Ordinal Rank and the Structure of Ability Peer Effects *

Marco Bertoni [a, c], Roberto Nisticò [b, c]

[a] Department of Economics and Management "Marco Fanno", University of Padova and HEDG, York

[b] Department of Economics and Statistics, University of Naples "Federico II" and CSEF

[c] IZA, Institute of Labor Economics, Bonn

* Marco Bertoni (corresponding author): University of Padova, Department of Economics and Management "Marco Fanno" Via del Santo 33, 35123 Padova, Italy. Phone: +39 049 8274002. Email: marco.bertoni@unipd.it. Roberto Nisticò: University of Naples Federico II, Department of Economics and Statistics, and CSEF Via Cintia, Complesso Monte Sant'Angelo, 80126 Naples, Italy. Email: roberto.nistico@unina.it. We thank the editor (Sandip Sukhtankar), three anonymous referees, the audiences at the 2020 EALE/SOLE, at seminars at the University of Milan and Padova, and Massimo Anelli, Giorgio Brunello, Scott Carrell, Marta De Philippis, Ingo Isphording, Brian Jacob, Monica Langella, Sandra McNally, Konrad Mierendorff, Richard Murphy, Cheti Nicoletti, Matthias Parey, Caterina Pavese, Lorenzo Rocco, Uta Schönberg, Annalisa Scognamiglio, Vincenzo Scoppa, Olmo Silva, Michela Tincani and Felix Weinhardt for helpful comments. We are especially grateful to Adam Booij, Edwin Leuven and Hessel Oosterbeek for allowing us to use their original code and data, made available through the Review of Economic Studies data archive, and 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.

Abstract

Exposure to high-ability peers entails positive learning externalities, but it also decreases students' academic self-concept because of lower ordinal ability rank. We show that, as a result, the linear-in-means parameter identifies a composite (i.e., reduced form) effect. We illustrate the empirical relevance of this issue using data from two experiments that randomly assign students to groups. We find that the structural effect of mean peer ability estimated by a model that includes rank is much larger than the reduced form effect obtained when rank is omitted. This finding also holds in non-linear and heterogeneous peer effects models and helps clarify the mechanisms behind the effects of ability tracking policies.

Keywords: rank effects; peer effects; omitted variables bias; ability tracking.

JEL codes: I21; I24; J24.

1. Introduction

Establishing the existence and the magnitude of ability peer effects has been a central concern in education. To this aim, most researchers have adopted the linear-in-means model popularised by Manski, 1993, which relates individual achievement to the average pre-determined ability of peers. Its extensive application has led to the conclusion that ability peer effects on academic outcomes exist but are small in magnitude.[1]

Motivated by the possibility of obtaining gains from re-mixing students, that would not be feasible if peer effects were linear-in-means, researchers have investigated whether a non-linear and heterogeneous structure for peer effects might exist (Hoxby and Weingarth, 2005). There is by now a consensus that low- and high-ability peers in a class differentially impact on high- and low-ability individuals, and that the dispersion of peer ability also matters for achievement (Lyle, 2009, Lavy et al., 2012, Carrell et al., 2013, and Feld and Zoelitz, 2017). Even when these more comprehensive structures are adopted, the available empirical evidence still points to small effects of peer ability on learning outcomes.

Could these small effects be the result of omitted variables bias? The omission of a peer effect that consistently works in the opposite direction as mean peer ability would result in an underestimation of mean peer ability effects, as the linear-in-means parameter would capture a reduced-form combination of the two countervailing peer effects. In this paper, we argue that ordinal ability rank within groups could be this missing peer effect.

A growing literature in economics documents that ordinal ability rank has a positive causal impact on educational achievement (Elsner and Isphording, 2017; Cicala et al., 2017; Denning

[1] The review of the literature carried out by Sacerdote, 2014, suggests that other outcomes, such as misbehaviour or career choices, are instead more affected by peers' behaviour.

et al., 2018; Delaney and Devereux, 2021; Murphy and Weinhardt, 2020; Elsner et al., 2021).[2] There are several reasons why rank may matter for achievement. First, the so-called "big-fish-little-pond" hypothesis (Marsh and Parker, 1984) states that students of the same prior ability shall have lower (higher) academic self-concept and motivation when surrounded by classmates of high (low) average ability. Moreover, as reviewed by Delaney and Devereux, 2022, parents and teachers may respond to information about ordinal rank and encourage higher-ranked students, thereby also affecting achievement. Finally, variation in rank across subjects may alter students' perceptions about their comparative advantages and affect their specialization choices at later educational stages (Goulas et al., 2022).[3]

Importantly, for given individual ability, students have lower rank when surrounded by abler peers, and vice-versa. Therefore, exposure to high-ability peers implies a trade-off between the positive learning externality effect and the negative motivational effect implied by lower ordinal rank. Because of this negative correlation, the linear-in-means parameter identifies a composite (i.e., reduced form) effect of mean peer ability and rank.

While the mechanical correlation between rank and mean peer ability has been discussed by studies on rank effects as well as by studies using lotteries or GPA cut-offs to uncover the effects of getting into a better school (Cullen et al., 2006; Duflo et al., 2011; Pop-Eleches and Urquiola, 2013; Abdulkadiroğlu et al., 2014; Ribas et al., 2020), we are the first to unpack the reduced form effect into its structural components.

Importantly, the "reduced form" interpretation of the linear-in-means model holds not only when omitting rank, but also when other features of the distribution of peer ability are omitted.

---

[2] Comi et al., 2021, report rank effects on violence at school, Comi et al., 2019, investigate rank effects on conscientiousness, and Elsner and Isphording, 2018, estimate rank effects on risky behaviors.

[3] Tincani, 2017, discusses models where rank in achievement enters students' utility function (a rank "concern").

However, the mechanical negative correlation between rank and mean peer ability makes rank the prime candidate to illustrate how large this misconception can be and provides an intuitive reason why the existing literature has failed to empirically detect large effects of mean peer ability.

Our empirical analysis proceeds as follows. We use data from two experiments where students were randomly assigned to classes, thereby ruling out concerns related with endogenous student sorting. These were carried out in Kenyan primary schools (Duflo et al., 2011, 2015) and at the University of Amsterdam (Booij et al., 2017).[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. Next, we show that ordinal rank has a casual impact on achievement, that is large in magnitude and statistically significant. Finally, we compare the linear-in-means parameter when rank is excluded and included in the model, and show that, in the former case, this parameter is attenuated by the omission of rank. The magnitude of the bias is sizable: once rank is included in the linear-in-means specification the effect of mean peer ability moves from 0.345SD ($p>0.1$) to 0.433SD ($p<0.05$) in the Kenyan setup and from 0.048SD ($p>0.1$) to 0.148SD ($p<0.01$) in the Dutch one. We assess the significance of the omitted variables bias using the generalized Hausman test developed by Pei et al., 2019.

We also show that this identification problem holds true even when we depart from the linear-in-means model and: *i)* allow additional features of the distribution of peer ability – such as the share of high- or low-ability peers (Carrell et al., 2013) and the standard deviation of peer ability (Lyle, 2009) – to affect outcomes; *ii)* consider heterogeneous effects of peer

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

composition by own ability. These extensions underline that our main result does not strictly depend on the functional form adopted for the structure of peer effects.

A crucial point concerns the assumptions needed to identify in a single model the effects of rank, own ability, and mean peer ability, as these three effects cannot be jointly and non-parametrically identified. Following Murphy and Weinhardt, 2020, the identification of the rank effect can be achieved by controlling for individual ability directly and by accounting for group ability composition with group-fixed-effects. This design exploits between-groups variation across pupils with the same distance from average group ability but different distance from average group rank.

However, the impact of mean peer ability cannot be identified with groups fixed effects. Consequently, in our specification we control for the distribution of peer ability parametrically. Following the literature on non-linear and heterogeneous peer effects, in our preferred model we include the mean and standard deviation of peer composition as well as the interactions between these moments and own ability. The assumption underlying this specification is that higher order moments of the peer ability distribution are either irrelevant for achievement or uncorrelated with rank, so that their omission does not bias the effect of rank. We also verify the appropriateness of our functional form restriction by testing the equality of the rank effect estimated with the parametric controls for peer ability vis-à-vis with group fixed effects, giving up on identifying the peer ability parameters. Reassuringly, we fail to reject the null of this Hausman test.

In the final part of the paper, we illustrate how our main result contributes to explain the mechanisms behind ability tracking policies. We unpack the overall achievement effect of alternative grouping policies into a rank effect and a peer composition effect and show that rank and peer effects contribute in opposite directions to generate outcomes for low- and high-ability students. For instance, as we move from ability mixing to two-way tracking, students at the bottom of the ability distribution lose out in terms of average ability of peers but gain in terms of within-group ordinal ability rank. Therefore, if ordinal rank within groups enhances student achievement, tracking helps low-ability students, but it may harm the high achievers. This finding is relevant for the choice of the assignment policy in programs targeting different ability groups.

2. Experimental setups and data

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

The Kenyan experiment took place in May 2005 and involved 121 primary schools with one first-grade section only. The intervention provided resources to hire one additional teacher and create a second section. In 61 randomly selected "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. We focus on the remaining 60 "non-tracking schools", where students were unconditionally randomized into the two sections.[5]

Our outcome variables for this experiment are standardized test scores in math and literacy administered 18 months after the initial assignment, as well as their sum (total score). The observed baseline GPA_i is obtained from grades assigned by teachers in each school. To alleviate comparability issues and control for the stratified assignment of students to classes, we standardize this score by school and include school fixed effects in our regressions.

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

The experiment involved about 2,000 students starting the BA in economics and business at the University of Amsterdam in September 2009-11. In this program, roughly 60% of the total teaching time takes place into tutorial groups of about 40 students, whose composition is fixed

[5] Since assignment to classes within schools strictly depends on individual ability, there is no random assignment to classes within tracking schools and we do not use them in our analysis.

throughout the first year. The experiment allocated students to groups applying a conditional randomization aimed at achieving a wide support of ability composition. To this end, a different share of students from three end-of-secondary school GPA categories (GPA<6.5, 6.5≤GPA<7, GPA≥7) was assigned to each group.

In this setup, we measure student performance using three different outcome variables. First, our main outcome is the number of credits attained throughout the first year. The maximum number of credits attainable is 60 but only close to 20% of students reach this target and the average is close to 30, with a standard deviation of 23. Throughout the analysis, we standardize credits to have zero mean and unit standard deviation within cohort. Second, we consider average grade at the exams completed during the first year, that we also standardize within cohort. As on average students complete only 7 out of 13 exams that are scheduled for the first year, the validity of this otherwise commonly used performance measure is debatable in this context, because of self-selection issues. Finally, our third outcome is a "dropout" dummy, that in compliance with the University of Amsterdam's policies for enrolment in the second year of the program is equal to one if a student failed to complete at least 45 out of 60 credits during the first year, and zero otherwise. The choice of this variable is policy-motivated, as this is a core performance measure adopted by the University of Amsterdam. As in Booij et al., 2017, we mimic the conditional randomization embedded in the assignment mechanism by including in our regressions a set of randomization controls (a saturated set of own GPA category, advanced math, and cohort-dummies, interacted with application order).

2.3. Defining peer composition and rank

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 pre-determined GPA, computed leaving out individual i. We measure students' rank as their percentile rank in the baseline GPA distribution within groups. Since groups have different size, we follow Murphy and Weinhardt, 2020, and normalize 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. In case of ties, we assign the lower rank to all students. Still, results are robust when we use the average or the highest rank. Descriptive statistics for all variables used in each experiment are in Table A1 in Appendix A.

Figure 1 reports the variation in $RANK_{ig}$ and GPA_i in the two studies. Panel a reports the raw data and panel b reports the distribution of $RANK_{ig}$ by decile of GPA_i. These graphs illustrate that the two setups are complement in terms of the variation they generate. On the one hand, the unconditional randomization in the Kenyan study provides *very local* variation in rank for given ability: the peer group almost does not change, while students marginally change their ordinal rank. On the other hand, the conditional randomization in the Dutch experiment generates large variability in $RANK_{ig}$ *throughout* the distribution of GPA_i.

A potential limitation about both setups concerns students' information about their rank. Yu, 2020, as well as Megalokonomou and Zhangg, 2022, show evidence of a strong and positive correlation between objective and self-perceived ability rank in Chinese middle schools. We lack the data on self-perceived ability rank needed to show this correlation in our setups. In this sense, we view our estimates as the reduced form effects that would be obtained by instrumenting perceived with objective rank.

3. Empirical methodology

3.1. Identification

The “standard” linear-in-means peer effect model estimated in the literature relates academic achievement y_{ig} of individual i assigned to group g to $\overline{GPA}_{-i}$, GPA_i and a vector of additional covariates X_{ig} as follows:

$$y_{ig} = \gamma_0^S + \gamma_1^S \, \overline{GPA}_{-i} + \gamma_2^S GPA_i + X'_{ig}\gamma_3^{\text{S}} + v_{ig} \tag{1}$$

Identification of the effect of $\overline{GPA}_{-i}$ is achieved by exploiting the (conditional) random assignment of students to groups, that eliminates concerns about endogenous sorting of students. This comes with an implicit assumption on the structure of peer effects, i.e., that no other feature of the distribution of peers’ baseline score that is excluded from Eq. (1) generates omitted variables bias (i.e. it correlates with both y_{ig} and $\overline{GPA}_{-i}$). In the framework introduced by Manski, 1993, this boils down to assuming the absence of exogenous peer effects.[6]

While several other studies (Lyle, 2009, Duflo et al., 2011, Carrell et al., 2013, Feld and Zoelitz, 2017, Booij et al., 2017) have allowed for non-linear and heterogeneous effects of peers’ ability composition by own ability, no previous study has considered ordinal ability rank as a potential source of omitted variables bias. The key contribution of this paper is to quantify the implications of its omission for the identification of ability peer effects. On top of the theoretical reasons why rank should matter for education production, described in the introduction, its stark negative correlation with mean peer ability makes it the prime candidate to neatly show how the omission of relevant features of the structure of peer effects affects the identification of the effects of the included features.

For ease of exposition, we focus on the linear-in-means model in Eq. (1). However, in Section 5 we show that similar considerations also hold in models that allow for non-linear and

[6] Our model regresses students’ outcomes on the average baseline scores of peers and on ordinal rank computed on the basis of the baseline score. Since the baseline score predates group assignment, our estimates are immune to the reflection problem discussed by Manski, 1993, as well as to the exclusion bias presented by Caeyers and Fafchamps, 2016.

heterogeneous peer effects. Our "enriched" model, including both $\overline{GPA}_{-i}$ and $RANK_{ig}$, is specified as follows:

$$y_{ig} = \gamma_0^E + \gamma_1^E \overline{GPA}_{-i} + \beta_1 \, RANK_{ig} + \gamma_2^E GPA_i + X'_{ig}\gamma_3^E + \varepsilon_{ig} \qquad (2)$$

Hence, the error term in Eq. (1) can be rewritten as $v_{ig} = \beta_1 \, RANK_{ig} + \varepsilon_{ig}$. If we specify the linear regression that relates $RANK_{ig}$ to $\overline{GPA}_{-i}$ as

$$RANK_{ig} = \delta_0 + \delta_1 \overline{GPA}_{-i} + \delta_2 GPA_i + X'_{ig}\delta_3 + \xi_{ig} \, , \qquad (3)$$

then the bias on γ_1^S in Eq. (1) due to the omission of $RANK_{ig}$ is equal to $\beta_1\delta_1$ (see Angrist and Pischke, 2009, and Pei et al., 2019). The evidence from both our experiments suggests that β_1 is positive and δ_1 is negative. Therefore, the coefficient γ_1^S estimated in the standard linear-in-means model for peer effects suffers from attenuation bias.

This argument rests upon the assumption that we can identify the rank effect. To this aim, we propose the following thought experiment. Suppose that two students with the same baseline score – Marco and Roberto – are randomly assigned to classes with the same distribution of peers' ability. Then, a random perturbation to the baseline score of Marco's closest peer in the within-class ability distribution – Anna – is applied, small enough to leave the ability distribution in the two classes still comparable, but at the same time large enough to make Anna and Marco exchange rank. The rank effect is identified by comparing the outcomes of Marco and Roberto in the two classes.

Besides clarifying what is the causal parameter we are after, this highlights that $RANK_{ig}$ is a function of individuals' own ability and of the distribution of ability in their group, raising two important considerations about the identification of rank effects.

First, it is impossible to change the rank of one subject in a group without altering the rank of at least another subject in the group. This suggests a violation of the stable unit-treatment value assumption, SUTVA, prescribing that "the observation on one unit [is] unaffected by the particular assignment of treatments to the other units" (Cox, 1958).[7] We interpret SUTVA as requiring that potential achievements of Marco and Roberto depend on their own rank, but are independent of the rank of their groupmates. As a result, we can only give a causal interpretation to the difference in rank between Marco and Roberto that is obtained by randomly perturbing Anna's ability under the assumption that the resulting change in Anna's rank has no direct impact on Marco's achievement.[8]

Second, there would be no variation left to identify the effect of $RANK_{ig}$ if we controlled for all interactions between dummies for the position of student i in the GPA distribution and dummies for the share of pupils in student i's group that belong to each position of the GPA distribution. Therefore, disentangling the effect of $RANK_{ig}$, GPA_i, and features of the distribution of GPA_{-i} requires a model that flexibly controls for (higher-order) distributional features, interactions and nonlinearities while leaving enough variation to identify the effect of $RANK_{ig}$.

We achieve this by controlling for the mean and standard deviation of group peer ability. The validity of our specification rests upon the assumption that the omission of higher-order

[7] SUTVA also requires that the treatment, a change in rank from r to r', is well-defined. However, the same rank could be achieved under different underlying distributions of peer ability. We abstract from this issue but notice that similar concerns could be raised in other common problems in labour economics. For example, the effect of being first vs. second-born could depend on family size.

[8] This holds even in studies that estimate the effect of rank within given peer groups on outcomes measured after exposure to new peers at later educational stages (Murphy and Weinhardt, 2020). Denning et al., 2021, propose an alternative thought experiment and define the treatment as the allocation of Marco and Roberto to two classes sampled from the same population, so that in expectation they have the same distribution of prior achievement, but differ in realization due to sampling variability. The two ideal experiments require comparable assumptions in terms of both the information about own and peers' ability and the exclusion restriction on other peers' rank changes.

moments of the ability distribution does not lead to a biased rank effect, as they are either irrelevant for achievement or uncorrelated with rank.

An alternative to this specification would be to use group fixed effects. In Appendix B, where we present all our robustness tests, we use a Hausman test to verify that, in both setups, the models with and without group fixed effects deliver statistically indistinguishable estimates of the rank effect. This strongly corroborates the parametric functional form choices for the peer ability distribution adopted in our main specification.

Irrespective of whether we control for the mean and standard deviation of group peer ability or we use groups fixed effects, following Booij et al., 2017, we also include interactions of features of the distribution of peer ability and individual ability. Their inclusion allows for heterogeneous effects of the ability distribution that could otherwise be captured by the rank effects. This approach is also championed by Denning et al., 2021, in their analysis of the long-run effects of rank using data from Texas.

How do our assumptions differ with respect to other studies on peer and rank effects? On the one hand, the literature seeking to identify potentially non-linear and heterogeneous peer effects reaches identification of interactions between own ability and peer ability by omitting rank from the structure of peer effects (for instance, this is stated explicitly in Duflo et al., 2011, see footnote 19, page 1763). On the other hand, the literature on rank effects includes group fixed effects to account for the non-random sorting of students into groups and to eliminate confounders that vary at the group level, such as mean peer ability. In addition, given that features of the group ability distribution cannot arbitrarily interact with all possible features of individuals, both literatures impose different functional form restrictions (see also the discussion in Denning et al., 2021, section 3).[9]

[9] An exception in this sense is Hoxby and Weingarth, 2005, who use a very large datasets and adopt an admirably flexible specification for non-linear and heterogeneous peer effects.

We sit in the middle. First, unlike the literature on peer effects, we allow rank to impact achievement. Second, differently from the rank literature, we do not rely on group fixed effects to control for sorting because students are randomly assigned to groups. Third, we rely on parametric assumptions that have been commonly adopted in both literatures (see e.g. Carrell et al., 2013; Booij et al., 2017; Denning et al, 2021) to model the ability composition across groups and control for non-linear and heterogeneous peer effects. The robustness of our results across specifications lends empirical support to the validity of our approach.

3.2. Estimation

We take Eq. (2) to the data by estimating with Ordinary Least Squares (OLS) the following education production function:

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

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, while ε_{ig} is an error term. We cluster standard errors at the school level in the Kenyan experiment and at the tutorial group level in the Dutch one, respectively. The empirical specifications of Eqs. (1) and (3) are similar, but in Eq. (1) $RANK_{ig}$ is omitted from the model and in Eq. (3) it is the dependent variable.

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 amounts to jointly estimating the linear-in-means models without and with $RANK_{ig}$ – Eqs. (1) and (2) above – using seemingly unrelated estimation and testing the following null hypothesis: $H0: \gamma_S - \gamma_E = 0$.

In Section 5 we also show the relevance of this omitted variables bias in models that allow for more elaborated structures with heterogeneous and non-linear effects of peer composition.

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

We present our main results in Table 1. Columns (1)-(3) are for the total score of the Kenyan experiment, while Columns (4)-(6) are for the number credits in the Dutch one. Columns (1) and (4) report the estimates of Eq. (3), describing the relationship between ordinal rank and peers' mean ability, conditional on own ability. Columns (2) and (5) report the estimates of the standard linear-in-means model described in Eq. (1), while Columns (3) and (6) report the estimates of Eq. (2), which include also $RANK_{ig}$.

First, results in Columns (1) and (4) confirm that, for given individual ability, the higher is peers' mean ability, the lower is ordinal rank. Therefore, parameter δ_1 in Eq. (3) is negative. We also notice that the R-squared for this model is higher in the Kenyan case than in the Dutch one. In the latter setup, the conditional random assignment generates larger variability in rank for given individual ability and peers' mean ability than in the former one.

Second, results in Columns (3) and (6) show that rank has a positive effect on achievement, and hence that parameter β_1 in Eq. (2) is positive. In the Kenyan experiment, moving from the bottom to the top of the within-group ability distribution increases the total score by slightly more than 40 percent of a standard deviation.[10] The rank effect is positive and large (0.3 SD) also in the Dutch experiment.

The comparison of the coefficients of $\overline{GPA}_{-i}$ in Columns (2) and (3) as well as (5) and (6) confirms that the omission of $RANK_{ig}$ generates a severe attenuation of the effect of peer

[10] The size of this effect is comparable to the one estimated by Cicala et al., 2017, on the same data. It implies that a 1 SD increase in rank – equal to 0.3 – increases the total score by 0.12 SD. Similarly, for the Dutch experiment, a 1 SD increase in rank increases credits by 0.12 SD.

composition. As we proved in Section 3.1, the bias is exactly equal to the product between the effect of $RANK_{ig}$ on the outcome – reported in Columns (3) and (6) – and the effect of $\overline{GPA}_{-i}$ on $RANK_{ig}$ – reported in Columns (1) and (4). The p-values of the Hausman test for coefficient comparison across Columns (2) and (3) as well as (5) and (6), reported in brackets, confirm the statistical relevance of the bias, that is also large in magnitude. For the Kenyan experiment, the effect of $\overline{GPA}_{-i}$ moves from 0.345SD (p>0.1) to 0.433SD (p<0.05), while in the Dutch setup it passes from 0.048SD (p>0.1) to 0.148SD (p<0.01).[11]

In Table 2 we replicate Table 1 using the other available outcomes. In the Kenyan experiment, we detect no significant rank effect for the literacy score, while rank has a very large impact on the math score (0.846SD). Carneiro et al., 2022, find similar asymmetric effects of rank on math and literacy using data from Ecuador.[12] Consistently, the omission of rank is empirically relevant only for the math score: it is this result that is driving the effects on the total score reported in Table 1. Therefore, we will hereafter focus only on the math score.

A few potential mechanisms could explain the absence of rank effects on literacy. First, as argued for instance by Fryer, 2017, Charity, 2004, and Rickford, 1999, it may be hard to alter literacy scores for pupils who speak at home a different language than the one used in schools. As shown by Berthet, 2020, this issue is especially relevant in Kenya, where many households do not use Kiswahili and English – the official languages used in schools – at home. The Kenyan government recently addressed this matter by promoting the development of mother tongue learning materials in schools (Nyariki, 2020). Second, the literature in educational

[11] The effect of $\overline{GPA}_{-i}$ in Column (2) is significant at the 5 percent level with an estimated standard error of 0.150 if we cluster standard errors by class instead of by school (see Duflo et al., 2011). The significance of the Hausman test is unchanged.

[12] Duflo et al., 2015, find that the introduction of extra teachers and of school-based management training in Kenya did not significantly improve literacy scores while it affected math scores.

psychology stresses that academic self-concept may be subject-specific (Marsch et al., 1988) and there is some evidence that the correlation between self-concept and academic achievement is larger for math than for literacy (Booth and Gerard, 2011; Susperreguy et al., 2018; Álvarez and Szücs, 2022). We measure prior achievement with GPA, and in the data the correlation between GPA and the math score is larger than the one with GPA and the literacy score (0.5 vs. 0.4). As a result, ordinal rank based on GPA may be more related to rank in math than rank in literacy. Third, it may be that competitive pupils react to rank more than non-competitive ones, and prior studies have shown that competitiveness is correlated with achievement in math (Buser et al., 2014, Buser et al., 2017; Yagasaki and Nakamuro, 2018).

For the Dutch experiment, we find that higher rank implies a significantly lower dropout probability. The rank effect on average grade is positive, but imprecisely estimated. In both cases the attenuation bias in the linear-in-means parameter due to the omission of $RANK_{ig}$ is sizeable in magnitude, although only significant for the dropout probability.

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. Indeed, the peer composition effects that we obtain when rank is omitted are wholly comparable to other studies in the literature. For instance, the effect on math scores in the Kenyan study in Column (2) is comparable in magnitude to the effect obtained by Imberman et al., 2012, on data from primary school students in Louisiana (0.33SD), as well as to the one estimated by Hoxby, 2000, on primary school students in Texas (0.3-0.5SD). Similarly, the effect on credits in the Dutch study in Column (5) is comparable to the effect obtained by Jackson, 2013, on the number of tests passed by high school students in Trinidad (0.03-0.09SD). In both our experimental setups, however, the magnitude of the peer composition effects increases substantially once rank is accounted for.

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

The standard linear-in-means model discussed so far is very simple and intuitive, but also restrictive. In this Section we extend our results on the bias in the identification of ability peer effects due to the omission of rank in more general models that allow for non-linear and heterogeneous peer effects. We follow two complementary modelling approaches that have gained popularity in the literature.

Our first approach follows Booij et al., 2017,'s original analysis of the Dutch experiment. We consider models that include $SD(GPA_{-i})$, the interaction term $\overline{GPA}_{-i} \times SD(GPA_{-i})$, and the interaction terms between the peer composition variables and GPA_i. This richer model allows the dispersion of peer ability to affect outcomes (see Lyle, 2009), considers the possibility of rank concerns in the utility function (Tincani, 2017), and permits that the mean and the SD of peer GPA are not perfect substitutes in shaping student performance. It also accommodates the evidence that ability peer effects are heterogeneous by individual ability.

The estimates in Table 3 show that our main result on the attenuation bias of the effect of mean peer ability holds across both experiments even when we adopt this richer peer effects structure.

In addition, we do not find evidence of heterogeneous and non-linear peer effects in the Kenyan experiment, confirming that the linear-in-means model approximates well the data generating process in this setup. On the contrary, the larger variation in peer composition generated in the Dutch experiment makes this richer structure better than the linear-in-means model at explaining the data. In this setup, the presence of significant interactions between $\overline{GPA}_{-i}$, $SD(GPA_{-i})$ and GPA_i makes it such that 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 Appendix A reports the heterogeneous effects of $\overline{GPA}_{-i}$ for different levels of $SD(GPA_{-i})$ and GPA_i when rank is omitted and included. As the Figure shows, the omission of $RANK_{ig}$ changes the effect of $\overline{GPA}_{-i}$ only for students in groups at the bottom of the $SD(GPA_{-i})$ distribution, that is, in homogeneous groups.[13]

We also note that in both setups the effect of $RANK_{ig}$ is very stable when additional peer composition variables are flexibly included in the model, alleviating the second identification concern discussed in Section 3.1. If anything, in the Dutch setup this effect almost doubles in magnitude. This happens because the relationship between $SD(GPA_{-i})$ and both $RANK_{ig}$ and y_{ig} is not linear in GPA_i, and failure to account for this heterogeneity results in misspecification bias.[14]

Our second approach follows Carrell et al., 2013, and models non-linear and heterogeneous peer effects in a less parametric fashion by using the shares of low (bottom 25%) and high (top 25%) ability students in the group, and by interacting these measures by tertiles of individual ability. Results in Table 4 show that, especially in the Kenyan setup, students are harmed by low (vs. median) achieving peers. Moreover, in both setups the effects of low-achieving peers are significantly larger in magnitude when $RANK_{ig}$ is included in the model.

Overall, the evidence provided in this Section suggests that our results on omitted variable bias do not depend on the functional form adopted to model non-linear and heterogeneous peer effects.

---

[13] The p-value of the Hausman test reported in Table 3, column 6, for the difference in the coefficients on the interaction term $\overline{GPA}_{-i} \times SD(GPA_{-i})$ when rank is excluded (column 5) and included in the model (column 6) is equal to 0.009, confirming the statistical significance of this difference.

[14] Consider a normal distribution for GPA_i, and apply a mean-preserving spread to a group's ability. Compared to subjects in another identical group, whose ability was not altered, treated subjects with GPA_i below median would gain ranks, and those above median would lose. While the non-linear relationship between $SD(GPA_{-i})$ and $RANK_{ig}$ is mechanical, and is present also in the Kenyan experiment, its consequences for omitted variable bias are only relevant if the effect of $SD(GPA_{-i})$ on y_{ig} is nonlinear, a finding that holds only in the Dutch experiment.

6. Robustness tests

Our empirical strategy rests on several assumptions, that we test at length in Appendix B. We provide evidence on the robustness of our results along the following dimensions:

i) modelling group ability composition parametrically vs. using group fixed effects;

ii) assessing the presence of heterogeneous teaching styles by student rank;

iii) changing the functional form for the relationship between GPA_i and the outcomes;

iv) adopting a non-linear specification for the effect of $RANK_{ig}$ (see Gill et al., 2018);

v) addressing multiplicative measurement error in the baseline score used to compute $RANK_{ig}$, as done by Murphy and Weinhardt, 2020;

vi) adopting alternative definitions of rank and different sample selection criteria.

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.[15] As pointed out by Hoxby and Weingarth, 2005, the linear-in-means model has the property that all allocations of peers are equally beneficial in aggregate, as in a zero-sum game. However, there are no trivial implications for group assignments when one departs from a linear-in-means peer effects model to employ a more flexible specification.

[15] This analysis is not feasible in the Kenyan setup, as we can only estimate the peer effects model exploiting the local random variation in ability composition generated by the unconditional randomization to classes within non-tracking schools. We could carry out additional policy simulations (and unpack their effects into a rank and a peer composition effect) only if we were willing to extrapolate the structure of peer effects estimated within non-tracking schools beyond the support of variation in peer ability composition observed in this sample. However, this would require an assumption of homogeneity and linearity of peer effects that cannot be verified empirically beyond the observed support of ability composition.

We analyse the implications of ability tracking for achievement using the non-linear and heterogeneous peer effects specification augmented with rank reported in Table 3, Column (6). In particular, this specification allows $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$ to be imperfect substitutes in affecting student achievement by including an interaction term between the two. As a result, different grouping strategies will reflect different trade-offs between $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$. The interactions of $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$ with GPA_i generate further trade-offs.

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.

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 3, Column (6).[16] We do so for the whole population and by tertile of GPA_i.

Table 5 shows the estimated tracking effects on credits. Results in Columns (1)-(3) 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 (1), (4), (7), (10), while rank and peer effects are shown in Columns (2), (5), (8), (11) and (3), (6), (9), (12), respectively.

Results in Columns (1), (4), (7), (10) are very similar, though larger in magnitude to those reported in Columns (1) to (4) of Table 5 in Booij et al., 2017. This is expected, given the omitted variable bias discussed above, and confirms their two main findings:

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

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

The replication of this analysis with our rank-augmented production function highlights that while all types of tracking benefit students on average, different types of students will be differently affected by different types of tracking due to the effect of rank.

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 (as it would be for $\overline{GPA}_{-i}$ in a linear-in-means model). However, the estimates in Columns (2), (5), (8), (11)

[16] The (conditional) average treatment effects of tracking are computed as $(\bar{x}_{track} - \bar{x}_{mix})\hat{\beta}$ while the standard errors as $\sqrt{(\bar{x}_{track} - \bar{x}_{mix})' V(\hat{\beta})(\bar{x}_{track} - \bar{x}_{mix})}$ where $\bar{x}_{track}$ and $\bar{x}_{mix}$ 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 3, column (6).

and (3), (6), (9), (12) 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 is related to the assumptions about the functional form for the education production function. This is especially relevant for the estimation of counterfactual policy effects by ability. For instance, while all peer composition variables enter in the model with a flexible functional form, we assume that rank effects are linear and homogenous. In Table A2 and A3 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. For both experiments we fail to detect significant heterogeneous rank effects. 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, 2020.

8. Conclusions and Discussion

This paper contributes to the literature on the role of peers in education by showing that exposure to high-ability peers implies a trade-off between a positive learning externality effect and a negative motivational effect implied by lower ordinal rank. As a result, the linear-in-means parameter identifies a composite (i.e., reduced form) effect. Although this "reduced form" interpretation of the linear-in-means model applies even when other features of the distribution of peer ability are omitted, we specifically focus on rank to show how large this misconception can be because of the negative mechanical correlation between rank and mean peer ability.

We use data from two randomized experiments carried out in Kenyan primary schools (Duflo et al., 2011) and at the University of Amsterdam (Booij et al., 2017) to assess the extent to which omitting rank from the structure of peer effects impacts the estimates of the effect of mean peer ability in the standard linear-in-means model. We find that the omission of rank can attenuate the estimated effect of mean peer ability by more than 100 percent, and that this bias is statistically significant. We also confirm our main result in more general models that allow for non-linear and heterogeneous effects of peers' ability.

By stressing the importance of ordinal rank, our findings warn that the exclusion restriction behind the linear-in-means model can be easily violated and reinforce the consensus on the multifaceted nature of the structure of peer effects, thereby remarking that the linear-in-means parameter should be viewed as a "reduced-form" measure of the ability peer effect.

The consistency of our results across two very diverse institutional settings (Kenya and the Netherlands) and educational levels (primary and tertiary education), as well as under different assumptions on the structure of peer effects, is reassuring in terms of external validity.[17]

Furthermore, the joint determination of rank and peer composition at the time of group assignment is embedded in the mechanics of all peer effects studies. Therefore, our findings have implications beyond the economics of education (see Cornelissen, 2016, for a review of the evidence on peer effects in the workplace).

[17] The division of a cohort of students in two classes carried out in each school involved in the Kenyan experiment implies that classes are small by Kenyan standard – down from 90 pupils to around 40 – and at the same time larger than the average class sizes observed in developed countries. This can be an issue for the external validity of our findings in this setup.

Our results also provide key insights to understand the effectiveness of ability tracking policies. We use the large support of ability configurations in the Dutch data and a flexible education production function to unpack the overall achievement effect of a battery of grouping scenarios into two components: a rank effect and a peer composition effect. Our analysis indicates that rank and peer effects affect the outcomes of low- and high-ability students in opposite directions.

Although this analysis is based on some functional form assumptions, the findings have 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 ability tracking. Under this assignment, 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 including 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. They suggest that the big-fish-little-pond 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" schools. 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

- Abdulkadiroğlu, A., Angrist, J., & Pathak, P. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica, 82*(1), 137-196.
- Álvarez, C., & Szücs, D. (2022). The relation of school achievement with self-esteem and bullying in Chilean children. *Current Psychology*, 1-15.
- Angrist, J. D., & Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton university press.
- Berthet, V. J. (2020). *Reading achievement in Kenya: the language factor*. Doctoral Thesis in Economics. University of Cape Town. http://hdl.handle.net/11427/32475
- 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.
- Booth, M. Z., & Gerard, J. M. (2011). Self-esteem and academic achievement: a comparative study of adolescent students in England and the United States. *Compare: A Journal of Comparative and International Education,* 41(5), 629-648.
- Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *Quarterly Journal of Economics*, 129(3), 1409-1447.
- Buser, T., Peter, N., & Wolter, S. C. (2017). Gender, competitiveness, and study choices in high school: Evidence from Switzerland. *American Economic Review*, 107(5), 125-30.
- Caeyers, B., & Fafchamps, M. (2016). *Exclusion Bias in the Estimation of Peer Effects*. NBER Working Paper 22565.
- Carneiro, P., Aguayo, Y. C., Salvati, F., & Schady, N. (2022). *The effect of classroom rank on learning throughout elementary school: experimental evidence from Ecuador.* Mimeo, University College London.

- 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.
- Charity, A. H., Scarborough, H. S., & Griffin, D. M. (2004). Familiarity with school English in African American children and its relation to early reading achievement. *Child Development*, 75(5), 1340-1356.
- 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.
- Comi, S., Origo, F., Pagani, L., & Tonello, M. (2021). Last and furious: Relative position and violence at school. Forthcoming, *Journal of Economic Behaviour & Organization.*
- Comi, S., Origo, F., & Pagani, L. (2019). The Effect of School Rank on Personality Traits. Forthcoming, *Journal of Human Resources.*
- 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.
- Cox, D. R. (1958). *Planning of Experiments.* John Wiley.
- Cullen, J. B., Jacob, B. A., & Levitt, S. (2006). The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica*, 74(5), 1191-1230.
- Delaney, J. M., & Devereux, P. J. (2021). High school rank in math and English and the gender gap in STEM. *Labour Economics*, 69, 101969.
- Delaney, J. & Devereux, P. J. (2022). *Rank Effects in Education. What Do We Know So Far?* IZA Discussion Paper No. 15128.
- Denning, J. T., Murphy, R., & Weinhardt, F. (2021). Class rank and long-run outcomes. Forthcoming, *Review of Economics and Statistics.*

- 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.
- Duflo, E., Dupas, P., & Kremer, M. (2015). School Governance, Pupil-Teacher Ratios, and Teacher Incentives: Experimental Evidence from Kenyan Primary Schools. *Journal of Public Economics*, 123, 92-110.
- 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. (2021). Achievement Rank Affects Performance and Major Choice in College. *Economic Journal*, ueab034.
- 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.
- Fryer, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In: *Handbook of economic field experiments* (Vol. 2, pp. 95-322). North-Holland.
- Gill, D., Kissová, Z., Lee, J., & Prowse, V. (2019). First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Management Science*, 65(2), 494-507.
- Goulas, S., Griselda, S., & Megalokonomou, R. (2022). Comparative advantage and gender gap in STEM. *Journal of Human Resources*, 0320-10781R2.
- Hoxby, C. (2000). *Peer effects in the classroom: learning from gender and race variation*. NBER Working Paper 7867.

- Hoxby, C., & Weingarth, G. (2005). *Taking race out of the equation: school reassignment and the structure of peer effects.* Working Paper, Harvard University, Cambridge: MA.
- Imberman S. A., Kugler A.D., & Sacerdote B.I. (2012). Katrina's children: evidence on the structure of peer effects from hurricane evacuees. *American Economic Review,* 102:2048–82.
- Jackson, K. (2013). Can higher-achieving peers explain the benefits to attending selective schools? Evidence from Trinidad and Tobago. *Journal of Public Economics*, 108:63–77.
- 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.
- Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of educational psychology*, 80(3), 366.
- Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, 47(1), 213–231.
- Megalokonomou, R., & Zhang, Y. (2022). *How Good Am I? Effects and Mechanisms Behind Salient Ranks.* IZA Discussion Paper N. 15604.
- Murphy, R., & Weinhardt, F. (2020). Top of the Class: The Importance of Ordinal Rank. Forthcoming, *Review of Economic Studies.*

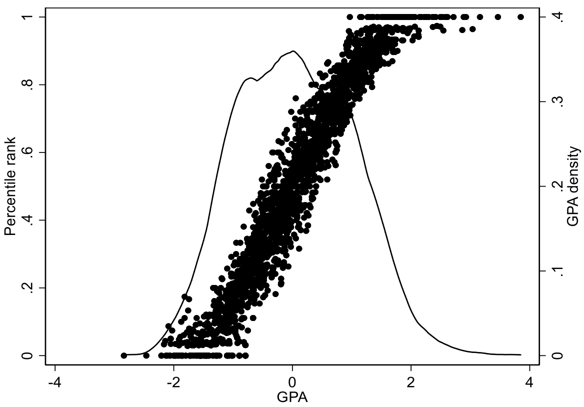
- Nyariki, L., 2020. Kenya: The Key Role of National Languages in Education and Training. Available online at https://www.globalpartnership.org/blog/kenya-key-role-national-languages-education-and-training.
- Pei, Z., Pischke, J. S., & Schwandt, H. (2019). Poorly Measured Confounders Are More Useful on the Left Than on the Right. *Journal of Business & Economic Statistics*, 37(2), 205-216.
- Pop-Eleches, C., & Urquiola, M. (2013). Going to a better school: Effects and behavioral responses. *American Economic Review*, *103*(4), 1289-1324.
- Ribas, R. P., Sampaio, B., & Trevisan, G. (2020). Short-and long-term effects of class assignment: Evidence from a flagship university in Brazil. *Labour Economics*, *64*, 101835.
- Rickford, J. R. (1999). *African American vernacular English: Features, evolution, educational implications.* Wiley-Blackwell.
- Sacerdote, B. (2014). Experimental and Quasi-experimental Analysis of Peer Effects: Two Steps Forward? *Annual Review of Economics*, 6(1), 253-272.
- Susperreguy, M. I., Davis-Kean, P. E., Duckworth, K., & Chen, M. (2018). Self-concept predicts academic achievement across levels of the achievement distribution: Domain specificity for math and reading. *Child Development*, 89(6), 2196-2214.
- Tincani, M. (2017). *Heterogeneous Peer Effects and Rank Concerns: Theory and Evidence.* CESifo Working Paper No. 6331.
- Yagasaki, M., & Nakamuro, M. (2018). *Competitiveness, Risk Attitudes, and the Gender Gap in Math Achievement.* RIETI Discussion Paper Series 18-E-066, University of Tokyo.
- Yu, H. (2020). Am I the Big Fish? The Effect of Ordinal Rank on Student Academic Performance in Middle School. *Journal of Economic Behavior & Organization*, 176 (2020), 18-41.

Figures and Tables

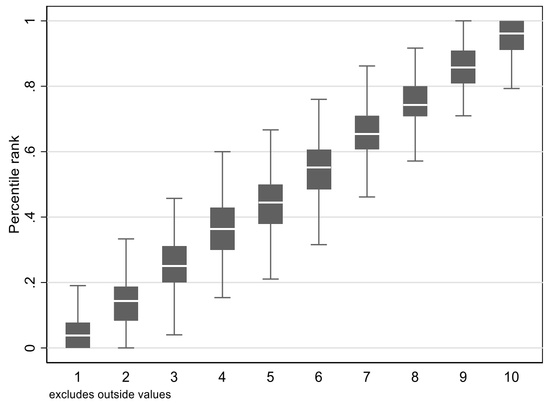
Figure 1. Variation in rank by level of GPA

Kenyan experiment

a. Raw data

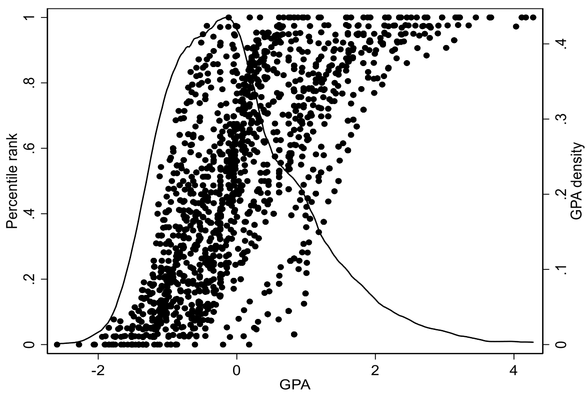


b. Boxplot by decile of GPA

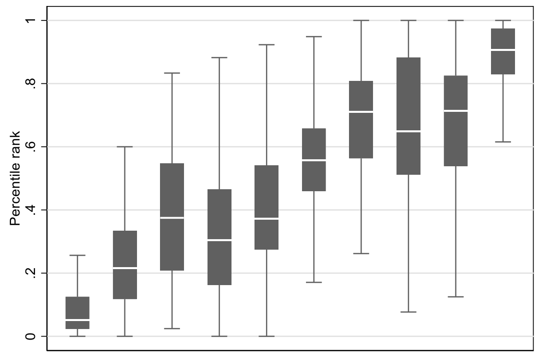


Dutch experiment

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.

Table 1. Rank and peer effects in the linear-in-means models - Kenyan and Dutch experiment

	(1)	(2)	(3)	(4)	(5)	(6)
Experimental setup		Kenyan			Dutch	
Dependent variable:	$RANK_{ig}$	Total score	Total score	$RANK_{ig}$	Credits	Credits
$RANK_{ig}$			0.403**			0.351**
			(0.185)			(0.140)
$\overline{GPA}_{-i}$	-0.219***	0.345	0.433**	-0.287***	0.048	0.148***
	(0.028)	(0.211)	(0.208)	(0.014)	(0.041)	(0.052)
			[0.051]			[0.012]
GPA_i	0.291***	0.507***	0.390***	0.352***	0.314***	0.191***
	(0.002)	(0.032)	(0.064)	(0.026)	(0.066)	(0.050)
			[0.027]			[0.015]
R-squared	0.926	0.357	0.362	0.801	0.266	0.269
Observations	2,188	2,188	2,188	1,876	1,876	1,876
Clusters	48	48	48	48	48	48

Notes: Each column reports the results from a different OLS regression. Dependent variable reported at the top of each column. For the Kenyan experiment, school fixed effects and background controls (gender, age, being assigned to the contract teacher) are included in all specifications. For the Dutch experiment, randomization and background controls (gender, age categorized in tertiles within cohort, previous attendance of a professional college before university enrolment) are included in all specifications. Standard errors clustered by school in the Kenyan experiment and by tutorial group in the Dutch experiment are reported in parenthesis. Hausman test p-values are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 2. Rank and peer effects in the linear-in-means models – Kenyan and Dutch experiment – alternative outcomes.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Experimental setup	Kenyan				Dutch			
Dependent variable:	Literacy score		Math score		Dropout		Avg. Grade	
$RANK_{ig}$		-0.056		0.846***		-0.267**		0.175
		(0.192)		(0.216)		(0.102)		(0.158)
$\overline{GPA}_{-i}$	0.291	0.279	0.324	0.509**	-0.042	-0.142***	0.124**	0.189***
	(0.184)	(0.189)	(0.226)	(0.214)	(0.027)	(0.047)	(0.047)	(0.069)
		[0.766]		[0.001]		[0.008]		[0.261]
GPA_i	0.413***	0.430***	0.496***	0.250***	-0.168***	-0.070*	0.489***	0.426***
	(0.037)	(0.072)	(0.026)	(0.067)	(0.018)	(0.036)	(0.035)	(0.068)
		[0.769]		[<0.001]		[0.007]		[0.264]
Observations	2,189	2,189	2,188	2,188	1,876	1,876	1,753	1,753
Clusters	48	48	48	48	48	48	48	48

Notes: see Table 1. The number of observations changes because of missing values in the math score for one student in the Kenyan setup, and in the average grade for students who did not complete at least one first-year exam in the Dutch setup.

Table 3. Rank and peer effects in non-linear and heterogeneous models - Kenyan and Dutch experiment

	(1)	(2)	(3)	(4)	(5)	(6)
Experimental setup		Kenyan			Dutch	
Dependent variable	$RANK_{ig}$	Math score	Math score	$RANK_{ig}$	Credits	Credits
$RANK_{ig}$			0.921***			0.559***
			(0.238)			(0.178)
$\overline{GPA}_{-i}$	-0.239***	0.431	0.651**	-0.376***	0.148***	0.358***
	(0.032)	(0.273)	(0.262)	(0.013)	(0.052)	(0.086)
			[0.001]			[0.002]
$SD(GPA_{-i})$	-0.026	-0.241	-0.218	0.027	-0.185**	-0.200**
	(0.031)	(0.252)	(0.248)	(0.022)	(0.082)	(0.079)
			[0.397]			[0.235]
$\overline{GPA}_{-i} \times SD(GPA_{-i})$	-2.436*	13.816	16.059*	0.377***	0.343*	0.132
	(1.353)	(9.504)	(9.349)	(0.051)	(0.190)	(0.183)
			[0.095]			[0.009]
GPA_i	0.292***	0.493***	0.223***	0.365***	0.350***	0.145**
	(0.002)	(0.026)	(0.072)	(0.013)	(0.035)	(0.065)
			[<0.001]			[0.002]
$GPA_i \times \overline{GPA}_{-i}$	0.044**	-0.056	-0.096	0.025**	-0.117***	-0.131***
	(0.022)	(0.171)	(0.176)	(0.011)	(0.042)	(0.040)
			[0.096]			[0.069]
$GPA_i \times SD(GPA_{-i})$	-0.268***	0.329	0.575*	-0.356***	0.104	0.303***
	(0.038)	(0.316)	(0.340)	(0.027)	(0.075)	(0.091)
			[0.001]			[0.003]
$GPA_i \times \overline{GPA}_{-i} \times SD(GPA_{-i})$	0.405	-1.243	-1.616	0.160***	-0.287**	-0.376***
	(0.363)	(2.397)	(2.234)	(0.045)	(0.138)	(0.135)
			[0.314]			[0.034]
R-squared	0.930	0.358	0.364	0.864	0.271	0.274
Observations	2,188	2,188	2,188	1,876	1,876	1,876
Clusters	48	48	48	48	48	48

Notes: see Table 1 for details. $\overline{GPA}_{-i}$ and $SD(GPA_{-i})$ are re-centred to have zero mean in the estimation sample to ease the interpretation of interaction terms.

Table 4. Rank and peer effects from alternative heterogeneous-peer-effects models – Kenyan and Dutch experiment

	(1)	(2)	(3)	(4)	(5)	(6)
Experimental setup		Kenyan			Dutch	
Dependent variable	$RANK_{ig}$	Math score	Math score	$RANK_{ig}$	Credits	Credits
$RANK_{ig}$			0.641**			0.368*
			(0.286)			(0.191)
Fraction of Low-GPA peers × Low GPA_i	0.557***	-1.368**	-1.725***	0.583***	-0.074	-0.288
	(0.076)	(0.565)	(0.559)	(0.025)	(0.188)	(0.218)
			[0.018]			[0.054]
Fraction of Low-GPA peers × Middle GPA_i	0.465***	-1.009*	-1.308**	0.715***	-0.151	-0.413*
	(0.067)	(0.558)	(0.573)	(0.029)	(0.222)	(0.235)
			[0.033]			[0.058]
Fraction of Low-GPA peers × High GPA_i	-0.057	-0.352	-0.316	-0.088**	-0.256	-0.224
	(0.054)	(0.712)	(0.724)	(0.043)	(0.281)	(0.277)
			[0.377]			[0.160]
Fraction of High-GPA peers × Low GPA_i	-0.020	-1.082	-1.070	-0.038	0.122	0.136
	(0.075)	(0.777)	(0.786)	(0.027)	(0.251)	(0.252)
			[0.796]			[0.228]
Fraction of High-GPA peers × Middle GPA_i	-0.456***	-0.243	0.050	-0.450***	0.054	0.220
	(0.101)	(0.850)	(0.843)	(0.028)	(0.264)	(0.308)
			[0.051]			[0.053]
Fraction of High-GPA peers × High GPA_i	-0.695***	-0.741	-0.295	-0.665***	-0.062	0.182
	(0.057)	(0.743)	(0.810)	(0.037)	(0.182)	(0.201)
			[0.025]			[0.056]
GPA_i	0.213***	0.370***	0.233***	0.284***	0.263***	0.158***
	(0.009)	(0.043)	(0.075)	(0.026)	(0.042)	(0.052)
			[0.020]			[0.058]
Low GPA_i (1st Tertile)	-0.240***	0.136	0.289	-0.183***	-0.167	-0.099
	(0.037)	(0.325)	(0.311)	(0.023)	(0.150)	(0.166)
			[0.032]			[0.052]
High GPA_i (3rd Tertile)	0.279***	0.149	-0.030	0.314***	0.109	-0.007
	(0.035)	(0.326)	(0.353)	(0.023)	(0.132)	(0.139)
			[0.032]			[0.057]
R-squared	0.947	0.364	0.366	0.869	0.269	0.270
Observations	2,188	2,188	2,188	1,876	1,876	1,876
Clusters	48	48	48	48	48	48

Notes: see Table 1 for details.

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

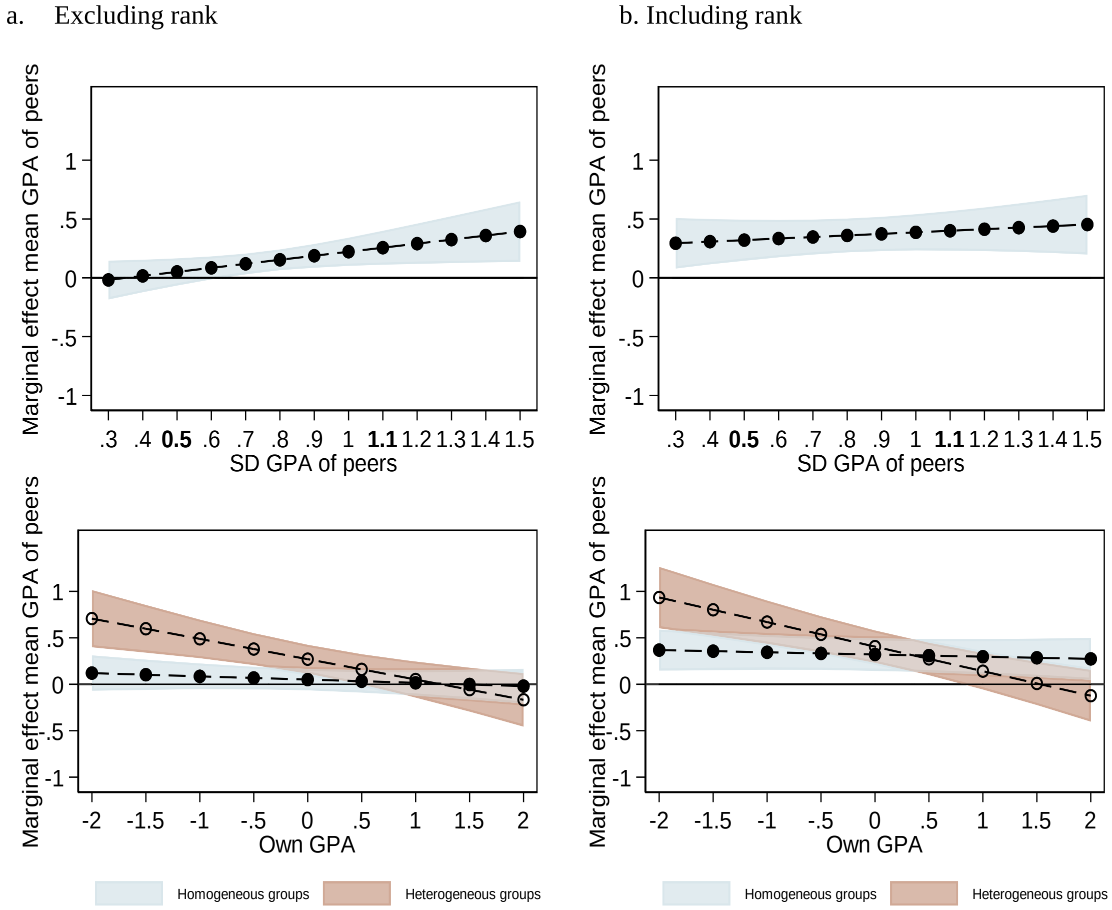
					Student GPA category								
		ATE			L(B)			M			H(A)		
		Total	Rank	Peer	Total	Rank	Peer	Total	Rank	Peer	Total	Rank	Peer
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Two-way tracking	{B},{A}	0.103***	-0.002***	0.105***	0.157***	0.138***	0.019				0.050	-0.141***	0.191***
		(0.028)	(0.000)	(0.028)	(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 3, Column (6). 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.

Appendix – For Online Publication Only.

A. Additional Figures and Tables

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



Notes: The Figure reports the marginal effects of $\overline{GPA}_{-i}$ by $SD(GPA_{-i})$ and by GPA_i when including $RANK_{ig}$ in the model, respectively. Marginal effects based on the estimates from Table 3, Column (5) and (6), respectively.

Table A1. 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
Share of male students	0.477	0.500
Age at test	9.188	1.469
Share of students 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
Share of students who 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
Share of male students	0.733	0.443
Share of student whose age is in youngest third of the distribution	0.333	0.472
Share of student whose age is in oldest third of the distribution	0.329	0.470
Share of students with a professional college degree	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 A2. 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 1, Column (3). 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 A3. 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 (6). 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.

B. Robustness tests

This Appendix presents a comprehensive set of robustness test to our main results. Results in Section 5 in the main text show that there is evidence of non-linear and heterogeneous peer effects only in the Dutch setup, while the linear-in means model is flexible enough to represent the Kenyan data. Hence, we present all our robustness tests using the linear-in-means model for the Kenyan experiment and the flexible specification of Column (6) in Table 3 for the Dutch one. We present each robustness test in a dedicated subsection.

B1. Group fixed effects.

Identification of the rank effect in our main specification relies on the assumption that the chosen parametric specification of the peer ability distribution is correctly specified. As mentioned in Section 3.1, an alternative would be to use group fixed effects, that non-parametrically control for group-specific characteristics – e.g., size, teacher effects, atmosphere – as well as for mean peer ability. The inclusion of group fixed effects also allows to control for any rank-preserving class-specific shock (such as teaching style or leniency). Murphy and Weinhardt, 2020, highlight that this specification 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. As illustrated by Elsner et al., 2021, in this specification rank effects are identified by comparing students with the same GPA across *all* groups after controlling for mean differences across groups. By subtracting the group mean from each variable, the within-groups estimator does not change the *shape* of the ability distribution while at the same time eliminating differences in mean ability across groups. As remarked in the recent review of the literature on rank effects by Delaney and Devereux, 2022, the assumption embedded in the specification with group fixed effects is that differences in the error term across combinations of ability and groups can be summarized by an additive group effect and an additive ability effect. These effects eliminate potential confounders that vary by group. Following Booij et al., 2017, with respect to the existing literature we enrich the specification with group fixed effects for the Dutch setup by including interactions of features of the distribution of peer ability and individual ability. This allows for heterogeneous effects of the ability distribution that could be captured by the rank effects if omitted.[18]

Column (1) in Table B1 shows that, in both experiments, the effect of $RANK_{ig}$ is very stable when we include group fixed effects.

We can verify that the models with and without group fixed effects deliver statistically indistinguishable estimates for the rank effect using a Hausman test. We fail to reject the null with a p-value of 0.1 for the Dutch experiment (the rank effects are equal to 0.503 vs. 0.559) and 0.65 for the Kenyan one (rank effects: 0.817 vs. 0.846). This result strongly corroborates the parametric functional form choices we adopted in our main specification.

B2. Heterogeneous teaching styles by student rank

A remaining identification concern is that there may be within-class unobservables that correlate with rank and impact on achievement. The leading one could be a heterogeneous effect of teachers with respect to students' rank (homogeneous teacher effects would be captured by group fixed effects). In such case, the rank effect would be an artifact of uncontrolled heterogeneous teaching styles.

The Dutch experiment provides us with survey data on students' perceptions on teaching. The survey was carried out three months after the beginning of the academic year. 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.

[18] This approach is also championed by Denning et al., 2021, in their analysis of the long-run effects of rank using data from Texas.

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 (6) of Table 3, are reported in Table B2 below.

On the one hand, the results confirm the absence of significant rank effects on the teaching perceptions of students. 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.

Regarding the Kenyan experiment, Duflo et al., 2011, report evidence of a positive effect of mean peer ability at the bottom and the top of the baseline ability distribution. According to their theoretical model, this is consistent with the presence of direct effects of peers and indirect effects that go through teachers' effort and ability targeting choices. However, the absence of data on students' perceptions about the targeting of teaching prevents us from providing direct evidence on the relevance of heterogeneous teaching styles for the identification of rank effects in this setup. To the extent that teacher influence is stronger for younger student than for older ones, the different age of the students involved in the Kenyan and the Dutch studies could contribute to explain the difference in findings about teacher effects.

B3. Different functional forms for the relationship between GPA_i and the outcomes

Spurious rank effects could also emerge if there are specification errors, for instance in the choice of the functional form for the relationship between GPA_i and the outcomes. In Columns (2) to (7) of Table B1 we verify that the linear functional form 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 remains positive and large.[19] For the Dutch experiment, however, it falls below the minimum effect that we can significantly detect

[19] This pattern of results is consistent with the ones reported, for instance, by Murphy and Weinhardt, 2020, that estimate rank effects in large administrative data using very flexible specifications for previous achievement (high-order polynomials, dummies for deciles or ventiles, non-parametric controls for each level). As in our case, the rank effects they estimate decrease in magnitude when extremely flexible specifications are adopted. Still, given that they use much larger datasets than us, their estimated effects remain highly significant.

with the sample size we have at hand when we use the fourth order polynomial.[20] Eventually, the experiment was not designed to estimate rank effects, and we have to make do with limited power for this purpose. Similar conclusions also hold when it comes to the attenuation bias in the effect of $\overline{GPA}_{-i}$ on achievement.

B4. Non-linear specifications for the effect of $RANK_{ig}$

The linear specification for the effect of $RANK_{ig}$ could also be restrictive (see Gill et al., 2018). In Figure B1 we report estimates from a non-linear specification with distinct effects for each ventile of the distribution of $RANK_{ig}$. In both setups we find limited evidence of non-linear effects. Reassuringly, the p-values of the Hausman tests for the attenuation bias in the effect of $\overline{GPA}_{-i}$ on achievement are equal to 0.002 in the Kenyan case and to 0.005 in the Dutch experiment, corroborating our main result.

B5. Multiplicative measurement error in the baseline score used to compute $RANK_{ig}$

Murphy and Weinhardt, 2020, point out that multiplicative measurement error in the baseline score used to compute $RANK_{ig}$ could generate a spurious rank effect. To deal with this potential issue, we follow their approach and use the percentile (instead of the level) of the baseline score both as a control and as the basis to compute $RANK_{ig}$. To ease the comparison with Murphy and Weinhardt, 2020, we transform the outcome variables in percentiles as well.

Results are in Table B1, Columns (8) and (9). We find a rank effect of 14.6 percentiles (0.50SD) for the Kenyan experiment (Panel A) and 10.9 percentiles (0.39SD) for the Dutch one (Panel B). If we further scale these effects by the standard deviation of $RANK_{ig}$, we obtain an effect of 0.15SD or 4.37 percentiles in the Kenyan experiment and 0.12SD or 3.25 percentiles in the

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

Dutch experiment. For the Dutch experiment the rank effect is significant at the 10 percent level, while it is not significant in the Kenyan one, where the correlation between $RANK_{ig}$ and the percentiles (instead of the level) of GPA_i is very high. This generates a problem of multicollinearity, as witnessed by the large standard error on GPA_i in Column (9), that is 5 times larger than in Column (8). Needless to say, a larger sample would help to enhance precision.

Still, our main result is not undermined by these concerns. In both experiments, the effect of $\overline{GPA}_{-i}$ substantially increases as we include $RANK_{ig}$. For the Dutch experiment this difference is significant at the 10 percent level, while it is not significant for the Kenyan experiment. We see this as a limitation in terms of statistical validity rather than an identification issue.

B7. Alternative definitions of rank and sample selection

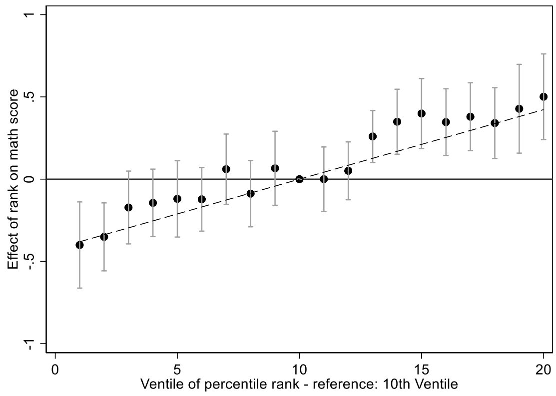
We present additional robustness tests in Table B3. We replicate the specification of Table 1, Columns (2) and (3) in Panel A (Kenyan experiment), and the specification of Table 3, Columns (5) and (6) in Panel B (Dutch experiment).

As explained in Section 2.3, 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, 2020. However, an equally plausible assumption could have been to assign the average or the higher rank. Columns (1) – (4) of Table B3 show that our results are robust even in those cases.

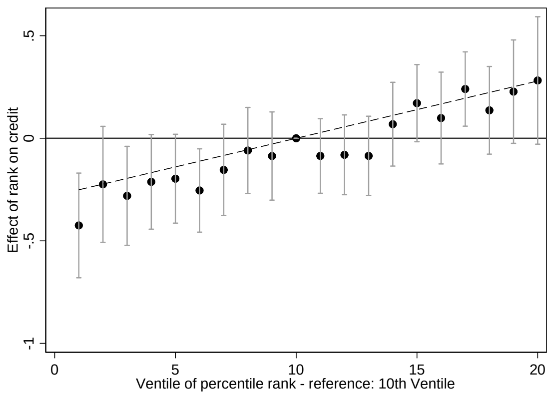
In addition, Figure B1 highlights that, for the Dutch experiment, the $RANK_{ig}$ effect seems larger at the very bottom of the support of its distribution. Speculatively, one reason why the effect at the very bottom is most pronounced could capture a stigma effect of being “the worst”. In Columns (5) – (8) of Table B3 we show the results that we obtain when we drop students in the top and bottom 1 percent of the overall distribution of GPA_i and $RANK_{ig}$, for whom rank could be more salient. Reassuringly, our main result holds in both cases.

Figure B1. 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 1, column (3) for the Kenyan experiment and in Table 3, Column (6) 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 B1. Main robustness tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Using Group FE	Quadratic GPA_i		Cubic GPA_i		Quartic GPA_i		$RANK_{ig}$ and $\overline{GPA}_{-i}$ based on the percentiles of GPA_i	
Panel A. Kenyan experiment									
Dep. Var.	Math score	Math score		Math score		Math score		Math score percentile	
$RANK_{ig}$	0.817***		0.817***		0.737***		0.696***		14.556
	(0.225)		(0.225)		(0.258)		(0.258)		(16.831)
$\overline{GPA}_{-i}$		0.320	0.501**	0.331	0.487**	0.333	0.479**	0.046	0.207
		(0.225)	(0.213)	(0.227)	(0.212)	(0.227)	(0.211)	(0.204)	(0.260)
			[0.003]		[0.015]		[0.021]		[0.323]
GPA_i	0.249***	0.502***	0.261***	0.556***	0.300***	0.579***	0.326***	0.514***	0.352**
	(0.070)	(0.027)	(0.071)	(0.035)	(0.094)	(0.042)	(0.100)	(0.028)	(0.168)
Observations	2,188	2,188	2,188	2,188	2,188	2,188	2,188	2,188	2,188
Clusters	48	48	48	48	48	48	48	48	48
Panel B. Dutch experiment									
Dep. Var.	Credits	Credits		Credits		Credits		Credits percentile by cohort	
$RANK_{ig}$	0.503***		0.501**		0.439*		0.328		10.953*
	(0.175)		(0.227)		(0.222)		(0.226)		(6.404)
$\overline{GPA}_{-i}$		0.145***	0.336***	0.153***	0.318***	0.148***	0.272***	0.191**	0.314**
		(0.050)	(0.097)	(0.050)	(0.094)	(0.050)	(0.094)	(0.092)	(0.125)
			[0.025]		[0.047]		[0.140]		[0.072]
GPA_i	0.149**	0.393***	0.179*	0.451***	0.253**	0.536***	0.374***	0.322***	0.171*
	(0.068)	(0.043)	(0.100)	(0.051)	(0.103)	(0.071)	(0.129)	(0.039)	(0.099)
			[0.026]		[0.047]		[0.142]		[0.084]
Observations	1,876	1,876	1,876	1,876	1,876	1,876	1,876	1,876	1,876
Clusters	48	48	48	48	48	48	48	48	48

Notes: we adopt the specification of Table 1, Columns (2) and (3) in Panel A, and the specification of Table 3, Columns (5) and (6) in Panel B. In Columns (8) and (9) we use the percentiles of GPA_i both as a control and to compute rank and peer composition, while the dependent variable is also percentilized.

Table B2. 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)
Observations	1,342	1,342	1,342	1,342	1,342	1,342
Clusters	47	47	47	47	47	47

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. Standard errors clustered by tutorial group are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table B3. Alternative definitions of rank and sample selection

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Using alternative definitions of $RANK_{ig}$ to break ties				Excluding obs. with GPA_i in top and bottom 1%		Excluding obs. with $RANK_{ig}$ in top and bottom 1%	
	Mean Rank	Mean Rank	Max Rank	Max Rank				
Panel A. Kenyan experiment								
$RANK$		0.785***		0.678***		0.746***		0.601**
		(0.218)		(0.220)		(0.254)		(0.241)
$\overline{GPA}_{-i}$	0.324	0.493**	0.324	0.468**	0.298	0.457**	0.371	0.505**
	(0.226)	(0.212)	(0.226)	(0.213)	(0.231)	(0.218)	(0.247)	(0.234)
		[0.003]		[0.009]		[0.013]		[0.026]
GPA_i	0.496***	0.269***	0.496***	0.302***	0.511***	0.283***	0.509***	0.329***
	(0.026)	(0.067)	(0.026)	(0.068)	(0.030)	(0.085)	(0.029)	(0.078)
		[<0.001]		[0.002]		[0.003]		[0.011]
Observations	2,188	2,188	2,188	2,188	2,144	2,144	2,017	2,017
Clusters	48	48	48	48	48	48	48	48
Panel B. Dutch experiment								
$RANK_{ig}$		0.494**		0.402**		0.485**		0.418*
		(0.188)		(0.192)		(0.201)		(0.219)
$\overline{GPA}_{-i}$	0.148***	0.335***	0.148***	0.301***	0.156***	0.340***	0.106	0.285**
	(0.052)	(0.090)	(0.052)	(0.092)	(0.053)	(0.096)	(0.068)	(0.123)
		[0.009]		[0.036]		[0.016]		[0.052]
GPA_i	0.350***	0.171**	0.350***	0.206***	0.388***	0.193**	0.380***	0.202**
	(0.035)	(0.068)	(0.035)	(0.070)	(0.040)	(0.082)	(0.046)	(0.100)
		[0.008]		[0.033]		[0.015]		[0.050]
Observations	1,876	1,876	1,876	1,876	1,838	1,838	1,772	1,772
Clusters	48	48	48	48	48	48	48	48

Notes: we adopt the specification of Table 1, Columns (2) and (3) in Panel A, and the specification of Table 3, Columns (5) and (6) in Panel B.