Journal of Economic Behavior & Organization 100 (2014) 44–63

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

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

Awards unbundled: Evidence from a natural field experiment☆

Nava Ashraf[a,*], Oriana Bandiera[b], Scott S. Lee[c]

[a] HBS and NBER, United States
[b] Department of Economics and STICERD, LSE, United Kingdom
[c] Harvard Medical School and HBS, United States

ARTICLE INFO

Article history:
Received 21 September 2013
Received in revised form
17 December 2013
Accepted 1 January 2014
Available online 29 January 2014

JEL classification:
D84
D83
J33
M52
O15

Keywords:
Awards
Social comparison
Optimal expectations
Incentives

ABSTRACT

Organizations often use non-monetary awards to incentivize performance. Awards may affect behavior through several mechanisms: by conferring employer recognition, by enhancing social visibility, and by facilitating social comparison. In a nationwide health worker training program in Zambia, we design a field experiment to unbundle these mechanisms. We find that employer recognition and social visibility increase performance while social comparison reduces it, especially for low-ability trainees. These effects appear when treatments are announced and persist through training. The findings are consistent with a model of optimal expectations in which low-ability individuals exert low effort in order to avoid information about their relative ability.

1. Introduction

> What are the advantages which we propose by that great purpose of human life which we call bettering our condition? To be observed, to be attended to, to be taken notice of with sympathy, complacency, and approbation, are all the advantages which we can propose to derive from it.
> —Adam Smith, "Of the Origin of Ambition, and of the Distinction of Ranks," The Theory of Moral Sentiments (1759)

☆ This research is the fruit of a collaboration with the Ministry of Health, Zambia, to whose staff we owe a tremendous debt, especially Mutinta Musonda and Miriam Libetwa. We also thank Alice Nyirenda and the teachers of the CHA Training School for their patience and valuable feedback. Madeleen Husselman, Mardieh Dennis, Manpreet Singh, Katherine Otto, Johann Blauth, Alister Kandyata, Sandy Tsai, Elena Moroz, and, especially, Kristin Johnson provided skilled research assistance. We thank Charles Angelucci, Francesca Gino, Ed Glaeser, Daisuke Hirata, Kelsey Jack, Bryce Millett Steinberg, Emily Oster, Michele Rigolizzo, participants at the Stanford Institute for Theoretical Economics, Psychology and Economics session, two anonymous reviewers, and the Editor, William Neilson, for very helpful input. Financial support by the International Growth Centre is gratefully acknowledged.

* Corresponding author.
E-mail addresses: nashraf@hbs.edu (N. Ashraf), o.bandiera@lse.ac.uk (O. Bandiera), ssl@mail.harvard.edu (S.S. Lee).

* Corresponding author.
E-mail addresses: nashraf@hbs.edu (N. Ashraf), o.bandiera@lse.ac.uk (O. Bandiera), ssl@mail.harvard.edu (S.S. Lee).

0167-2681/$ – see front matter
http://dx.doi.org/10.1016/j.jebo.2014.01.001

The innate human desire for approbation might make status awards a cost-effective tool to incentivize good performance (Besley and Ghatak, 2008; Moldovanu et al., 2007). Awards can motivate employees to exert effort in order to gain recognition and visibility, both of which are free for the employer to bestow but valuable to the employee. However, given that awards derive their value from their scarcity, they inevitably facilitate social comparisons, which might be *demotivating* to employees.[1]

Our goal in this paper is to "unbundle" awards—that is, to provide evidence on the mechanisms that underlie their effectiveness and potential harm. We conduct a natural field experiment to separately identify channels through which awards can affect behavior, unbundling the effect of social comparison through the (private or public) disclosure of rank information, from the effect of employer recognition and social visibility.

We study the effect of awards in the context of a nationwide training program for health workers in Zambia. Our agents are 314 health workers recruited from 162 rural communities and brought to professional school for a one-year training program aimed at teaching community-based health care. After training, trainees will be employed by the Ministry of Health and deployed to their communities of origin, where they will become the first point of contact for health services. Incentivizing learning is key in this context because trainees have no previous medical training; thus, the skills they learn will determine their effectiveness in the field.[2]

During the training program, trainees take courses on several topics, on which they are tested at baseline (at the beginning of the year) and at the end of each course. The field experiment randomly allocates trainees to two broad classes of treatments (in addition to control): those that only provide information on trainees' relative performance, and those that also offer awards. After each exam, trainees in the control group receive a letter from the school reporting their absolute score and their value added, measured as improvement over their baseline score for the given course. Trainees in the "private social comparison" treatment (T1) receive the same letter with added information on their rank in the class distribution of both absolute score and value added. Trainees in the "public social comparison" treatment (T2) receive the same letter as in the previous treatment as well as the names of the top four performers in the class (top two by absolute score and top two by value added).

The third and fourth treatments add awards to rank information. Awards are given to the trainees with the top two scores and those with the top two most improved scores (from baseline). The latter ensures that weaker trainees have a chance to win and are therefore motivated by the award. In the "employer recognition award" treatment (T3), the top four performers receive a congratulatory letter from the Ministry of Health. In the "social visibility award" treatment (T4), one of the top four performers is randomly selected to be featured in an interview, which is printed along with the candidate's photo in a newsletter distributed back to their community of origin. Under a linearity assumption the difference between each of the award treatments and the "public social comparison" treatment isolates the effect of awards from the effect of the social comparisons they inevitably create.

Our setting has three key features that make it ideal for the purpose of this experiment. First, since trainees take four courses during the study period and treatments are announced at the beginning of the first course, we can assess whether they change their behavior in anticipation of receiving rank information and awards or only after these have been provided. Second, during training, the performance of the health workers is measured by an institution (the school) that is different from their employer (the Ministry), and the health workers are physically removed from their home communities. This allows us to separate the effect of information on relative performance (provided by the school) from that of the employer's recognition and from visibility to one's social circle (the home community). In most settings, the employer measures and provides information on performance, such that the provision of information necessarily entails some recognition. The fact that trainees are distant from their communities is similarly useful, as no treatment other than the social visibility award can be used to enhance visibility within their social circles. In most settings in which agents are co-located with their social network, any treatment that reveals their rank in the distribution could potentially be used to enhance visibility.

Third, performance in this setting is uni-dimensional (trainees are solely meant to attend classes and study the topics on the syllabus), and thus not subject to a multitasking problem in the face of additional incentives.[3] Moreover, performance can be measured objectively and precisely by test scores. Value added in test scores is a good measure of learning, as is often the case when knowledge at baseline is very limited (Muralidharan and Sundararaman, 2011).

[1] Lazear (1989) describes the tradeoff in relative performance evaluation: it could motivate employees to work harder, but could also create an excessively competitive work environment and decrease employee morale. Major et al. (1991)'s review of the literature in social psychology provides evidence on the demotivating effect of social comparisons. A related literature in management emphasizes the importance of concealing relative performance information to improve employee motivation (Milkovich and Newman, 1996).

[2] A number of field experiments have evaluated the effect of financial incentives on student learning; the evidence of their efficacy is mixed (Fryer, 2011; Angrist and Lavy, 2009; Kremer et al., 2009; Leuven et al., 2010). In particular, we know of two experimental studies that examine the impact of rank disclosure on academic performance (Azmat and Iriberri, 2010; Tran and Zeckhauser, 2012). These studies fundamentally differ from ours, however, in that neither announces the rank treatment in advance of disclosing rank information. Thus, the studies capture only the *ex post* effect of rank disclosure, whereas we critically distinguish the *ex ante* anticipatory effect from the *ex post* information effect.

[3] The training curriculum comprises both a classroom component, which the school's exams directly measure, and a practical component, in which trainees visit local field sites to apply knowledge and skills gained in the classroom. While trainees' performance during practicals was not separately evaluated, it is important to note that the practical component was explicitly designed to reinforce concepts learned in the classroom, and thus effort across both venues would be expected to affect exam performance. The exams were designed to assess both theoretical knowledge and practical skills.

The analysis reveals that social comparison and awards have opposite effects on performance. Compared to trainees in the control group, the "private social comparison" treatment significantly reduces test scores by 0.31 standard deviations, and the "public social comparison" treatment reduces it by 0.38 standard deviations. Importantly, the two social comparison treatments reduce performance as soon as they are announced—i.e. before trainees get the first letter with their rank information. A likely explanation is that individuals value the *belief* that they have high relative ability, and the anticipatory utility this provides. They may thus prefer to exert low effort in order to decrease the informativeness of the ranking signal. This is akin to refusing to take a medical test for a disease, so as to justify holding an optimistic belief about one's health status (Oster et al., 2013), and is consistent with the literature on belief utility and information avoidance (Bénabou and Tirole, 2002; Köszegi, 2002). In our context, the negative effect due to information avoidance seems to dominate the potential positive effect of competition among trainees (Charness and Grosskopf, 2001; Freeman and Gelber, 2010).[4]

Adding awards to rank information significantly improves performance. Compared to trainees in the "public social comparison" treatment, mean scores of trainees in the "employer recognition award" and "social visibility" treatments are 0.38 and 0.44 standard deviations higher, respectively. Recognition from one's employer can increase performance if agents have career concerns (Dewatripont et al., 1999) or preferences for reciprocity (Fehr and Schmidt, 1999). The net effect of either type of award is nil: because the positive effects of employer recognition and social visibility are nullified by the negative effect of providing information on relative ranks, trainees in the two award treatments perform as well as trainees in control.[5]

Quantile treatment estimates show that both the negative effect of social comparison and the positive effects of recognition and visibility are stronger on the left tail of the conditional productivity distribution, and both are zero at the top two deciles. In line with this, we also find that these negative and positive effects are stronger for low-ability trainees, and zero for high-ability trainees, where ability is measured by baseline test score. The fact that the negative effect of ranking is stronger for the weakest trainees is intuitive, as these are more likely to receive a negative signal about their skills.[6] That the positive effects of recognition and visibility are also stronger for the weakest trainees may be due to the fact that, since awards are given to trainees with the highest value added (the "most improved"), those who start at the bottom have a better chance to win. That the effects are zero for the top two deciles is consistent with the fact that scores are capped, such that the highest performing trainees have little room for improvement. Evidence from settings with no cap on performance suggests that, in contrast, awards are most effective at the top of the distribution (Nalbantian and Schotter, 1997; Müller and Schotter, 2010; Bandiera et al., 2013; Leuven et al., 2010).

An important implication of these findings is that, due to the negative effect of social comparison, awards might increase the dispersion of performance by weakening the weakest. In our setting, this may matter just as much as—and perhaps more than—mean effects. Health workers wield considerable power to influence the utility of their patients, especially for the worse (e.g., missed diagnoses, incorrectly dosed medications, wrong-site amputations). Both because of the potential for harm and the government's mandate to ensure equity of services across populations, the distribution of performance during training in this field experiment is crucial. The findings thus suggest caution in using mechanisms that facilitate interpersonal comparisons in contexts in which worsening performance at the bottom of the distribution is costly. This is particularly germane in the policy domains of public service delivery, such as in health and education, where the use of awards is increasingly common (Mathauer and Imhoff, 2006; Ashraf et al., 2013), but where distributional effects on agents' performance could have severe welfare consequences for those they are serving.

The paper is organized as follows. Section 2 describes the context and the experiment. Section 3 presents the empirical analysis. Section 4 interprets the findings in light of optimal expectations theory (Brunnermeier and Parker, 2005; Oster et al., 2013). Section 5 concludes.

2. Experimental design

2.1. Context

In 2010, the Government of Zambia (GOZ) launched a national effort to create a new civil service cadre called the Community Health Assistant (CHA). In the program's first year, GOZ sought to recruit, train, and deploy approximately 300 Community Health Assistants across seven of Zambia's nine provinces.[7] Within these seven provinces, based on population

[4] The empirical evidence on the *ex post* effect of providing rank information is markedly mixed. Tran and Zeckhauser (2012), Azmat and Iriberri (2010), and Bradler et al. (2013) show positive effects of rank information on performance, while Bandiera et al. (2013) and Barankay (2012) find that this information reduces productivity.

[5] This echoes the findings of Bandiera et al. (2013) who show the impact of receiving information on relative rankings can offset the positive impact of monetary prizes. In their setting, however, the effect is driven by changes in team composition.

[6] Tran and Zeckhauser (2012) find that private rank disclosure motivates high-ability more so than low-ability trainees, whereas Azmat and Iriberri (2010) find uniformly positive effects across the distribution.

[7] This is the first generation of community health workers trained by the Government of Zambia. Although this paper does not evaluate the efficacy of community health workers, they have been shown to improve health outcomes in randomized controlled trials in other countries (Baqui et al., 2008; Bhandari et al., 2011; Spencer et al., 2011).

density, GOZ chose the 48 most rural of the 58 constituent districts, and across these, GOZ identified 165 underserved communities, each with an average population of 3500 individuals (Government of Zambia, 2011). From each community, the intention was to recruit two Community Health Assistants. The recruitment and selection process occurred at the community level, with on-the-ground implementation coordinated by district health officials.

In total, 314 individuals accepted GOZ's training offer and moved to Ndola, Zambia's second-largest city, to join a newly established training school. The training program lasted one year and was structured in a modular format (see Fig. A.1).[8] At the beginning of the training, all trainees took a baseline exam which covered all the material that would be subsequently taught during the training year. After each course, trainees took an exam covering the material taught in that course. During this time, the trainees engaged only in attending classes and studying for their exams. The training school was divided into five classrooms, each accommodating roughly 60 trainees. The school was led by a principal and staffed by ten full-time teachers. The trainees were not formally paid during the training year, but their tuition and room and board were covered by the Ministry. In addition, the participants were aware that wages upon completing training would be the same for all CHAs, and that opportunities for promotion would be available. The program is thus effectively training "on the job" and career concerns were likely at play.

Once deployed to the field, the CHAs were to routinely visit households and provide a variety of home-based services: basic medical care to any sick persons, health education and counseling, and referrals to nearby health facilities as needed. Two key features of the job illuminate how critical the training period was to subsequent performance. First, CHAs were expected to provide a very broad scope of health services to all age groups. Second, they were to do so with a great deal of autonomy. In contrast to nurses, whose job it classically is to implement a physician's orders, and who typically are not trained to diagnose and treat illnesses, the CHA is more like a physician, making decisions autonomously without direct supervision. The human capital required to perform such varied activities is substantial, and the one-year training was consequently critical.

2.2. Performance measurement

Since the trainees' only task during training was to attend classes and study the material taught therein, we measure performance by exam test scores. All exams were based on multiple choice questions created by external medical advisors based on the content of the official training textbook.[9] Grading was done electronically by the research team. After each course exam, each student's completed exam was returned to him or her, along with an answer key.

Several measures were taken to prevent cheating or gaming. First, all exams were administered under timed, proctored conditions. Second, each exam had four versions, in which the order of the answer choices for each question was randomly varied. The exam versions were distributed within a classroom in sequential fashion, such that no two neighboring trainees had the same exam version. Even if they were alike, the exam version was indicated discreetly in the lower corner of the exam, such that a student attempting to cheat by copying a neighbor's answers would have difficulty determining whether the neighbor's exam version was the same as hers.

2.3. Experimental design

We worked with the school administrators to randomly allocate trainees to five groups (one control and four treatments) of approximately 60 individuals each, stratified by average baseline exam score and other potential determinants of performance.[10] To minimize contamination across treatments, all trainees in a given treatment group were assigned to the same classroom, and classrooms were kept together for the entire duration of the experiment that lasted nine consecutive months. For each course, each classroom was co-taught by two teachers. The teachers rotated and were assigned to different classrooms after each course, using a schedule that was determined by the principal and by the researchers with the aim of ensuring even coverage of teachers across classrooms. Teachers and trainees in all groups used the same textbook that was developed by GOZ for the CHA training.

The experimental treatments are as follows (see Fig. 1 for a schematic diagram). After each exam, trainees in the control group receive a letter from the school reporting their absolute score and their value added over their baseline sub-score for the relevant course content after each exam (see Appendix Fig. A.2). Trainees in the "private social comparison" treatment (T1) receive the same letter from the school with added information on their rank in the distribution of both absolute score and value added (see Appendix Fig. A.3). Trainees in the "public social comparison" treatment (T2) receive the same letter from the school with added information on their rank and the names of the top four performers in the class—that is, the top

[8] The training curriculum was designed in the months leading up to the training launch through a consultative process led by the Zambian Ministry of Health, with input from health educators, clinicians, and public health and development practitioners.

[9] Multiple choice questions are a standard question type on exams in Zambian secondary and tertiary education.

[10] We used the "T-min-max" method to balance the classrooms on gender, baseline exam score, any previous health experience, employment status, and district-level recruitment strategy (different districts advertised the CHA position with different emphases on social vs. private benefits). For a discussion of this randomization method, see Bruhn and Mckenzie (2009).

	Components				
	After each course exam, each trainee is given a private letter that indicates his or her **absolute score** on the exam as well as his or her **value added** with respect to his or her course sub-score in the baseline exam.	The letter reports the trainee's **rank** within the classroom on absolute score and value added, respectively.	The two trainees with the highest exam score and the two trainees with the highest value added are **publicly named in the letters** given to all trainees within the classroom.	The four top-performing trainees are given a **personalized letter from the program director at the Ministry of Health** congratulating them on their achievement.	One of the four top-performing trainees is randomly chosen to be **profiled in a newsletter** that is sent to all of the trainees' communities of origin.
C: Control	✔				
T1: Private Social Comparison	✔	✔			
T2: T1 + Public Social Comparison	✔	✔	✔		
T3: T2 + Employer Recognition Award	✔	✔	✔	✔	
T4: T2 + Social Visibility Award	✔	✔	✔		✔

Fig. 1. Schematic of experimental treatment conditions.

two by absolute score and the top two by value added (see Appendix Fig. A.4).[11] We include the "most improved" category based on theoretical and empirical evidence that multiple prizes at different points in the distribution are more motivating across the distribution than a single prize (Moldovanu and Sela, 2001; Freeman and Gelber, 2010; Bradler et al., 2013). Importantly, trainees did not know that they could have won "value added" prizes when they took the baseline test, so there is no scope for gaming by obtaining a low score at baseline.

Trainees in the "employer recognition award" treatment (T3) receive the same letter as those in the "public social comparison" treatment, and, in addition, the top four performers receive a congratulatory letter from the Ministry of Health, handwritten and signed by the director of the department that runs the CHA program (see Appendix Fig. A.5 for an example). Trainees in the "social visibility award" treatment (T4) receive the same letter as those in the "public social comparison" treatment, and, in addition, one of the top four performers is randomly selected to be featured in an interview that is printed together with the candidate's photo in a newsletter that is distributed back to their community of origin (see Appendix Fig. A.6 for an example).

The timeline of the experiment is as follows. Trainees took four sequential courses during the experimental period (covering 9.5 out of the 12 months of training) and sat exams at the end of each course. Courses varied in duration from two weeks (course 2) to four months (course 1). At the beginning of the first course, trainees were told the content of the letters that they would receive after each exam, and a reminder was delivered toward the middle of the same course. In an assessment given to all trainees after the reminder announcement, 79% of CHAs responded to the question, "How clear do you find the information presented in the letters?," with "very clear" and an additional 14% with "somewhat clear." In addition, trainees in the two awards treatments were shown sample employer recognition letters and community newsletters, respectively.

Table 1 shows balance across the five groups on a number of variables including baseline exam score, English exam score, gender, age, health experience and employment status. Only 3 out of 24 (12%) pairwise differences are different from zero

[11] The design decision to make comparisons only at the very top public was done in consultation with senior teachers who were concerned about severe demotivating effects of publicizing the lowest performers. Although ranking throughout the entire distribution is publicly displayed in many professional schools in Zambia, there is no consensus on whether this is helpful or harmful.

Table 1
Balance across treatments.

	C: Control	T1: private social comparison	Test of equality: C = T1	T2: T1 + public social comparison	Test of equality: C = T2	T3: T2 + employer recognition award	Test of equality: C = T3	T4: T2 + social visibility award	Test of equality: C = T4	Joint test of equality: C = T1 = T2 = T3 = T4
Baseline test score (0–100)	0.610	0.600	$p = 0.541$	0.598	$p = 0.463$	0.609	$p = 0.936$	0.613	$p = 0.864$	$p = 0.865$
	(0.088)	(0.090)		(0.094)		(0.087)		(0.084)		
English test score (0–100)	39.97	44.56	$p = 0.048$	42.81	$p = 0.219$	40.1	$p = 0.944$	43.26	$p = 0.155$	$p = 0.193$
	(11.89)	(11.07)		(15.70)		(10.70)		(13.62)		
Number of O-levels	5.97	5.93	$p = 0.898$	5.89	$p = 0.807$	5.91	$p = 0.847$	5.86	$p = 0.720$	$p = 0.993$
	(1.36)	(1.56)		(1.57)		(1.68)		(1.67)		
Age (years)	26.24	26.02	$p = 0.843$	27.01	$p = 0.466$	27.33	$p = 0.313$	28.47	$p = 0.040$	$p = 0.169$
	(5.73)	(5.60)		(5.95)		(6.05)		(6.60)		
Gender (=1 if female)	0.459	0.583	$p = 0.174$	0.484	$p = 0.784$	0.484	$p = 0.784$	0.516	$p = 0.529$	$p = 0.695$
	(0.502)	(0.497)		(0.504)		(0.504)		(0.504)		
Unemployed at time of application (=1 if yes)	0.705	0.639	$p = 0.460$	0.565	$p = 0.112$	0.565	$p = 0.112$	0.565	$p = 0.112$	$p = 0.393$
	(0.460)	(0.484)		(0.500)		(0.500)		(0.500)		
Work experience in the health sector (=1 if yes)	0.361	0.344	$p = 0.855$	0.484	$p = 0.168$	0.452	$p = 0.309$	0.516	$p = 0.082$	$p = 0.221$
	(0.484)	(0.479)		(0.504)		(0.502)		(0.504)		
District used community recruitment (=1 if yes)	0.475	0.525	$p = 0.589$	0.500	$p = 0.786$	0.419	$p = 0.537$	0.484	$p = 0.926$	$p = 0.828$
	(0.503)	(0.504)		(0.504)		(0.497)		(0.504)		
Number of trainees	61	61		62		62		62		

Notes: Table reports means and standard deviations (in parentheses) for the five experimental conditions, as well as tests of equality for each control-treatment pair. Baseline test score is the score (0–100) on a baseline test administered at the beginning of the training program to measure trainees pre-existing knowledge of the materials taught during the training program. English test score is the score (0–100) on a test administered at the beginning of the training program to measure trainees' knowledge of English. Unemployed at time of application is a self-reported variable that equals 1 if the trainee had no full- or part-time employment when they applied for the CHA position. Work experience is a self-reported variable that equals 1 if the trainee has held a paid or unpaid position in the health sector in the past. District used community recruitment equals 1 if trainees come from a district where the main mission of the CHA job was advertised as "serving the community;" it equals 0 if trainees come from a district where the main mission of the CHA job was advertised as "fostering your career."

at the 10% level, as expected by chance. During the course of the training year, 6 of 314 (1.9%) trainees dropped out of the program. These trainees were distributed across four of the five treatment conditions (one in the control group, one in T2, three in T3, and one in T4) and have been excluded from our analysis. Due to the very low rate of attrition, any differences between the attrited and the non-attrited, or between treatment conditions as a result of the attrition, are indistinguishable from random error.

3. Analysis

3.1. Methodology

During the experimental study period, trainees attend four courses and take an exam at the end of each. Trainees in all treatments take the same courses and complete the same exams at the same time.

To evaluate treatment effects we estimate the following model using panel data at the trainee-course level:

$$score_{ic} = \alpha + \sum_{t=1}^{4} \beta_t T_i^t + \gamma B_{ic} + X_i \delta + I_{ic} \eta + \varepsilon_{ic} \tag{1}$$

where $score_{ic}$ is trainee i's test score in course c normalized by the mean and standard deviation of test scores for the same course in the control group; treatment effects are thus measured in standard deviation units. B_{ic} is trainee i's score in the baseline exam content for course c that was administered at the start of the training program. The difference between the "post" exam score and the baseline score is used to measure the value added that is reported in the treatment letters and to rank trainees for the "most improved" awards. X_i are individual characteristics that include all stratification variables as well as age and trainee i's test score for an English language test. I_{ic} are teacher-specific traits such as teacher ratings (as reported anonymously by the trainees) and expertise in the subject matter of course c. Since all trainees take the same courses at the same time, different trainees have different teachers for the same course. The teacher rotation schedule was determined by the principal and by the researchers with the aim of ensuring even coverage of teachers across classes.

Standard errors are clustered at the trainee level, as trainee-specific unobservables in the error term create correlation within trainee. Since courses are of different durations and trainees have more time to exert learning effort the longer the course is, we weight observations by course duration.

The parameters β_t measure the causal effect of treatment t vis-à-vis the control group under the identifying assumption of no contamination across treatments. Contamination can occur if the response to treatment j is affected by the knowledge that treatment k exists. For instance, trainees might respond differently to being given information on relative rankings if they know that other trainees are also getting employer recognition awards while they are not. To minimize the risk of contamination, we allocate trainees in different treatments to different classes and keep classes together for the duration of the experiment. Trainees were told that other classrooms may receive different types of letters with their exam scores, as this was the pilot year of the government program and different classrooms were trying different things. Trainees were not told that this was a research experiment, thus mitigating potential experimenter demand effects. Reassuringly, no student ever complained or raised the issue of different treatments for the entire duration of the experiment. Despite these precautions, trainees in the non-award treatments could have come across the award recognition letter and community newsletter given to trainees in other treatments. This, however, could have occurred only after the awards were distributed—that is, after the first test. To provide evidence on the practical relevance of contamination, we estimate Eq. (1) using scores from the first exam only, which was taken before trainees could have seen letters given to their colleagues in other treatments.

To separate the different mechanisms through which awards can affect performance, treatments are designed to be cumulative so that the "public social comparisons" treatment (T2) also contains information about relative ranks (T1), and the two awards treatments (T3 and T4) contain the same information as T2. Under the assumption that the effect of each component does not interact with the effect of the others, the net effect of each additional component can be identified by the appropriate linear combination of β_t estimators. For instance, the net effect of employer recognition is given by $\beta_3 - \beta_2$. We report all relevant linear combinations at the foot of each table.

3.2. Average treatment effects

Table 2 reports the estimates of Eq. (1). Test scores are normalized by the mean and standard deviation of the control group for each test so that treatment effects are measured in standard deviation units. Columns (1) and (2) estimate average treatment effects with and without teacher characteristics (average teacher rating and whether at least one of the teachers had specific expertise in the subject matter).

Three findings are of note. First, trainees in treatments 1 (private social comparison) and 2 (public social comparison) perform significantly worse than trainees in the control group. Estimates from the baseline specification in Column (1) show that giving private information about relative rankings lowers performance by 0.31 standard deviations, while rank information combined with a public list of the top four performers lowers performance by 0.38 standard deviations. By contrast, trainees in treatments 3 and 4, where top performers receive awards either in the form of a letter from the Ministry of Health or a profile in the organization's newsletter, performed the same as trainees in the control group.

Table 2
Average treatment effects.

	(1)	(2)	(3)	(4)	(5)
	Baseline	Teacher controls	Heterogeneous effects by exam period		
			First exam	Following exams	Test of equality
T1: private social comparison	−0.308*	−0.351**	−0.287*	−0.322*	$p = 0.844$
	(0.161)	(0.171)	(0.168)	(0.194)	
T2: T1 + public social comparison	−0.379***	−0.409***	−0.456***	−0.324**	$p = 0.378$
	(0.145)	(0.155)	(0.172)	(0.156)	
T3: T2 + employer recognition award	0.005	−0.044	0.128	−0.080	$p = 0.123$
	(0.133)	(0.174)	(0.165)	(0.136)	
T4: T2 + social visibility award	0.064	0.122	−0.140	0.112	$p = 0.087$
	(0.135)	(0.148)	(0.165)	(0.143)	
Net effect of public social comparison (T2 – T1)	−0.071	−0.058	−0.170	−0.003	$p = 0.381$
	(0.166)	(0.164)	(0.174)	(0.201)	
Net effect of employer recognition award (T3 – T2)	0.384***	0.365**	0.585***	0.244*	$p = 0.026$
	(0.141)	(0.161)	(0.175)	(0.148)	
Net effect of social visibility award (T4 – T2)	0.443***	0.531***	0.316*	0.436***	$p = 0.458$
	(0.141)	(0.151)	(0.170)	(0.153)	
Trainee controls	Yes	Yes		Yes	
Teacher controls	No	Yes		No	
Number of clusters (trainees)	307	307		307	
Number of observations (trainee-courses)	1149	850		1149	
Adjusted R^2	0.213	0.212		0.224	

Notes: Table reports OLS estimates, weighted by course duration, with standard errors clustered at the trainee level in parentheses. Dependent variable is normalized exam score, normalized by the mean and standard deviation of the control group for each exam. Trainee controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy. Teacher controls include: average teacher rating and whether at least one of the teachers had specific expertise in the subject matter. Both variables are defined at the course-treatment level. Columns (2) and (3) report coefficients estimated in the same regression where we include all treatments interacted with an indicator variable that takes value 0 in the first period and 1 thereafter.

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Importantly, since trainees can do very little other than studying during the program, we can rule out that the drop in exam performance is associated with an increase in effort devoted to other tasks. This strongly suggests that there is no compensating effort that can make up for negative performance during the training period.

Second, we can identify the net effect of each additional treatment component by differencing out common elements under the independence assumption discussed above. This exercise reveals that the difference between T2 and T1 is small and not statistically significant, suggesting that making the list of top performers public does not motivate individuals to work harder. This might be driven by the possibility that other trainees are not the natural peer group to whom these individuals compare themselves. Alternatively, given that in T1, individual ranks were printed on a letter, top performers in T1 could have easily made themselves known even if the list was not public.

Similarly, we can identify the effect of employer recognition and social visibility by differencing out the effect of providing rank information. Column (1) of Table 2 shows that both mechanisms have a strong positive effect on performance. The difference between T3 and T2 is 0.38 standard deviations, whereas the difference between T4 and T2 is 0.44 standard deviations. Both effects are precisely estimated.

Third, the estimated treatment effects are not sensitive to the inclusion of teacher controls (i.e., average teacher rating in students' evaluations and whether at least one of the teachers had specific expertise in the subject matter). This allays the concern that differences among treatments could be due to correlated unobservables at the class-course level. Since including teacher controls reduces the sample as these were not collected for course 2, we use the specification without teacher controls in what follows.

Taken together, the findings indicate that the three mechanisms described at the outset of this paper are relevant in this setting, but their signs and magnitudes differ: social comparisons (whether private or public) weaken motivation whereas employer recognition and social visibility strengthen it.[12] Thus, depending on the size of these opposing effects in a given context, the net effect of relative performance-based awards is uncertain.

[12] It could be argued that social visibility (in our case, visibility in one's home community) is a more powerful motivator in this setting; indeed, while the estimated effects are similar, the underlying treatment strength is different, as all top four performers receive the employer recognition award, whereas only one of them is randomly selected for the social visibility award.

3.3. Timing

Since treatments are announced at the beginning of course 1, we are able to identify whether their effect differs before and after the first round of letters containing rank information and awards are handed out. This test can help us shed light on why individuals change their behavior in response to rank information. For instance, if individuals are uncertain about the production function of test scores and use rank information to update on the marginal return of their effort, we would expect them to respond differentially before and after the information is provided. In contrast, if individuals know the production function and intrinsically care about their performance rank, we would expect their responses to be the same before and after.

To test whether responses vary before and after rank information is first disclosed, we estimate Eq. (1) augmented by the interactions of all treatment indicators with an indicator variable that takes value 0 for the first test (i.e., before treatment letters were distributed) and 1 for the following tests. Columns (3) and (4) of Table 2 report the estimated treatment effects on performance in the first test and performance in the following tests, respectively. Column (5) reports the *p*-value of the null hypothesis that these are equal.

The findings show that social comparisons, whether private or public, lead to a similar reduction in performance in the first and subsequent tests. Thus, this effect cannot be driven by "demotivation" in the traditional sense, or updating on the marginal return to effort, since trainees' performance dropped *in anticipation* of receiving rank information. Likewise, the findings are at odds with the assumption that trainees have competitive preferences (Charness and Grosskopf, 2001; Freeman and Gelber, 2010); otherwise, we should have observed an *increase* in effort in anticipation of receiving rank information. In Section 4, we describe a model of optimal expectations that is consistent with this pattern, in which trainees choose low effort (i.e., self-sabotage) to avoid rank information that delivers a signal about their relative ability.

In addition, columns (3) and (4) cast doubt on the relevance of the concern that responses to treatments 1 and 2 might be contaminated by the awareness of the other two treatments. Indeed, we show that providing rank information with or without a public list of top performers reduces performance by the same amount even before trainees in these treatments were likely to have become aware of the other treatments.[13]

The findings also show that both the employer recognition award and the social visibility award are effective at increasing performance conditional on rank information both before and after the first exam. The effect of the recognition award becomes weaker after the first exam, possibly because individuals revise their chance of winning downwards once the first round of rank information is revealed. We do not, however, find a similar pattern with respect to the social visibility award. We will return to this below when we allow responses to differ by baseline ability levels.

3.4. Distributional effects

Awards are likely to affect individuals at different points in the performance distribution differently, as the incentive power of awards should be stronger for those who have a greater chance of winning the award—in our case, more able trainees and those who have more potential to improve. Similarly, the effect of social comparison is likely to depend on whether, given their knowledge of their own ability and expectation about others', individuals expect to be ranked high or low.

To provide evidence on these distributional issues we estimate quantile treatment effects at each decile. Fig. 2 shows this graphically, and Appendix Table A.1 reports the regression coefficients by decile. The estimated treatment effect at each decile is the difference in conditional test score between two statistical trainees—one in the treatment group and one in the control group—both positioned at the same decile of the distribution of test scores within her group. Fig. 2 shows that both the negative effect of relative rankings and the positive effects of recognition and visibility are stronger on the left tail of the conditional test score distribution, and they gradually diminish to zero at the top two deciles. Standard errors reported in Appendix Table A.1 show that all effects are statistically different from zero until the seventh decile. Taken together, these findings indicate that information on relative ranks, with or without a public component, increases the dispersion of performance by reducing performance on the left tail.

To provide further evidence on this issue, Table 3 allows treatments to have heterogeneous effects by trainees' ability, measured by their score in the baseline exam. The estimates show that both the negative effect of relative rankings and the positive effects of recognition and visibility are stronger for low-ability trainees and zero for high-ability trainees. The fact that the negative effect of relative rankings is stronger for the weakest trainees is intuitive, as these are more likely to receive a negative signal about their skills. That the positive effects of recognition and visibility are also stronger for the weakest trainees is presumably due to the fact that, since awards are given to trainees with the highest value added (the "most improved"), those who start at the bottom have a better chance to win. That the effects are zero for the top two deciles is consistent with the fact that scores are capped, such that the very best trainees have little room for improvement.[14]

---

[13] It was of course possible for trainees in the award treatments to tell their colleagues in other treatments before the first exam, and before the awards were distributed. However, these claims should have been more credible and hence contamination stronger after the awards letters became potentially visible to all. The fact that treatment effects are stable throughout courses casts doubt on this.

[14] Across the entire sample, mean absolute post-test score in the highest decile is 90.0% and in the second-highest decile is 86.5%. The single highest absolute score across all exams was 96.1%. As our team, in consultation with medical training experts, wrote the exams to ensure quality and precision,

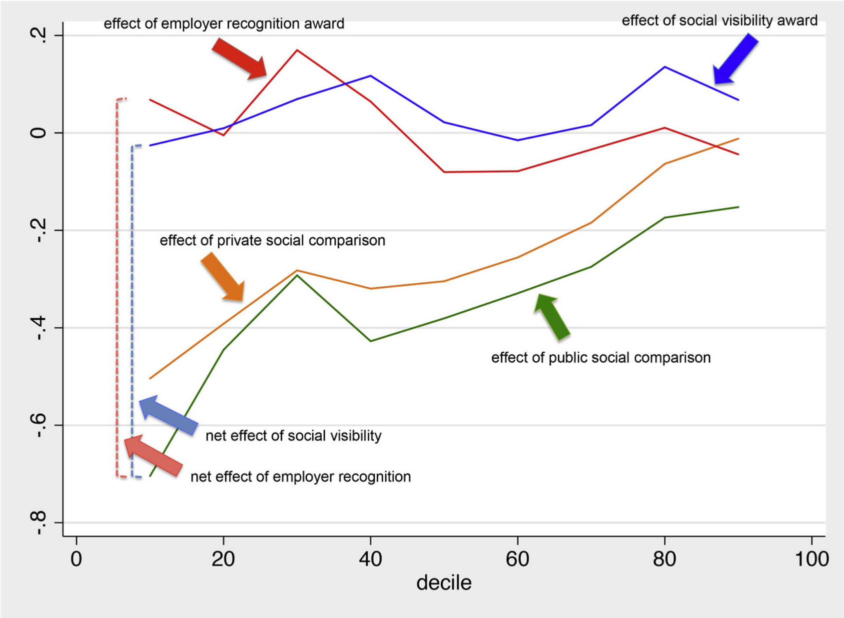


Fig. 2. Quantile treatment effects. *Notes*: Each line connects treatment effects of each treatment estimated at each decile. Point estimates and standard errors are reported in Table A.1. Individual controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy.

Table 3
Heterogeneous treatment effects by baseline test score.

	(1) First tercile of baseline test score	(2) Second tercile of baseline test score	(3) Third tercile of baseline test score
T1: private social comparison	−0.672**	−0.233	0.010
	(0.298)	(0.271)	(0.213)
T2: T1 + public social comparison	−0.812***	−0.179	−0.113
	(0.260)	(0.199)	(0.203)
T3: T2 + employer recognition award	−0.280	0.157	−0.005
	(0.216)	(0.250)	(0.159)
T4: T2 + social visibility award	−0.102	0.204	−0.031
	(0.309)	(0.223)	(0.177)
Net effect of public social comparison (T2 – T1)	−0.140	0.054	−0.153
	(0.334)	(0.243)	(0.214)
Net effect of employer recognition award (T3 – T2)	0.532**	0.336	0.107
	(0.255)	(0.214)	(0.166)
Net effect of social visibility award (T4 – T2)	0.710**	0.383**	0.082
	(0.342)	(0.183)	(0.173)
Trainee controls	Yes	Yes	Yes
Number of clusters (trainees)	92	107	107
Number of observations (trainee-courses)	350	401	398
Adjusted R^2	0.186	0.079	0.157

Notes: Table reports OLS estimates, weighted by course duration, with standard errors clustered at the trainee level in parentheses. Dependent variable is normalized exam score, normalized by the mean and standard deviation of the control group for each exam. Individual controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy.

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Finally, Appendix Table A.2 allows treatments to have heterogeneous effects by trainees' ability and exam timing. This confirms that most effects are driven by trainees at the bottom of the ability distribution and that, as discussed above, rank information leads to a similar reduction in performance in the first and subsequent tests. The results also show some evidence that the positive effect of awards becomes weaker after the first exam for trainees in the bottom tercile while it becomes stronger for those in the middle tercile. These findings should, however, be taken with caution as samples are small and tests have low power.

4. Interpretation

Our results suggest that the prospect of receiving information about one's rank in the distribution makes trainees exert lower effort, whereas the possibility of receiving employer recognition or improving one's visibility in the community makes them exert higher effort. In this section, we attempt to interpret these findings in a unified utility maximization framework.

The fact that individuals put in more effort when it is announced that doing so might earn them an award is intuitive. The main challenge is to explain why individuals, especially those with low ability, reduce effort in the rank information treatments even before being told their rank. A growing literature models the possibility that information lowers utility leading to information avoidance, at the cost of taking worse decisions (Oster et al., 2013; Brunnermeier and Parker, 2005; Köszegi, 2006; Stone, 2004; McManus and Rao, 2013). In these models, agents choose their beliefs optimally to maximize their lifetime utility, including an interim period of anticipatory utility arising from the belief. This implies that individuals may take actions to avoid a precise signal that has the potential to threaten their belief, for instance by choosing not to take a medical test.

In our setting, all agents know their individual baseline test scores and are told their individual exam scores in all treatment groups (including control). This, combined with the fact that individuals have a relatively good sense of their own individual ability since they all enter as adults, allows us to assume all that all individuals know their own ability. What individuals do not know is the ability of others—a reasonable assumption in our setting given that trainees entered the training school from an extremely diverse set of geographic and skill backgrounds. Realistically, individuals have a noisy expectation over their peers' ability.

In this setting, rank information (as provided in all treatment groups but not in control) increases the precision of the estimate of each individual's expectation over his peers' average ability. This would not affect behavior in standard models where agents only care about the expected value, but it might matter if agents prefer a fuzzy signal which can support optimistic beliefs about their relative ability.

We assume that agents are risk-neutral, and individual i with ability level a_i chooses effort e_i to maximize his expected utility:

$$\beta_i^t t(a_i, e_i) + \beta_i^{sc}(a_i - a_{-i}^*)(1 + T_i^{sc}\sigma(e_i)) + \beta_i^a T_i^A p(a_i, e_i, b_i)A - d(e_i) \tag{2}$$

The first term captures the effect of effort on learning proxied by test scores, which might provide utility either directly as individuals care about learning or through future wages. $\beta_i^t > 0$ is the weight individual i puts on learning, and learning is a function of effort and ability with $t_e > 0$, $t_a > 0$, $t_{ee} < 0$. We assume that individuals know their own ability, as, again, individuals know their baseline scores and their absolute test scores in control and treatment alike.

The second term captures the utility deriving from social comparisons to which individual i gives weight $\beta_i^{sc} > 0$. Social comparisons enter additively as in Kandel and Lazear (1992), and we assume that individuals care about being of higher ability than their peers rather than having higher test scores *per se*. We assume that individuals enjoy social comparisons when their own ability is higher than the average of their peers a_{-i}^*, and the effect is larger the larger is the ability gap. Conversely, individuals suffer from social comparisons when their own ability is lower than the average of their peers, and the effect is larger the larger is the ability gap.

Finally, as discussed above, we assume that when individuals are given information on test score ranks, they can obtain more precise information about their relative ability. Thus, $T_i^{sc} = 1$ if individual i is in one of the four treatment groups that provide rank information on test scores and 0 in control. In particular, we assume that whether test score ranks give information on ability ranks depends positively on effort through the "signal" function $\sigma(e_i)$ with $\sigma_e > 0$. Intuitively, since test scores are a function of ability and effort, receiving a low rank when exerting low effort could still allow an individual to retain the option of believing that he is of truly high relative ability, but receiving a low rank with high effort could not.[15] The "signal" function $\sigma(e_i)$ is a reduced form representation of "choosing beliefs" in the spirit of Yariv (2002), Köszegi (2006), Oster et al. (2013) and Brunnermeier and Parker (2005).[16]

we purposely included extremely difficult questions making scores of 100% very difficult. We thus take the evidence on top decile scores as support for the existence of capping, although we cannot rule out the possibility that other effects (such as an aversion to being singled out among peers, due to being taxed by expectations of assistance) may have also been occurring at the top decile.

[15] This holds regardless of whether i expects others to exert high or low effort. If he chooses high effort and expects others to do so as well, a low rank in test scores implies a low rank in ability. If he expects others to choose low effort, a low rank in test scores implies a low rank in ability *a fortiori*.

[16] Similarly, Bénabou and Tirole (2002) consider a model where agents can manipulate their interim belief through the choice of information structure. They show that less information can be preferable as it can weaken the time inconsistency problem and induce more effort in the future. We do not assume

The third term captures the utility deriving from award A, which in our setting is either employer recognition or social visibility. Thus, $T_i^A = 1$ if individual i is in one of the two treatment groups that provide awards, and 0 otherwise. Note that $T_i^A = 1 \Rightarrow T_i^{SC} = 1$ but not vice versa. The probability of winning the award $p(a_i, e_i, b_i)$ depends on effort, ability, and the baseline score b_i, with $p_e > 0$, $p_{ee} < 0$, $p_a > 0$, and $p_b < 0$. The latter captures that "most improved" awards are more easily obtainable by those who did poorly in the baseline test.

Finally, $d(e)$ is the disutility of effort, with $d_e > 0$, $d_{ee} > 0$ as is standard.

Maximizing Eq. (2) with respect to e_i yields:

$$\begin{cases} e_i^C \quad s.t. \quad \beta_i^t t_e(a_i, e_i^c) - d_e(e_i^c) = 0 & \text{if } T_i^{SC} = T_i^A = 0 \\ e_i^{SC} \quad s.t. \quad \beta_i^t t_e(a_i, e_i^{SC}) + \beta_i^{sc}\sigma_e(e_i^{SC})(a_i - a_{-i}^*) - d_e(e_i^{SC}) = 0 & \text{if } T_i^{SC} = 1, T_i^A = 0 \\ e_i^A \quad s.t. \quad \beta_i^t t_e(a_i, e_i^A) + \beta_i^{sc}\sigma_e(e_i^A)(a_i - a_{-i}^*) + \beta_i^a p_e(a_i, e_{i,}^A b_i)A - d_e(e_i^A) = 0 & \text{if } T_i^{SC} = T_i^A = 1 \end{cases}$$

The first order conditions for effort in the three cases inform the comparison of performance in each treatment vs. the control group and across treatments. These comparisons map into the empirical findings as follows.

First, $e_i^C > e_i^{SC}$ *iff* $a_i < a_{-i}^*$; namely, providing rank information reduces effort for individuals whose ability is lower than their expectation of their peers' ability. Intuitively, all other things equal, effort provides them with something undesirable (a precise private signal of their ranking in the class). This is related to the psychology literature on "self-handicapping," which can take the form of withdrawing effort in performance settings where there is potential for self-image-damaging feedback (Jones and Berglas, 1978; Berglas, 1985).[17] In line with this, a qualitative survey we administered before implementing the treatments reveals that 43% of the trainees in the bottom quartile of baseline scores did not want to know their relative rank in the class, while only 24% of those in the top quartile said the same.

These findings are similar to the behavioral model in Oster et al. (2013), in which people at risk for Huntington's disease prefer not to be undergo a test to learn whether they have the genetic mutation that causes the disease because the anticipatory utility of believing they might not get sick outweighs the costs of potentially distorted actions.

Second $e_i^A > e_i^{SC}$ for all i; namely, providing awards in addition to rank information can only increase effort as long as awards are valuable, since $\beta_i^a p_e(a_i, e_i^A, b_i)A > 0$. The strength of the effect depends on ability and baseline scores, both of which determine the marginal return to effort $p_e(a_i, e_{i,}^A b_i)$.

Third, $e_i^A > e_i^C$ for all i such that $a_i \geq a_{-i}^*$, as both the second (rank information) and third (award) term in the first order condition are weakly positive, thus increasing the marginal return to effort. In contrast, $e_i^A \lesseqgtr e_i^C$ for all i such that $a_i < a_{-i}^*$ since the second (rank information) term is negative and the third (award) term is positive. Thus, providing awards might reduce effort compared to the control group that receives no awards or rank information if the response to rank information is stronger than the response to the award itself.

5. Conclusion

Our results suggest that awards can have a negative effect on performance as they facilitate social comparison, even though they have a positive effect through employer recognition and enhanced social visibility. In our context, the negative effect of rank information on learning was large enough to undo the positive effects of awards. Since learning is directly related to future labor productivity, the distortion in effort is practically significant.

While the relative magnitudes of these effects are likely to be context-specific, the possible negative effects of rank information and social comparisons are important to consider in the overall productivity effects and design of performance awards. In particular, we show that individuals may even reduce effort in anticipation of learning rank information. This is consistent with a model in which individuals have preferences over their self-perceived ability ranking, and thus engage in information avoidance to be able to retain a positive view of themselves. In effect, in this model, agents self-sabotage in order to avoid the signal contained in ranking information so as not to have to update their beliefs about their own relative ability. As the risk of a low-rank signal is greater in lower ability ranges, this self-handicapping is worse at the bottom of the distribution.

From an employer or policy maker perspective, the cost of an incentive structure that differentially affects the lower tail depends on the nature of the production function. It is particularly costly when there are complementarities in production or when the performance of the lower tail is critical to the principal's goal, as in our setting where the government wants to ensure equitable provision of health services to remote rural areas. In domains such as innovation in science and finance, the effect on productivity of the upper tail of the distribution might be most important. But in domains such as health services

time inconsistency in our setting, which would exacerbate our effect as it would further discount the future impacts of taking distorted actions for interim belief utility.

[17] This allows the trainee to manipulate the attribution of failure to himself (Kelley, 1971). If failure occurs, the extent to which the outcome is attributed to his lack of ability is discounted because of the equally likely reason for failure: decreased effort. However, if the trainee obtains a high grade on the exam, the attribution to ability is strengthened because the decreased learning effort made it even more difficult to do well.

delivery where the potential for harm is high, it is critical to employ incentives that are not detrimental to the lower tail of the distribution. This depends in part, as well, on the ease of exit and entry; if the goal is to induce the lower tail to withdraw effort, and potentially exit, then providing rank information could be a highly effective means to do so. It is left to future research to unbundle these effects across cultures and professional sectors.

Appendix A.

See Figs. A.1–A.6 and Tables A.1 and A.2.

Course	Module	Teaching Days	Topics
0	1: Health care system in Zambia	4	• Organization and functions of the health care system in Zambia • Roles and responsibilities of a Community Health Assistant • Code of conduct • Gender equality • Community mobilization and networking
	2: Behavioral sciences	19	• Introduction to psychology • Mental health and common psychiatric conditions • Introduction to sociology • Family and community
	3: Health promotion	9	• Introduction to health promotion • Communication skills
	4: Environmental health	31	• General principles of infection prevention • Water supply • Excreta disposal • Solid waste management • Food hygiene and safety • Housing and health • Insect and rodent control
1	5: Epidemiology	34	• Infectious disease epidemiology • Epidemic investigation and management • Epidemiological surveillance • Data collection
	6: Reproductive and child health	35	• Introduction to reproductive health • Introduction to child health • School health services
2	7: Anatomy and physiology	20	• Introduction to the human body • Musculoskeletal system • Cardiovascular system • Respiratory system • Digestive system • Urinary system • Special senses
3	8: Basic procedures	20	• Occupational safety and health • Lifting and moving patients • History and physical exam skills • Assessment of hygiene, nutrition, physical activity, pain, and vital signs • Wound care • Palliative care
	9: Common medical conditions	38	• Common conditions (malaria, diarrhea, respiratory infection, HIV, tuberculosis, anemia, etc.) • Oral health
4	10: Diagnostic procedures	4	• Malaria rapid diagnostic testing • HIV testing • Sputum collection for TN testing
	11: First aid	9	• Principles of first aid • Bandaging • Lifting and moving patients • Cardio-pulmonary resuscitation • Handling of selected emergencies (toxic ingestion, bites and stings, fractures, burns, drowning, foreign body ingestion)

Fig. A.1. Outline of one-year community health assistant training curriculum.

NDOLA COMMUNITY HEALTH ASSISTANT TRAINING SCHOOL

Ndola 18th July 2011

Dear **John Banda,**

Please find below your scores on the **Module 4** exam:

- Theoretical: **76%**
- Improvement from baseline exam: **20%** points improvement (76%-56% on baseline exam)

Yours sincerely,

Mrs Nyirenda
Training coordinator

Fig. A.2. Control group.

NDOLA COMMUNITY HEALTH ASSISTANT TRAINING SCHOOL

Ndola 18th July 2011

Dear **John Banda,**

Please find below your scores on the **Module 4** exam:

- Theoretical: **76%**
- Improvement from baseline exam: **20%** points improvement (76%-56% on baseline exam)

Within your class you were:

- **3rd** out of 60 students on the Theoretical.
- **13th** out of 60 students in terms of most improved from the baseline exam.

Yours sincerely,

Mrs Nyirenda
Training coordinator

Fig. A.3. Treatment #1.

NDOLA COMMUNITY HEALTH ASSISTANT TRAINING SCHOOL

Ndola 18th July 2011

Dear **John Banda,**

Please find below your scores on the **Module 4** exam:

- Theoretical: **76%**
- Improvement from baseline exam: **20%** points improvement (76%-56% on baseline exam)

Within your class you were:

- **3rd** out of 60 students on the Theoretical.
- **13th** out of 60 students in terms of most improved from the baseline exam.

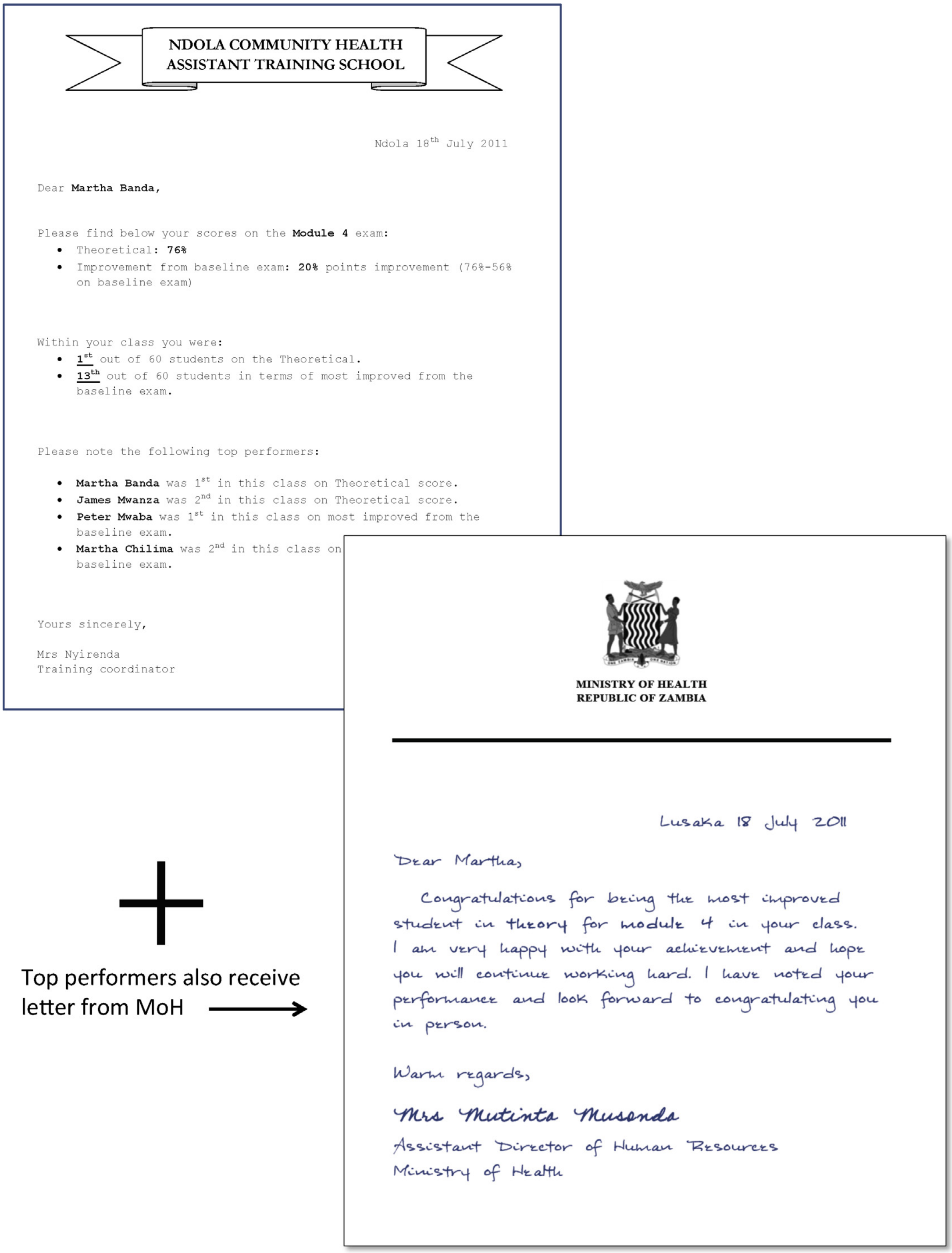
Please note the following top performers:

- **Mary Phiri** was 1st in this class on Theoretical score.
- **James Mwanza** was 2nd in this class on Theoretical score.
- **Peter Mwaba** was 1st in this class on most improved from the baseline exam.
- **Martha Chilima** was 2nd in this class on most improved from the baseline exam.

Yours sincerely,

Mrs Nyirenda
Training coordinator

Fig. A.4. Treatment #2.



NDOLA COMMUNITY HEALTH ASSISTANT TRAINING SCHOOL

Ndola 18th July 2011

Dear **Martha Banda**,

Please find below your scores on the **Module 4** exam:

- Theoretical: **76%**
- Improvement from baseline exam: **20%** points improvement (76%-56% on baseline exam)

Within your class you were:

- 1st out of 60 students on the Theoretical.
- 13th out of 60 students in terms of most improved from the baseline exam.

Please note the following top performers:

- **Martha Banda** was 1st in this class on Theoretical score.
- **James Mwanza** was 2nd in this class on Theoretical score.
- **Peter Mwaba** was 1st in this class on most improved from the baseline exam.
- **Martha Chilima** was 2nd in this class on baseline exam.

Yours sincerely,

Mrs Nyirenda
Training coordinator

MINISTRY OF HEALTH
REPUBLIC OF ZAMBIA

Lusaka 18 July 2011

Dear Martha,

Congratulations for being the most improved student in theory for module 4 in your class. I am very happy with your achievement and hope you will continue working hard. I have noted your performance and look forward to congratulating you in person.

Warm regards,

Mrs Mutinta Musonda
Assistant Director of Human Resources
Ministry of Health

Fig. A.5. Treatment #3.

NDOLA COMMUNITY HEALTH
ASSISTANT TRAINING SCHOOL

Ndola 18th July 2011

Dear **Martha Banda**,

Please find below your scores on the **Module 4** exam:

- Theoretical: **76%**
- Improvement from baseline exam: **20%** points improvement (76%-56% on baseline exam)

Within your class you were:

- **1st** out of 60 students on the Theoretical.
- **13th** out of 60 students in terms of most improved from the baseline exam.

Please note the following top performers:

- **Martha Banda** was 1st in this class on Theoretical score.
- **James Mwanza** was 2nd in this class on Theoretical score.
- **Peter Mwaba** was 1st in this class on most improved from the baseline exam.
- **Martha Chilima** was 2nd in this class on most improved from the baseline exam.

Yours sincerely,

Mrs Nyirenda
Training coordinator

One of top performers also featured in a newsletter sent to their community

CHA NEWSLETTER | Issue 1 2

"Why I want to be a CHA"

By Martha Banda

Martha Banda, age 27 years

LET'S GET TO KNOW MARTHA!

Martha Banda is a 27-year-old CHA student representing Lubanda Health Post, Itezhi-tezhi District, Southern Province. Her counterpart at this health post is Andrew Phiri.

She is married and a mother of two children. She completed secondary school at Kabulonga Girls High School in 2001. Before joining the CHA program, Martha was volunteering as an HIV/AIDS educator in her community.

*In this section, we wish to provide an opportunity for CHA students to be known to the community. **Martha Banda** from Lubanda Health Post shares her thoughts on why she wants to be a Community Health Assistant.*

Zambia has faced many challenges in its health sector but I have become inspired by the fact that the government wants to train community health workers. I joined the program because of my passion for patients, my love for others, and because I want to reduce the gap between my people and the hospital. I plan to encourage my community to use healthy resources like boreholes for clean drinking water and community resources such as the health post to learn more about nutritious diets.

I like telling people about health issues and I want to be a Community Health Assistant so that I can help my friends in my community, and so we can work together. I like caring for people who are sick – cooking for them, washing them, and giving them words of encouragement.

My first step to being a Community Health Assistant, after finishing my training, will be to gain the respect of my community, first and foremost. I will start by talking with my village headman and, after getting his blessing, I will begin to call the community for meetings where I plan to give health talks on how to care for children under-five and educate the community on treating their water with chlorine prior to drinking it. I am looking forward to serving my community!

Fig. A.6. Treatment #4.

Table A.1
Quantile treatment effects.

Quantile	(1) 10th	(2) 20th	(3) 30th	(4) 40th	(5) 50th	(6) 60th	(7) 70th	(8) 80th	(9) 90th
T1: private social comparison	−0.504***	−0.392***	−0.282*	−0.320***	−0.304**	−0.256**	−0.184**	−0.0635	−0.0118
	(0.176)	(0.142)	(0.158)	(0.115)	(0.135)	(0.105)	(0.0875)	(0.117)	(0.124)
T2: T1 + public social comparison	−0.704***	−0.445***	−0.292*	−0.428***	−0.381***	−0.329***	−0.275***	−0.174	−0.152
	(0.171)	(0.140)	(0.156)	(0.114)	(0.133)	(0.103)	(0.0857)	(0.114)	(0.115)
T3: T2 + employer recognition award	0.0681	−0.00494	0.170	0.0644	−0.0805	−0.0786	−0.0341	0.0105	−0.0440
	(0.157)	(0.137)	(0.152)	(0.113)	(0.133)	(0.103)	(0.0862)	(0.117)	(0.117)
T4: T2 + social visibility award	−0.0257	0.00956	0.0694	0.117	0.0218	−0.0151	0.0161	0.136	0.0677
	(0.168)	(0.153)	(0.172)	(0.126)	(0.147)	(0.115)	(0.0965)	(0.130)	(0.135)
Net effect of public social comparison (T2 – T1)	−0.200	−0.053	−0.010	−0.108	−0.076	−0.074	−0.090	−0.110	−0.140
	(0.180)	(0.141)	(0.156)	(0.114)	(0.133)	(0.104)	(0.087)	(0.117)	(0.128)
Net effect of employer recognition award (T3 – T2)	0.772***	0.441***	0.462***	0.492***	0.300**	0.251**	0.241***	0.184	0.108
	(0.164)	(0.137)	(0.152)	(0.111)	(0.131)	(0.102)	(0.086)	(0.116)	(0.118)
Net effect of social visibility award (T4 – T2)	0.679***	0.455***	0.362**	0.545***	0.402***	0.314***	0.291***	0.309	0.220
	(0.172)	(0.149)	(0.169)	(0.124)	(0.145)	(0.114)	(0.097)	(0.131)	(0.138)
Number of observations (trainee-courses)	1149	1149	1149	1149	1149	1149	1149	1149	1149
Adjusted R^2	0.1528	0.1392	0.1317	0.1335	0.1336	0.1341	0.1363	0.1283	0.1183

Notes: Table reports OLS estimates, weighted by course duration, with standard errors clustered at the trainee level in parentheses. Dependent variable is normalized exam score, normalized by the mean and standard deviation of the control group for each exam. Individual controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A.2
Heterogeneous treatment effect by baseline test score and period.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Bottom tercile of baseline test score				Middle tercile of baseline test score				Top tercile of baseline test score			
	All periods	Heterogeneous effects by exam period			All periods	Heterogeneous effects by exam period			All periods	Heterogeneous effects by exam period		
		First period	Following periods	Test of equality: (2) = (3)		First period	Following periods	Test of equality: (6) = (7)		First period	Following periods	Test of equality: (10) = (11)
Private social comparison	−0.672**	−0.578**	−0.737*	p = 0.656	−0.233	−0.119	−0.316	p = 0.519	0.010	−0.018	0.081	p = 0.618
	(0.298)	(0.263)	(0.394)		(0.271)	(0.286)	(0.328)		(0.213)	(0.233)	(0.234)	
T2: T1 + public social comparison	−0.812***	−1.065***	−0.640**	p = 0.124	−0.179	−0.258	−0.126	p = 0.562	−0.113	0.059	−0.225	p = 0.186
	(0.260)	(0.261)	(0.314)		(0.199)	(0.256)	(0.209)		(0.203)	(0.226)	(0.230)	
T3: T2 + employer recognition award	−0.280	−0.146	−0.372	p = 0.335	0.157	0.075	0.213	p = 0.600	−0.005	0.210	−0.154	p = 0.060
	(0.216)	(0.249)	(0.244)		(0.250)	(0.329)	(0.246)		(0.159)	(0.207)	(0.170)	
T4: T2 + social visibility award	−0.102	−0.102	−0.150	p = 0.891	0.204	−0.108	0.286	p = 0.064	−0.031	−0.203	0.028	p = 0.294
	(0.309)	(0.391)	(0.336)		(0.223)	(0.275)	(0.226)		(0.177)	(0.234)	(0.191)	
Net effect of public social comparison (T2 − T1)	−0.140	−0.487	0.097	p = 0.115	0.054	−0.138	0.190	p = 0.303	−0.153	0.076	−0.306	p = 0.079
	(0.334)	(0.298)	(0.428)		(0.243)	(0.243)	(0.320)		(0.214)	(0.229)	(0.247)	
Net effect of employer recognition award (T3 − T2)	0.532**	0.919***	0.268	p = 0.007	0.336	0.333	0.339	p = 0.982	0.107	0.152	0.071	p = 0.705
	(0.255)	(0.278)	(0.284)		(0.214)	(0.300)	(0.220)		(0.166)	(0.200)	(0.195)	
Net effect of social visibility award (T4 − T2)	0.710**	0.963**	0.490	p = 0.182	0.383**	0.150	0.412**	p = 0.223	0.082	−0.261	0.253	p = 0.032
	(0.342)	(0.408)	(0.375)		(0.183)	(0.222)	(0.202)		(0.173)	(0.226)	(0.200)	
Trainee controls	Yes		Yes		Yes		Yes		Yes		Yes	
Number of clusters (trainees)	92		92		107		107		107		107	
Number of observations (trainee-courses)	350		350		401		401		398		398	
Adjusted R^2	0.186		0.228		0.079		0.111		0.157		0.207	

Notes: Table reports OLS estimates, weighted by course duration, with standard errors clustered at the trainee level in parentheses. Dependent variable is normalized exam score, normalized by the mean and standard deviation of the control group for each exam. Individual controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy. Columns (2) and (3), (6) and (7), and (10) and (11) report coefficients estimated in the same regression where we include all treatments interacted with an indicator variable that takes value 0 in the first period and 1 thereafter.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

References

Angrist, J., Lavy, V., 2009. The effects of high stakes high school achievement awards: evidence from a randomized trial. The American Economic Review 99 (4), 1384–1414.
Ashraf, N., Bandiera, O., Jack, K., 2013. No margin, no mission? A field experiment on incentives for public service delivery. Manuscript.
Azmat, G., Iriberri, N., 2010. The importance of relative performance feedback information: evidence from a natural experiment using high school students. Journal of Public Economics 94 (7), 435–452.
Bandiera, O., Barankay, I., Rasul, I., 2013. Team incentives: evidence from a firm level experiment. Journal of the European Economic Association 11 (5), 1079–1114.
Baqui, A., Kumar, V., Mohanty, S., Kumar, A., Misra, R.P., Santosham, M., Awasthi, S., 2008. Effect of community-based behaviour change management on neonatal mortality in Shivgarh, Uttar Pradesh, India: a cluster-randomised controlled trial. The Lancet 372 (9644), 1151–1162.
Barankay, I., 2012. Rank incentives: evidence from a randomized workplace experiment. Working Paper. Wharton School, University of Pennsylvania.
Bénabou, R., Tirole, J., 2002. Self-confidence and personal motivation. The Quarterly Journal of Economics 117 (3), 871–915.
Berglas, S., 1985. Self-handicapping and self-handicappers: a cognitive/attributional model of interpersonal self-protective behavior. Perspectives in Personality 1, 235–270.
Besley, T., Ghatak, M., 2008. Status incentives. The American Economic Review 98 (2), 206–211.
Bhandari, N., Paul, V.K., Sachdev, H.S., Mavalankar, D., Ramachandran, P., Sankar, M.J., Sreenivas, V., 2011. Reproductive health, child health and nutrition in India: meeting the challenge. The Lancet 377, 332–349.
Bradler, C., Dur, R., Neckermann, S., Non, A., 2013. Employee recognition and performance: a field experiment. In: ZEW Discussion Paper No. 13-017.
Bruhn, M., Mckenzie, D., 2009. In pursuit of balance: randomization in practice in development field experiments. American Economic Journal: Applied Economics 1 (4), 200–232.
Brunnermeier, M., Parker, J.A., 2005. Optimal expectations. The American Economic Review 95 (4), 1092–1118.
Charness, G., Grosskopf, B., 2001. Relative payoffs and happiness: an experimental study. Journal of Economic Behavior and Organization 45, 301–328.
Dewatripont, M., Jewitt, I., Tirole, J., 1999. The economics of career concerns. Part I: Comparing information structures. Review of Economic Studies 66, 183–198.
Fehr, E., Schmidt, K., 1999. A theory of fairness, competition and cooperation. Quarterly Journal of Economics 114 (3), 817–868.
Freeman, R., Gelber, A., 2010. Prize structure and information in tournaments: experimental evidence. American Economic Journal: Applied Economics 2 (1), 149–164.
Fryer, R., 2011. Financial incentives and student achievement: evidence from randomized trials. The Quarterly Journal of Economics 126, 1755–1798.
Government of Zambia, 2011. National human resources for health strategic plan 2011–2015. Technical report. Ministry of Health.
Jones, E., Berglas, S., 1978. Control of attributions about the self through self-handicapping strategies: the appeal of alcohol and the role of underachievement. Personality and Social Psychology Bulletin 4 (2), 200–206.
Kandel, E., Lazear, E., 1992. Peer pressure and partnerships. Journal of Political Economy 100 (4), 801–817.
Kelley, H., 1971. Attribution in Social Interaction. General Learning Press, Morristown, NJ.
Köszegi, B., 2002. Emotional agency. The Quarterly Journal of Economics 121 (1), 121–155.
Köszegi, B., 2006. Ego utility, overconfidence, and task choice. Journal of the European Economic Association 4 (4), 673–707.
Kremer, M., Miguel, E., Thornton, R., 2009. Incentives to learn. The Review of Economics and Statistics 91 (3), 437–456.
Lazear, E., 1989. Pay equality and industrial politics. Journal of Political Economy 97 (3), 561–580.
Leuven, E., Oosterbeek, H., van der Klaauw, B., 2010. The effect of financial rewards on students' achievement: evidence from a randomized experiment. Journal of the European Economic Association 8 (6), 1243–1265.
Major, B., Testa, M., Bylsma, W.H., 1991. Social Comparison: Contemporary Theory and Research. Erlbaum, Hillsdale, NJ.
Mathauer, I., Imhoff, I., 2006. Health worker motivation in Africa: the role of non-financial incentives and human resource management tools. Human Resources for Health 4 (1), 24.
McManus, T., Rao, J., 2013. Signaling smarts: Revealed preferences for self and social perceptions of intelligence. Working paper.
Milkovich, G., Newman, J., 1996. Compensation, 4th ed. Irwin, Homewood, IL.
Moldovanu, B., Sela, A., 2001. The optimal allocation of prizes in contests. The American Economic Review 91, 542–558.
Moldovanu, B., Sela, A., Shi, X., 2007. Contests for status. Journal of Political Economy 115, 338–363.
Müller, W., Schotter, A., 2010. Workaholics and dropouts in organizations. Journal of the European Economic Association 8 (4), 717–743.
Muralidharan, K., Sundararaman, V., 2011. Teacher performance pay: experimental evidence from India. Journal of Political Economy 119 (1), 39–77.
Nalbantian, H., Schotter, A., 1997. Productivity under group incentives: an experimental study. The American Economic Review 87 (3), 314–341.
Oster, E., Shoulson, I., Dorsey, E.R., 2013. Optimal expectations and limited medical testing: evidence from Huntington disease. American Economic Review 103 (2).
Smith, A., 1759. The Theory of Moral Sentiments. Liberty Fund, Indianapolis.
Spencer, M., Rosland, A.-M., Kieffer, E.C., Sinco, B., Valerio, M., Palmisano, G., Anderson, M., Guzman, J.R., Heisler, M., 2011. Effectiveness of a community health worker intervention among African American and Latino adults with type 2 diabetes: a randomized controlled trial. American Journal of Public Health 101 (12), 2253–2260.
Stone, R., 2004. Self-handicapping, tournaments, and ego utility. Unpublished manuscript.
Tran, A., Zeckhauser, R., 2012. Rank as an inherent incentive: evidence from a field experiment. Journal of Public Economics 96, 645–650.
Yariv, L., 2002. I'll see it when I believe it—a simple model of cognitive consistency. In: Cowles Foundation Discussion Paper No. 1352.