WORKING PAPER · NO. 2022-111

The Impact of Team Incentives on Performance in Graduate School: Evidence from Two Pilot RCTs

John A. List and Rohen Shah

AUGUST 2022

Becker Friedman Institute
FOR ECONOMICS AT UCHICAGO

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

THE IMPACT OF TEAM INCENTIVES ON PERFORMANCE IN GRADUATE SCHOOL:
EVIDENCE FROM TWO PILOT RCTS

John A. List
Rohen Shah

August 2022

This study was registered on the AEA RCT Registry (AEARCTR-0009477) and was approved by the Social and Behavioral Sciences IRB from the University of Chicago (IRB21-0580). We would like to thank Uditi Karna and Charles Shi for outstanding research assistance.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w30374.ack

The Impact of Team Incentives on Performance in Graduate School: Evidence from Two Pilot RCTs
John A. List and Rohen Shah
August 2022
JEL No. C9,C93,D79,I2,J3

ABSTRACT

In organizations, teams are ubiquitous. “Weakest Link” and “Best Shot” are incentive schemes that tie a group member’s compensation to the output of their group’s least and most productive member, respectively. In this paper, we test the impact of these incentive schemes by conducting two pilot RCTs (one in-person, one online), which included more than 250 graduate students in a graduate math class. Students were placed in study groups of three or four students, and then groups were randomized to either control, Weakest Link, or Best Shot incentives. We find evidence that such incentive approaches can affect test scores, both in-person and online.

John A. List
Department of Economics
University of Chicago
1126 East 59th
Chicago, IL 60637
and NBER
jlist@uchicago.edu

Rohen Shah
University of Chicago
1307 E. 60th Street
Chicago, IL 60637
shahr@uchicago.edu

1. Introduction

Work in teams is an increasingly common business practice (Lazear & Shaw, 2007), and it has become commonplace for team productivity to be changed through incentives (Haeckl et al., 2018). However, teams may include free-rider behavior, especially if people do not have freedom to opt in and out of a team (Riedl et al., 2016). Theoretically, we would expect a *Weakest Link* incentive design, wherein a group is compensated based on the least productive member, to raise the floor of the minimum performer and a *Best Shot* incentive design, where a group is compensated based on the most productive member, to raise the ceiling of the maximum performer (Harrison & Hirshleifer, 1989).

We use a field experiment in a math refresher course for incoming master's students at a large private university to explore whether such incentive structures impact group performance in this manner. The course had two sections: one online and one in-person. While there were no graded assignments during this three-week course, students took a 100-point exam at the end of the course, which they were required to pass for entrance. Students were assigned to work in informal study groups of three or four students, and then each group was randomly assigned to either a control, Weakest Link, or Best Shot incentive regime ($15 Amazon gift card).

Consistent with theory, we find that students randomly assigned to the Weakest Link incentive had a higher group minimum exam score relative to both control and Best Shot, but only for the online section. However, we also found that students randomly assigned to the Best Shot incentive had a higher group minimum performance relative to control in both the online and in-person sections. While this finding is surprising at first, it is related to work showing the multiplicity of equilibria in Best Shot and Weakest Link games (Chowdhury & Topolyan, 2016). In particular, the theory shows that beliefs are an important driver of effort, and with the right set of beliefs results such as ours can arise. For example, if students believe themselves to be the star of their group, then a Best Shot incentive would induce most people to put in high effort, even those at the bottom. Depending on how universal this mindset is, we may expect Best Shot to induce a higher group minimum as well. In this manner, our results highlight the importance of understanding beliefs and how they interact with marginal incentives.

Our work contributes to the economics literature on group productivity and incentives (Hossain & List, 2012; Weidmann & Deming, 2021). It also adds to the literature on financial incentives used in educational settings to increase student effort and performance (Gneezy et al., 2019; Levitt et al., 2016; List et al., 2018). Finally, our study adds to the literature on student learning during the COVID-19 pandemic (Orlov et al., 2021), as well as research exploring differences between online and in-person instruction (Bettinger et al., 2016; Bettinger et al., 2017; Kraft et al., 2022).

2. Experimental Design

Our field experiment took place during a three-week math refresher course for incoming master's students in public policy at a large private university in the U.S. during the summer of 2021. The course covers algebra and calculus, and incoming students have a wide range of mathematical backgrounds. A total of 391 students enrolled in the course, of which 225 were in an in-person section and 166 were in an online section, taught synchronously via zoom. Students chose whether to sign up for the in-person or online section.

The in-person section primarily consisted of students from the U.S., while the online section primarily consisted of international students. This was largely due to pandemic-related restrictions on international travel. This is an important consideration, as we can cleanly compare online versus in-person results only if we make the strong assumption that domestic and international students have a common treatment effect. To avoid confusion, and to make it clear that these are samples drawn from two different populations, we split the data into RCT 1 (in-person) and RCT 2 (online).

On the first day of class, students were presented with information about the study and given a chance to opt-in, thus our work should be considered a framed field experiment. Students were told that their participation in the study would involve being assigned an informal study group, taking two brief surveys, and a chance to earn financial incentives. A total of 259 students (66%) participated in the study, of which 176 were in-person and 83 were online. All students who participated received a $15 Amazon gift card and took a brief survey with a demographic questionnaire.

Students who participated were placed into groups of three or four students, which gave a total of 72 groups. Each group was then randomized into either control, Weakest Link, or Best Shot conditions. On the second day of the class, students received an email containing the names and emails of their group members, with all three or four members cc'ed in the email.

The email informed students that at the end of the course, they would have a chance to win an additional $15 Amazon gift card. Each student in the control group was told that the probability of winning this prize is 10%. Each student in the Weakest Link group was told that their probability of winning the prize would be equal to the exam score of the lowest performing member of their group. Likewise, each student in the Best Shot group was told that their probability of winning the prize would be equal to the exam score of the highest performing member of their group. The recruitment scripts and emails are in Appendix A.

Tables 1 and 2 show descriptive statistics for RCT 1 and 2, respectively. There is no significant difference in gender composition among treatment conditions in both RCTs. The Best Shot group is slightly younger than the other groups in RCT 1, and there is no significant difference in age across treatment conditions in RCT 2. Note that there are no qualitative changes to the main results presented in Section 3 if these covariates are included in the regressions.

Tables 1 and 2 also show that there was no significant difference in the final exam scores across treatment conditions in both RCTs. The exam had a left skew, with most students scoring above a 90. Additionally, about a quarter of participating students did not take the survey at the end of the course, and this attrition rate for each treatment condition is shown in Tables 1 and 2 as well. Despite this attrition, we are able to recover an internally valid estimate of the intent-to-treat effects, since we had access to exam scores for all students regardless of whether they took the second survey.

Table 1: Descriptive Statistics and Balance for RCT 1 (In-Person)

	Control (N=59)	Weakest Link (N=58)	Best Shot (N=59)	F Statistic p-value
Female	57.6%	59.7%	67.2%	0.5362
Age (Years)	27.0 (SD: 4.0)	26.2 (SD: 3.5)	24.8 (SD: 2.7)	0.0021***
Exam Score	89.3 (SD: 13.2)	88.1 (SD: 14.5)	90.9 (SD: 8.7)	0.4580
Attrition Rate	15.3%	24.1%	25.4%	0.3482

Note. The F-Statistic p-value column represents the p-value on a joint hypothesis test with a null hypothesis of equal means across treatment conditions. *** p<.01, ** p<.05, * p<.10.

Table 2: Descriptive Statistics and Balance for RCT 2 (Online)

	Control (N=29)	Weakest Link (N=27)	Best Shot (N=27)	F Statistic p-value
Female	65.5%	53.8%	50.0%	0.4870
Age (Years)	25.6 (SD: 3.9)	25.0 (SD: 2.9)	25.7 (SD: 4.4)	0.7438
Exam Score	89.4 (SD: 12.2)	93.7 (SD: 5.2)	90.8 (SD: 9.0)	0.2171
Attrition Rate	24.1%	25.9%	37.0%	0.5303

Note. The F-Statistic p-value column represents the p-value on a joint hypothesis test with a null hypothesis of equal means across treatment conditions. *** p<.01, ** p<.05, * p<.10.

3. Results

To measure treatment effects, we estimate the following model separately for online and in-person students:

$$Y_i = \alpha + \beta\ W_i + \delta\ B_i + \varepsilon_i$$

where Y_i represents either individual i's exam score (Model 1), individual i's group's maximum exam score (Model 2), or individual i's group's minimum exam score (Model 3). W_i and B_i are indicators for assignment to Weakest Link and Best Shot treatments respectively, and ε_i is the error term. Tables 3 and 4 summarize the OLS regression results for RCT 1 (in-person) and RCT 2 (online students), respectively.

Table 3: Treatment Effect for RCT 1 (In-Person Students)

	(1) Score	(2) Group Max	(3) Group Min
Weakest Link	-1.210	0.172	-3.906
	(2.564)	(0.417)	(3.302)
Best Shot	1.644	0.568	6.695***
	(2.064)	(0.364)	(2.483)
Constant	89.31***	97.47***	75.45***
	(1.722)	(0.304)	(2.155)
Observations	176	176	176

Note. Robust standard errors are in parentheses. *** p<.01, ** p<.05, * p<.1

Table 4: Treatment Effect for RCT 2 (Online Students)

	(1) Score	(2) Group Max	(3) Group Min
Weakest Link	4.360*	-0.204	12.91***
	(2.482)	(0.541)	(2.740)
Best Shot	1.471	-1.148	7.280**
	(2.863)	(1.018)	(3.169)
Constant	89.36***	98***	75.28***
	(2.274)	(0.452)	(2.507)
Observations	83	83	83

Note. Robust standard errors are in parentheses. *** p<.01, ** p<.05, * p<.1

A first result in the tables is that neither treatment has an effect on a group's maximum score. This is likely due to the fact that the exam scores are high on average (median of 94), capped at 100, and are highly left-skewed. Almost every group had at least one person with a perfect score, with the control group having an average group maximum of 98. This suggests that there is little room for increasing the maximum score in our experimental context. Thus, our focus will be on the group minimum (Model 3).

We test whether the coefficients in Model 3 are different across RCT 1 and RCT 2 using a Chow test. The test indicates that there is no difference in the Best Shot coefficient (p=0.8986), but there is a difference in the Weakest Link coefficient

(p=0.0003). Figure 1 visually displays the results of Model 3, showing the Group Minimum scores by treatment status and course format.

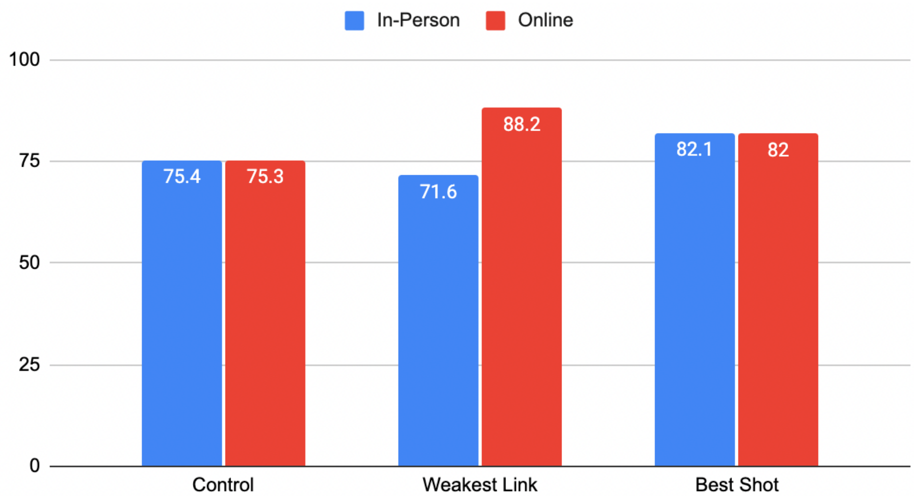

Figure 1: Group's Minimum Score by Treatment Status

For both the control and Best Shot conditions, there is no difference between the performances of the online and in-person sections. While we don't know whether participating in a study group in and of itself was helpful to students, the fact that we find no difference in control group performance for the two sections might indicate that any impact of working in a group does not depend on course format. The treatment effect on group minimum of Best Shot relative to control is about 0.45 standard deviations in both sections, which is consonant with the hypothesis that this type of incentive is less affected by course format.

In the online section, the Weakest Link incentive increased a group's minimum score by 0.86 standard deviations relative to control. This is a large effect and according to a Chow Test is significantly larger than the effect of the Best Shot treatment at conventional levels. However, there was no significant difference between control and Weakest Link in the in-person section. This might indicate the limited generalizability of this incentive structure in achieving its aim.

4. Discussion

This study piloted an exploration of team-based incentives. As the pedagogical approach worldwide is continuously moving online, we present two RCTs: one team-based incentive scheme in-person, and one online. Because the nature of the populations is different across the two settings, we cannot pinpoint exactly why we observe differences between in-person and online, but we do find interesting treatment effects. In particular, two main themes emerge. First, Weakest Link incentives may have limited generalizability. They had a large and significant impact on the intended outcome, but only for primarily international students interacting with each other virtually. There was no significant impact of the Weakest Link incentive for U.S. students who mostly interacted with each other in-person.

Second, we learn that there are multiple approaches to raise the floor of performance, thereby decreasing inequity within a group's performance. In addition to the Weakest Link incentive in the online setting, we found that the Best Shot incentive boosted the group minimum performance in both settings, at odds with our priors. While the primary intention of the Best Shot incentive was to increase the group's maximum, our evidence shows that it can increase a group's minimum performance. This is consistent with the theoretical equilibrium where most agents incentivized by Best Shot put in high effort, which can result in a higher group minimum as well. In this case, belief uncertainty of your personal ranking is helpful. More research is warranted as our interpretation is certainly ad hoc and post data collection.

Appendix A

Recruitment script and experimental instructions are in the online appendix here.

References

Bettinger, E., Fox, L., Loeb, S., & Taylor, E. S. (2017). Virtual classrooms: How online college courses affect student success. *American Economic Review*, *107*(9), 2855–2875.

Bettinger, E., Liu, J., & Loeb, S. (2016). Connections matter: How interactive peers affect students in online college courses. *Journal of Policy Analysis and Management*, *35*(4), 932–954.

Chowdhury, S. M., & Topolyan, I. (2016). Best-shot versus weakest-link in political lobbying: An application of group all-pay auction. *Social Choice and Welfare*, *47*(4), 959–971.

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., & Xu, Y. (2019). Measuring success in education: The role of effort on the test itself. *American Economic Review: Insights*, *1*(3), 291–308.

Haeckl, S., Sausgruber, R., & Tyran, J.-R. (2018). Work Motivation and Teams. *Available at SSRN 3239948.*

Harrison, G. W., & Hirshleifer, J. (1989). An experimental evaluation of weakest link/best shot models of public goods. *Journal of Political Economy*, *97*(1), 201–225.

Hossain, T., & List, J. A. (2012). The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science*, *58*(12), 2151–2167.

Lazear, E. P., & Shaw, K. L. (2007). Personnel economics: The economist's view of human resources. *Journal of Economic Perspectives*, *21*(4), 91–114.

Kraft, M.A. List, J.A., Livingston, J.A., & Sadoff, S. (2022). "Online Tutoring by College Volunteers: Experimental Evidence from a Pilot Program," AEA Papers and Proceedings, American Economic Association, vol. 112, pages 614-618, May.

Levitt, S. D., List, J. A., & Sadoff, S. (2016). The effect of performance-based incentives on educational achievement: Evidence from a randomized experiment. *National Bureau of Economics Research. Working Paper 22107.*

List, J. A., Livingston, J. A., & Neckermann, S. (2018). Do financial incentives crowd out intrinsic motivation to perform on standardized tests? *Economics of Education Review*, *66*, 125–136.

Orlov, G., McKee, D., Berry, J., Boyle, A., DiCiccio, T., Ransom, T., Rees-Jones, A., & Stoye, J. (2021). Learning during the COVID-19 pandemic: It is not who you teach, but how you teach. *Economics Letters*, *202*, 109812.

Riedl, A., Rohde, I. M., & Strobel, M. (2016). Efficient coordination in weakest-link games. *The Review of Economic Studies*, *83*(2), 737–767.

Weidmann, B., & Deming, D. J. (2021). Team players: How social skills improve team performance. *Econometrica*, *89*(6), 2637–2657.