Learning and Individual Differences xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Slowing the hare: Quick finishers and class performance on standardized tests

Chandler Phelps [a], Joseph Price [b,*]

[a] Australia National University, Australia
[b] Brigham Young University, United States

ARTICLE INFO

Article history:
Received 24 February 2015
Received in revised form 4 November 2015
Accepted 11 August 2016
Available online xxxx

Keywords:
Math performance
Field experiment
Distraction
Social rewards

ABSTRACT

We conduct a randomized field experiment in which elementary school students participated in a series of competitions using standardized math test problems. Prizes were rewarded to the top scorers in each round. Students in the treatment group were allowed to raise their hands when they finish a round to get their time recorded, while the students in the control group were asked to quietly check their answers until the round time had expired. We find that students in the timed group finish the test too quickly and as a result perform 0.32 of a standard deviation lower than those in the control group. This gap is even larger for boys. These results are consistent with a model in which students exchange higher performance for the social rewards of completing the task faster than their classmates.

1. Introduction

Student performance on a standardized test may be affected by factors completely unrelated to their understanding on the subject, the quality of teacher instruction, or the student's own preparation for the test. There are a number of small changes schools can implement on the day of the test that can boost student performance including providing breakfast (Meyers et al., 1989), altering the time of day the test occurs (Callan, 1999), or offering small rewards to students to do well on the test (Braun, Kirsch, & Yamamoto, 2011; Brown & Walberg, 1993; Levitt, List, Neckermann, & Sadoff, 2012). Identifying small factors that influence test scores can help schools to remove barriers that might prevent students from demonstrating their actual ability on a standardized test.

We examine a common way that tests are administered in schools that can have a large impact on student performance. With the increase in the use of computer-based testing, students frequently go to a school computer lab to take their standardized tests. A common feature of this arrangement is to allow students to return to the classroom as soon as they have completed the test. Leaving the room provides a clear signal to other students about how quickly each student completes the test. This can create a competing dilemma for students between taking as much time as possible to do well on the test and being perceived by other students as being fast at math.

We conduct an experiment that mimics the key features of this dilemma that students face between performance and social rewards. Our experiment involved elementary students competing for cash prizes in a set of short math competitions using the same type of questions they would experience on their standardized tests that year. After random assignment into treatment and control groups, we imposed the following conditions: the students in the treatment group were allowed to raise their hands when they finish a round to get their time recorded, while the students in the control group were asked to quietly check their answers until the round time had expired. In the treatment group, the time recorded played a role in breaking round ties at the end of the competition.

We find that students in the treatment group performed worse than the control group, and this difference increased as rounds progressed. Total scores in the treatment group were 0.32 of a standard deviation lower than those in the control group. Girls were hardly affected by the treatment, but boys performed much worse in the treatment group, with a drop in their performance of 0.66 of a standard deviation. Although high ability students in the treatment group did experience a performance drop, the effect of the treatment was the largest for the students least likely to do well on the test.

Building on work by Spinath, Freudenthaler, and Neubauer (2010), we discuss mechanisms that plausibly explain the differences by gender. As evidence that the performance drop cannot be attributed to the tiebreak incentive, we show that a) students are better off using nearly the full time on each round, b) students seem to understand that this is the optimal strategy, and c) students gradually deviate from this strategy as they compete to finish first. We conclude that

* Corresponding author at: 162 FOB, Brigham Young University, Provo, 84602, UT, United States.
E-mail address: joe_price@byu.edu (J. Price).

http://dx.doi.org/10.1016/j.lindif.2016.08.005
1041-6080/

students are adversely affected by the chance to let their peers know that they have finished. Lastly, we discuss possible policy implications and suggest strategies that teachers can implement to avoid a competing social incentive in their classrooms.

2. Background

From a policy perspective, a main area of concern about tests is to ensure that they measure what they are intended to measure while minimizing bias. The Ohio Department of Education established various standards for fair testing practices, including a requirement to "enact procedures that help to ensure that differences in performance are related primarily to the skills under assessment rather than to irrelevant factors," to prevent unfairness based on differences across gender, racial, or ethnic backgrounds (Ohio Department of Education, 2014).

A concern for individual educators and school administrators is to find ways to help their students perform better on standardized tests. Sometimes their efforts to appear excellent are aimed at manipulating results dishonestly (see Cullen & Reback, 2006), but sometimes altering the environment and circumstances of student test taking can result in significant differences in performance. In addition to improving teaching methods, many changes can be made to the circumstances surrounding tests. For example, because students have different time-of-day preferences, some schools schedule classes and tests so that as many students as possible will be at their most alert during their most important classes (Callan, 1999; Klavas, 1994). School breakfast programs have been shown to improve test scores and school attendance rates (Meyers et al., 1989). Interestingly, one school system that experienced remarkable improvements in student grade point average attributed the academic progress to hiring more cafeteria and custodial staff—a cleaner facility and better cafeteria may have led to better performance (O'Grady, 2014).

Strategies or policies implemented by teachers and administrators ought to consider the incentives and goals of students in addition to administrative goals. In addition to obvious academic achievement goals, most students are motivated to a great degree by wanting to display socially desirable attributes ranging from intelligence to apathy by their academic habits and performance. Ryan, Hicks, and Midgley (1997) explore a setting in which students place social goals with negative academic effects ahead of academic goals. In a study to determine what prevents a student from publicly asking a teacher for help, the results consistently show that when students pursue goals maintaining a certain image in front of others, either an image of academic ability or social reputation, the need for help is construed as a threat to self-worth. They highlight that some of these goals include seeking popularity, measuring academic achievement relative to others, and wanting approval from the teacher.

One specific type of social goal, called a "superiority" goal by Ford (1992), can lead students to seek to differentiate themselves from their peers by excellence in a chosen endeavor. Even a superiority goal can motivate either high or low achievement. If, for example, the goal is to prove superior indifference toward academics (if academic excellence is discouraged by a peer group), then poor performance could achieve that superiority goal. In our study, we find that a policy that encourages students to finish tests early fosters an environment where a possible social goal impedes academic achievement. When testing policies create incentives to complete the test quickly, students make more careless errors and overall perform worse than they would if they used all available testing time.

We add to existing literature by analyzing a common testing situation where students are allowed to finish early. How this type of testing situation affects performance by gender and by student ability can help educators perfect their testing environments so that tests accurately measure ability, and bias by ability and gender is minimized.

3. Method

We examine data from a field experiment in which elementary school students competed against their classmates to solve standardized math test problems. In the experiment, students were called upon to complete a total of 50 math problems split up into five 10-question rounds. Students were given 5 min to complete each round. All students were given the same tests each round and the tests were designed to have similar difficulty across rounds. To incentivize student effort, students were informed that cash prizes would be allocated via raffle once the students' tests were graded. Emphasis was placed on the random element of the prize assignment, ensuring the students that even a low-skilled student had a chance at a prize so long as they earned at least one raffle ticket.

The subject pool consisted of 86 fourth grade students all from the same elementary school in Utah County, Utah. The students were individually randomized into either the treatment or control group and were moved to separate rooms where researchers administered the tests. Before the start of each round, the researchers announced the prize structure in terms of number of raffle tickets. The top scores from each round were given the most raffle tickets, then the next highest were given fewer raffle tickets, and so on. Both the treatment and control groups' tests were conducted in this manner. In the control group, tiebreaks for each round were resolved via randomization. In the treatment group, students were informed that tiebreaks would be rewarded to the student who used the least amount of time to finish the round. When a student finished the round, they raised their hand so that the researcher could tell them their time, which the student then recorded on their test.

This tiebreak incentive to finish quickly is potentially a confounding factor in identifying social incentives. If we find that the treatment group performs worse than the control group, these results could be explained in at least the following three ways. First, if the tiebreak gives students a large enough edge in earning tickets, then students could profitably trade time for tiebreaks, thus decreasing their performance. Second, if trading time for tiebreaks is an unprofitable exchange, students could fail to understand this and thus adopt a faulty strategy that decreases their performance. Third, if trading time for tiebreaks is unprofitable and students understand that they are better off using the whole time, then social rewards can explain the worse performance. To attribute the results to the third explanation, we provide evidence that students could only profit from finishing early if they finish with less than 15 s left in the round, then we show that students seem to initially understand that using all or nearly all the time is the optimal strategy, but they deviate from this strategy as the rounds progress.

4. Results

We present descriptive statistics in Table 1 by round, and show evidence of the randomization of the groups. The treatment group performed worse than the control group across all rounds and the difference is statistically significant for rounds 3 and 5. Although males perform slightly better than females overall, this difference is not statistically significant.

Table 2 provides the main results of our regression analysis. The outcome variable is the number of questions that the student answered correctly and has been normalized to have a standard deviation of one. In column 1, we find that the treatment group experienced performance that was 0.32 of a standard deviation lower than the control group. To put this difference in perspective, the gap between average math test scores of white and black students at the end of first grade is about 0.7 of a standard deviation (Fryer & Levitt, 2004). In column 2, we find that the gap between the treatment and control groups increases over the five rounds of competition, with the gap widening by 0.11 of a standard deviation each round. In this regression, we center

Please cite this article as: Phelps, C., & Price, J., Slowing the hare: Quick finishers and class performance on standardized tests, *Learning and Individual Differences* (2016), http://dx.doi.org/10.1016/j.lindif.2016.08.005