I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12789

Ordinal Rank and Peer Composition: Two Sides of the Same Coin?

Marco Bertoni
Roberto Nisticò

NOVEMBER 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12789

Ordinal Rank and Peer Composition: Two Sides of the Same Coin?

Marco Bertoni
University of Padua, HEDG and IZA

Roberto Nisticò
University of Naples "Federico II", CSEF and IZA

NOVEMBER 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Ordinal Rank and Peer Composition: Two Sides of the Same Coin?*

We use data from two experiments that randomly assign students to groups to show that, so long as ordinal rank has a causal effect on educational achievement, estimates of the effects of peer ability composition obtained from models that omit rank are downward biased. This finding holds both in the standard linear-in-means model as well as in models that allow for non-linear and heterogeneous peer effects, and contributes to explain why previous studies have detected only modest effects of peer ability on achievement. We also illustrate how this finding helps understand the mechanisms behind the effects of ability tracking policies.

JEL Classification: I21, I24, J24

Keywords: rank effects, peer effects, omitted variables bias, ability tracking

Corresponding author:
Roberto Nisticò
University of Naples Federico II
Department of Economics and Statistics
Sant'Angelo, 80126 Naples
Italy
E-mail: roberto.nistico@unina.it

* This is a revised version of a paper titled "Rank Concerns, Peer Composition, and Ability Tracking in University. Evidence from a Randomized Experiment". We thank the audience at seminars at the University of Milan and Padova, and Massimo Anelli, Giorgio Brunello, Scott Carrell, Marta De Philippis, Ingo Isphording, Monica Langella, Sandra McNally, Konrad Mierendorff, Richard Murphy, Cheti Nicoletti, Matthias Parey, Lorenzo Rocco, Uta Schönberg, Annalisa Scognamiglio, Vincenzo Scoppa, Olmo Silva, Michela Tincani and Felix Weinhardt for helpful comments. Adam Booij, Edwin Leuven and Hessel Oosterbeek deserve special thanks for their many helpful comments on the previous version of this paper and for allowing us to use their original code and data, made available through the Review of Economic Studies data archive. We are also grateful to Esther Duflo, Pascaline Dupas and Michael Kremer for letting us use their original code and data, made available through the American Economic Review data archive. Marco Bertoni acknowledges funding from a CARIPARO Foundation Starting Grant. The usual disclaimer applies.

1. Introduction

A long-standing debate in economics 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 Carrell et al., 2013), there is by now a consensus on their modest size and their heterogeneous and non-linear nature (see Lavy et al., 2012, Sacerdote, 2014; and Feld and Zoelitz, 2017), that makes it hard to evaluate their implications from a general perspective. On the other hand, within homogeneous groups student interactions and involvement should be easier (see Booij et al., 2017) and teaching more effective (see Duflo et al., 2011), favouring ability tracking.[1]

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. In fact, different group assignments will change at the same time not only the whole ability distribution across groups, but also students' ordinal ability rank within groups.

Let us illustrate this issue in Figure 1 with a simple example. We assume ability is uniformly distributed between 0 and 1, and report the hypothetical ability distribution within groups when students are assigned to groups using ability mixing (i.e. randomly drawing from the whole ability distribution) or two-way tracking (i.e. separating students with ability below and above median). The two vertical lines respectively indicate the position of students at the 25th and 75th percentile of the overall ability distribution in the population – low and high achievers.

[1] Tracking systems may embed also a risk of misallocation (see Dustmann et al., 2017). If the assignment happens early on, information on students' ability may be noisy and incomplete.

As noted by many in the literature, the distribution of peer ability composition faced by high and low achievers under mixing and tracking is very different. In the former case, high and low achievers will both face the same mean peer ability.[2] In the latter, mean peer ability will be high for the high achievers and low for the low achievers. Yet, what has been overlooked so far is that different assignment rules also affect the ordinal ability rank within groups of low and high achievers. Under mixing, low achievers have low rank and high achievers have high rank within groups. Under tracking, instead, they are both in the median ordinal rank position within their group.

Eventually, ordinal rank and peer composition are two sides of the same coin: for given individual ability, students have lower rank when surrounded by abler peers, and vice-versa. In turn, a recent literature has shown that ordinal ability rank has a positive causal effect on achievement (see Murphy and Weinhardt, 2019, Elsner and Isphording, 2017 and 2018, Elsner et al., 2018, Cicala et al., 2017, Denning et al., 2018).[3] If this is the case, then the omission of rank from the education production function will cause an attenuation bias on the estimated effect of peers' mean ability – the key parameter in the linear-in-means peer effect model. This result may contribute to explain why, despite much interest and investigation on the topic, previous studies have detected only modest effects of peer ability on educational achievement (see Sacerdote, 2014, for a review).

This paper makes two contributions to the literature on peer and rank effects. First, we focus on the linear-in-means model of peer effects and show that the omission of rank from the

[2] Under tracking there is also less variability in ability.

[3] Using survey data, previous studies have shown that the main channel behind the estimated rank effects on achievement is that rank improves non-cognitive skills such as self-confidence (Murphy and Weinhardt, 2019) and conscientiousness (Pagani et al., 2019). A parallel literature (see e.g. Tincani, 2017, and Tincani and Mierendorff, 2018) discusses how models where rank in achievement enters students' utility function (a rank "concern") are able to rationalise peer effects on achievement. While our paper investigates the effect of pre-determined ability rank, an input in the education production function, we are silent on the role played by students' preferences for rank in achievement, an outcome of education production.

education production function generates an attenuation bias in the effect of peers' mean ability. We do so using data from two randomized experiments carried out in a large set of primary schools in Kenya (see Dulfo et al., 2011) and at the University of Amsterdam (see Booij et al., 2017), where students were randomly assigned to classes.[4]

In both setups, we first show that the higher is peers' mean ability, the lower is one's ordinal rank conditional on own ability. We then use the standard linear-in-means model for peer effects and show that rank has a positive causal effect on achievement. Therefore, its omission generates an attenuation bias on the effect of mean peer ability.

We assess the significance of the omitted variables bias using the generalized Hausman test developed by Pei et al., 2019, and reject the null whenever we detect a statistically significant rank effect. The magnitude of the bias is not negligible either. Depending on the setup and the outcome variable, it can even be larger than 100% of the effect of mean peer ability in the mis-specified model. In both setups this is enough to render statistically significant the otherwise small and insignificant effect of mean peer ability.

Our analysis is not limited to the linear-in-means model. In fact, we show that the omission of rank is relevant also when using more flexible peer effect models that allow other features of the distribution of peer ability such as the standard deviation (Lyle, 2009) to affect outcomes, as well as for heterogeneities of peer composition effects by own ability.

The two experimental setups we use complement each other not only in terms of the level of education (primary vs. tertiary) and the country (developing vs. developed) considered, but also in terms of the experimental variation in group ability composition that they generate. On the one hand, in the Kenyan experiment primary school students are assigned to classes

[4] Cicala et al., 2017, also estimate rank effects in the Kenyan experiment, but do not test for omitted variables bias in the standard linear-in-means peer effects model.

using ability mixing, and we observe very local variation in rank for given ability.[5] This is ideal to alleviate the concern that uncontrolled differences in peer composition between groups could confound the identification of the effect of rank. However, it also provides limited variability to carry out counterfactual simulations of the effects of different grouping policies without running in the risk of extrapolating outside of the observed support of the data. On the other hand, the Dutch experiment randomly assigned students to tutorial groups using an algorithm that artificially generated an unprecedentedly wide support of group ability configurations. While this setup requires careful modelling of peer composition to estimate rank effects, it is also very useful to credibly carry out policy simulations.

In fact, our second contribution is centred on the Dutch experiment, where we observe a much wider support of peer ability configurations. We use a flexible education production function to predict 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 directions to generate outcomes for low- and high-ability students.

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. For instance, as we move from a system based on ability mixing to two-way tracking, students at the bottom of the ability distribution will lose out in terms of average ability of peers, but at the same time they will gain in terms of within-group ordinal ability rank. Therefore, if ordinal rank within classes enhances student achievement, tracking will help low-ability students, but it may harm the high achievers. We also discuss how rank and peer

[5] Duflo et al., 2011, assigned students to classes using either ability mixing or two-way tracking. As in their original work, we estimate peer composition effects on the former group of classes, where there is local random variation in peer quality.

composition effects contribute to explain the overall zero effect for low achievers of the "track middle" assignment policy proposed by Carrell et al., 2013.

Finally, our results from survey data on student perceptions in the Dutch experiment show little evidence of both rank effects and teacher responses to group ability composition. If anything, low-ability students feel less involved when surrounded by high-ability peers, contributing to explain why they gain from ability tracking.

The paper unfolds as follows. Section 2 presents the two experimental setups we use, and Section 3 describes the data and the experimental variation. Section 4 illustrates the empirical methodology. Section 5 presents our main results on omitted variables bias in the linear-in-means model, and Section 6 extends them to models that allow for heterogeneous and non-linear peer effects. We describe out results on ability tracking in Section 7, while Section 8 is devoted to robustness tests and extensions. Conclusions follow thereafter.

2. Experimental setups

Our analysis exploits data from two randomized experiments carried out by Duflo et al., 2011, in a set of primary schools in Kenya and by Booij et al., 2017, at the University of Amsterdam. We here give only a brief description of the key features of each experiment, and refer the interested reader to the original papers for more details.

2.1. The Kenyan experiment - Duflo et al., 2011

The Extra Teacher Provision (ETP) experiment took place in Kenia in May 2005 and involved 121 primary schools. These were schools with one first-grade section only, and the intervention provided them with extra resources to hire one additional teacher and create a second section. In total, 61 of these schools were randomly selected as "tracking schools": students with average baseline exam score above the median were grouped in one section, while those below the median were assigned to the other section (i.e. they applied the two-way tracking system described above). Instead, students in the remaining 60 “non-tracking schools” were unconditionally randomized into the two sections (i.e. ability mixing).

Once students have been assigned to sections, within each school each section was also randomly assigned to either a contract teacher or a civil-service teacher. Whereas parents could ask for a reassignment of their children in a different section (a very rare event), the main reason why there could have been no compliance with the initial assignment is the absence of the teacher, which sometimes resulted in the two sections being combined. Following Duflo et al., 2011, we construct all rank and peer composition variables on the basis of the initial (random) assignment to classes. The experiment lasted for 18 months and at the end of the program a standardized math and literacy test was run in all schools. The test was administered and then graded blindly by trained examiners. In each school, the test was administered to a maximum of 60 randomly selected students only (30 per section).

2.2. The Dutch Experiment - Booij et al., 2017

The experiment involved about 2,000 students starting the 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. Ability is measured in terms of the Grade Point Average (GPA) 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.

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 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.

3. Data and experimental variation

Table 1 reports descriptive statistics for the main variables in the two studies we consider in our analysis. In the Kenyan experiment, the available background controls are assignment to a contract vs. civil-service teacher, age and gender. We also observe baseline GPA_i, that is standardized to have zero mean and unit standard deviation in the full initial sample including non-tracking schools and students without endline test scores. The baseline score is not comparable across schools, as it is obtained from grades assigned by teachers in each school. For this reason, and to control for the stratified assignment of students to classes and of classes to contract teachers within schools, we include school fixed effects in all our regressions. Our main outcome variable for this experiment is the score at the math test administered at the end of the ETP program. As additional outcomes, we also consider the literacy score and the total score (sum of math and literacy).[6] All scores are standardized to have zero mean and unit standard deviation in the full sample.

In the Dutch experiment, the data contains information on the exact individual end-of-secondary school GPA_i - standardized to have zero mean and unit standard deviation within

[6] We prefer to focus on the math score for external validity reasons, as literacy results may be more specific to the context and language of the country.

cohort - on gender, age (categorized in tertiles within cohort), previous attendance of a professional college before university enrolment, as well as on the randomization controls (see above - descriptive statistics not reported). Our main outcome variable in this setup 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. We also show results on two additional measures of achievement. On the one hand, we consider the average grade at the exams completed during the first year. 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. We standardize both the number of credits and average grade to have zero mean and unit standard deviation within cohort. On the other hand, our third outcome is a "dropout" dummy, that is a core performance measure adopted by the University of Amsterdam and is equal to one if a student failed to complete at least 45 out of 60 credits during the first year, and zero otherwise. As shown in Table 1, only slightly more than half of the students in our sample pass the threshold for admission to the second year. Additionally, the Dutch data also provide us with survey data on students' perceptions about teaching and the learning environment, that we use to provide some additional evidence on teacher responses to group ability composition and on students' interactions within groups.

In both experiments, we describe the ability composition of a student's peer group with the mean – $\overline{GPA}_{-i}$ – and the standard deviation – $SD(GPA_{-i})$ – of their end-of-secondary-school GPA. These are constructed leaving out individual i. We measure students' rank as their percentile rank in the baseline GPA distribution within groups. Since groups have different size, as done by Murphy and Weinhardt, 2019, we normalize the raw ordinal rank by group size as follows:

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

where n_{ig} is the ordinal rank of individual i assigned to group g, and N_g is group size.[7] Table 1 shows that the average individual rank in our sample is 0.500 in the Kenyan experiment, and 0.486 in the Dutch one.[8]

Figure 2 describes the variation in peer ability configurations available in the two studies, and clearly highlights how the conditional assignment in the Dutch experiment considerably broadened the support relative to the unconditional assignment carried out in the Kenyan experiment. Figure 3 portrays the relationship between $RANK_{ig}$ and GPA_i in the two setups we analyse. Panel a reports the raw data and Panel b reports box-whisker plots of the distribution of $RANK_{ig}$ by decile of GPA_i (see Elsner and Isphording, 2017).

Taken together, these graphs illustrate well the reason why we believe that the two setups are complements in terms of the variation in peer composition they generate. On the one hand, the unconditional randomization carried out in the Kenyan study provides *very local* variation in rank for given ability. This setup is therefore very close to the ideal experiment one would want to carry out to identify rank effects, where the peer group almost does not change, while

[7] In case of ties in ability within groups, we follow Murphy and Weinhardt, 2019, and assign the lower rank to all students. Results are robust when we use the average or the highest rank. This definition is slightly different from the one used by Cicala et al., 2017, in their analysis of the Kenyan experiment, as they do not subtract 1 from the numerator and denominator of this formula and use average rank to break ties. Results are unaffected by these choices. In addition, for the Dutch case we follow the authors' original calculations for peer composition and – unlike in the previous version of this paper, see Bertoni and Nisticò, 2018 – we compute $RANK_{ig}$ before dropping 10 students that are part of the original peer group but moved to the fiscal economics track in the second year, and are dropped from the experiment. Results are again unaffected by this choice.

[8] A potential issue about both setups concerns students' information about their ability rank. Pagani et al., 2019, and Yu, 2019, show evidence of a very strong and positive correlation between objective and self-perceived ability rank in very different setups (Italian primary schools and Chinese middle schools). We do not have data to show this correlation in our setups. In this regard, we view our estimates as the reduced form or ITT effects that would be obtained when instrumenting perceived with objective rank.

students marginally change their individual rank for given individual ability.[9] However, the support of ability configurations that can be used to estimate the effects of counterfactual group assignment policies in the Kenyan experiment is undoubtedly limited.

On the other hand, the conditional randomization in the Dutch experiment generates 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 her group even if she ranks only at the 35th percentile of the overall GPA_i distribution. We will exploit this to carry out counterfactual policy simulations without facing the risk of extrapolating outside the support of ability configurations that is observed in the data. At the same time, this highlights a potential limitation of the Dutch experiment to identify rank effects.[10] In fact, the large variation in rank for given ability stems from differences in the ability configurations of peers between groups – quite the opposite of the ideal experiment described above.

Therefore, there might be a concern that uncontrolled variation in peer group composition may confound the identification of the rank effect in this setup. We address this issue in two standard ways in the literature on rank effects: we directly and flexibly control for several features of the peer ability distribution (see Table 4 and Table A3 in the Appendix) or include group fixed effects (see Table 6 and Table 7). All in all, we believe that the consistent evidence of positive rank effects across the two studies helps us to alleviate this concern.

[9] A general concern about identification of rank effect could be related to potential violations of the stable unit treatment value assumption (Rubin, 1980). As the same rank could be achieved under different underlying distributions of peer ability, one may argue that the treatment (a change in rank from *r* to *r'*) is not well-defined. We abstract from this issue in our analysis. However, we notice that similar concerns could be raised in other common problems in labour economics. For example, a change in average peer ability from the median level to the first quartile could be a badly defined treatment if the within-group dispersion of peer ability differs among groups with the same average ability. Similarly, being first vs. second-born could hide different treatments depending on family size.

[10] We thank Booij, Leuven and Oosterbeek for point this out to us.

4. Empirical methodology

4.1. Identification

The key contribution of this paper is to show that rank and other features of the distribution of peer ability within groups, such as its mean, are jointly determined at the time of assignment of students to groups, and to highlight the implications of this problem for identification of ability peer effects.[11]

Roughly speaking, the literature on peer effects has so far assumed that – by virtue of the random (or quasi-random, depending on the setup) assignment of students to groups – a causal effect of peer composition can be simply obtained by comparing average outcomes of students in differently composed peer groups.[12] However, this does not hold true anymore in presence of a causal effect of rank on achievement and of interdependence between rank and peer composition.[13]

To illustrate, we frame the identification problem as one of omitted variable bias. To ease the exposition, let us focus on a very simple linear-in-means model relating individual outcomes to $\overline{GPA}_{-i}$, and assume that all other relevant controls – including GPA_i – have been partialled out. To begin, let us describe the relationship between $RANK_{ig}$ and $\overline{GPA}_{-i}$ as:

$$RANK_{ig} = \delta\overline{GPA}_{-i} + \xi_{ig} \quad (1)$$

The standard "short" linear-in-means peer effect model estimated in the literature is the following:

[11] Manski, 1993, calls this a problem of "contextual" or "exogenous" peer effects.
[12] We are simplifying matters here. For instance, the empirical analysis of both Booij et al., 2017, and Duflo et al., 2011, allow for heterogeneous peer ability composition effects by own ability.
[13] Similarly, failure to properly control for the distribution of peer ability would lead to biased estimates of rank effects in presence of large variation in peer composition across groups. Previous studies on rank effects have address this identification issue either by directly controlling for peer composition or by including group fixed effects (see for instance Murphy and Weinhardt, 2019).

$$y_{ig} = \gamma_s \overline{GPA}_{-i} + v_{ig}, \text{ with } v_{ig} = \beta\, RANK_{ig} + \varepsilon_{ig} \qquad (2)$$

Our "long" enriched education production function, including both $\overline{GPA}_{-i}$ and $RANK_{ig}$, is specified as follows:

$$y_{ig} = \gamma_L \overline{GPA}_{-i} + \beta\, RANK_{ig} + \varepsilon_{ig} \qquad (3)$$

Following Angrist and Pischke, 2009, and Pei et al., 2019, it is easy to show that the bias on γ_s in (2) due to the omission of $RANK_{ig}$ is equal to $\beta\delta$. The evidence from both our experiments suggests that β is positive and δ is negative. Therefore, the coefficient γ_s estimated in the standard linear-in-means model for peer effects suffers from attenuation bias. As we will show, similar considerations also hold in more complex models that allow for non-linear and heterogeneous peer effects, as well as for the effect of individual GPA_i.

4.2. Estimation

In our main analysis we discuss the empirical relevance of the omitted variables bias described above for the standard linear-in-means peer effects model. Therefore, we take Eq. (3) to the data by estimating with Ordinary Least Squares (OLS) the following education production function:

$$y_{ig} = \alpha + \beta\, RANK_{ig} + \gamma_1 \overline{GPA}_{-ig} + \gamma_2 GPA_i + X_i'\phi + \varepsilon_{ig} \qquad (4)$$

where y_{ig} is the outcome of student *i* in group *g*; $RANK_{ig}$ is student *i*'s percentile rank within the assigned group g, $\overline{GPA}_{-i}$ measures the mean of peers' ability (excluding individual *i*), and GPA_i is individual ability. The vector of covariates X_i includes school fixed effects (randomization controls) for the Kenyan (Dutch) experiment, and individual background controls. In addition, ε_{ig} is an error term. We allow for correlation among the outcomes of students that share a common learning environment by clustering standard errors, respectively, at the school level in the Kenyan experiment and at the tutorial group level in the Dutch one. The empirical specifications of Eqs. (1) and (2) are similar, but in the former case $RANK_{ig}$ is the dependent variable, and in the latter it is instead omitted from the model.

We test for the significance of the bias in the coefficient of $\overline{GPA}_{-i}$ due to the omission of $RANK_{ig}$ using the generalized Hausman test developed by Pei et al., 2019. This simply amounts to jointly estimating the linear-in-means peer effect models with and without $RANK_{ig}$ – Eqs. (2) and (3) above – using seemingly unrelated estimation, and testing the following null hypothesis: $H0: \gamma_s - \gamma_L = 0$.

In an extension to our main result for the linear-in-means model, we follow the recent literature on peer effects (see Sacerdote, 2014, for a review) and discuss omitted variables bias in models that allow for heterogeneities and non-linearities in the effects of peer composition. As a matter of fact, a more flexible peer effects specification turns out to be supported by the data in the Dutch experiment but not in the Kenyan one. We refer to Section 6 for further details.

Finally, the ample support of group configurations in the Dutch data will be helpful for the last part of our analysis, where we use the flexibility granted by our education production function to estimate student outcomes under different assignments, and unpack the contribution of rank and peer composition effects to generate tracking effects. We refer to Section 7 for details on estimation.

5. Main results: the linear-in-means model

Our main results are presented in Table 2 and Table 3 for the Kenyan and the Dutch experiment, respectively. The specification adopted is the one described in Eq. (4) above. Column (1) in each Table reports the estimates of Eq. (1), describing the relationship between ordinal rank and peers' mean ability, conditional on own ability. Column (2a) reports the estimates of the standard linear-in-means peer effects model described in Eq. (2), while

Column (2b) reports the estimates of Eq. (3), which enriches the specification by including also $RANK_{ig}$ among the inputs of the education production function.

First, results in Column (1) in both Table 2 and Table 3 confirm that, for given individual ability, the higher is peers' mean ability, the lower is ordinal rank. Therefore, parameter δ in Eq. (1) is negative. It is also interesting to notice that the R-squared for this model is higher in the Kenyan case than in the Dutch one. In the latter setup, the large variation in ability composition across groups makes it such that there is still large variability in rank for given individual ability and peers' mean ability, that is due to idiosyncratic variability in the ability distributions across small groups.

Second, in both Tables results from Column (2b) show that rank has a positive effect on achievement, highlighting that parameter β in Eq. (3) is positive. In the Kenyan experiment the rank effect is very large in magnitude. Moving from the bottom to the top of the within-group ability distribution increases the score in the math test by slightly more than 80 percent of a standard deviation. The rank effect is positive and of large magnitude (0.3 SD) also in the Dutch experiment.

At this stage, the key result of the paper follows very intuitively. The comparison of the coefficients of $\overline{GPA}_{-i}$ in Columns (2a) and (2b) confirms that the omission of $RANK_{ig}$ generates a severe attenuation of the effect of peer composition.[14] The p-values of the Hausman test for coefficient comparison across Columns (2a) and (2b), reported in brackets, confirm the statistical relevance of the bias, that is also of substantial magnitude. For the Kenyan experiment, the effect of $\overline{GPA}_{-i}$ almost doubles, while for the Dutch one it becomes almost three times as large. In both cases, the inclusion of $RANK_{ig}$ is sufficient to render the

[14] Besides rounding issues, the bias is exactly equal to the product between the effect of $RANK_{ig}$ on the outcome – reported in Column (2b) – and the effect of $\overline{GPA}_{-i}$ on $RANK_{ig}$ – reported in Column (1).

effect of $\overline{GPA}_{-i}$ statistically significant at the 5% level in the Kenyan experiment and at the 1% in the Dutch one.[15]

Despite much interest on the topic, the existing literature has failed to empirically detect large ability peer effects, especially using linear-in-means models (see Sacerdote, 2014). By highlighting this attenuation bias, our results provide one simple explanation behind this seemingly puzzling result.

In addition, the results in Tables 2 and 3 also show that the omitted variable bias is problematic not only for the estimation of the effect of $\overline{GPA}_{-i}$, but also for other variables included in the model. For instance, given the positive correlation between GPA_i and $RANK_{ig}$ reported in Column (1) of Tables 2 and 3, the coefficient on GPA_i in Eq. (2), reported in Column (2a), is overestimated.

Finally, Appendix Tables A1 and A2 report results for the additional outcomes available in the Kenyan and the Dutch experiment, respectively. In the Kenyan experiment we detect no evidence of peer or rank effects for literacy. Nonetheless, rank and peer effects on the total score are still positive and significant, and there is significant evidence of omitted variables bias when rank is omitted. For the Dutch experiment we find results for the dropout probability that are comparable to the ones for credits, and no effect on the average grade.

6. Extensions: non-linear and heterogeneous peer effects models

The standard linear-in-means peer effect model discussed so far is very simple and intuitive, but it could also be restrictive. For instance, it assumes that the only feature of the peer ability

[15] The effect of $\overline{GPA}_{-i}$ in Column (2a) of Table 2 is significant at the 5 percent level with an estimated standard error of 0.160 if we cluster standard errors by class instead of by school (see Duflo et al., 2011). The significance of the Hausman test is unchanged.

distribution that affects individual outcomes is $\overline{GPA}_{-i}$, and that the effect of $\overline{GPA}_{-i}$ is homogeneous across the distribution of individual ability.

In what follows we extend our results on the bias in the identification of peer effects due to the omission of rank in more general models that allow for non-linear and heterogeneous peer effects. As done by Booij et al., 2017, in their original analysis of the Dutch experiment, we will consider more flexible peer effects models, and progressively include in the model:

(i) $SD(GPA_{-i})$ – the standard deviation of peer ability – to allow the dispersion of peer ability to affect outcomes (see Lyle, 2009) and to consider the possibility of rank concerns in the utility function (Tincani, 2015). In addition, so long as $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$ are correlated, the inclusion of the latter also serves to avoid omitted variables bias in the estimation of the effect of the former;

(ii) the interaction term $\overline{GPA}_{-i} \times SD(GPA_{-i})$, to allow for the possibility that the mean and the SD of peer GPA are not perfect substitutes in shaping student performance;

(iii) interaction terms between the peer composition variables and GPA_i, to accommodate the abundant evidence from the empirical literature that ability peer effects are heterogeneous by individual ability.

On the one hand, using these rich specifications we fail to find evidence of heterogeneous and non-linear peer effects in the Kenyan experiment, and confirm that the linear-in-means model is a good approximation of the data generating process in that setup. This evidence is likely due to the very local variation in ability composition across groups. We report these estimates in the Appendix, Table A3.

On the other hand, results for the Dutch experiment are reported in Table 4 following the same structure of Table 3, and deliver several insights. First, our main result on the omitted variables bias in the effect of $\overline{GPA}_{-i}$ when rank is excluded from the model holds across all specifications. Of course, in models with interactions of $\overline{GPA}_{-i}$ with $SD(GPA_{-i})$ and GPA_i the coefficient on $\overline{GPA}_{-i}$ only identifies the effect for subjects with a value of the interacting variables that is equal to zero. Figure A1 in the Appendix reports heterogeneous effects of $\overline{GPA}_{-i}$ for different levels of $SD(GPA_{-i})$ and GPA_i using the specifications in Columns (6a) and (6b), that respectively omit and include rank. It shows that the omitted variables bias is especially salient in homogeneous groups - where the effects of $\overline{GPA}_{-i}$ turns to be significant once $RANK_{ig}$ is included in the model.

Second, results in Columns (6a) and (6b) also highlight that the inclusion of $RANK_{ig}$ steepens the gradient in the effect of $SD(GPA_{-i})$ by GPA_i. As shown in Figure A2 in the Appendix - that reports heterogeneous effects of $SD(GPA_{-i})$ for different levels of $\overline{GPA}_{-i}$ and GPA_i - this effect is especially salient in low ability groups.

It also worth noticing that the effect of $RANK_{ig}$ is very stable when additional peer composition variables are flexibly included in the model, alleviating the concern discussed above about the identification of the effect of $RANK_{ig}$ in presence of large variation in peer composition across groups. If anything, this effect almost doubles in magnitude when we include the interaction terms between the peer composition variables and GPA_i, in Column (6b). This happens because – as shown in Columns (5), (6a) and (6b) – the relationship between $SD(GPA_{-i})$ and both $RANK_{ig}$ and achievement is not linear in GPA_i, and failure to account for this heterogeneity results in mis-specification bias.[16] For this reason, we will use the model in Column (6b) to carry out the counterfactual simulations in the next Section.

[16] In fact, the inclusion of this interaction in the simple model in Column (2b) is sufficient to generate this result. To get the intuition behind the non-linear relation between $SD(GPA_{-i})$ and $RANK_{ig}$, consider a normal distribution for GPA_i, and apply a mean-preserving spread. Those with GPA_i below median will gain ranks, and those above median will lose. The inclusion of the other interaction terms only increase precision in the estimation of the effect of $RANK_{ig}$, as they allow for a better description of the underlying data generating

Finally, it is also insightful to see that - as witnessed by the smaller-than-one R-squared reported in odd-numbered columns - there is still a large share of variation in $RANK_{ig}$ that owes to idiosyncratic variability in the ability distributions across small groups. This variation is not mechanically explained by the peer composition variables included in the model, and allows us to separately identify rank and peer composition effects.

7. Implications for Ability Tracking

The ample support of group ability configurations in the Dutch experiments permits us to assess the relative contribution of rank and peer composition effects in explaining the educational effects at the student population level of different group assignment policies. We will compare ability mixing with five alternative grouping configurations:

(i) *Two-way tracking*: high-ability (GPA above median) and low-ability (GPA below median) students are grouped separately.

(ii) *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.

(iii) *Track low*: bottom-ability students are grouped together, while middle- and top-ability students are mixed.

(iv) *Track middle*: middle-ability students are grouped together, while bottom- and top-ability students are mixed.

(v) *Track high*: top-ability students are grouped together, while bottom- and middle-ability students are mixed.

---

process, but do not change the point estimate much. On the one hand, the non-linear relationship between $SD(GPA_{-i})$ and $RANK_{ig}$ is mechanical, and in fact it shows up also in the Kenyan experiment - see Column (5) of Table A3 in the Appendix. On the other hand, its consequences for omitted variable bias are only relevant if the effect of $SD(GPA_{-i})$ on achievement is nonlinear, a finding that holds only in the Dutch experiment.

Relative to the original work of 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 ability composition 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. 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 4, Column (6b).[17] We do so for the whole population and by tertile of GPA_i.

Table 5 shows the estimated tracking effects on credits. The estimated tracking effects on the other two outcomes, dropout and average grade, are qualitatively similar to those in Table 5 and are available from the authors.

Results in Columns (1a)-(1c) are for the whole population, while results in the following columns split students by tertile of GPA_i (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:

[17] 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 4, column (6b).

(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 second 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. Since an increase in rank for one individual is offset by a decrease in rank for another, it is not surprising that the rank effect in the whole population is always close to zero. However, the estimates in Columns (1b), (2b), (3b), (4b) and (1c), (2c), (3c), (4c) suggest that rank and peer effects work in opposite directions in the production of outcomes by student ability category (Low, Middle, High), and that different assignments generate different winners and losers in terms of rank.

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 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 abler peers negatively affects their ordinal rank within groups, thereby harming their outcomes.

Our results also provide new evidence about the “track middle” option that Carrell et al., 2013, viewed as optimal on the basis of their estimates on pre-treatment data. As found by both Carrell et al., 2013, and Booij et al., 2017, 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 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 a rank effect, while the overall zero effect for high-ability students hides a negative rank effect and a positive effect of peer ability composition.

A potential concern about this exercise could be related to our asymmetric treatment of heterogeneities in rank and peer composition effects. While all peer composition variables enter in the model with a very flexible functional form, we assume that rank effects are linear and homogenous. This could be especially relevant for the estimation of counterfactual policy effects by ability. In the Online Appendix we present heterogeneous effects of rank by own ability and features of the group composition such as the level and heterogeneity of its ability composition and its size, and by students’ gender. All in all, in both experiments we fail to detect significant heterogeneous rank effects, supporting our baseline specification. The only exception concerns gender heterogeneities in the Dutch data, where we find that the rank effect is larger and only significant for males. This finding is in line with Murphy and Weinhardt, 2019, and is also consistent with the literature on heterogeneous gender attitudes towards competitiveness (see e.g. Gneezy et al., 2003).

8. Robustness tests and additional results from survey data

Before moving to the conclusion it is useful to discuss some robustness tests and some extensions to the results presented so far. The results in Section 6 show substantial evidence on non-linear and heterogeneous peer effects in the Dutch setup, while this is not the case in the Kenyan one. Hence, we will present all our robustness tests using the linear-in-means model for the Kenyan experiment, and the flexible specification of Column (6b) in Table 4 for the Dutch one. In addition, to save space we only show robustness tests for our main outcomes. Results for the other outcomes are available from the authors.

To begin with, as the attenuation of ability peer effects is only present so long as rank affects achievement, it is important to show that the estimated rank effects are robust to several sensitivities. First, Column (1) in Tables 6 and 7 shows that, in both experiments, the effect of $RANK_{ig}$ is very stable when we include group fixed effects. As discussed above, identification of rank effects would in principle require holding the peer group constant. In our main specifications we have done so by controlling for features of the distribution of peer ability within groups. Yet, there may always be a concern that rank is picking up some remaining heterogeneities or non-linearities. The inclusion of group fixed effects allows to control for further unobserved group characteristics - such as group size, teacher effects and group atmosphere - as well as for higher-order moments of the ability distribution within groups.[18] However, the fixed effects absorb the peer composition variables, and as a result their coefficient cannot be estimated. Since a key contribution of this paper is to show how the inclusion of rank affects these coefficients, we prefer not to include them in our baseline specification. Still, the similarity of results between the two approaches in both experiments seems reassuring.

Second, an additional concern could be that rank effects are the results of a specification error, especially as far as the choice of the functional form for the relationship between GPA_i and outcomes is concerned. In Columns (2b), (3b) and (4b) of Tables 6 and 7 we verify that the linear functional form that we have so far used for GPA_i is not overly restrictive by progressively adding second, third, and fourth order terms in GPA_i. Although the coefficient on $RANK_{ig}$ shrinks somewhat, for both experiments it continues to stay positive and large in magnitude. For the Dutch experiment, however, it falls below the (undoubtedly large) minimum effect that we can significantly detect with the sample size we have at hand when we use the fourth order polynomial.[19] Similar conclusions also hold when it comes to the bias in the effect of $\overline{GPA}_{-i}$. These results highlights the limits in terms of statistical validity that we face for the estimation of rank effects. Eventually, the experiment was not designed to estimate rank effects, and we have to make do with limited power for this purpose.

As a matter of fact, the linear specification for the effect of $RANK_{ig}$ could also be restrictive. In Figure 4 we report rank effects together with their 90% confidence intervals obtained from a demanding specification that allows for non-linear effects of $RANK_{ig}$ – by means of dummies for each ventile of its distribution except for the 10th, that we take as the reference group. This specification challenges both the statistical validity of the design and the amount

[18] Eventually, as highlighted by Murphy and Weinhardt, 2019, this boils down to between-group comparison of students with the same GPA relative to group mean, but different rankings due to differences in the distribution of GPA across groups.

[19] We stop at the fourth grade because the point estimates do not shrink further if we include higher order terms.

of identifying variation present in the data. Still, it is useful to highlight that the estimated rank effect from the baseline model with a linear specification - also reported - is not overly restrictive, at least in statistical terms. In addition, in the Dutch experiment the $RANK_{ig}$ effect seems larger at the very bottom of the support of its distribution.[20] Speculatively, one reason why the effect at the very bottom is most pronounced could be related to stigma effects related to being "the worst".

Third, as explained in Section 3, in defining $RANK_{ig}$ we have arbitrarily chosen to break ties by assigning the lower rank to all students, as done by Murphy and Weinhardt, 2019. However, an equally plausible assumption could have been to assign the average or higher rank. Tables A4 and A5 in the Appendix show that results are stable even in those cases.

Fourth, further checks that we have carried out on the Kenyan and Dutch experiments are respectively reported in Tables A6 and A7. In particular, Columns (1b) and (2b) show results when we drop the top 1 and bottom 1 students by group, for whom rank may be especially salient. The resulting rank effects are unchanged with respect to our main results, and so are the conclusions on omitted variable bias. We obtain results that are at least qualitatively in line with our baseline also when we drop students in the top and bottom 1 or 5 percent of the overall distribution of GPA_i – Columns (3b) and (4b) – and $RANK_{ig}$ – Columns (5b) and (6b). For the most extreme trimmings, results are more robust in the Kenyan experiment than in the Dutch one. However, this is not so surprising in the light of the findings reported in Figure 4, that show how the rank effect in the Dutch experiment is more pronounced in the bottom tail of the $RANK_{ig}$ distribution. In addition, it is worth noticing that in those cases the magnitude and significance of the peer composition effects is reduced as well.

[20] That is to say, moving from the 10^{th} to the 1^{st} ventile of $RANK_{ig}$ is proportionally more detrimental for achievement than moving from the 10^{th} to the 2^{nd} ventile.

Fifth, in Tables A8 and A9 we discuss omitted variables bias due to the omission of $RANK_{ig}$ when we use an alternative specification for heterogeneous and non-linear peer effects with respect to the one presented in Table A3 and Table 4 for the Dutch and the Kenyan experiment, respectively. Following Carrell et al., 2013, we use the shares of low and high (vs. median) ability students in the group,[21] and also interact these measures by tertiles of individual ability. Similar specifications have been adopted by Lavy et al., 2012, as well as by Feld and Zoelitz, 2017.

Results show that in both experiments, on average, students are harmed by low (vs. median) achieving peers, and that this effect is significantly larger in magnitude when $RANK_{ig}$ is included in the model. On the other hand, the effect of high (vs. median) achieving peers is never statistically significant, and in the Kenyan experiment it is even negative in sign. When we allow for heterogeneous effects by tertiles of individual ability, we see that in the Kenyan experiment the positive effect of $\overline{GPA}_{-i}$ detected by the linear-in-means model is mostly due to low and middle achieving students being harmed by low- (vs. median) ability peers. Results are less clear-cut in the Dutch case. However, in both setups we see that the inclusion of $RANK_{ig}$ leads us to estimate peer effects of larger magnitude, confirming that omitted variable bias is relevant even when considering a different way to model non-linear and heterogeneous peer effects.

Finally, 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 about their perceptions on teaching and on the learning environment. In the Online Appendix we show detailed results on rank and peer effects in these data. Our main findings can be summarized as follows:

[21] As in Carrell et al., 2013, we define these as the shares of students in the top and bottom quartile of ability.

i) rank effects are small and not significant;

ii) according to students' perceptions, teachers are not very responsive to group ability composition;

iii) low-ability students are less likely to feel involved in the class when surrounded by peers of higher ability, all the more so the more the group is heterogeneous, contributing to explain why ability tracking is especially helpful for them.

Conclusions

Economists have long been arguing about the relevance of ability peer effects to determine educational outcomes, but so far the data have provided evidence of modest effects of peers' ability on individual achievement (see Sacerdote, 2014, for a review). In this paper, we argue that one reason behind this result could be the omission of ordinal ability rank within group from the inputs of the education production function.

We provide evidence based on data from two randomized experiments carried out in a set of Kenyan primary schools by Duflo et al., 2011, and at the University of Amsterdam by Booij et al., 2017, respectively. While in both experiments students are randomly assigned to classes, the former provides very local variation in group ability composition, while the latter was artificially crafted to span a very broad support.

In spite of these differences, we find consistent evidence that rank positively affects student achievement. We also show that, in both setups, rank and mean peer ability are negatively related for given individual ability. We then assess the extent to which omitting rank from the inputs in the production function biases the estimates of the effect of mean peer ability in the standard linear-in-means peer effects model. In both experiments we find that the omission of rank substantially attenuates the estimated effect of mean peer ability by more than 100 percent. Using the generalized Hausman test developed by Pei et al., 2019, we show that this bias is also statistically significant. In an extension to our main result we also characterise the consequences of the omission of rank in more general models that allow for non-linear and heterogeneous effects of peers' ability on individual outcomes.

The consistency of our results across two very diverse institutional settings (Kenya and the Netherlands) and educational levels (primary and tertiary education) seems reassuring also as far as external validity and generalizability is concerned. As a matter of fact, the joint determination of rank and peer composition at the time of group assignment, and the consequent omitted variable bias are embedded in the mechanics of all peer effects studies. Therefore, our findings have implications beyond the economics of education (see for instance Cornelissen, 2016, for a review of the evidence on peer effects in the workplace and Cornelissen et al., 2017, for a leading example).

Moreover, we provide novel evidence that ordinal ability rank within groups is an important mechanism behind the effect of ability grouping policies. Using the large support of ability configurations present in the Dutch data, we unpack 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 composition effects contribute in opposite directions in the production of outcomes for low and high achievers. For instance, when switching from ability mixing to 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 ordinal ability rank within groups. In addition, we show that the overall zero effect on low achievers of the "track middle" assignment proposed by Carrell et al., 2013, is the result of a positive rank effect and a negative peer composition effect of similar magnitude.

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 a rank effect. This has relevant implications for student assignment. On the one hand, policy makers interested in improving the outcomes of low achievers (such as in the case of remedial programs) should favour three-way tracking or equivalently a "track low" system. As we show, under these mechanisms low achievers benefit substantially from increased rank, while they do not suffer negative peer composition effects. On the other hand, the adoption of ability mixing in programs intended to be beneficial for high achievers (i.e. excellence programs or elite schools) would minimize the negative rank effect generated by tracking for these students. At the limit, this would suggest to include low achievers in excellence programs with the only aim of improving the rank of high achievers.

Finally, our results on ability tracking also speak to the literature on school choice. In fact, they suggest that the "big fish in a small pond" rank effect that could motivate the choice of "low tier" schools shall be weighed against the positive externalities that would instead support the choice of "top tier" ones. To gauge this trade-off, parents should be able to carry out an exercise that is not different from our counterfactual simulation. In our view, however, this task requires more information and processing abilities than the ones available to the average family, highlighting the importance of providing tailored information to guide families in this process.

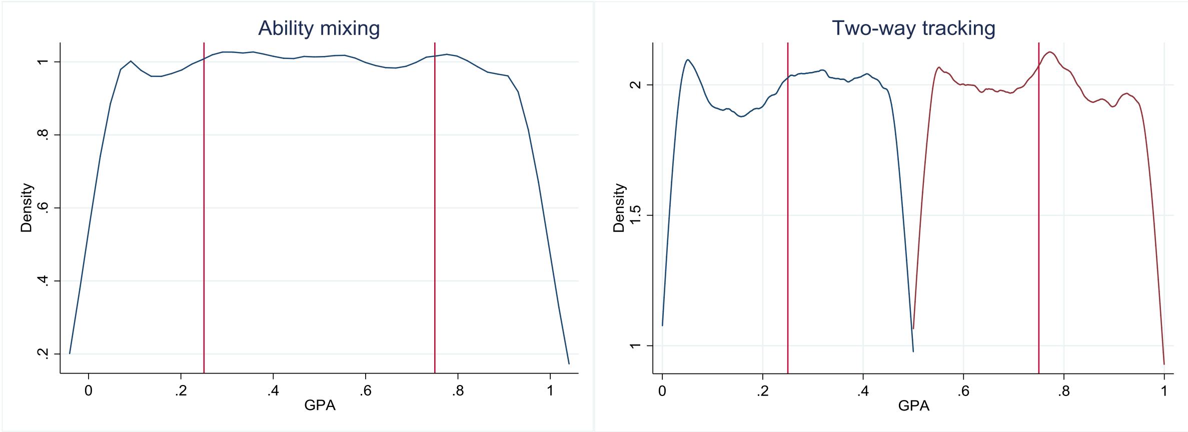
References

- Angrist, J. D., & Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton university press.
- Bertoni, M. & Nisticò, R. (2018). *Rank Concerns, Peer Effects, and Ability Tracking in University: Evidence from a Randomized Experiment*. CSEF Working Paper No. 525.
- 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*, 16(4), 983-1020.
- Cornelissen, T.(2016). Do Social Interactions in the Workplace Lead to Productivity Spillover Among Co-Workers? *IZA World of Labor*: 314 doi: 10.15185/izawol.314.
- Cornelissen, T., Dustmann, C., & Schönberg, U. (2017). Peer Effects in the Workplace. *American Economic Review*, 107(2), 425-56.
- Denning, J. T., & Murphy, R., and Weinhardt, F. (2018). *Class Rank and Long Run Outcomes*. IZA Discussion Paper No. 11808.
- 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.
- Dustmann, C., Puhani, P. A., & Schönberg, U. (2017). The Long-Term Effects of Early Track Choice. *Economic Journal*, 127(603), 1348-1380.
- 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. *Journal of Human Resources*, 53(2), 356-381.
- Elsner, B., & Isphording, I. E., & Zőlitz, U. (2018). *Achievement Rank Affects Performance and Major Choice*. University of Zurich Economics Working paper No. 300.
- Feld, J., & Zölitz, U. (2017). Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects. *Journal of Labor Economics*, 35(2), 387-428.

- Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *Quarterly Journal of Economics*, 118(3), 1049-1074.
- Lavy, V., Silva, O., & Weinhardt, F. (2012). The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools. *Journal of Labor Economics*, 30(2), 367-414.
- Lyle, D. S. (2009). The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point. *American Economic Journal: Applied Economics*, 1, 69–84.
- Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60(3), 531-542.
- Murphy, R., & Weinhardt, F. (2019). Top of the Class: The Importance of Ordinal Rank. Forthcoming, *Review of Economic Studies.*
- Pagani, L., Comi, S., & Origo, F. (2019). The Effect of School Rank on Personality Traits. Forthcoming, *Journal of Human Resources*..
- Pei, Z., Pischke, J. S., & Schwandt, H. (2019). Poorly Measured Confounders Are More Useful on the Left Than on the Right. *Journal of Business & Economic Statistics*, 37(2), 205-216.
- Rubin, D. B. (1980). Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. *Journal of the American Statistical Association*, 75 (371), 591-593.
- 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 No. 6331.
- Tincani, M., & Mierendorff, K. (2018). *Explaining Experimental Evidence on Ability Peer Effects with a Theory of Rank Concerns*. Mimeo, UCL.
- Yu, H. (2019). *Am I the Big Fish? The Effect of Ordinal Rank on Student Academic Performance in Middle School*. Mimeo, Texas A&M University.

Figures and Tables

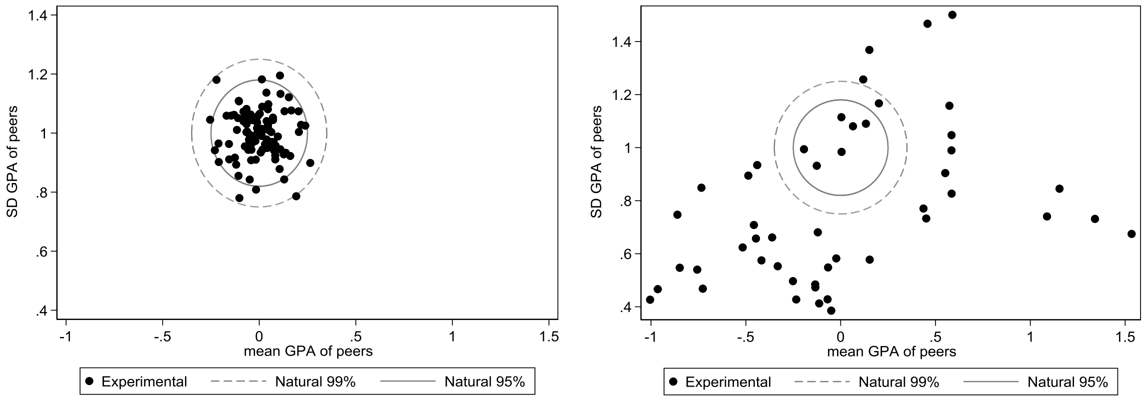
Figure 1. Rank and peer composition as two sides of the same coin



Notes: the panels represent the hypothetical distribution of ability (GPA) in groups formed under ability mixing and two-way tracking. Ability is drawn from a uniform distribution with support (0,1). Under ability mixing, all pupils in a grade are pooled together. Pupils are split by ability (above-below the 0.5 median) under two-way tracking. The vertical lines are for pupils at the 25th and 75th percentile of the overall ability distribution.

Figure 2. Variation in mean and SD of peers' ability

a. Kenyan experiment b. Dutch experiment

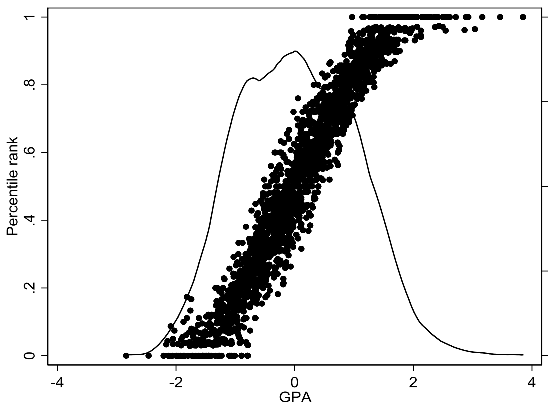


Note: Each dot in the graph represents one group. The dashed (solid) circle represents the area where 99% (95%) of the groups would be located when students would be randomly assigned to groups (this is the situation in the Kenyan experiment) and when the composition of the groups would not be manipulated (contrary to what happens instead in the Dutch experiment).

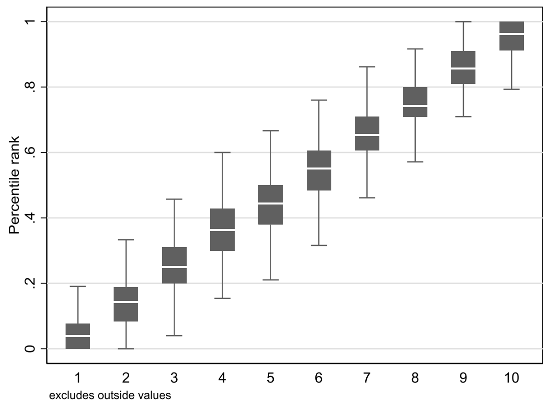
Figure 3. Variation in rank by level of GPA

Kenyan experiment

a. Raw data

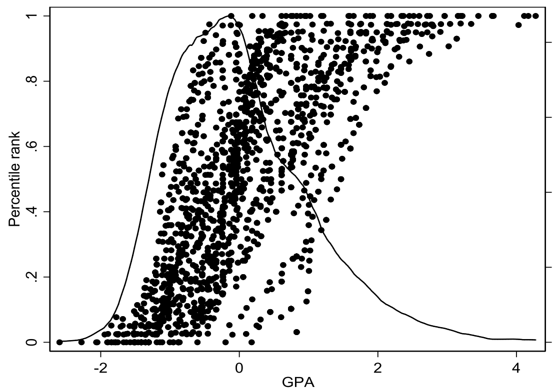


b. Boxplot by decile of GPA

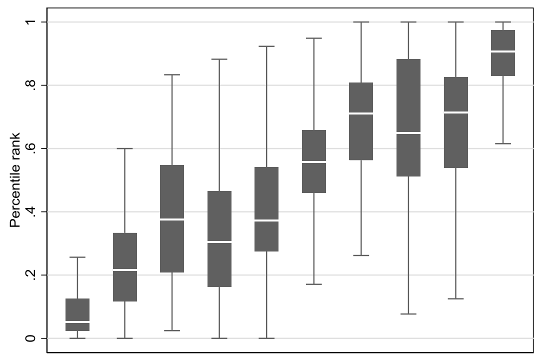


Dutch experiment

a. Raw data



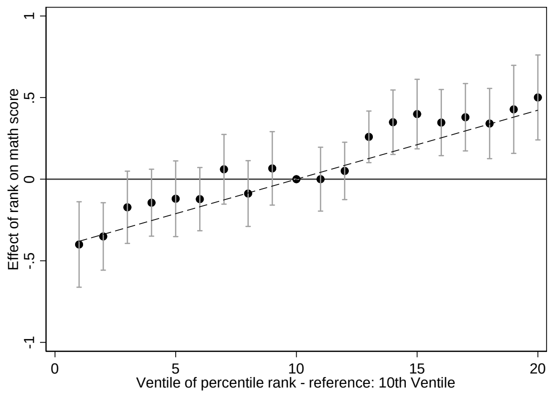
b. Boxplot by decile of GPA



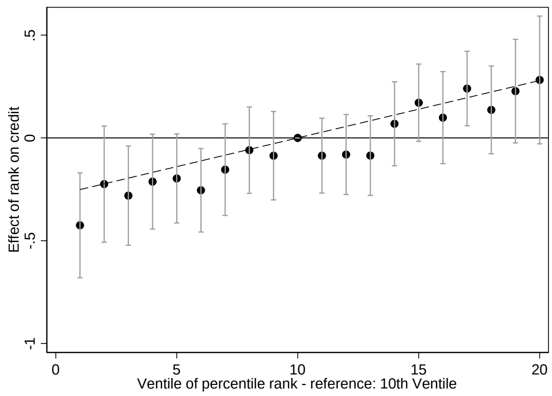
Notes: Upper panels are for the Kenyan experiment while lower panels for the Dutch one. Panel a. reports the joint distribution of rank and GPA. The estimated density of GPA is overlaid. Panel b. reports the box-plot of rank by decile of GPA. Number of observations: 2,188 in the Kenyan experiment and 1,876 in the Dutch one.

Figure 4. Non-linear rank effects on math score (Kenyan experiment) and on credits (Dutch experiment), controlling for linear GPA_i

a. Kenyan experiment



b. Dutch experiment



Notes: Each panel reports the estimated rank effects by ventile (reference: 10^{th} ventile) and their 90% confidence interval. Estimates are based on the specification used in Table 2, column (2b) for the Kenyan experiment and in Table 3, Column (2b) in the Dutch experiment, respectively, but using dummies for each ventile of $RANK_{ig}$ and linear trends in GPA_i. The estimated rank effect is also reported with a dashed line. The dependent variable is math score in the Kenyan experiment and credits in the Dutch experiment.

Table 1. Descriptive statistics

	(1) Mean	(2) Std. Dev.
Panel A: Kenyan experiment		
Outcomes:		
Total score (standardized in full sample)	0.014	0.999
Math score (standardized in full sample)	-0.011	0.988
Literacy score (standardized in full sample)	0.032	1.006
Explanatory variables:		
$RANK_{ig}$	0.500	0.300
GPA_i (standardized by school in the full sample)	0.045	0.978
$\overline{GPA}_{-i}$	-0.001	0.107
$SD(GPA_{-i})$	0.999	0.082
Male	0.477	0.500
Age at test	9.188	1.469
Assigned to contract teacher	0.517	0.500
Panel B: Dutch experiment		
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	0.487	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 in the Kenyan experiment is 2,189, except for the total and math score that are only available for 2,188 students. The number of observations in the Dutch experiment 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. Rank and peer effects on math score from linear-in-means models - Kenyan experiment

	(1)	(2a)	(2b)
Dependent variable:	$RANK_{ig}$	Math score	Math score
$RANK_{ig}$			0.846***
			(0.216)
$\overline{GPA}_{-i}$	-0.219***	0.324	0.509**
	(0.028)	(0.226)	(0.214)
			[0.001]
GPA_i	0.291***	0.496***	0.250***
	(0.002)	(0.026)	(0.067)
			[<0.001]
R-squared	0.926	0.357	0.362

Notes: Each column reports the results from a different OLS regression. Dependent variable is reported at the top of each column. School fixed effects and background controls are included in all specifications. Background controls are: gender, age, being assigned to the contract teacher. Number of observations: 2,188. Standard errors clustered by school are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 3. Rank and peer effects on credits from linear-in-means models - Dutch experiment

	(1)	(2a)	(2b)
Dependent variable:	$RANK_{ig}$	Credits	Credits
$RANK_{ig}$			0.351**
			(0.140)
$\overline{GPA}_{-i}$	-0.287***	0.048	0.148***
	(0.014)	(0.041)	(0.052)
			[0.012]
GPA_i	0.352***	0.314***	0.191***
	(0.026)	(0.066)	(0.050)
			[0.015]
R-squared	0.801	0.266	0.269

Notes: Each column reports the results from a different OLS regression. Dependent variable is reported at the top of each column. Randomization and background controls are included in all specifications. The peer variable $\overline{GPA}_{-i}$ is re-centred to have zero means. Number of observations: 1,876. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 4. Rank and peer effects on credits from heterogeneous-peer-effects models - Dutch experiment

	(1)	(2a)	(2b)	(3)	(4a)	(4b)	(5)	(6a)	(6b)
Dependent variable	$RANK_{ig}$	Credits	Credits	$RANK_{ig}$	Credits	Credits	$RANK_{ig}$	Credits	Credits
$RANK_{ig}$			0.373**			0.358**			0.559***
			(0.141)			(0.138)			(0.178)
$\overline{GPA}_{-i}$	-0.302***	0.070	0.182***	-0.296***	0.095**	0.201***	-0.376***	0.148***	0.358***
	(0.015)	(0.043)	(0.059)	(0.015)	(0.046)	(0.062)	(0.013)	(0.052)	(0.086)
			[0.009]			[0.010]			[0.002]
$SD(GPA_{-i})$	0.065***	-0.095	-0.119	0.059***	-0.121*	-0.142**	0.027	-0.185**	-0.200**
	(0.013)	(0.073)	(0.077)	(0.015)	(0.063)	(0.067)	(0.022)	(0.082)	(0.079)
			[0.018]			[0.028]			[0.235]
$\overline{GPA}_{-i} \times SD(GPA_{-i})$				0.093**	0.423**	0.390**	0.377***	0.343*	0.132
				(0.039)	(0.176)	(0.177)	(0.051)	(0.190)	(0.183)
						[0.107]			[0.009]
GPA_i	0.351***	0.314***	0.183***	0.352***	0.317***	0.191***	0.365***	0.350***	0.145**
	(0.026)	(0.034)	(0.049)	(0.026)	(0.034)	(0.049)	(0.013)	(0.035)	(0.065)
			[0.011]			[0.012]			[0.002]
$GPA_i \times \overline{GPA}_{-i}$							0.025**	-0.117***	-0.131***
							(0.011)	(0.042)	(0.040)
									[0.069]
$GPA_i \times SD(GPA_{-i})$							-0.356***	0.104	0.303***
							(0.027)	(0.075)	(0.091)
									[0.003]
$GPA_i \times \overline{GPA}_{-i} \times SD(GPA_{-i})$							0.160***	-0.287**	-0.376***
							(0.045)	(0.138)	(0.135)
									[0.034]
R-squared	0.803	0.267	0.269	0.804	0.268	0.270	0.864	0.271	0.274

Notes: Each column reports the results from a different OLS regression. Dependent variable is reported at the top of each column. Randomization and background controls are included in all specifications. The peer variables $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$ are re-centred to have zero means. Number of observations: 1,876. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 5. Estimated tracking effects on credits compared to mixing. Total effects and unpacking rank and peer effects - Dutch experiment

					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}	0.103***	-0.002***	0.105***	0.157***	0.138***	0.019				0.050	-0.141***	0.191***
		(0.028)	(0.000)	(0.028)	(0.040)	(0.044)	(0.048)				(0.040)	(0.045)	(0.059)
Three-way tracking	{L}, {M}, {H}	0.147***	-0.003***	0.150***	0.267***	0.183***	0.084	0.147***	-0.004***	0.151***	0.027	-0.188***	0.215**
		(0.036)	(0.001)	(0.037)	(0.072)	(0.058)	(0.077)	(0.056)	(0.001)	(0.056)	(0.055)	(0.060)	(0.084)
Track low	{L}, {M, H}	0.128***	-0.002***	0.130***	0.267***	0.183***	0.084	0.090**	-0.141***	0.231***	0.027	-0.047***	0.074*
		(0.031)	(0.000)	(0.031)	(0.072)	(0.058)	(0.077)	(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***	-0.004***	0.151***	-0.011	-0.048***	0.036
		(0.011)	(0.000)	(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}	0.064***	-0.002***	0.066***	0.094***	0.046***	0.048	0.073**	0.139***	-0.066	0.027	-0.188***	0.215**
		(0.024)	(0.000)	(0.024)	(0.028)	(0.015)	(0.030)	(0.035)	(0.044)	(0.051)	(0.055)	(0.060)	(0.084)

Notes: The table reports (conditional) average treatment effects of different tracking configurations relative to mixing based on the estimates from Table 4, Column (6b). Dependent variable is number of credits collected in the first year. 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 6. Robustness results. Rank and peer effects on math score when accounting for group fixed effects and flexible GPA - Kenyan experiment

	(1)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
	Using			Accounting for flexible GPA_i			
	Group FE	Quadratic	Quadratic	Cubic	Cubic	Quartic	Quartic
$RANK_{ig}$	0.817***		0.817***		0.737***		0.696***
	(0.225)		(0.225)		(0.258)		(0.258)
$\overline{GPA}_{-i}$		0.320	0.501**	0.331	0.487**	0.333	0.479**
		(0.225)	(0.213)	(0.227)	(0.212)	(0.227)	(0.211)
			[0.003]		[0.015]		[0.021]
GPA_i	0.249***	0.502***	0.261***	0.556***	0.300***	0.579***	0.326***
	(0.070)	(0.027)	(0.071)	(0.035)	(0.094)	(0.042)	(0.100)
			[<0.001]		[0.003]		[0.007]

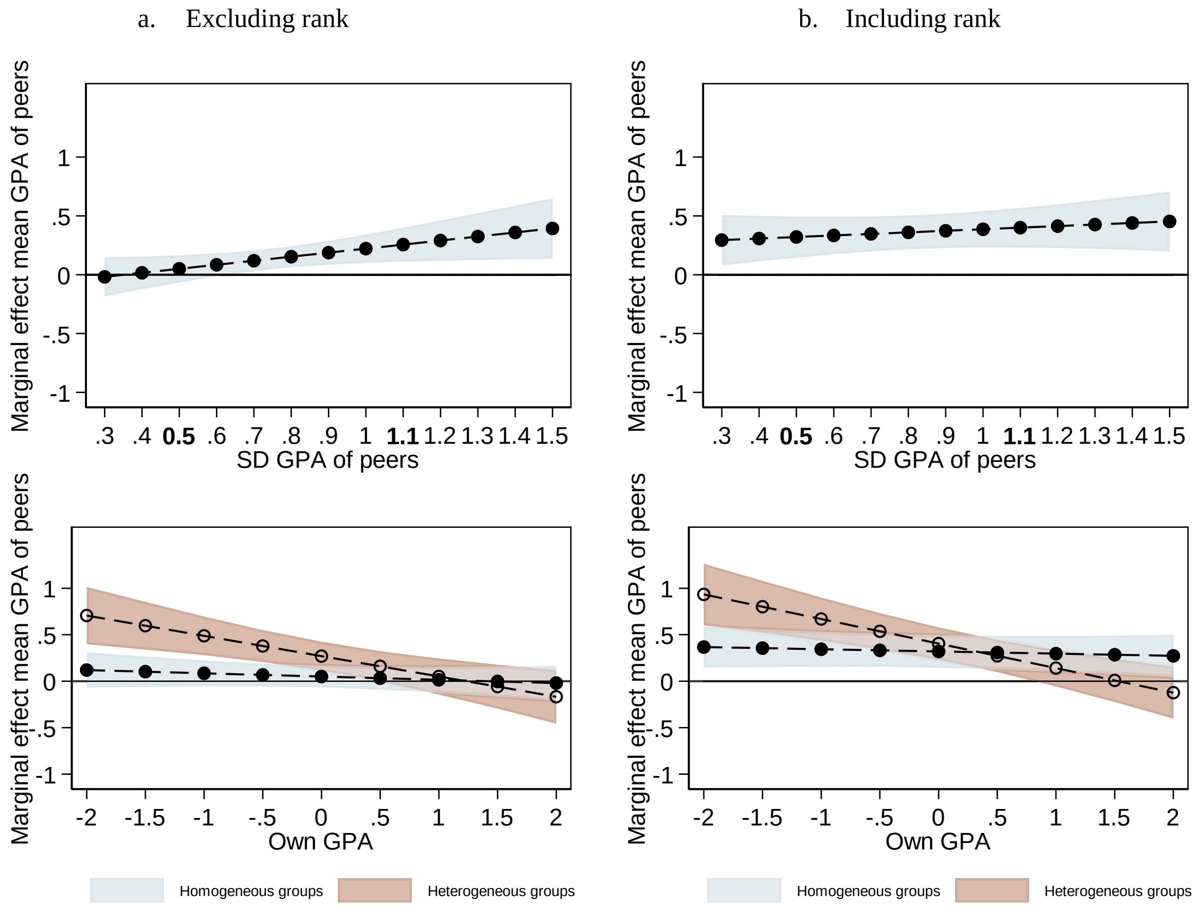
Notes: Each column reports the results from a different OLS regression. Dependent variable is math score. School fixed effects and background controls are included in all specifications. Background controls are: gender, age, being assigned to the contract teacher. Number of observations: 2,188. Standard errors clustered by school are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 7. Robustness results. Rank and peer effects on credits when accounting for group fixed effects and flexible GPA - Dutch experiment

	(1)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
	Using			Accounting for flexible GPA_i			
	Group FE	Quadratic	Quadratic	Cubic	Cubic	Quartic	Quartic
$RANK_{ig}$	0.503***		0.501**		0.439*		0.328
	(0.175)		(0.227)		(0.222)		(0.226)
$\overline{GPA}_{-i}$		0.145***	0.336***	0.153***	0.318***	0.148***	0.272***
		(0.050)	(0.097)	(0.050)	(0.094)	(0.050)	(0.094)
			[0.025]		[0.047]		[0.140]
$SD(GPA_{-i})$		-0.143*	-0.186**	-0.153*	-0.189**	-0.144*	-0.173*
		(0.083)	(0.083)	(0.084)	(0.083)	(0.085)	(0.086)
			[0.030]		[0.045]		[0.118]
$\overline{GPA}_{-i} \times SD(GPA_{-i})$		0.268	0.132	0.243	0.128	0.205	0.125
		(0.183)	(0.182)	(0.190)	(0.185)	(0.185)	(0.185)
			[0.038]		[0.066]		[0.176]
GPA_i	0.149**	0.393***	0.179*	0.451***	0.253**	0.536***	0.374***
	(0.068)	(0.043)	(0.100)	(0.051)	(0.103)	(0.071)	(0.129)
			[0.026]		[0.047]		[0.142]
$GPA_i \times \overline{GPA}_{-i}$	-0.122**	-0.092**	-0.122**	-0.122**	-0.134***	-0.096**	-0.116**
	(0.047)	(0.044)	(0.046)	(0.046)	(0.046)	(0.046)	(0.048)
			[0.029]		[0.056]		[0.132]
$GPA_i \times SD(GPA_{-i})$	0.255**	0.173**	0.303***	0.193**	0.303***	0.181**	0.265***
	(0.103)	(0.080)	(0.091)	(0.078)	(0.086)	(0.080)	(0.088)
			[0.032]		[0.054]		[0.157]
$GPA_i \times \overline{GPA}_{-i} \times SD(GPA_{-i})$	-0.196	-0.356**	-0.387***	-0.326**	-0.359***	-0.308**	-0.335**
	(0.197)	(0.134)	(0.132)	(0.136)	(0.133)	(0.145)	(0.141)
			[0.187]		[0.223]		[0.291]

Notes: Each column reports the results from a different OLS regression. Dependent variable is number of credits collected in the first year. The outcome is standardized to have zero mean and unit standard deviation. Randomization and background controls are included in all specifications. The peer variables $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$ are re-centred to have zero means. Number of observations: 1,876. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Figure A1. Average marginal effects of mean peer ability on number of credits

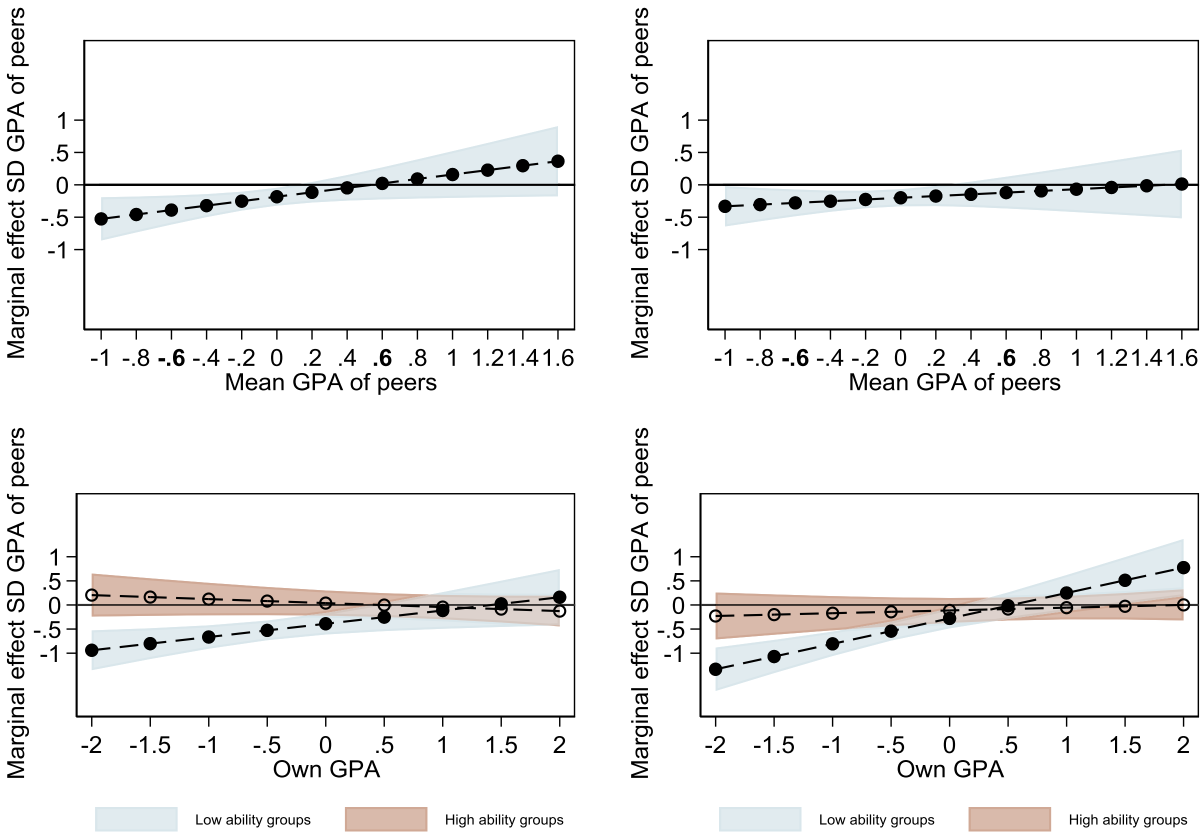


Notes: the figure replicates the left column of Figure 3 in Booij et al. (2017). It shows the marginal effects of $\overline{GPA}_{-i}$ by $SD(GPA_{-i})$ and by GPA_i in the original specification of Booij et al. (2017) and when including $RANK_{ig}$ in the model, respectively. Marginal effects based on the estimates from Table 4, Column (6a) and (6b), respectively.

Figure A2. Average marginal effects of peer ability's heterogeneity on number of credits

a. Excluding rank

b. Including rank



Notes: the figure replicates the right column of Figure 3 in Booij et al. (2017). It shows the marginal effects of $SD(GPA_{-i})$ by $\overline{GPA}_{-i}$ and by GPA_i in the original specification of Booij et al. (2017) and when including $RANK_{ig}$ in the model, respectively. Marginal effects based on the estimates from Table 4, Column (6a) and (6b), respectively.

Table A1. Main results. Rank and peer effects on literacy and total score from linear-in-means models - Kenyan experiment

	(1a)	(1b)	(2a)	(2b)
Dependent variable:	Literacy score	Literacy score	Total score	Total score
$RANK_{ig}$		-0.056		0.403**
		(0.192)		(0.185)
$\overline{GPA}_{-i}$	0.291	0.279	0.345	0.433**
	(0.184)	(0.189)	(0.211)	(0.208)
		[0.766]		[0.051]
GPA_i	0.413***	0.430***	0.507***	0.390***
	(0.037)	(0.072)	(0.032)	(0.064)
		[0.769]		[0.027]
Observations	2,189	2,189	2,188	2,188

Notes: Each column reports the results from a different OLS regression. Dependent variable is reported at the top of each column. School fixed effects and background controls are included in all specifications. Background controls are: gender, age, being assigned to the contract teacher. Standard errors clustered by school are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A2. Main results. Rank and peer effects on dropout and average grade from linear-in-means models - Dutch experiment

	(1a)	(1b)	(2a)	(2b)
Dependent variable:	Dropout	Dropout	Avg grade	Avg grade
$RANK_{ig}$		-0.210***		0.014
		(0.073)		(0.121)
$\overline{GPA}_{-i}$	-0.006	-0.067**	0.022	0.026
	(0.019)	(0.028)	(0.037)	(0.047)
		[0.004]		[0.910]
GPA_i	-0.150***	-0.076***	0.458***	0.453***
	(0.018)	(0.027)	(0.031)	(0.049)
		[0.004]		[0.910]
Observations	1,876	1,876	1,753	1,753

Notes: Each column reports the results from a different OLS regression. Dependent variable is reported at the top of each column. The outcome is standardized to have zero mean and unit standard deviation. Randomization and background controls are included in all specifications. The peer variable $\overline{GPA}_{-i}$ is re-centred to have zero means. Number of observations: 1,876. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A3. Rank and peer effects on math score from heterogeneous-peer-effects models - Kenyan experiment

	(1)	(2a)	(2b)	(3)	(4a)	(4b)	(5)	(6a)	(6b)
	$RANK_{ig}$	Math score	Math score	$RANK_{ig}$	Math score	Math score	$RANK_{ig}$	Math score	Math score
$RANK_{ig}$			0.839***			0.826***			0.921***
			(0.217)			(0.225)			(0.238)
$\overline{GPA}_{-i}$	-0.219***	0.322	0.507**	-0.196***	0.379	0.541**	-0.239***	0.431	0.651**
	(0.027)	(0.225)	(0.214)	(0.029)	(0.252)	(0.237)	(0.032)	(0.273)	(0.262)
			[0.001]			[0.002]			[0.001]
$SD(GPA_{-i})$	-0.052*	-0.160	-0.116	-0.064**	-0.187	-0.135	-0.026	-0.241	-0.218
	(0.029)	(0.231)	(0.227)	(0.030)	(0.240)	(0.238)	(0.031)	(0.252)	(0.248)
			[0.087]			[0.062]			[0.397]
$\overline{GPA}_{-i} \times SD(GPA_{-i})$				3.041**	7.395	4.884	-2.436*	13.816	16.059*
				(1.479)	(6.551)	(7.298)	(1.353)	(9.504)	(9.349)
						[0.091]			[0.095]
GPA_i	0.291***	0.496***	0.252***	0.290***	0.495***	0.255***	0.292***	0.493***	0.223***
	(0.002)	(0.026)	(0.067)	(0.002)	(0.026)	(0.068)	(0.002)	(0.026)	(0.072)
			[<0.001]			[<0.001]			[<0.001]
$GPA_i \times \overline{GPA}_{-i}$							0.044**	-0.056	-0.096
							(0.022)	(0.171)	(0.176)
									[0.096]
$GPA_i \times SD(GPA_{-i})$							-0.268***	0.329	0.575*
							(0.038)	(0.316)	(0.340)
									[0.001]
$GPA_i \times \overline{GPA}_{-i} \times SD(GPA_{-i})$							0.405	-1.243	-1.616
							(0.363)	(2.397)	(2.234)
									[0.314]
R-squared	0.926	0.357	0.362	0.927	0.358	0.361	0.930	0.358	0.364

Notes: Each column reports the results from a different OLS regression. Dependent variable is reported at the top of each column. School fixed effects and background controls are included in all specifications. Background controls are: gender, age, being assigned to the contract teacher. Number of observations: 2,188. Standard errors clustered by school are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A4. Robustness results. Rank and peer effects on math score when using alternative definitions of rank to break ties - Kenyan experiment

	(1a)	(1b)	(2a)	(2b)
	Using alternative definitions of $RANK_{ig}$			
	Mean Rank	Mean Rank	Max Rank	Max Rank
$RANK_{ig}$		0.785***		0.678***
		(0.218)		(0.220)
$\overline{GPA}_{-i}$	0.324	0.493**	0.324	0.468**
	(0.226)	(0.212)	(0.226)	(0.213)
		[0.003]		[0.009]
GPA_i	0.496***	0.269***	0.496***	0.302***
	(0.026)	(0.067)	(0.026)	(0.068)
		[<0.001]		[0.002]

Notes: Each column reports the results from a different OLS regression. Dependent variable is math score. School fixed effects and background controls are included in all specifications. Background controls are: gender, age, being assigned to the contract teacher. Number of observations: 2,188. Standard errors clustered by school are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A5. Robustness results. Rank and peer effects on credits when using alternative definitions of rank to break ties - Dutch experiment

	(1a)	(1b)	(2a)	(2b)
	Using alternative definitions of $RANK_{ig}$			
	Mean Rank	Mean Rank	Max Rank	Max Rank
$RANK_{ig}$		0.494**		0.402**
		(0.188)		(0.192)
$\overline{GPA}_{-i}$	0.148***	0.335***	0.148***	0.301***
	(0.052)	(0.090)	(0.052)	(0.092)
		[0.009]		[0.036]
$SD(GPA_{-i})$	-0.185**	-0.194**	-0.185**	-0.189**
	(0.082)	(0.079)	(0.082)	(0.080)
		[0.418]		[0.664]
$\overline{GPA}_{-i} \times SD(GPA_{-i})$	0.343*	0.151	0.343*	0.183
	(0.190)	(0.184)	(0.190)	(0.185)
		[0.021]		[0.052]
GPA_i	0.350***	0.171**	0.350***	0.206***
	(0.035)	(0.068)	(0.035)	(0.070)
		[0.008]		[0.033]
$GPA_i \times \overline{GPA}_{-i}$	-0.117***	-0.131***	-0.117***	-0.131***
	(0.042)	(0.040)	(0.042)	(0.040)
		[0.074]		[0.108]
$GPA_i \times SD(GPA_{-i})$	0.104	0.280***	0.104	0.247**
	(0.075)	(0.095)	(0.075)	(0.098)
		[0.011]		[0.039]
$GPA_i \times \overline{GPA}_{-i} \times SD(GPA_{-i})$	-0.287**	-0.365**	-0.287**	-0.351**
	(0.138)	(0.136)	(0.138)	(0.138)
		[0.053]		[0.093]

Notes: Each column reports the results from a different OLS regression. Dependent variable is number of credits collected in the first year. The outcome is standardized to have zero mean and unit standard deviation. Randomization and background controls are included in all specifications. The peer variables $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$ are re-centred to have zero means. Number of observations: 1,876. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A6. Further robustness results. Rank and peer effects on math score when trimming the sample - Kenyan experiment

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)	(6a)	(6b)
	Excluding top or bottom students by group				Excluding obs. with GPA_i in top and bottom				Excluding obs. with $RANK_{ig}$ in top and bottom			
	Top 1	Top 1	Bottom 1	Bottom 1	1%	1%	5%	5%	1%	1%	5%	5%
$RANK_{ig}$		0.769***		0.777***		0.713***		0.695**		0.846***		0.530**
		(0.244)		(0.223)		(0.255)		(0.272)		(0.216)		(0.249)
$\overline{GPA}_{-i}$	0.318	0.492**	0.363	0.537**	0.295	0.446**	0.330	0.485**	0.324	0.509**	0.362	0.484**
	(0.233)	(0.223)	(0.236)	(0.223)	(0.229)	(0.217)	(0.245)	(0.229)	(0.226)	(0.214)	(0.253)	(0.240)
		[0.006]		[0.003]		[0.017]		[0.029]		[0.001]		[0.049]
GPA_i	0.505***	0.274***	0.501***	0.274***	0.513***	0.296***	0.537***	0.310***	0.496***	0.250***	0.509***	0.349***
	(0.028)	(0.077)	(0.027)	(0.070)	(0.030)	(0.085)	(0.033)	(0.095)	(0.026)	(0.067)	(0.029)	(0.079)
		[0.001]		[<0.001]		[0.005]		[0.010]		[<0.001]		[0.030]
Observations	2,109	2,109	2,119	2,119	2,147	2,147	1,982	1,982	2,189	2,189	1,982	1,982

Notes: Each column reports the results from a different OLS regression. Dependent variable is math score. School fixed effects and background controls are included in all specifications. Background controls are: gender, age, being assigned to the contract teacher. Standard errors clustered by school are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A7. Further robustness results. Rank and peer effects on credits when trimming the sample - Dutch experiment

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)	(6a)	(6b)
	Excluding top or bottom students by group				Excluding obs. with GPA_i in top and bottom				Excluding obs. with $RANK_{ig}$ in top and bottom			
	Top 1	Top 1	Bottom 1	Bottom 1	1%	1%	5%	5%	1%	1%	5%	5%
$RANK_{ig}$		0.550**		0.515***		0.485**		0.145		0.418*		0.027
		(0.215)		(0.180)		(0.201)		(0.251)		(0.219)		(0.243)
$\overline{GPA}_{-i}$	0.128**	0.350***	0.141**	0.345***	0.156***	0.340***	0.090	0.150	0.106	0.285**	0.081	0.093
	(0.056)	(0.104)	(0.066)	(0.104)	(0.053)	(0.096)	(0.063)	(0.120)	(0.068)	(0.123)	(0.065)	(0.114)
		[0.009]		[0.004]		[0.016]		[0.556]		[0.052]		[0.912]
$SD(GPA_{-i})$	-0.193**	-0.193**	-0.188**	-0.212**	-0.173**	-0.192**	-0.124	-0.135	-0.180*	-0.189*	-0.151	-0.152
	(0.087)	(0.083)	(0.093)	(0.090)	(0.084)	(0.081)	(0.093)	(0.097)	(0.096)	(0.094)	(0.096)	(0.095)
		[0.996]		[0.066]		[0.094]		[0.542]		[0.387]		[0.912]
$\overline{GPA}_{-i} \times SD(GPA_{-i})$	0.341	0.115	0.306	0.104	0.356*	0.201	0.407*	0.365	0.377	0.185	0.341	0.330
	(0.212)	(0.220)	(0.194)	(0.182)	(0.184)	(0.177)	(0.234)	(0.245)	(0.227)	(0.238)	(0.209)	(0.235)
		[0.013]		[0.016]		[0.044]		[0.566]		[0.057]		[0.912]
GPA_i	0.381***	0.154	0.340***	0.150**	0.388***	0.193**	0.451***	0.386***	0.380***	0.202**	0.412***	0.400***
	(0.043)	(0.094)	(0.038)	(0.068)	(0.040)	(0.082)	(0.051)	(0.121)	(0.046)	(0.100)	(0.049)	(0.114)
		[0.009]		[0.004]		[0.015]		[0.559]		[0.050]		[0.912]
$GPA_i \times \overline{GPA}_{-i}$	-0.117**	-0.119**	-0.122**	-0.142***	-0.123**	-0.135***	-0.091	-0.096	-0.118**	-0.125**	-0.104*	-0.104*
	(0.052)	(0.051)	(0.045)	(0.044)	(0.046)	(0.046)	(0.064)	(0.066)	(0.057)	(0.057)	(0.056)	(0.056)
		[0.784]		[0.033]		[0.142]		[0.554]		[0.365]		[0.914]
$GPA_i \times SD(GPA_{-i})$	0.064	0.280**	0.070	0.270**	0.123	0.281**	0.037	0.084	-0.019	0.170	-0.025	-0.012
	(0.086)	(0.118)	(0.091)	(0.113)	(0.089)	(0.109)	(0.122)	(0.147)	(0.103)	(0.146)	(0.099)	(0.141)
		[0.010]		[0.006]		[0.023]		[0.559]		[0.051]		[0.912]
$GPA_i \times \overline{GPA}_{-i} \times SD(GPA_{-i})$	-0.308**	-0.386**	-0.243*	-0.343**	-0.339**	-0.405**	-0.413*	-0.417**	-0.269*	-0.346**	-0.353**	-0.357**
	(0.153)	(0.155)	(0.110)	(0.141)	(0.157)	(0.159)	(0.206)	(0.204)	(0.155)	(0.159)	(0.139)	(0.141)
		[0.029]		[0.036]		[0.115]		[0.747]		[0.074]		[0.912]
Observations	1,828	1,828	1,828	1,828	1,838	1,838	1,661	1,661	1,772	1,772	1,691	1,691

Notes: Each column reports the results from a different OLS regression. Dependent variable is number of credits collected in the first year. The outcome is standardized to have zero mean and unit standard deviation. Randomization and background controls are included in all specifications. The peer variables $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$ are re-centred to have zero means. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A8. Rank and peer effects on math score from alternative heterogeneous-peer-effects models - Kenyan experiment

	(1)	(2a)	(2b)	(3)	(4a)	(4b)
Dependent variable	$RANK_{ig}$	Math score	Math score	$RANK_{ig}$	Math score	Math score
$RANK_{ig}$			0.518*			0.641**
			(0.272)			(0.286)
GPA_i	0.214***	0.370***	0.259***	0.213***	0.370***	0.233***
	(0.009)	(0.043)	(0.073)	(0.009)	(0.043)	(0.075)
			[0.048]			[0.020]
Low GPA_i (1^{st} Tertile)	-0.103***	-0.169***	-0.116*	-0.240***	0.136	0.289
	(0.016)	(0.061)	(0.068)	(0.037)	(0.325)	(0.311)
			[0.092]			[0.032]
High GPA_i (3^{rd} Tertile)	0.094***	0.175**	0.127*	0.279***	0.149	-0.030
	(0.011)	(0.068)	(0.074)	(0.035)	(0.326)	(0.353)
			[0.068]			[0.032]
Fraction of Low-GPA	0.310***	-0.876*	-1.037**			
peers	(0.049)	(0.474)	(0.476)			
			[0.052]			
Fraction of High-GPA	-0.410***	-0.666	-0.454			
peers	(0.057)	(0.553)	(0.585)			
			[0.071]			
Low GPA_i × Fraction of Low-GPA peers				0.557***	-1.368**	-1.725***
				(0.076)	(0.565)	(0.559)
						[0.018]
Middle GPA_i × Fraction of Low-GPA peers				0.465***	-1.009*	-1.308**
				(0.067)	(0.558)	(0.573)
						[0.033]
High GPA_i × Fraction of Low-GPA peers				-0.057	-0.352	-0.316
				(0.054)	(0.712)	(0.724)
						[0.377]
Low GPA_i × Fraction of High-GPA peers				-0.020	-1.082	-1.070
				(0.075)	(0.777)	(0.786)
						[0.796]
Middle GPA_i × Fraction of High-GPA peers				-0.456***	-0.243	0.050
				(0.101)	(0.850)	(0.843)
						[0.051]
High GPA_i × Fraction of High-GPA peers				-0.695***	-0.741	-0.295
				(0.057)	(0.743)	(0.810)
						[0.025]
R-squared	0.943	0.363	0.365	0.947	0.364	0.366

Notes: Each column reports the results from a different OLS regression. Dependent variable is reported at the top of each column. School fixed effects and background controls are included in all specifications. Background controls are: gender, age, being assigned to the contract teacher. Number of observations: 2,188. Standard errors clustered by school are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A9. Rank and peer effects on credits from alternative heterogeneous-peer-effects models - Dutch experiment

	(1)	(2a)	(2b)	(3)	(4a)	(4b)
Dependent variable	$RANK_{ig}$	Credits	Credits	$RANK_{ig}$	Credits	Credits
$RANK_{ig}$			0.326*			0.368*
			(0.164)			(0.191)
GPA_i	0.278***	0.260***	0.169***	0.284***	0.263***	0.158***
	(0.007)	(0.039)	(0.049)	(0.026)	(0.042)	(0.052)
			[0.264]			[0.058]
Low GPA_i (1^{st} Tertile)	-0.158***	-0.137*	-0.085	-0.183***	-0.167	-0.099
	(0.010)	(0.077)	(0.085)	(0.023)	(0.150)	(0.166)
			[0.050]			[0.052]
High GPA_i (3^{rd} Tertile)	0.125***	0.066	0.025	0.314***	0.109	-0.007
	(0.011)	(0.103)	(0.107)	(0.023)	(0.132)	(0.139)
			[0.066]			[0.057]
Fraction of Low-GPA	0.434***	-0.149	-0.291*			
Peers	(0.021)	(0.142)	(0.150)			
			[0.042]			
Fraction of High-GPA	-0.391***	0.025	0.153			
Peers	(0.018)	(0.111)	(0.134)			
			[0.054]			
Low GPA_i × Fraction of Low-GPA peers				0.583***	-0.074	-0.288
				(0.025)	(0.188)	(0.218)
						[0.054]
Middle GPA_i × Fraction of Low-GPA peers				0.715***	-0.151	-0.413*
				(0.029)	(0.222)	(0.235)
						[0.058]
High GPA_i × Fraction of Low-GPA peers				-0.088**	-0.256	-0.224
				(0.043)	(0.281)	(0.277)
						[0.160]
Low GPA_i × Fraction of High-GPA peers				-0.038	0.122	0.136
				(0.027)	(0.251)	(0.252)
						[0.228]
Middle GPA_i × Fraction of High-GPA peers				-0.450***	0.054	0.220
				(0.028)	(0.264)	(0.308)
						[0.053]
High GPA_i × Fraction of High-GPA peers				-0.665***	-0.062	0.182
				(0.037)	(0.182)	(0.201)
						[0.056]
R-squared	0.835	0.268	0.270	0.869	0.269	0.270

Notes: Each column reports the results from a different OLS regression. Dependent variable is reported at the top of each column. Randomization and background controls are included in all specifications. Number of observations: 1,876. Standard errors clustered by tutorial group are reported in parenthesis. Number of clusters: 48. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Online Appendix to Ordinal Rank and Peer Composition: Two Sides of the Same Coin?

1. Heterogeneous rank effects

In this section we investigate how the estimated rank effects varies with respect to features of the group – size, mean and heterogeneity in prior achievement – and two individual background characteristics - prior achievement and gender.

Tables B1 and B2 below report heterogeneous effects of rank on math score in the Kenyan experiment and on credits in the Dutch experiment, respectively, using the specifications in Column (2b) of Table 2 and Table 3.

First, in Column (1) we explore whether $RANK_{ig}$ differentially affects student performance depending on group peer ability composition, i.e., when we distinguish groups with $\overline{GPA}_{-i}$ above and below median. In both experiments, we do not detect heterogeneous effects by the level of mean peer ability.

Second, to investigate whether rank effects differ depending on the distance in GPA among groupmates, Column (2) reports heterogeneous effects by the $SD(GPA_{-i})$, distinguishing between groups with $SD(GPA_{-i})$ above and below median. Like Elsner and Isphording, 2017, we do not find any evidence of heterogeneous effects along this margin in either experiment.

Third, Column (3) shows that the effect of rank is not significantly different when we distinguish between students with GPA_i above and below the median, although the point estimate is larger for the former group than for the latter. And this holds true in both Table B1 and Table B2.

Fourth, in Column (4) 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 14 and 39 in the Kenyan experiment and between 32 and 45 in the Dutch one. The median size is 28 in the Kenyan experiment and 40 in the Dutch one. Results for both experiments show that the effect of rank does not vary with group size.

Finally, in Column (5) we assess the heterogeneity of the effect of rank with respect to gender. Results 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, though for the Dutch experiment only. This finding is in line with Murphy and Weinhardt, 2019, who also find heterogeneous effects by gender of primary school ability rank among English secondary school students, and is also consistent with the literature on heterogeneous gender attitudes towards competitiveness (see e.g. Gneezy et al., 2003).

Table B1. Heterogeneous rank effects on math score - Kenyan experiment

	(1)	(2)	(3)	(4)	(5)
$RANK_{ig}$ (a)	0.906***	0.799***	0.790***	0.857***	0.802***
	(0.229)	(0.211)	(0.218)	(0.222)	(0.240)
$RANK_{ig} \times\ \overline{GPA}_{-i}$ $above\ median\ (b)$	-0.106				
	(0.139)				
$RANK_{ig} \times\ SD(GPA)\ above\ median\ (b)$		0.183			
		(0.136)			
$RANK_{ig} \times GPA_i\ above\ median\ (b)$			0.064		
			(0.200)		
$RANK_{ig} \times GROUP\ SIZE_i\ above\ median\ (b)$				0.020	
				(0.133)	
$RANK_{ig} \times MALE_i\ (b)$					0.084
					(0.121)
$(a) + (b)$	0.801***	0.982***	0.855***	0.877***	0.886***
	(0.223)	(0.263)	(0.265)	(0.221)	(0.204)

Notes: the table reports heterogeneous effects of rank. Each column reports the results from a different OLS regression. Estimates based on the specification used in Table 2, Column (2b). Dependent variable is math score. 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: 2,188. Standard errors clustered by school are reported in parenthesis. Number of clusters: 48. *** p<0.01, ** p<0.05, * p<0.1.

Table B2. Heterogeneous rank effects on credits - Dutch experiment

	(1)	(2)	(3)	(4)	(5)
$RANK_{ig}$ (a)	0.573***	0.543***	0.417**	0.630***	0.331
	(0.171)	(0.196)	(0.205)	(0.189)	(0.208)
$RANK_{ig} \times\ \overline{GPA}_{-i}$ $above\ median\ (b)$	-0.025				
	(0.161)				
$RANK_{ig} \times\ SD(GPA)\ above\ median\ (b)$		0.044			
		(0.184)			
$RANK_{ig} \times GPA_i\ above\ median\ (b)$			0.113		
			(0.182)		
$RANK_{ig} \times GROUP\ SIZE_i\ above\ median\ (b)$				-0.214	
				(0.134)	
$RANK_{ig} \times MALE_i\ (b)$					0.282**
					(0.134)
$(a) + (b)$	0.548**	0.586***	0.530**	0.416**	0.613***
	(0.228)	(0.206)	(0.246)	(0.189)	(0.179)

Notes: the table reports heterogeneous effects of rank. Each column reports the results from a different OLS regression. Estimates based on the specification used in Table 3, Column (2b). Dependent variable is number of credits collected in the first year. 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.

2. Survey data

Booij et al., 2017, investigated the mechanisms behind peer effects using a survey administered to 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 content of the questionnaire changed slightly between years. The response rate was close to 70%, and Booij et al., 2017, show that survey response was unrelated to the ability composition of tutorial groups.

As done by Booij et al., 2017, we study the mechanisms behind both rank and peer effects on six index variables that summarize the 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 specification in Column (6b) of Table 4, are reported in Table B3 below.

On the one hand, all our estimates of rank effects are too imprecise to be significant, but reassuringly they have the expected sign. 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, and to benefit from learning from others in the tutorial group.

On the other hand, given the insignificance of rank effects, the coefficients related to the peer variables are in line with the ones estimated by Booij et al., 2017. The results on peer composition effects suggest that - at least as far as this is revealed by student perceptions - teachers are not very responsive to group ability composition, while there is evidence that low-ability students are less likely to feel involved in the class when surrounded by peers of higher ability, all the more so the more the group is heterogeneous. This evidence is also in line with findings by Feld and Zoelitz, 2017.

Table B3. Rank and peer effects on survey data on teaching style and learning environment - Dutch experiment

	(1) Too slow	(2) Too fast	(3) Stimulating	(4) Conducive	(5) Interactive	(6) Involved
$RANK_{ig}$	0.205	-0.269	0.025	-0.291	0.062	0.066
	(0.260)	(0.222)	(0.276)	(0.228)	(0.199)	(0.187)
$\overline{GPA}_{-i}$	0.060	-0.131	-0.015	-0.169	0.076	0.091
	(0.096)	(0.119)	(0.116)	(0.122)	(0.109)	(0.119)
$SD(GPA_{-i})$	0.023	-0.034	-0.269	0.101	-0.152	-0.230
	(0.141)	(0.125)	(0.206)	(0.150)	(0.143)	(0.145)
$\overline{GPA}_{-i} \times SD(GPA_{-i})$	-0.142	0.321	-0.370	0.150	0.392	0.299
	(0.312)	(0.333)	(0.346)	(0.323)	(0.270)	(0.287)
GPA_i	0.031	-0.024	-0.013	0.126	-0.007	0.053
	(0.131)	(0.082)	(0.106)	(0.081)	(0.082)	(0.075)
$GPA_i \times \overline{GPA}_{-i}$	0.005	0.023	-0.098	0.030	-0.025	-0.137**
	(0.081)	(0.054)	(0.076)	(0.057)	(0.069)	(0.058)
$GPA_i \times SD(GPA_{-i})$	0.233	-0.049	0.069	-0.048	0.145	0.075
	(0.166)	(0.158)	(0.174)	(0.153)	(0.120)	(0.121)
$GPA_i \times \overline{GPA}_{-i} \times SD(GPA_{-i})$	-0.375	0.087	-0.286	0.133	-0.103	-0.472**
	(0.323)	(0.257)	(0.355)	(0.307)	(0.240)	(0.192)

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. 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.