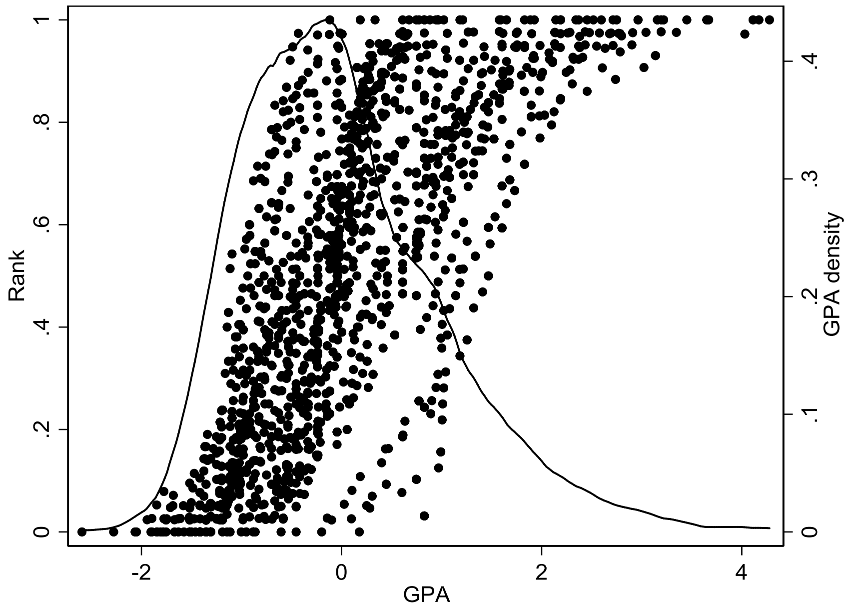
Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: two steps forward? *Annual Review of Economics, 6(1), 253-272.*

Tincani, M. (2017). *Heterogeneous peer effects and rank concerns: Theory and evidence.* CESifo working paper n.6331.

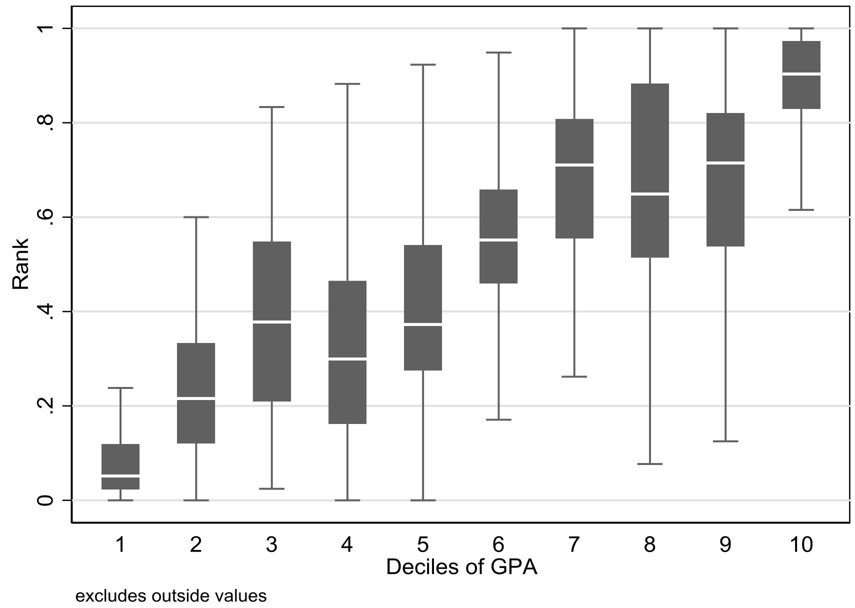
Figures and Tables

Figure 1. Variation in rank by level of GPA

a. Raw data



b. Boxplot by decile of GPA



Notes: Panel a. reports the joint distribution of rank and GPA. The estimated density of GPA is overlaid. Panel b. reports the box-plot of rank by decile of GPA. Number of observations: 1,876.

Table 1. Descriptive statistics

	(1) Mean	(2) Std. Dev.
Outcomes:		
Credits collected in the first year (standardized by cohort)	0	1
Average grade in the first year (standardized by cohort)	0	1
Dropout at the end of first year	0.488	0.500
Explanatory variables:		
$RANK_{ig}$	0.486	0.298
GPA_i (standardized by cohort)	0	1
$\overline{GPA}_{-i}$	-0.004	0.580
$SD(GPA_{-i})$	0.785	0.289
Male	0.733	0.443
Age in youngest third of the distribution	0.333	0.472
Age in oldest third of the distribution	0.329	0.470
Professional college	0.056	0.207

Notes: the number of observations is 1,876 for all variables except for average grade, which is only available for 1,753 students who completed some exams. Dropout is a dummy variable for having collected less than 45/60 credits in the first year. Professional college is a dummy for entering university after enrolment in professional college.

Table 2. Main results. Rank and peer effects on number of credits collected.

	(1)	(2)	(3)	(4)
$RANK_{ig}$	-	0.564***	0.508***	0.529***
		(0.179)	(0.175)	(0.174)
$\overline{GPA}_{-i}$	0.148***	0.360***	-	-
	(0.052)	(0.086)		
$SD(GPA_{-i})$	-0.185**	-0.200**	-	-
	(0.082)	(0.079)		
$\overline{GPA}_{-i} \times SD(GPA_{-i})$	0.343*	0.132	-	-
	(0.190)	(0.183)		
GPA_i	0.350***	0.143**	0.147**	0.166**
	(0.035)	(0.065)	(0.068)	(0.067)
$GPA_i \times \overline{GPA}_{-i}$	-0.117***	-0.130***	-0.122**	-0.123**
	(0.042)	(0.040)	(0.047)	(0.048)
$GPA_i \times SD(GPA_{-i})$	0.104	0.305***	0.257**	0.242**
	(0.075)	(0.091)	(0.103)	(0.103)
$GPA_i \times \overline{GPA}_{-i} \times SD(GPA_{-i})$	-0.287**	-0.376***	-0.196	-0.206
	(0.138)	(0.135)	(0.197)	(0.201)
Controls				
Tutorial group fixed effects	No	No	Yes	Yes
Randomization controls	Yes	Yes	Yes	Yes
Background controls	Yes	Yes	Yes	No
F test (p-value)				
Peer (and rank) variables = 0	0.003	<0.001	0.008	0.005
AIC	4824	4815	4774	4800
BIC	5085	5075	5007	5011

Notes: Each column reports the results from a different OLS regression. Dependent variable: number of collected credit points in the first year. The outcome is standardized to have zero mean and unit standard deviation. Randomization controls are a saturated set of own GPA category, advanced math, and cohort-dummies, interacted with application order. Background controls are: male, being in the youngest third of the age distribution, being in the oldest third of the age distribution, professional college. The peer variables $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$ are re-centred to have zero means. $\overline{GPA}_{-i}$, $SD(GPA_{-i})$ and $\overline{GPA}_{-i} \times SD(GPA_{-i})$ are not included in columns (3) and (4) due to collinearity with the tutorial group fixed effects. Number of observations: 1,876. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. *** p<0.01, ** p<0.05, * p<0.1.

Table 3. Heterogeneous effects of rank on number of credits collected

	(1)	(2)	(3)	(4)
$RANK_{ig}$ (a)	0.326	0.270	0.611***	0.504**
	(0.205)	(0.207)	(0.185)	(0.192)
$RANK_{ig} \times GPA_i$ *above median* (b)	0.223			
	(0.227)			
$RANK_{ig} \times MALE_i$ (b)		0.296**		
		(0.141)		
$RANK_{ig} \times GROUP\ SIZE_i$ *above median* (b)			-0.289*	
			(0.164)	
$RANK_{ig} \times SD(GPA)$ *above median* (b)				0.066
				(0.190)
$(a) + (b)$	0.549**	0.566***	0.321	0.570***
	(0.253)	(0.178)	(0.203)	(0.199)

Notes: the table reports heterogeneous effects of rank on credits by own ability, gender, tutorial group size and tutorial group heterogeneity. Each column reports the results from a different OLS regression. Estimates based on the specification used in Table 2, column (3). Each column reports the linear effect of rank, the interaction term between rank and the dummy variable for the category of interest, and the linear combination of the two. The dummy variable for the category of interest is also included among the controls. Number of observations: 1,876. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. *** p<0.01, ** p<0.05, * p<0.1.

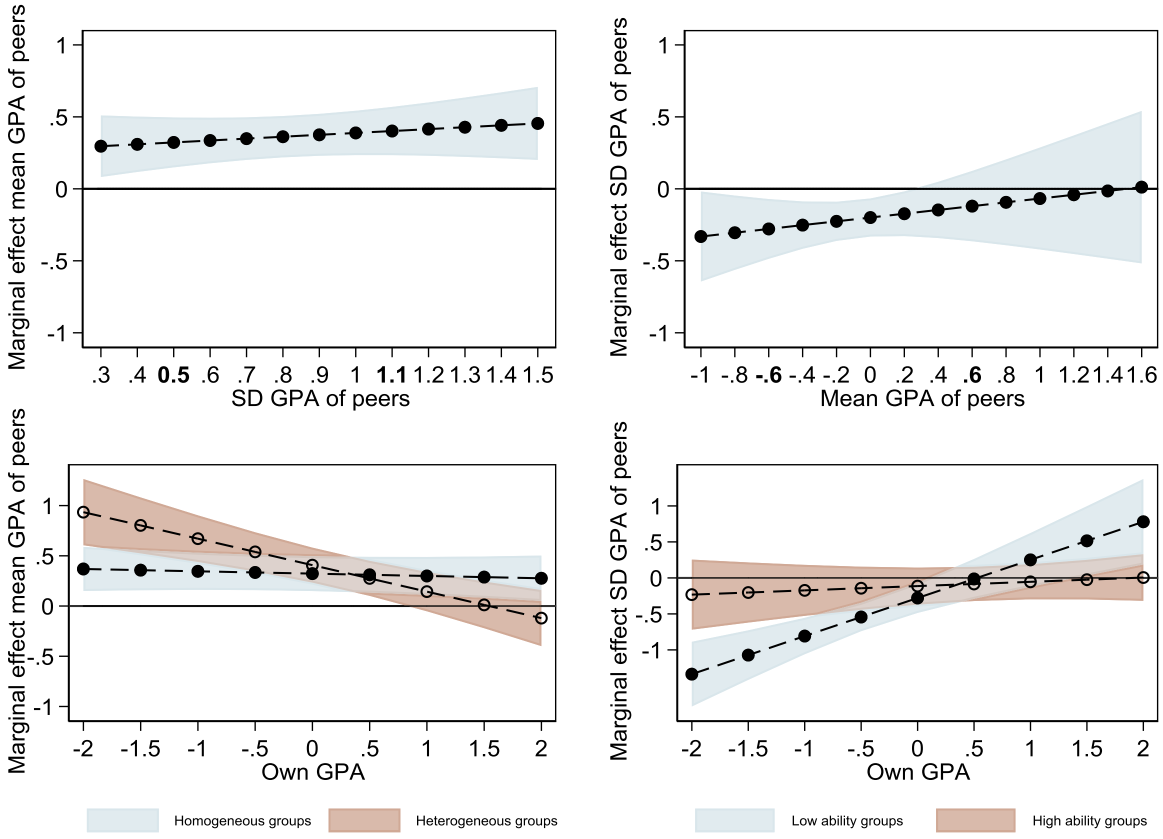
Table 4. Estimated tracking effects on first-year credits compared to mixing. Total effects and unpacking rank and peer effects.

					Student GPA category								
			ATE			L(B)			M			H(A)	
		Total	Rank	Peer	Total	Rank	Peer	Total	Rank	Peer	Total	Rank	Peer
		(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)
Two-way tracking	{B},{A}	0.104***	-0.001***	0.105***	0.157***	0.139***	0.018				0.050	-0.142***	0.192***
		(0.028)	(0.000)	(0.029)	(0.040)	(0.044)	(0.048)				(0.040)	(0.045)	(0.060)
Three-way tracking	{L}, {M}, {H}	0.147***	-0.004***	0.151***	0.268***	0.184***	0.084	0.147***	-0.003***	0.150***	0.028	-0.190***	0.218**
		(0.037)	(0.001)	(0.037)	(0.072)	(0.058)	(0.076)	(0.056)	(0.001)	(0.056)	(0.055)	(0.060)	(0.084)
Track low	{L}, {M, H}	0.128***	-0.002***	0.130***	0.268***	0.184***	0.084	0.090**	-0.142***	0.232***	0.027	-0.048***	0.074*
		(0.031)	(0.001)	(0.031)	(0.072)	(0.058)	(0.076)	(0.037)	(0.045)	(0.050)	(0.032)	(0.015)	(0.038)
Track middle	{M}, {L, H}	0.042***	-0.002***	0.044***	-0.009	0.046***	-0.055**	0.147***	-0.003***	0.150***	-0.011	-0.047***	0.037
		(0.011)	(0.001)	(0.011)	(0.023)	(0.015)	(0.024)	(0.056)	(0.001)	(0.056)	(0.023)	(0.015)	(0.028)
Track high	{L, M}, {H}	0.065***	-0.001***	0.066***	0.094***	0.046***	0.048	0.073**	0.140***	-0.067	0.028	-0.190***	0.218**
		(0.024)	(0.000)	(0.024)	(0.028)	(0.015)	(0.029)	(0.035)	(0.044)	(0.051)	(0.055)	(0.060)	(0.084)

Notes: The table reports (conditional) average treatment effects on credits of different tracking configurations relative to mixing based on the estimates from Table 2, column (2). Total effects obtained by changing both peer characteristics and rank as we move from ability mixing to tracking. Rank (Peer) effects are obtained by holding peer characteristics (rank) fixed and moving rank (peer characteristics) as we move from ability mixing to tracking. Student GPA groups are L(ow), M(iddle), H(igh) in case of three-way tracking, and for two-way tracking B(elow) and A(bove). The curly brackets indicate the grouping of GPA groups. Number of observations: 1,876. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. *** p<0.01, ** p<0.05, * p<0.1.

Appendix – for On-line publication only

Figure A1. Average marginal effects of peer group composition on number of credits – interaction effects conditional on $RANK_{ig}$



Notes: the figure replicates Figure 3 in Booij et al. (2017), including $RANK_{ig}$ among the controls. Marginal effects based on the estimates from Table 2, column (2).

Table A1. Rank and peer effects on average grade and dropout.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable		Average Grade			Dropout	
$RANK_{ig}$	-	0.181	0.154	-	-0.267**	-0.248**
		(0.158)	(0.172)		(0.102)	(0.104)
$\overline{GPA}_{-i}$	0.124**	0.192***	-	-0.042	-0.142***	-
	(0.047)	(0.070)		(0.027)	(0.047)	
$SD(GPA_{-i})$	-0.203***	-0.208***	-	0.066	0.073	-
	(0.074)	(0.071)		(0.045)	(0.044)	
$\overline{GPA}_{-i} \times SD(GPA_{-i})$	-0.019	-0.086	-	-0.294***	-0.194*	-
	(0.200)	(0.211)		(0.090)	(0.100)	
GPA_i	0.489***	0.424***	0.427***	-0.168***	-0.070*	-0.071*
	(0.035)	(0.069)	(0.076)	(0.018)	(0.036)	(0.038)
$GPA_i \times \overline{GPA}_{-i}$	-0.109***	-0.114***	-0.080**	0.042*	0.049**	0.050**
	(0.035)	(0.034)	(0.034)	(0.021)	(0.020)	(0.023)
$GPA_i \times SD(GPA_{-i})$	0.162*	0.226**	0.189	-0.008	-0.103*	-0.085
	(0.081)	(0.102)	(0.132)	(0.038)	(0.055)	(0.065)
$GPA_i \times \overline{GPA}_{-i} \times SD(GPA_{-i})$	-0.339**	-0.367**	-0.257	0.134*	0.176**	0.097
	(0.143)	(0.145)	(0.232)	(0.074)	(0.080)	(0.111)
Observations	1,753	1,753	1,753	1,876	1,876	1,876
Controls						
Tutorial group fixed effects	No	No	Yes	No	No	Yes
Randomization controls	Yes	Yes	Yes	Yes	Yes	Yes
Background controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each column reports the results from a different OLS regression. Dependent variable reported at the top of each column. The outcome is standardized to have zero mean and unit standard deviation. Randomization controls are a saturated set of own GPA category, advanced math, and cohort-dummies, interacted with application order. Background controls are: male, being in the youngest third of the age distribution, being in the oldest third of the age distribution, professional college. The peer variables $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$ are re-centred to have zero means. $\overline{GPA}_{-i}$, $SD(GPA_{-i})$ and $\overline{GPA}_{-i} \times SD(GPA_{-i})$ are not included in columns (3) and (6) due to collinearity with the tutorial group fixed effects. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. *** p<0.01, ** p<0.05, * p<0.1.

Table A2. Main results. Instrumenting rank and peer composition with beginning-of-the-year rank and peer composition (including no-shows)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Credits		Average Grade		Dropout	
Estimation method	OLS	IV	OLS	IV	OLS	IV
$RANK_{ig}$	0.508***	0.559***	0.154	0.179	-0.248**	-0.290**
	(0.175)	(0.194)	(0.172)	(0.166)	(0.104)	(0.108)
Observations	1,876	1,876	1,753	1,753	1,876	1,876
First stage F statistic		47.08		52.72		47.08

Notes: Each column reports the results from a different OLS or TSLS regression. Dependent variable stated at the top of each column. Estimates based on the specification used in Table 2, column (3). In the IV regressions, we instrument actual rank with beginning-of-year rank (including no-shows) and the peer characteristics interacted with individual ability with the beginning-of-year peer characteristics (including no-shows) interacted with individual ability. The Kleibergen-Paap weak identification first stage F statistic is also reported. Number of observations: 1,876. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. *** p<0.01, ** p<0.05, * p<0.1.

Table A3. Main results. Robustness tests: accounting for non-linearities in rank

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable		Credits			Average grade			Dropout	
$RANK_{ig}$	0.508***	0.450		0.154	-0.020		-0.248**	-0.259*	
	(0.175)	(0.275)		(0.172)	(0.299)		(0.104)	(0.152)	
$RANK_{ig}^2$		0.069			0.206			0.012	
		(0.267)			(0.267)			(0.135)	
$RANK_{ig}(2nd\ Tertile)$			0.132**			0.013			-0.074**
			(0.060)			(0.062)			(0.032)
$RANK_{ig}(3rd\ Tertile)$			0.277***			0.153*			-0.141***
			(0.079)			(0.082)			(0.044)
Observations	1,876	1,876	1,876	1,753	1,753	1,753	1,876	1,876	1,876
AIC	4775	4777	4784	4256	4257	4253	2250	2252	2249
BIC	5007	5015	5013	4485	4492	4489	2482	2490	2488

Notes: Each column reports the results from a different OLS regression, using a different functional form for $RANK_{ig}$. Dependent variable reported at the top of each column. Estimates based on the specification used in Table 2, column (3). Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. *** p<0.01, ** p<0.05, * p<0.1.

Table A4. Main results. Robustness tests: accounting for non-linearities in GPA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable		Credits			Average grade			Dropout	
$RANK_{ig}$	0.508***	0.485**	0.451**	0.154	0.310	0.174	-0.248**	-0.214*	-0.195*
	(0.175)	(0.229)	(0.185)	(0.172)	(0.210)	(0.157)	(0.104)	(0.123)	(0.111)
Observations	1,876	1,876	1,876	1,753	1,753	1,753	1,876	1,876	1876
GPA_i	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GPA_i^2	No	Yes	No	No	Yes	No	No	Yes	No
$GPA_i(2nd\ tertile)$	No	No	Yes	No	No	Yes	No	No	Yes
$GPA_i(3rd\ tertile)$	No	No	Yes	No	No	Yes	No	No	Yes
AIC	4775	4777	4778	4256	4255	4258	2250	2251	2251
BIC	5007	5015	5022	4485	4491	4499	2482	2490	2494

Notes: Each column reports the results from a different OLS regression, using a different functional form for GPA_i. Dependent variable reported at the top of each column. Estimates based on the specification used in Table 2, column (3). Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. *** p<0.01, ** p<0.05, * p<0.1.

Table A5. Main results. Additional robustness tests

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
	Credits		Average grade		Dropout	
	No 2009 cohort	No outliers in GPA_i	No 2009 cohort	No outliers in GPA_i	No 2009 cohort	No outliers in GPA_i
$RANK_{ig}$	0.594**	0.446**	0.382*	0.092	-0.248**	-0.214*
	(0.232)	(0.190)	(0.216)	(0.172)	(0.104)	(0.123)
Observations	1,270	1,840	1,180	1,721	1,270	1,840

Notes: Each column reports the results from a different OLS regression. Dependent variable reported at the top of each column. Estimates based on the specification used in Table 2, column (3). Estimates in columns (1), (3) and (5) are without the 2009 cohort; estimates in columns (2), (4) and (6) are without outlier students in the top and bottom 1% of the distribution of GPA_i. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. *** p<0.01, ** p<0.05, * p<0.1.

Table A6. Heterogeneity in the effects of rank on average grade and dropout

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	Average grade				Dropout			
$RANK_{ig}$ (a)	-0.032	-0.101	0.226	0.214	-0.145	-0.087	-0.264**	-0.233**
	(0.186)	(0.212)	(0.171)	(0.168)	(0.115)	(0.118)	(0.106)	(0.103)
$RANK_{ig} \times HIGH\ GPA_i$ (b)	0.306*				-0.048			
	(0.176)				(0.117)			
$RANK_{ig} \times MALE_i$ (b)		0.323**				-0.202***		
		(0.140)				(0.065)		
$RANK_{ig} \times HIGH\ GROUP\ SIZE_i$ (b)			-0.200				0.044	
			(0.131)				(0.085)	
$RANK_{ig} \times HIGH\ SD(GPA_{-i})$ (b)				-0.082				-0.058
				(0.159)				(0.106)
$(a) + (b)$	0.274	0.222	0.0257	0.133	-0.193	-0.288***	-0.220*	-0.291**
	(0.204)	(0.169)	(0.195)	(0.201)	(0.151)	(0.105)	(0.120)	(0.129)
Observations	1,753	1,753	1,753	1,753	1,876	1,876	1,876	1,876

Notes: the table reports heterogeneous effects of rank on average grade and dropout by own ability, gender, tutorial group size and tutorial group ability heterogeneity. Each column reports the results from a different OLS regression. Estimates based on the specification used in Table 2, column (3). Each column reports the linear effect of rank, the interaction term between rank and the dummy variable for the category of interest, and the linear combination of the two. The dummy variable for the category of interest is also included among the controls. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. *** p<0.01, ** p<0.05, * p<0.1.

Table A7. Estimated tracking effects on first-year average grade compared to mixing. Total effects and unpacking rank and peer effects.

					Student GPA category								
			ATE			L(B)			M			H(A)	
		Total	Rank	Peer	Total	Rank	Peer	Total	Rank	Peer	Total	Rank	Peer
		(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)
Two-way tracking	{B},{A}	0.073**	-0.001	0.074**	0.108*	0.045	0.063				0.038	-0.046	0.084*
		(0.032)	(0.000)	(0.032)	(0.056)	(0.039)	(0.065)				(0.037)	(0.040)	(0.043)
Three-way tracking	{L}, {M}, {H}	0.117***	-0.001	0.118***	0.175*	0.059	0.116	0.154***	-0.002	0.156***	0.021	-0.061	0.082
		(0.040)	(0.001)	(0.040)	(0.101)	(0.052)	(0.106)	(0.054)	(0.001)	(0.054)	(0.047)	(0.053)	(0.063)
Track low	{L}, {M, H}	0.088**	-0.001	0.089**	0.175*	0.059	0.116	0.095**	-0.046	0.141***	-0.006	-0.016	0.010
		(0.037)	(0.000)	(0.037)	(0.101)	(0.052)	(0.106)	(0.037)	(0.040)	(0.040)	(0.028)	(0.013)	(0.029)
Track middle	{M}, {L, H}	0.031***	-0.000	0.031***	-0.046**	0.015	-0.061**	0.154***	-0.002	0.156***	-0.016	-0.015	-0.001
		(0.011)	(0.000)	(0.011)	(0.023)	(0.013)	(0.024)	(0.054)	(0.001)	(0.054)	(0.022)	(0.013)	(0.028)
Track high	{L, M}, {H}	0.053**	-0.000	0.053**	0.097**	0.014	0.083*	0.040	0.045	-0.005	0.021	-0.061	0.082
		(0.023)	(0.000)	(0.023)	(0.045)	(0.013)	(0.046)	(0.032)	(0.039)	(0.055)	(0.047)	(0.053)	(0.063)

Notes: The table reports (conditional) average treatment effects on average grade of different tracking configurations relative to mixing based on the estimates from Appendix Table A1, column (2). Total effects obtained by changing both peer characteristics and rank as we move from ability mixing to tracking. Rank (Peer) effects are obtained by holding peer characteristics (rank) fixed and moving rank (peer characteristics) as we move from ability mixing to tracking. Student GPA groups are L(ow), M(iddle), H(igh) in case of three-way tracking, and for two-way tracking B(elow) and A(bove). The curly brackets indicate the grouping of GPA groups. Number of observations: 1,753. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. *** p<0.01, ** p<0.05, * p<0.1.

Table A8. Estimated tracking effects on dropout probability compared to mixing. Total effects and unpacking rank and peer effects.

					Student GPA category								
			ATE			L(B)			M			H(A)	
		Total	Rank	Peer	Total	Rank	Peer	Total	Rank	Peer	Total	Rank	Peer
		(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)
Two-way tracking	{B},{A}	-0.061 ***	0.001 ***	-0.062 ***	-0.109 ***	-0.065 ***	-0.044				-0.013	0.067 ***	-0.080 ***
		(0.022)	(0.000)	(0.022)	(0.035)	(0.025)	(0.040)				(0.022)	(0.026)	(0.030)
Three-way tracking	{L}, {M}, {H}	-0.082 ***	0.001 ***	-0.083 ***	-0.184 ***	-0.087 ***	-0.097	-0.062 *	0.002 ***	-0.064 *	0.002	0.090 ***	-0.088 **
		(0.029)	(0.001)	(0.029)	(0.058)	(0.033)	(0.061)	(0.035)	(0.001)	(0.035)	(0.027)	(0.034)	(0.043)
Track low	{L}, {M, H}	-0.075 ***	0.001 ***	-0.076 ***	-0.184 ***	-0.087 ***	-0.097	-0.027	0.067 ***	-0.094 ***	-0.014	0.022 ***	-0.036 **
		(0.024)	(0.000)	(0.024)	(0.058)	(0.033)	(0.061)	(0.021)	(0.026)	(0.028)	(0.016)	(0.009)	(0.018)
Track middle	{M}, {L, H}	-0.022 ***	0.000 ***	-0.022 ***	-0.010	-0.022 ***	0.012	-0.062 *	0.002 ***	-0.064 *	0.007	0.022 ***	-0.015
		(0.007)	(0.000)	(0.007)	(0.010)	(0.008)	(0.011)	(0.035)	(0.001)	(0.035)	(0.012)	(0.009)	(0.016)
Track high	{L, M}, {H}	-0.038 **	0.000 ***	-0.038 **	-0.057 **	-0.022 ***	-0.035	-0.058 ***	-0.066 ***	0.008	0.002	0.090 ***	-0.088 **
		(0.017)	(0.000)	(0.017)	(0.026)	(0.008)	(0.027)	(0.022)	(0.025)	(0.035)	(0.027)	(0.034)	(0.043)

Notes: The table reports (conditional) average treatment effects on dropout of different tracking configurations relative to mixing based on the estimates from Appendix Table A1, column (5). Total effects obtained by changing both peer characteristics and rank as we move from ability mixing to tracking. Rank (Peer) effects are obtained by holding peer characteristics (rank) fixed and moving rank (peer characteristics) as we move from ability mixing to tracking. Student GPA groups are L(ow), M(iddle), H(igh) in case of three-way tracking, and for two-way tracking B(elow) and A(bove). The curly brackets indicate the grouping of GPA groups. Number of observations: 1,876. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. *** p<0.01, ** p<0.05, * p<0.1.

Table A9. Mechanisms. Peer and rank effects on survey data on teaching style and learning environment.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Too slow		Too fast		Stimulating		Conducive		Interactive		Involved	
$RANK_{ig}$	0.207	0.221	-0.286	-0.328	-0.003	-0.027	0.207	0.221	-0.286	-0.328	-0.003	-0.027
	(0.262)	(0.270)	(0.223)	(0.231)	(0.273)	(0.297)	(0.262)	(0.270)	(0.223)	(0.231)	(0.273)	(0.297)
$\overline{GPA}_{-i}$	0.061	-	-0.137	-	-0.026	-	0.061	-	-0.137	-	-0.026	-
	(0.097)		(0.119)		(0.115)		(0.097)		(0.119)		(0.115)	
$SD(GPA_{-i})$	0.023	-	-0.035	-	-0.268	-	0.023	-	-0.035	-	-0.268	-
	(0.141)		(0.125)		(0.206)		(0.141)		(0.125)		(0.206)	
$\overline{GPA}_{-i} \times SD(GPA_{-i})$	-0.143	-	0.327	-	-0.359	-	-0.143	-	0.327	-	-0.359	-
	(0.312)		(0.333)		(0.344)		(0.312)		(0.333)		(0.344)	
GPA_i	0.030	0.009	-0.018	0.012	-0.004	0.012	0.030	0.119	-0.018	-0.019	-0.004	0.080
	(0.131)	(0.138)	(0.082)	(0.088)	(0.105)	(0.120)	(0.131)	(0.094)	(0.082)	(0.078)	(0.105)	(0.080)
$GPA_i \times \overline{GPA}_{-i}$	0.006	0.048	0.024	0.047	-0.097	-0.174**	0.006	0.071	0.024	0.074	-0.097	-0.209***
	(0.081)	(0.095)	(0.054)	(0.061)	(0.076)	(0.076)	(0.081)	(0.054)	(0.054)	(0.087)	(0.076)	(0.054)
$GPA_i \times SD(GPA_{-i})$	0.234	0.217	-0.055	-0.150	0.058	0.098	0.234	0.018	-0.055	0.150	0.058	0.102
	(0.167)	(0.206)	(0.158)	(0.161)	(0.173)	(0.203)	(0.167)	(0.147)	(0.158)	(0.142)	(0.173)	(0.118)
$GPA_i \times \overline{GPA}_{-i} \times SD(GPA_{-i})$	-0.376	-0.271	0.090	0.277	-0.280	-0.339	-0.376	-0.277	0.090	-0.412	-0.280	-0.591**
	(0.323)	(0.437)	(0.257)	(0.269)	(0.355)	(0.424)	(0.323)	(0.244)	(0.257)	(0.332)	(0.355)	(0.221)
Controls												
Tutorial group fixed effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Randomization controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Background controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each column reports the results from a different OLS regression. The dependent variables are the indexes constructed from the survey data, stated at the top of each column. Estimates in odd columns are based on the specification used in Table 2, column (2), while estimates in even columns follow the specification in Table 2, column (3). Number of observations: 1,342. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 47. *** p<0.01, ** p<0.05, * p<0.1.