



Gutenberg School of Management and Economics
& Research Unit “Interdisciplinary Public Policy”

Discussion Paper Series

*Pushing Through or Slacking Off?
Heterogeneity in the Reaction to Rank
Feedback*

Florian Hett, Felix Schmidt

March 2018

Discussion paper number 1806

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
wiwi.uni-mainz.de

Contact Details:

Florian Hett
Chair of Digital Economics
Johannes Gutenberg-University,
Jakob-Welder-Weg 9, 55128 Mainz
Germany
florian.hett@uni-mainz.de

Felix Schmidt
Chair of Public and Behavioral Economics
Johannes Gutenberg-University,
Jakob-Welder-Weg 9, 55128 Mainz
Germany
felix.schmidt@uni-mainz.de

Pushing Through or Slacking Off? Heterogeneity in the Reaction to Rank Feedback

Florian Hett¹
Felix Schmidt²

March 30, 2018

Abstract

This paper studies heterogeneity in the reaction to rank feedback. In a laboratory experiment, individuals take part in a series of dynamic real-effort contests with intermediate feedback. To solve the identification problem in estimating the causal effect of rank feedback on subsequent effort provision we implement a random multiplier in the first round of each contest. The realization of this multiplier then serves as a valid instrument for rank feedback. While rank feedback has a robust effect on subsequent effort provision on average, an explicit analysis of between-subject heterogeneity reveals that a substantial fraction of participants in fact react entirely opposite than the aggregated results indicate. We further show that this heterogeneity has consequences for overall outcomes, thereby arguing that heterogeneous sensitivities to rank feedback could have implications for the design of various policies in education and organizations.

Acknowledgments: The authors are thankful to participants of the Workshop on Microeconomics in Lueneburg for constructive feedback. We gratefully acknowledge financial support from the Research Center SAFE, funded by the State of Hessen research initiative LOEWE.

Keywords: heterogeneity, competitiveness; contest; rank feedback, relative performance evaluation

JEL-codes: C91

¹Goethe University Frankfurt, hett@econ.uni-frankfurt.de

²University of Mainz, Department of Economics, felix.schmidt@uni-mainz.de

1 Introduction

In many economically relevant domains it is relative rather than absolute performance that ultimately determines outcomes. Examples range from the educational system to work environments: In the professional sphere, performance rankings are often explicit elements of compensation schemes. They determine the winner among competing suppliers of products and services (think of league tables in investment banking or track records for fund managers) or implicitly matter through career concerns and promotions. In education, many systems rely on ‘curving’ as an element of grading schemes, recommendations of teachers or professors are essentially relative assessments, and the restricted nature of prestigious scholarships or program spots implies an allocation based on relative performance.

A notable feature of many such environments is that during the respective activities, individuals receive intermediate feedback on their relative performance, often in the form of rank information. This triggers the question if and how individuals react to such information: Do they condition their choice of effort on the current position in such rankings? If so, how? Do people at the top try harder or rather slack off? And do those at the bottom of the ranking give up or try everything to avoid ending up in their current position? Given this broad range of conceivable ways in which rank information could affect subsequent performance, it is not surprising that the existing theoretical and empirical literature as a whole provides only inconclusive guidance on which effects actually to expect. In this paper we therefore refrain from the idea of finding one general behavioral pattern for all individuals. Rather, we conjecture that given the huge variety of possible mechanisms, it is more likely that people do not homogeneously react to performance feedback but differ in their individual rank sensitivity in both magnitude as well as structure.

Consequently, the aim of this paper is to explicitly study the extent, structure, and consequences of heterogeneity in the reaction to rank feedback by answering the following research questions: Do people systematically differ in their reaction to rank feedback? If so, are there specific (measurable) characteristics associated with these differences? And how are differences in rank sensitivity linked to overall competitive outcomes? These answers have potential implications for policy design. In many institutions, competitive mechanisms explicitly or implicitly provide incentives for effort provision. If individuals differ in their reaction to feedback on their relative performance, this implies non-trivial allocative and distributional effects, as initial differences might either be amplified or dampened for different individuals. Variations in the competitive design then potentially shift individual outcomes not only with respect to obvious characteristics like ability, but also with respect to feedback sensitivity. Evidence regarding heterogeneity in the reaction to rank feedback might therefore help in assessing a broad variety of

institutional rules like relative grading in schools or peer evaluations in organizations.

Providing such evidence is challenging for at least three reasons: First, to assess differences in the reaction to rank feedback between individuals, one needs to credibly estimate these individual reactions. This requires substantial variation of rank at the individual level, i.e. within-subject. Second, there are obvious endogeneity problems in identifying the causal effect of rank feedback on subsequent performance: The rank a participant holds is likely correlated with ability, thereby potentially confounding the pure effect of relative performance feedback. Third, to investigate potential associations with individual characteristics one needs reliable corresponding measurement protocols actually capturing these individual traits. Our research design addresses these challenges using a laboratory experiment featuring the slider task (Gill and Prowse 2011). In the experiment, participants repeatedly compete in a series of real effort dynamic contests with intermediate feedback, thereby generating substantial amounts of within-subject data, allowing to estimate reactions to rank feedback at the individual level. As a source of exogenous variation, we randomly assign point multipliers in the first round of each contest, holding material incentives constant in expectation. This design feature allows using the realized random multiplier as a fully exogenous instrument for the rank of participants after the first round. To investigate potential associations of individual rank sensitivity with other characteristics, we employ established behavioral measures and survey items for competitiveness, social value orientation, and risk preferences.

Our results show that rank feedback significantly affects effort provision. In more detail, we find that the information to currently occupy a better rank on average leads to more effort and vice versa. Furthermore, we show that feedback does not only affect effort within a particular contest but its effects even spill over to subsequent contests with new competitors. While these results appear reasonable and are statistically robust, once we exploit our data in more detail we find that these aggregated patterns mask a profound degree of heterogeneity: Not only does the strength of the effect of rank feedback vary substantially between subjects. Rather, even the direction of the effect of rank feedback varies considerably. We also find that these differences in rank sensitivity matter for the overall performance in our experiment: Individuals that react positively to a better rank (and therefore negatively to a worse rank) win fewer and lose more contests and achieve overall worse outcomes. Finally, we find very low correlations of rank sensitivity and other common individual characteristics, i.e. competitiveness, ability, risk aversion, and other-regarding preferences. This result gives rise to the idea that rank sensitivity is an idiosyncratic individual trait rather than a mere artifact driven by other behavioral mechanisms.

Our paper contributes to the broad literature on behavior in contests and the role of feedback therein. For extensive corresponding reviews we refer to the survey articles by Dechenaux et al.

(2015) and Sheremeta (2013, forthcoming) as well as the detailed literature review in Gill et al. (forthcoming). Here, we limit ourselves to discussing the papers most closely related to our own: Dijk et al. (2014) investigate how relative performance-based incentive schemes and status concerns for higher rank influence portfolio choice in laboratory experiments. They find that both under-performers and over-performers adapt their portfolios to their current relative performance. In a field experiment featuring a sales tournament with retail stores, Delfgaauw et al. (2014) find that on average, tournament incentives do not lead to higher performance: They show that stores falling far behind do not respond to feedback, but performance increases once stores get within the reach of winning the bonus. In line with this, Casas-Arce and Martínez-Jerez (2009) also analyze a contest among retailers and show that effort decreases in ‘leading’ and ‘trailing distance’. Blanes i Vidal and Nossol (2011) find that if workers are given private feedback on their individual rank of pay as well as on their productivity this increases performance. Finally, Gill et al. (forthcoming) document a non-monotonous effect of rank feedback on subsequent effort provision, as their experimental subjects behave as ‘first-place loving’ and ‘last-place loathing’. While all of these papers analyze the effects of relative feedback on subsequent performance in one way or another, none of the above explicitly analyzes how these effects differ systematically between contestants.

Our paper is structured as follows. The next section describes our research design and lays out the detailed steps of our experiment and how they address our research objectives. Section three presents our empirical results. Finally, section four discusses our results, provides an outlook on promising next steps for future research, and concludes.

2 Research Design

Our objective is (i) measuring heterogeneity in the reaction to rank feedback and (ii) testing for its association with other individual characteristics. Doing the first requires sufficient within-subject data, featuring a substantial exogenous component in the variation of rank feedback provided to each individual. Regarding the second objective, one needs to rely on validated and tested measurement protocols. We address these requirements using a laboratory experiment where subjects participate in a series of contests within small groups. Each contest consists of several rounds including intermediate feedback. To induce exogenous variation in rank, we implement a random multiplier, which keeps material incentives fixed and only varies rank. To elicit potential factors associated with feedback sensitivity, we rely on established behavioral measures and survey items for competitiveness, social value orientation, and risk preferences.

The general course of our experiment is hence as follows: The experiment consists of three stages and a post experimental questionnaire. In the first stage we measure competitiveness us-

ing an adapted design based on Niederle and Vesterlund (2007). Stage two then represents the main part of the experiment, where subjects participate in five repetitions of a real effort contest with four rounds each, featuring the slider task by Gill and Prowse (2011). In stage three we elicit individuals' social value orientation (SVO) based on Murphy et al. (2011) and in the final questionnaire we include survey measures for risk preferences based on Falk et al. (2016).

The experiment took place in December 2017 and March 2018 in the Mainz Behavioral and Experimental Laboratory (MABELLA). The experiment was programmed using 'z-Tree' by Fischbacher (2007). Subjects were students from various disciplines at the Johannes Gutenberg-University Mainz, recruited via ORSEE (Greiner 2004). We conducted eight sessions stratified by gender with a total of 184 participants. An experimental session lasted about 90 minutes. We informed participants that one of the three stages will be randomly determined for payoff which led to average earnings of 14.14 Euros (about US \$16.65 at the time, including a show-up fee of 6 Euro).

2.1 Measuring Competitiveness

In the first stage of the experiment we measure attitudes towards competition using an adapted version of the established 'competitiveness' protocol introduced by Niederle and Vesterlund (2007). The main idea is to capture individual taste for competitive environments, that is settings with the explicit potential to win and lose. To do so, participants can choose between different compensation schemes in a real-effort task. The key difference between the payment options are whether remuneration is based on a piece rate or a tournament.

As we aim for measuring a general attitude towards competition, we choose to build this stage on a different real-effort task than the one being used in the subsequent contest stage. Thereby, we circumvent the potential problem that a correlation between our measure of competitiveness and subsequent competitive behavior simply reflects a certain task specificity. The task we use is therefore adding up sets of three two-digit numbers, as it has been previously used for example in Bartling et al. (2009) and Dohmen and Falk (2011). There are three periods: Period 1 serves as a trial period in which participants have 40 seconds to get familiar with the task and is hence not monetarily incentivized. In period 2 participants work for two minutes on the task under a piece rate which gives them 30 cents per correct answer. Performance in this task can hence be used as a proxy for general task ability.

Before performing the task again for two minutes in period three, participants now choose the compensation scheme for this upcoming round for ten different potential scenarios, out of which one is later randomly chosen and implemented for payment. The basic alternatives are being paid

according to a piece rate scheme including a fixed payment or being paid based on the outcome of a tournament. The conditions for this tournament are constant across all ten scenarios: Participants receive 12 Euro if they solve more exercises than a randomly chosen other participant did in Period 2 and nothing if they solve less. In case of a tie, the payoff is 6 Euro. To rule out that choices are driven by anticipated selection of opponents, participants can be matched with all other participants independent of their chosen payment scheme. In the piece rate scheme, the actual piece rate is also constant across all ten conditions, which is 30 cents for every correctly adjusted slider. However, the fixed payment systematically increases by increments of 60 cents, starting from 1.80 Euro in scenario 1.¹ Hence, the tournament scheme becomes less attractive in relative terms as the fixed payment increases. Accordingly, we interpret a later switching point from tournament to piece rate compensation as our revealed preference measure of individual competitiveness.

2.2 Dynamic Contests and Rank Feedback

Stage 2 represents the main part of the experiment and consists of a series of dynamic contests. To induce exogenous rank variation, we introduce a random point multiplier in the first round of each contest.

Dynamic Contests

The contest stage employs the slider task by Gill and Prowse (2011). In the slider task, participants receive a screen with a series of sliders in the range of 1 to 100. They then need to adjust as many sliders as possible to the exact value of 50.²

This stage is again divided into three sub-stages. First, participants get a trial period of 40 seconds in which they can get familiar with the task. Accordingly, this sub-stage is not incentivized. After that, participants work on the slider task again, now for a period of two minutes. Further, they now receive 12 cents for each correctly adjusted slider. Finally, they take part in a series of five dynamic contests, each against three other randomly selected participants. Each of these five contests consists of four rounds with one minute each.

After every round participants see an information screen that shows the points and ranks of all participants of their current contest.³ Points are calculated after every round by multiplying the number of correctly adjusted sliders with a randomly determined factor that is described below and added to the points already earned in previous rounds of the same contest. After the

¹See Figure A.1 in the Appendix.

²See Figure A.2 in the Appendix.

³See Figure A.3 in the Appendix.

fourth round of each contest, final ranks and points are displayed. After the end of the contest participants are randomly assigned to a new group of four in which they compete another four rounds. As we are interested in the pure effect of rank, payments in this stage are purely based upon absolute and not relative performance: For each point a participant receives 3 cents which adds up to 12 cents per slider. One out of the four contests is randomly selected to determine payments for stage 2 of the experiment.

Multipliers

Correctly adjusted sliders translate into points and therefore payments by means of a multiplication factor. For round two to four of every contest, this multiplier is fixed to the value of four, which implies that per correctly adjusted slider a participant receives 4 points and hence earns 12 cents. In contrast, in each first round of an individual contest the multiplier is randomly drawn from the set $\{1; 3; 5; 7\}$ without replacement. This implies that every participant within one contest has a different multiplier.⁴ Importantly, while participants know about the multipliers available and its random assignment, they do not learn about their own multiplier before they finished the first round. Hence, monetary incentives are constant across rounds, as in expectation each slider still yields four points and hence 12 cents.⁵ At the same time, the realization of the multiplier induces an exogenous variation in rank, which then allows to identify its causal effect on subsequent performance.

—Table 1 about here—

2.3 Measuring Social Value Orientation and Risk Preferences

Next to competitiveness, we consider social preferences to be a likely individual characteristic potentially associated with rank sensitivity. The intuition behind this conjecture is based on the ‘zero-sum’-nature of contests, i.e. the fact that increasing effort in order to affect one’s own rank imposes a negative externality on other participant’s rank.

To get a comprehensive measure of social preferences, we elicit the ‘social value orientation’ of participants (SVO), which provides a quite detailed assessment of an individual’s social preferences. We use the z-Tree implementation by Crosetto et al. (2012) which adapts the paper-based SVO slider measure by Murphy et al. (2011). Subjects have to make a series of allocation decisions with varying monetary payoffs between themselves and others. Finally, one of these

⁴Table 1 gives an overview over the contests.

⁵While participants are in principle able to infer their own and their contestants’ multipliers, we consider this rather unlikely given that ranks and points are only displayed for 10 seconds and because total points do not necessarily unambiguously identify a particular multiplier.

allocation decisions determines the monetary payment from this stage.

The last stage of the experiment comprises a questionnaire. Next to some standard sociodemographic variables like gender, age, and subject of study, we include the survey module measuring general and domain-specific risk preferences by Falk et al. (2016).

3 Results

We present the results of our experiment in two steps. First, we take an aggregate perspective and discuss how rank affects effort provision for the average ('representative') participant. Second, we then explicitly consider heterogeneous effects of rank feedback and look at the extent of heterogeneity as well as its consequences.

3.1 Aggregate Effects

—Table 3 about here—

Table 3 shows how participants react to variation in rank. We regress a participant's performance in a given round on her rank at the beginning of the respective round. Given that this correlation is likely affected by endogeneity problems, we make use of our experimental design by instrumenting rank with a series of dummy variables, one for each potential realization of the multiplier, i.e. 3, 5, or 7, with 1 being the omitted category. Observing each individual in different contests further allows including individual fixed effects in the regression.

In the first column of Table 3 we limit ourselves to looking at behavior in the second round of each contest, as we expect our instrument to be most relevant here. As each individual participates in five contests each, this yields an overall number of 920 observations. We find that increasing rank by one unit (which reflects a worse rank) leads to a significant decrease in effort in round 2 by 0.16 sliders (the standard deviation is 2.79⁶, which is also highly statistically significant).

In column (2) we expand our analysis to also cover behavior in round 3 and 4. As rank is based on cumulative points from all subsequent rounds, the realization of the multiplier in round 1 also affects rank at the beginning of round 3 and 4 and hence also qualifies as a valid instrument in those rounds. Adding round three and four of each contest, our sample triples in size and now consists of 2760 observations. While we see that the effect remains statistically significant (although now only at the 5%-level) it lowers slightly in magnitude. To summarize our first result,

⁶See the summary statistics in Table 2

we find that rank variation has a significant causal effect on effort provision.

Result 1 – Rank matters: *Individuals react to rank: On average, a better rank causes higher subsequent effort.*

—Table 4 about here—

In a next step we take a closer look to see whether this overall effect is driven by the reaction to a good rank, the reaction to a bad rank, or both. In Table 4 we now distinguish between good and bad ranks by including *Rank 1 (d)* (*Rank 4 (d)*), which is a dummy equal to one if the respective participant is ranked first (last) after the first round and zero otherwise. Column (1)-(3) look at effort provision only in second rounds whereas columns (4)-(6) consider all rounds for the analysis. Again, in all specifications rank at the beginning of each round is instrumented by dummies capturing the realization of the random multiplier.

Column (1) shows that being ranked first induces higher effort provision in round 2: Participants correctly adjust an additional half slider (0.51) when being ranked first as compared to holding any other rank. In contrast, column (2) shows that individuals reduce effort in response to occupying the last rank by almost one third of a slider (-0.3). In column (3) we include both regressors in one specification. While statistical significance unsurprisingly weakens, the direction of both effects remains the same, even though it diminishes in size.

By and large, these results are supported by column (4)-(6) which again also include behavior in round three and four. Compared to other positions, being ranked first causes more and being ranked last causes less correctly adjusted sliders. This indicates that the overall positive effect of a better rank on subsequent effort documented in Table 3 is not purely driven by a positive (negative) effect of having a particularly good (bad) rank. Taken together, on average rank feedback therefore amplifies initial performance differences, as better ranked individuals further increase, while worse ranked individuals further decrease their performance.

Result 2 – Rank matters ... at the top and at the bottom: *Rank feedback amplifies initial performance differences: Compared to intermediate ranks, the top rank causes higher and the last rank lower effort, respectively.*

—Table 5 about here—

So far we analyzed the role of rank feedback within a given contest. However, our experimental design even allows investigating the causal effect of rank within one contest on performance

in subsequent ones. Put differently, we can check whether rank feedback causes spillover effects. This would be particularly noteworthy as contestants are re-matched after each contest. In Table 5 we therefore regress effort provision in first rounds on the final rank of previous contests. Given that each individual participates in five contests this gives us four subsequent round 1 observations per individual and 736 observations in total. As in this setting similar endogeneity concerns as before might arise, we again employ our instrumental variable strategy and instrument the final rank of the previous contest with dummies capturing the realized multiplier in the previous contest's first round.

Column (1) in Table 5 shows that a worse final rank in the previous contest leads to statistically significantly lower effort provision in the first round of the subsequent contest. Interestingly, this spillover effect is numerically even larger than the effect within contests. The same holds true once we again disentangle the effect of first and last ranks in column (2) and (3): Similar to the effect within contests, also across contests holding the first rank increases effort whereas inhabiting the last rank reduces effort.

***Result 3 – Rank matters ... even across contests:** A better rank in previous contests has a positive causal effect on effort in subsequent contests.*

3.2 Heterogeneity

So far, our results show that rank feedback significantly affects effort provision: We found that on average a better rank leads to more and a worse rank to less subsequent effort. Furthermore, we saw that the effects of rank feedback even spill over to subsequent contests with new opponents. While these effects appear reasonable and are also statistically significant, the key question of this paper is whether such an aggregate perspective – that only considers average behavior – might not be misleading by disguising systematic between-subject heterogeneity. To assess this conjecture, we therefore now turn to investigate if and how rank sensitivity varies across participants.

We therefore estimate each participant's rank sensitivity individually, thus fully exploiting the within-subject variation of our data. The basic empirical strategy thereby mirrors the one from the aggregate perspective: We regress effort provision in rounds two to four on rank at the beginning of the respective round. Again, we make use of our instrumental variable estimation strategy to address potential endogeneity issues. As a result, instead of estimating a common beta for all individuals this approach yields an individual beta for each of the 181 participants

and hence an individual measure of rank feedback sensitivity.⁷

—Figure 1 about here—

Figure 1 depicts these individual sensitivities for all 184 participants in ascending order. The picture clearly reveals substantial heterogeneity in the reaction to rank feedback: Not only does the magnitude of the effect vary considerably between individuals. In fact, for a substantial share of the sample even the direction of the effect is fundamentally different than suggested by the aggregated analysis. While we do see that the majority of individual betas have a negative sign indicating a positive effort reaction to a better rank, the share of betas with a positive sign can hardly be ignored. For these individuals the effect of rank feedback is actually opposite to the one documented in the aggregated analysis: These participants decrease effort in reaction to a better rank and increase effort in reaction to a worse one.⁸

Result 4 – Heterogeneity of rank sensitivity: *Individual sensitivities to rank feedback vary substantially in magnitude and direction.*

—Table 6 about here—

The substantial degree of between-subject heterogeneity in rank feedback sensitivity naturally leads to the question whether it translates into differences in aggregate outcomes. We therefore look at participants' final rank as a measure of overall performance and how it is associated with individual rank feedback sensitivity.

Table 6 shows the corresponding results. In column (1) we see that individuals who increase (decrease) effort in reaction to better (worse) ranks perform overall worse. This result holds when we include ability (as measured by performance in the piece rate scheme), a gender dummy (Female (d)), and competitiveness in column (2). Column (3) and (4) show that this result is mostly driven by individuals who negatively react to a better rank (and hence positively to a worse rank), as they perform significantly better overall. In turn, participants who amplify initial differences, i.e. who increase effort in reaction to better ranks and decrease effort in reaction to worse ranks, perform overall worse, albeit not statistically significantly so.

⁷For 3 out of 184 participants it is not possible to compute the individual sensitivity due to a perfect collinearity of rank and multiplier.

⁸Note that in the figure as well as in the corresponding regressions a negative beta refers to what we call a 'positive' rank sensitivity, as a higher numerical rank reflects a worse ranking position.

Tables 7 and 8 confirm the negative effect of a positive rank sensitivity: By using dummies for winning or losing as dependent variables, it shows that individuals with a positive rank sensitivity win less and lose more often.

—Table 7 about here—

—Table 8 about here—

Result 5 – Heterogeneity of rank sensitivity affects overall performance: Individuals increasing (decreasing) effort in reaction to better (worse) ranks perform overall worse, win less and lose more often.

—Table 9 about here—

In a final step of our analysis, we ask whether individual feedback sensitivity is associated with other individual characteristics. As likely candidates we therefore test the explanatory power of measures for ability, competitiveness, risk aversion, social value orientation, and gender. Table 9 depicts the corresponding results. By and large, we find only very weak associations of the included variables with feedback sensitivity. In particular, the extremely low adjusted R^2 of these regressions document an almost complete lack of explanatory power.

While one needs to acknowledge that our analysis only features a rather limited list of potentially associated factors, the virtual absence of any explanatory power of those variables we include – and consider reasonable candidates to be correlated with rank sensitivity – indicates that rank sensitivity might actually be a quite idiosyncratic individual trait and not just a mere artifact driven by other behavioral mechanisms.

Result 6 – Rank sensitivity as an idiosyncratic, individual trait: Ability, gender, and common behavioral measures are not associated with rank sensitivity at the individual level.

4 Discussion and Conclusion

Using a laboratory experiment, we study how people differ in their reaction to rank feedback. While we document the average effect of rank feedback on subsequent effort provision to be statistically significant and robust, we also find that this aggregated view masks a substantial degree of between-subject heterogeneity. In fact, a significant share of our sample displays behavioral

patterns entirely opposite to those of the ‘representative’ participant. Further, we find that these differences in the reaction to rank feedback translate into differences in overall performance.

Given the pervasiveness and importance of contest structures for many economic outcomes, our results provide a fresh perspective to better assess the consequences of competitive mechanisms in real settings: Many aspects of different institutional frameworks are not just competitive in nature, but also fundamentally dynamic, i.e. feature repeated decisions including intermediate feedback on relative performance. In such settings, if individuals differ in their reaction to this intermediate feedback, this implies non-trivial allocative and distributional effects on final outcomes, as initial differences might either be amplified or dampened for different individuals.

As a result, institutional change that comes with changes in the competitive frame – like increased transparency, relative performance incentives, or ability tracking – might trigger unexpected heterogeneous effects and thus have unintended consequences. Future research should therefore test such policy changes regarding market and allocation mechanisms to inform actual policy design, for instance by using specific laboratory experimental research designs. Depending on their concrete structure, the corresponding results might help to better understand how to design educational systems and organizations to allow materializing the potential efficiency gains from competitive institutions in providing good incentives without having to bear their negative effects through allocative distortions and undesirable distributional effects.

References

- Bartling, B., E. Fehr, M. A. Marechal, and D. Schunk (2009). Egalitarianism and Competitiveness. *American Economic Review* 99(2), 93–98.
- Blanes i Vidal, J. and M. Nossol (2011). Tournaments Without Prizes: Evidence from Personnel Records. *Management Science* 57(10), 1721–1736.
- Casas-Arce, P. and F. A. Martínez-Jerez (2009). Relative Performance Compensation, Contests, and Dynamic Incentives. *Management Science* 55(8), 1306–1320.
- Crosetto, P., O. Weisel, and F. Winter (2012). A flexible z-Tree Implementation of the Social Value Orientation Slider Measure.
- Dechenaux, E., D. Kovenock, and R. M. Sheremeta (2015). A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics* 18(4), 609–669.
- Delfgaauw, J., R. Dur, A. Non, and W. Verbeke (2014). Dynamic incentive effects of relative performance pay: A field experiment. *Labour Economics* 28, 1–13.
- Dijk, O., M. Holmen, and M. Kirchler (2014). Rank matters—The impact of social competition on portfolio choice. *European Economic Review* 66, 97–110.
- Dohmen, T. and A. Falk (2011). Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. *American Economic Review* 101(2), 556–590.
- Falk, A., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2016). The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences. *Human Capital and Economic Opportunity Global Working Group Working Paper Series 003*.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2), 171–178.
- Gill, D., Z. Kissova, J. Lee, and V. L. Prowse (forthcoming). First-Place Loving and Last-Place Loathing: How Rank in the Distribution of Performance Affects Effort Provision. *Management Science*.
- Gill, D. and V. Prowse (2011). A Novel Computerized Real Effort Task Based on Sliders. *IZA Discussion Paper 5801*.
- Greiner, B. (2004). The Online Recruitment System ORSEE - A Guide for the Organization of Experiments in Economics. *Papers on Strategic Interaction 2003-10, Max Planck Institute of Economics, Strategic Interaction Group*.

- Murphy, R. O., K. A. Ackermann, and M. J. J. Handgraaf (2011). Measuring Social Value Orientation. *Judgment and Decision Making* 6(8), 771–781.
- Niederle, M. and L. Vesterlund (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics* 122(3), 1067–1101.
- Sheremeta, R. M. (2013). Overbidding and heterogeneous behavior in contest experiments. *Journal of Economic Surveys* 27(3), 491–514.
- Sheremeta, R. M. (forthcoming). Behavior in group contests: a review of experimental research. *Journal of Economic Surveys*.

Appendix

5 Table and Figures

Table 1: Overview Competitions

Contest number	1				...	5			
	Round	1	2	3		4	1	2	3
Multipliers	1					1			
	3	4				3	4		
	5		4			5		4	
	7			4		7			4

Table 2: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Female (d)	0.522	0.501	0	1	184
Competitiveness	3.993	2.772	0	10	151
Ability	6.853	2.348	0	13	184
SVO-Angle	36.592	6.255	18.435	48.621	184
Risk Aversion	5.473	2.288	0	10	184
Performance	9.68	2.786	0	22	3680

Note: *Female (d)* is a dummy which is equal to one if a participant is female and zero otherwise. *Competitiveness* presents the switching points from tournament to piece rate compensation. *Ability* captures the number of correctly solved math tasks under the piece rate scheme in stage 1. *SVO-Angle* presents the elicited ‘social value orientation’ of participants, which provides a quite detailed assessment of an individual’s social preferences. *Risk Aversion* is measured by the survey item: *Are you a person who is generally willing to take risks, or do you try to avoid taking risks?* Participants specify their level of agreement on a 11 points Likert scale ranging from “*I am not taking risk at all*” to “*I am very willing to take risk*”. *Performance* presents the number of correctly adjusted sliders in a round.

Table 3: Sensitivity to Multiplier Rank

	Correct Slider Round 2	Correct Slider Round 2-4
	(1)	(2)
Rank Round 1	-0.157*** (0.06)	
Rank Round 1-3		-0.090** (0.05)
Individual fixed-effects	yes	yes
Observations	920	2760
Adjusted R ²	0.510	0.537

Note: Coefficient estimates from ordinary least squares estimations with individual fixed effects. * significant at 10%; ** sign. at 5%; *** sign. at 1%. The dependent variable is the number of correctly adjusted sliders in round 2 (column 1) and rounds 2-4 (column 2). *Rank Round 1* presents the rank of a participant after round 1, instrumented with a series of dummy variables, one for each potential realization of the multiplier, i.e. 3, 5, or 7, with 1 being the omitted category. *Rank Round 1-3* presents the rank of a participant after round 1, 2 and 3. The ranks are instrumented with a series of dummy variables, one for each potential realization of the multiplier, i.e. 3, 5, or 7, with 1 being the omitted category. Standard errors clustered at the group level are in parentheses.

Table 4: Sensitivity to Multiplier Induced First and Last Rank

	Correct Slider Round 2			Correct Slider Round 2-4		
	(1)	(2)	(3)	(4)	(5)	(6)
Rank 1 (d)	0.514*** (0.17)		0.447** (0.19)			
Rank 4 (d)		-0.295** (0.14)	-0.123 (0.16)			
Rank 1 (d)				0.270** (0.13)		0.207 (0.16)
Rank 4 (d)					-0.227* (0.13)	-0.107 (0.16)
Individual fixed-effects	yes	yes	yes	yes	yes	yes
Observations	920	920	920	2760	2760	2760
Adjusted R ²	0.508	0.508	0.509	0.536	0.536	0.536

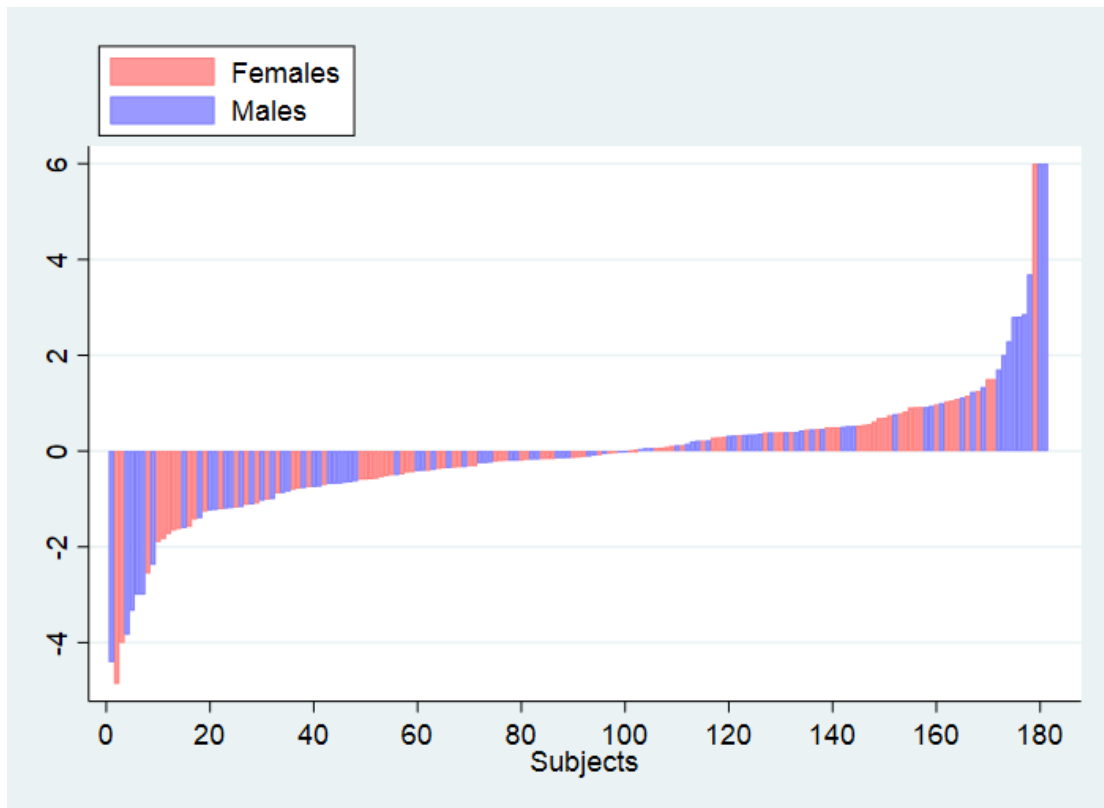
Note: Coefficient estimates from ordinary least squares estimations with individual fixed effects. * significant at 10%; ** sign. at 5%; *** sign. at 1%. The dependent variable is the number of correctly adjusted sliders in round 2 (column 1) and rounds 2-4 (column 2). *Rank 1 (d)* is a dummy which is equal to one if the respective participant is ranked first after a round and zero otherwise. *Rank 4 (d)* is a dummy which is equal to one if the respective participant is ranked last after a round and zero otherwise. All rank dummies are instrumented with a series of dummy variables, one for each potential realization of the multiplier, i.e. 3, 5, or 7, with 1 being the omitted category. Standard errors clustered at the group level are in parentheses.

Table 5: Sensitivity to Rank last Round last Competition

	Correct Slider Round 1			
	(1)	(2)	(3)	(4)
Rank after last round last competition	-0.242** (0.10)			
Rank 1 last round last competition (d)		0.684** (0.30)		-0.357 (0.88)
Rank 4 last round last competition (d)			-0.976** (0.40)	-1.381 (1.11)
Individual fixed-effects	yes	yes	yes	yes
Observations	736	736	736	736
Adjusted R ²	0.570	0.568	0.552	0.530

Note: Coefficient estimates from ordinary least squares estimations with individual fixed effects. * significant at 10%; ** sign. at 5%; *** sign. at 1%. The dependent variable is the number of correctly adjusted sliders in round 1. *Rank after last round last competition* presents the final rank a participant achieved in the foregoing competition. The rank is instrumented with a series of dummy variables, one for each potential realization of the multiplier, i.e. 3, 5, or 7, with 1 being the omitted category. *Rank 1 last round last competition (d)* is a dummy which is equal to one if the respective participant is ranked first at the end of the foregoing competition. *Rank 4 last round last competition (d)* is a dummy which is equal to one if the respective participant is ranked last at the end of the foregoing competition. Both rank dummies are instrumented with a series of dummy variables, one for each potential realization of the multiplier, i.e. 3, 5, or 7, with 1 being the omitted category. Given that only competitions 2-4 have a foregoing competition the number of observations is reduced to 736. Standard errors clustered at the group level are in parentheses.

Figure 1: Sensitivity to Rank



Notes: The figure displays individual sensitivity estimates from ordinary least squares estimations with individual fixed effects for 181 participants in ascending order. Individual sensitivities are estimated by regressing effort provision in rounds two to four on rank at the beginning of the respective round. Ranks are instrumented with a series of dummy variables, one for each potential realization of the multiplier, i.e. 3, 5, or 7, with 1 being the omitted category. For 3 participants it is not possible to compute individual sensitivity due to a lack of variation in rank.

Table 6: Sensitivity and Final Rank

	Final Rank Competition			
	(1)	(2)	(3)	(4)
Rank Sensitivity	-0.127*** (0.03)	-0.104*** (0.03)		
Lowest 25% rank betas (d)			0.182** (0.09)	0.059 (0.10)
Top 25% rank betas (d)			-0.211** (0.09)	-0.270*** (0.10)
Constant	2.502*** (0.01)	2.401*** (0.05)	2.518*** (0.04)	2.437*** (0.06)
Controls	no	yes	no	yes
Observations	905	745	905	745
Adjusted R ²	0.024	0.136	0.013	0.134

Note: Coefficient estimates from ordinary least squares estimations. * significant at 10%; ** sign. at 5%; *** sign. at 1%. The dependent variable is the final rank in a competition. *Rank sensitivity* presents the individual's estimated sensitivity by regressing effort provision in rounds two to four on rank at the beginning of the respective round. *Lowest 25% rank betas (d)* is a dummy which is equal to one if the respective participant individual sensitivity is among the lowest quartile of the sensitivity distribution and zero otherwise. *Top 25% rank betas (d)* is a dummy which is equal to one if the respective participant individual sensitivity is among the highest quartile of the sensitivity distribution and zero otherwise. Controls include: *Female (d)*, *Ability*, *Competitiveness*. *Female (d)* is a dummy which is equal to one if a participant is female and zero otherwise. *Ability* captures the number of correctly solved math tasks under the piece rate scheme in stage 1. *Competitiveness* presents the switching points from tournament to piece rate compensation. Columns (1) and (3) include five observations for all 181 participants we are able to estimate the individual sensitivity. Due to double switching points in the competitiveness measure we lose another 32 participants (160 observations) in column (2) and (4). Standard errors clustered at the group level are in parentheses.

Table 7: Sensitivity and Winning

	Winning Competition			
	(1)	(2)	(3)	(4)
Rank Sensitivity	0.032*** (0.01)	0.033*** (0.01)		
Lowest 25% rank betas (d)			-0.058* (0.03)	-0.021 (0.04)
Top 25% rank betas (d)			0.040 (0.04)	0.073* (0.04)
Constant	0.251*** (0.00)	0.290*** (0.02)	0.253*** (0.01)	0.281*** (0.02)
Controls	no	yes	no	yes
Observations	905	745	905	745
Adjusted R ²	0.009	0.073	0.004	0.069

Note: Coefficient estimates from ordinary least squares estimations. * significant at 10%; ** sign. at 5%; *** sign. at 1%. The dependent variable is a dummy which is equal to one if a participant won a competition and zero otherwise. *Rank sensitivity* presents the individual's estimated sensitivity by regressing effort provision in rounds two to four on rank at the beginning of the respective round. *Lowest 25% rank betas (d)* is a dummy which is equal to one if the respective participant individual sensitivity is among the lowest quartile of the sensitivity distribution and zero otherwise. *Top 25% rank betas (d)* is a dummy which is equal to one if the respective participant individual sensitivity is among the highest quartile of the sensitivity distribution and zero otherwise. Controls include: *Female (d)*, *Ability*, *Competitiveness*. *Female (d)* is a dummy which is equal to one if a participant is female and zero otherwise. *Ability* captures the number of correctly solved math tasks under the piece rate scheme in stage 1. *Competitiveness* presents the switching points from tournament to piece rate compensation. Columns (1) and (3) include five observations for all 181 participants we are able to estimate the individual sensitivity. Due to double switching points in the competitiveness measure we lose another 32 participants (160 observations) in column (2) and (4). Standard errors clustered at the group level are in parentheses.

Table 8: Sensitivity and Losing

	Losing Competition			
	(1)	(2)	(3)	(4)
Rank Sensitivity	-0.050*** (0.01)	-0.024** (0.01)		
Lowest 25% rank betas (d)			0.079** (0.04)	0.048 (0.04)
Top 25% rank betas (d)			-0.069** (0.03)	-0.044 (0.04)
Constant	0.250*** (0.00)	0.218*** (0.02)	0.251*** (0.01)	0.212*** (0.02)
Controls	no	yes	no	yes
Observations	905	745	905	745
Adjusted R ²	0.025	0.094	0.012	0.094

Note: Coefficient estimates from ordinary least squares estimations. * significant at 10%; ** sign. at 5%; *** sign. at 1%. The dependent variable is a dummy which is equal to one if a participant lost a competition and zero otherwise. *Rank sensitivity* presents the individual's estimated sensitivity by regressing effort provision in rounds two to four on rank at the beginning of the respective round. *Lowest 25% rank betas (d)* is a dummy which is equal to one if the respective participant individual sensitivity is among the lowest quartile of the sensitivity distribution and zero otherwise. *Top 25% rank betas (d)* is a dummy which is equal to one if the respective participant individual sensitivity is among the highest quartile of the sensitivity distribution and zero otherwise. Controls include: *Female (d)*, *Ability*, *Competitiveness*. *Female (d)* is a dummy which is equal to one if a participant is female and zero otherwise. *Ability* captures the number of correctly solved math tasks under the piece rate scheme in stage 1. *Competitiveness* presents the switching points from tournament to piece rate compensation. Columns (1) and (3) include five observations for all 181 participants we are able to estimate the individual sensitivity. Due to double switching points in the competitiveness measure we lose another 32 participants (160 observations) in column (2) and (4). Standard errors clustered at the group level are in parentheses.

Table 9: Correlations Rank Sensitivity

	Rank Beta					
	(1)	(2)	(3)	(4)	(5)	(6)
Ability	0.100** (0.04)					0.050 (0.05)
Female		-0.139 (0.21)				-0.218 (0.22)
Competitiveness			-0.013 (0.04)			-0.026 (0.04)
SVO-Angle				0.032* (0.02)		0.028* (0.02)
Risk Aversion					0.045 (0.05)	0.043 (0.05)
Observations	181	181	149	181	181	149
Adjusted R ²	0.023	-0.003	-0.006	0.015	-0.000	0.015

Note: Coefficient estimates from ordinary least squares estimations. * significant at 10%; ** sign. at 5%; *** sign. at 1%. The dependent variable is the individual's estimated sensitivity by regressing effort provision in rounds two to four on rank at the beginning of the respective round. *Ability* captures the number of correctly solved math tasks under the piece rate scheme in stage 1. *Female* (*d*) is a dummy which is equal to one if a participant is female and zero otherwise. *Competitiveness* presents the switching points from tournament to piece rate compensation. Columns (1) and (3) include five observations for all 181 participants we are able to estimate the individual sensitivity. *SVO-Angle* presents the elicited 'social value orientation' of participants, which provides a quite detailed assessment of an individual's social preferences. *Risk Aversion* is measured by the survey item: *Are you a person who is generally willing to take risks, or do you try to avoid taking risks?* Participants specify their level of agreement on a 11 points Likert scale ranging from "I am not taking risk at all" to "I am very willing to take risk". Standard errors are in parentheses.

6 Screenshots

Figure A.1: Measuring Competitiveness

Stufe:

Sie sehen unten 10 verschiedenen Entscheidungssituationen in denen Sie sich entweder für die Stückentlohnung oder die Wettbewerbsentlohnung entscheiden können.
 Sie müssen sich in allen 10 Situationen für eine der beiden Zahlungsarten entscheiden.
 Für den Gewinn des Wettbewerbs erhalten Sie 400 Punkte.

Verbleibende Zeit (sec): 59

Situation	Ihre Entscheidung	Fixes Grundgehalt plus Stückentlohnung
1	Wettbewerb <input type="radio"/> <input type="radio"/> Stückentlohnung	60 Punkte plus 10 Punkte für jede richtig gelöste Rechenaufgabe.
2	Wettbewerb <input type="radio"/> <input type="radio"/> Stückentlohnung	80 Punkte plus 10 Punkte für jede richtig gelöste Rechenaufgabe.
3	Wettbewerb <input type="radio"/> <input type="radio"/> Stückentlohnung	100 Punkte plus 10 Punkte für jede richtig gelöste Rechenaufgabe.
4	Wettbewerb <input type="radio"/> <input type="radio"/> Stückentlohnung	120 Punkte plus 10 Punkte für jede richtig gelöste Rechenaufgabe.
5	Wettbewerb <input type="radio"/> <input type="radio"/> Stückentlohnung	140 Punkte plus 10 Punkte für jede richtig gelöste Rechenaufgabe.
6	Wettbewerb <input type="radio"/> <input type="radio"/> Stückentlohnung	160 Punkte plus 10 Punkte für jede richtig gelöste Rechenaufgabe.
7	Wettbewerb <input type="radio"/> <input type="radio"/> Stückentlohnung	180 Punkte plus 10 Punkte für jede richtig gelöste Rechenaufgabe.
8	Wettbewerb <input type="radio"/> <input type="radio"/> Stückentlohnung	200 Punkte plus 10 Punkte für jede richtig gelöste Rechenaufgabe.
9	Wettbewerb <input type="radio"/> <input type="radio"/> Stückentlohnung	220 Punkte plus 10 Punkte für jede richtig gelöste Rechenaufgabe.
10	Wettbewerb <input type="radio"/> <input type="radio"/> Stückentlohnung	240 Punkte plus 10 Punkte für jede richtig gelöste Rechenaufgabe.

Bitte drücken Sie einmal auf "OK" nachdem Sie Ihre Entscheidungen getroffen haben.

OK

Figure A.2: Slider Task

Korrekt justierte Slider: 0

Figure A.3: Rank Feedback

Stufe: 2.2	Sie sind Gruppenmitglied 3	Verbleibende Zeit [sec]: 8
Nach Runde 1 in Wettbewerb 1 ergeben sich folgende Punkte und Ränge:		
Rang	Mitglied	Punkte
1	1	55
2	3	35
3	2	27
4	4	8