More effort or better technologies?
On the effect of relative-performance feedback*

Gwen-Jirō Clochard[†], Guillaume Hollard[‡], Julia Wirtz[§]

This version: May 2019

Abstract

Relative-performance feedback (RPF) allows one's own performance to be compared to that of others, for instance via rankings. It is traditionally assumed that RPF affects performance by changing the optimal level of effort, which we call the *effort channel*. Our first contribution is to propose an alternative way of modelling the effect of RPF on performance, namely the *technology channel*. "Room for improvement" (i.e. the possibility to improve upon existing technologies) is then a key driver of the relative contributions of the two channels. Our second contribution is to show that the variety of existing empirical results are strikingly easy to understand once room for improvement is taken into account. Our third contribution is to design an experiment in which the differences between the predictions of each channels are made salient. The empirical results provide clear support for the relevance of the technology channel.

Keywords: Relative performance feedback, tournament incentives, learning.

JEL Classification:D83, D84, D91

*We thank Nagore Iriberri, Pierre Boyer, Alfonso Montes Sanchez, Fabien Perez, Rémi Avignon, Yannick Guyonvarch, George Loewenstein, Phil Reny and Omar Sene for useful comments and discussions. We are also grateful to participants at the ESA AP Meeting 2019, the Bristol Workshop on Assessment and Feedback and various internal seminars. The data were collected by Mahmoud Farrokhi-Kashani during the course of his PhD thesis. For more details, see Farrokhi-Kashani (2012).

[†]Ecole polytechnique; gwen-jiro.clochard@polytechnique.edu.

[‡]Ecole polytechnique and CNRS; guillaume.hollard@polytechnique.edu.

[§]University of Bristol; j.wirtz@bristol.ac.uk.

1 Introduction and literature review

Relative-performance feedback (RPF henceforth) consists of information that enables individuals to compare their own performance to that of others, for instance by providing rankings or their position relative to the mean. RPF is ubiquitous in the economy: employers provide information on worker productivity, academics are able to compare themselves using publicly-available citation scores, the TripAdvisor website publishes a ranking of the best hotels, etc. What is less clear is the effect that RPF has on performance. It is traditionally assumed that RPF will affect performance by changing the optimal effort level, which we call the *effort channel*. For instance, in a tournament, if RPF reveals that the prize is within reach, the effort channel predicts that optimal effort will rise; conversely, for individuals who realise that they are at the bottom of the distribution, the effort channel predicts that RPF will result in less effort.

There exist, however, situations in which even poor performers do improve their performance after RPF. For instance, Azmat and Iriberri (2010) show that providing RPF to high-school students has a constant and positive effect *all along the performance distribution*, so that all students increase their effort. Blanes i Vidal and Nossol (2011) also find that providing RPF increases worker productivity at *all* prior productivity levels. The effort channel thus struggles to account for the effect of RPF on poor performers. A second limit of the effort channel relates to the evolution of the RPF effect over time. RPF is assumed to modify the marginal return to effort the first time it is provided, but to have no impact on performance after the initial round. However, Azmat and Iriberri (2016) find that the effect of RPF on performance evolves over time. For the effort channel to explain these phenomena, an ad-hoc term if often added to the utility function so that individuals are assumed to derive utility from their ranking *per se*. This requirement of a taste-for-ranking term in the utility function is, in our view, a weakness of the effort channel. The challenge is then to complement the effort channel by an alternative approach to help explain these results.

The existing theoretical literature on relative performance feedback in tournaments (Ederer, 2010; Aoyagi, 2010; Goltsman and Mukherjee, 2011; Gershkov and Perry, 2009) considers how this information affects incentives for effort.[1] All papers in this literature predict that the effect will not be unambiguously positive. Feedback will discourage effort, when the agent learns that effort is unlikely to improve their chances of winning. The idea that feedback could improve technologies rather than effort only appears in Wirtz (2016).

When agents have the option to choose and discard technologies, this will affect the riskiness of their output. This connects our work to the literature on risk-taking in contests (Hvide, 2002; Anderson and

[1] All examples are in two-player settings.

Cabral, 2007; Seel and Strack, 2013), who point out that agents who lag behind in a contest tend to benefit from increased risk-taking.

Finally our work relates to the literature on strategic experimentation. The closest work is Halac et al. (2017) who considers the optimal feedback and prize structure in an innovation contest. As in our model, feedback allows agents to learn about the quality of the technology they are pursuing as well as their chance of winning the contest. However, there is only one technology which all agents are using and which they get more and more pessimistic about. The agent only has the choice how much and how long to invest into the technology, but they cannot switch to a different one. This is also related to dynamic Bayesian Persuasion problems (Che and Horner, 2019; Kremer et al., 2014; Renault et al., 2017) where a principal uses information to influence the decision of short-lived agents. However, these papers do not consider competitive environments and there is no option of trying out different risky technologies.

The first contribution of this paper is to propose an alternative way of accounting for the effect of RPF on performance, which we call the *technology channel*. Elaborating on Wirtz (2016), we assume that the main choice to be made relates to technology (i.e. the way in which effort is transformed into performance) rather than effort itself. We assume that the relative effectiveness of the technologies available is not known with certainty. When an individual only knows her own performance, she cannot tell whether this reflects the effectiveness of her chosen technology or unobserved factors affecting all participants. In this context, RPF allows the individual and aggregate determinants to be disentangled.

Consider for instance a student who received a good mark in a test after studying with a group of friends the day before the test. She might be uncertain about whether working with friends (her technology) helped, or simply that the test was an easy one. If she learns her rank, in addition to her grade, and it turns out that she is well-ranked, she might infer that the change she introduced in her technology was beneficial and improved her performance. If, on the other hand, she learns that most of her peers did better than her, she may no longer want to work with friends for future tests, as it does not appear to have helped her performance. RPF helps identify better technologies, leading to increased productivity across rounds, even for low performers.

How relevant the RPF technology channel is in a particular situation depends on how much room for improvement there is. By room for improvement, we mean the possibility to improve upon existing technologies. When the room for improvement is low,[2] as is the case for running, it is unlikely that subjects can quickly change their technology. Here the effect of RPF on performance is expected to work mainly

[2] See Gneezy et al. (2003), Eriksson et al. (2009) and Kuhnen and Tymula (2012).

via effort. On the other hand, when there is considerable room for improvement,[3] for instance for exams in school, RPF conveys information about the relative productivity of the available technologies. Here RPF affects performance also via technology.

The second contribution of this paper is to reconsider the existing evidence in the light of room for improvement. Table 1 lists some of the important contributions to the literature on RPF. As can be seen in the last three columns, RPF appears to have mixed performance effects. However, the existing results are strikingly easy to classify once we allow for room for improvement (in the first column). Moreover, the existing evidence, once classified by room for improvement, appears in line with the predictions from the two channels: when there is little room for improvement, the empirical results are as in the effort channel (a small to detrimental effect at the bottom of the performance distribution), while the results when there is room for improvement are in line with the predictions of the technology channel (a shift in the entire distribution of performance).

The third contribution of this paper is to test the predictions of the technology channel using applicable data. We were fortunate enough to receive permission to set up an experiment with almost no restrictions in two Iranian high schools (similar permission was refused in France). In particular, students were not informed that they were taking part into an experiment, as the schools backed the experiment as part of regular training. Our objective was to design an experiment so that differences between the technology and effort channels were particularly sharp, i.e. regarding the effect of RPF on low performers and when RPF is repeatedly provided. We thus designed a four-round experiment so as to be able to assess whether the effect changes across rounds. To completely shut down the effort channel for low-ability individuals, we designed a tournament such that, after a few rounds, even performing at the best possible level will no longer suffice to win a prize. Under the effort channel, low-ability performers have zero incentive to increase their effort. Last, we needed to identify a task with great room for improvement, and for which performance is easy to measure and compare across individuals. We chose math tests. Our results indicate a continuous improvement of the effect of RPF over time, and a positive RPF effect at the bottom of the distribution. These empirical findings therefore provide a clear support for the existence of the technology channel.

The remainder of the paper is organised as follows. In Section 2, we present and compare the theoretical predictions of the effort and technology channels, in terms of how RPF can affect performance. In Section 3 we present the design and results of the experiment used to test the predictions from the technology channel. Last, Section 4 concludes.

[3] See Azmat and Iriberri (2010), Blanes i Vidal and Nossol (2011) and Tran and Zeckhauser (2012).

Table 1: Review of the empirical literature on the impact of performance feedback

Room for improvement	Paper	Task	Overall effect	Effect at the top	Effect at the bottom
No	Hannan et al. (2008)	Profit maximization	Piece-rate: increase Tournament: decrease	Increase under tournament	Decrease under tournament
No	Eriksson et al. (2009)	Adding numbers	No effect on average		
No	Fershtman and Gneezy (2011)	60m races		Positive	Negative (quit)
No	Barankay (2011)	Classifying images on Amazon Mechanical Turk	Negative		
No	Kuhnen and Tymula (2012)	Multiplication		Improvement over time	No improvement
No	Azmat and Iriberri (2016)	Adding numbers	Increase under piece-rate		
No	Gill et al. (2018)	Verbal and Adding numbers		Increase	Increase
No	Fischer and Wagner (2018)	Math exams	Negative if RPF late		
Yes	Azmat and Iriberri (2010)	Exams		Increase	Increase
Yes	Blanes i Vidal and Nossol (2011)	Work productivity		Increase	Increase
Yes	Tran and Zeckhauser (2012)	English tests		Increase	Increase (if public)
Yes	Bandiera et al. (2015)	University tests		Increase	Increase
Yes	Azmat et al. (2018)	Exams	Decrease for overconfident		
Yes	Fischer and Wagner (2018)	Math exams	Positive if RPF early		

2 Theoretical predictions

This section presents the theoretical frameworks of the two channels through which RPF can affect performance. The first assumes that RPF provides incentives to exert more effort, particularly through a taste-for-ranking, and the second that RPF allows individuals to identify better technologies. The two models, presented below, are polar cases. The effort model is at one extreme, assuming that all individuals have the same technology and decide how much effort to exert; the technology model, at the other extreme, assumes that all individuals exert the same effort but choose which technologies to use.

2.1 The effort channel

We consider a large number of individuals performing the same task at each period, over $T \geq 2$ periods. At each period, an individual produces an output $x_t = e_t + b_t + \varepsilon_t$, where e_t is individual effort provided at cost $c(e_t) = \frac{1}{2}e_t^2$, b_t a common shock (i.i.d, $b_t \sim N(0, \sigma^2)$) and ε_t is an individual error term (i.i.d., $\varepsilon_t \sim N(0, 1)$). The individual therefore faces two sources of uncertainty: that regarding the common shock and that in the error term.

We consider a tournament setting over the sum of total output, which is equivalent to a setting in which the individual wins a prize (normalised to 1) if the sum of her outputs, net of the common shocks, is greater than a given threshold s.

Individuals only care about winning: We first consider the case in which individuals do not care about their rank *per se*, which yields the following payoff:

$$P\left(\sum_{t=1}^{T}(e_t + \varepsilon_t) \geq s\right) - \sum_{t=1}^{T} c(e_t) \tag{1}$$

There are two groups. Individuals in the first group (*RPF group*) are informed each period about the value of their individual output and their individual rank, while those in the second group (*control group*) receive only the value of their individual output. For the description of the results we restrict ourselves to $T = 2$.

In the RPF group, the individual ranking suffices to entirely disentangle the two sources of uncertainty and infer the values of b_t and ε_t (as the grade x_t is known, e_t is a choice variable and $\varepsilon_t = x_t - e_t - b_t$). From Period 2 we can define $s' = s - e_1 - \varepsilon_1 - e_2$, and the winning probability becomes

$$P\left(\sum_{t=1}^{T}(e_t + \varepsilon_t) \geq s\right) = 1 - F_\varepsilon(s') \tag{2}$$

The first-order conditions thus provide the optimal effort in Period 2, and by backward induction the optimal effort in previous periods:

$$\begin{aligned} e_2^{RPF} &= f_\varepsilon(s') & (3) \\ e_1^{RPF} &= \mathbb{E}_{(\varepsilon_1,e_2)}[f_\varepsilon(s')] \\ &= \mathbb{E}_{(\varepsilon_1,e_2)}[e_2] & (4) \end{aligned}$$

As can be seen from Equation 4, effort in Period t equals the expected effort in subsequent periods, as the individual wishes to smooth effort across periods.

Deriving the first-order conditions for the control group (in which individuals receive only their individual performance) yields

$$e_1^C = \mathbb{E}_{(\varepsilon_1,e_2)}[f_\varepsilon(s')] \tag{5}$$

so that there is no difference between the RPF and control groups in Period 1. In the same fashion, we find:

$$e_2^C = E_{(\varepsilon_1|x_1)}\left[f_\varepsilon\left(s'\right)\right] \tag{6}$$

These equations indicate that there should be no average RPF treatment effect on performance. However, there could still be differences in individual effort between the control and the treatment groups, depending on the realisation of the error terms, but in expectation providing feedback information does not change performance.

In sum, RPF has no effect on performance, on average, while there is ample evidence that it does so in practice. One popular way of reconciling theory and evidence has been to add an *ad hoc* term to the utility function.

Individuals care about their rank *per se*: In this setup, the agent's payoff becomes

$$P\left(\sum_{t=1}^{T}(e_t+\varepsilon_t)\geq s\right)-\sum_{t=1}^{T}c(e_t)+g\left(r_t(e_t)\right)\mathbb{1}_{RPF} \tag{7}$$

with r_t being the rank of the individual in period t and g a "taste-for-rank" function. We assume only that the g function falls in the rank. This additional term increases the marginal return of effort, and so its optimal level. Introducing this additional term to the utility function therefore increases performance, with the size of the increase depending on the shape of the g function. The most common solution is to assume that the better the rank, the greater the marginal rank's utility, making the effect of RPF on performance stronger at the top of the distribution.

One interesting feature in this setting is that the additional term allows us to rationalise any type of distribution of improvements. It would equally be possible to assume that the effect of the taste-for-rank is stronger at the bottom of the distribution (people dislike being last, as in Kuziemko et al. (2014)), so that poor performers improve the most. If we relax the monotonicity of the g function, it is even possible to rationalise a negative effect of RPF on performance at any point of the distribution.

Another important feature of this model is that RPF shifts the distribution in the first period but has no further effect in the subsequent periods.

2.2 The technology channel

We now present our model in which RPF affects performance through the better identification of good *technologies*. As in the previous model, individuals carry out the same task over T periods, and play in a tournament. The model here differs in that output in Period t is now defined as

$$x_t = \sum_{1}^{t-1} \theta_r \cdot \mathbb{1}_{\text{Keep},r} + \theta_t \cdot \mathbb{1}_{\text{Experiment},t} + B_t \tag{8}$$

where θ_r is an individual technology drawn in the previous period r (i.i.d, $\theta_r \sim \mathcal{N}(0,1)$). For instance, in a school exam, technologies could represent preparation methods that either improve or worsen performance: taking notes on a computer (rather than on paper), meeting with classmates to work together, drinking coffee at breakfast, and studying late the night before can all affect performance. They can have either a positive or negative effect, which is ex ante uncertain. Overall productivity is the sum of all the preparation methods.

B_t is an aggregate shock that affects all individuals, and takes the form of a random walk with increments $b_t \sim \mathcal{N}(0, \sigma^2)$. The shape of the aggregate shock is convenient, as the individual only updates her beliefs about the value of the technology once, after the realisation of x_t. In the subsequent periods the belief will remain constant.

Individuals again face two sources of uncertainty: that regarding individual technologies and that from the common shock.

As in the effort model, individuals are paid if the sum of their output in all periods net of the common shock is above a threshold s. We consider the same two groups, the first in which individuals know both their rank and the value of their individual output (RPF group), and the second (control) group in which individuals only know the value of their output.

In each period, individuals have to make two decisions. They first need to decide whether they want to keep the technology drawn θ_t in the subsequent periods or drop it (**Keep** or **Drop**). If an individual decides to drop her technology, this is abandoned for all subsequent periods. Second, individuals decide whether they wish to experiment with new technology the next period, not knowing ex ante whether this will improve or reduce their output. They can also decide not to draw a technology and remain at their current level of output (**Experiment** or **Stop**). For illustrative purposes, we restrict the demonstration below to a setting with two periods ($T = 2$).

Optimal strategies: For individuals in the RPF group, learning about their rank allows them to perfectly identify the value of θ_1. They can therefore adopt the optimal strategy, i.e. **Keep** their technology if $\theta_1 \geq 0$, and decide to **Experiment** if $\theta_1 \leq \frac{s}{2}$. The optimal strategies are illustrated in Figure 1.

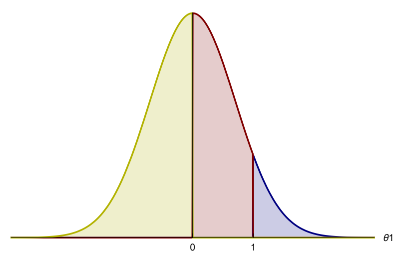


Figure 1: The optimal strategy in the feedback treatment, depending on the technology drawn in the first period ($s = 2$, the aggregate shock and level of uncertainty are irrelevant). The yellow area corresponds to the strategy (**Drop**, **Experiment**), that in red to (**Keep**, **Experiment**) and that in blue to (**Keep**, **Stop**).

Individuals in the control group do not however perfectly observe the value of the common shock, and can so only partially update their beliefs about the true value of their output. Using Bayes' rule, the updated belief becomes

$$\theta_1 | x_1 \sim \mathcal{N}\left(\frac{x_1}{1+\sigma^2}, \frac{\sigma^2}{1+\sigma^2}\right) \tag{9}$$

The estimated probability of reaching the threshold s, depending on decision strategy d_1 and denoting total output $y = \sum_{t=1}^{T}(x_t - B_t)$, becomes

$$
\begin{aligned}
\mathcal{P}\left[y \geq s|x_1, d_1\right] &= 1 - F_{y|x_1,d_1}(s) \\
&= 1 - \Phi\left(\frac{s - \mathbb{E}[y|x_1, d_1]}{\sqrt{Var[y|x_1, d_1]}}\right)
\end{aligned} \tag{10}
$$

From Equation 10, the probability of winning rises in $Var[y|x_1, d_1]$ when $\mathbb{E}[y|x_1, d_1] < s$. The **Experiment** strategy increases the variance without affecting the expectation of the output (recall $\theta \sim \mathcal{N}(0, 1)$). Individuals will therefore **Experiment** when their expected score is below the threshold, i.e. when $x_1 \leq (1 + \sigma^2)\frac{s}{2}$.

The decision to **Keep** or **Drop** the current technology in future periods affects both the variance and the expected value of the output $\mathbb{E}[y|x_1, d_1]$. Keeping the technology has a positive (expected) impact on $\mathbb{E}[y|x_1, d_1]$ if $x_1 \geq 0$. Keeping the technology always increases the variance, as there is uncertainty about the true value of θ_1. For $x_1 \geq 0$, individuals therefore always have an incentive to **Keep**. For $x_1 < 0$, there is a trade-off between a lower expectation and a larger variance. Individuals decide to **Keep** when the output in Period 1 is greater than a threshold $\underline{x_1}$ defined as

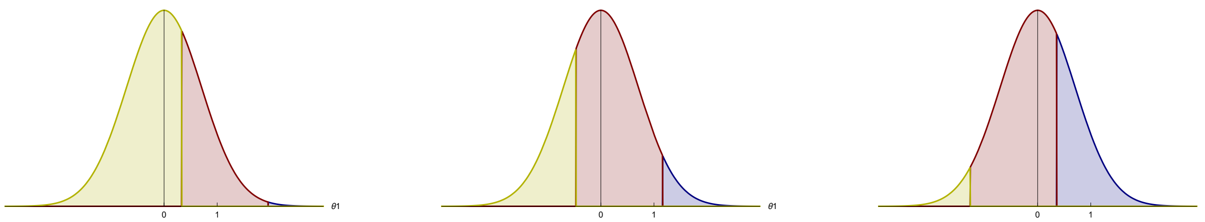
$$
\begin{aligned}
1 - \Phi\left(\frac{s - \mathbb{E}[y|x_1, (Keep, Experiment)]}{\sqrt{Var[y|x_1, (Keep, Experiment)]}}\right) &\geq 1 - \Phi\left(\frac{s - \mathbb{E}[y|x_1, (Drop, Experiment)]}{\sqrt{Var[y|x_1, (Drop, Experiment)]}}\right) \\
\Rightarrow \qquad x_1 &\geq \underline{x_1} = -\frac{s}{3}\left(\sigma^2 + \sqrt{(1 + 2\sigma^2)(1 + 5\sigma^2)} - 1\right)
\end{aligned} \tag{11}
$$

The threshold for keeping the technology falls in σ, so that the individual will decide to keep technologies that are worse in expectation when σ is large, as they believe that it is more likely that the shock reflects a bad common shock rather than poor technology. The threshold also falls in s, so that when s is high (i.e. a small share of individuals receive the prize) individuals retain potentially bad technologies, as the increased variance improves their probability of reaching the high s.

Effect of RPF: The differences between the two groups will depend on the sign of the common shock. Optimal strategies, depending on θ_1 and the value of the common shock, are illustrated in Figure 2. In the case of a negative aggregate shock (Figure 2a), individuals believe that their technology is worse than it actually is, and so are more likely to discard their technology even when θ_1 is higher than 0 (the yellow area) and more likely to continue experimenting even when they should stop (the red area).

If the common shock is neutral ($b_1 = 0$, Figure 2b), then RPF only affects the choice of strategies by increasing uncertainty. Individuals in the control group will **Keep** technologies with negative values, and will **Experiment** for values of θ_1 that are above $\frac{s}{2}$.

When the aggregate shock is positive (Figure 2c), individuals are on the contrary likely to believe that their technology is better than it actually is. The θ_1 threshold between (**Drop**, **Experiment**) and (**Keep**, **Experiment**) is negative, meaning that bad technologies are retained. Individuals are also more likely to stop experimenting, believing that their technology suffices to reach s. In this case their probability of winning can even drop to 0.



(a) Negative shock ($b_1 = -0.8$) (b) Neutral shock ($b_1 = 0$) (c) Positive shock ($b_1 = 0.8$)

Figure 2: The optimal strategy for the no-feedback treatment, depending on the sign of the common shock ($s = 2$ and $\sigma = 0.4$). The area in yellow corresponds to the strategy (**Drop**, **Experiment**), that in red to (**Keep**, **Experiment**) and that in blue to (**Keep**, **Stop**).

To summarise the predictions from the technology model, we can say that the distributions of strategies will depend on the value of the aggregate shock, but in any case, the uncertainty regarding the shock faced in the control group will reduce their average performance. We plot the resulting distributions in Period 2 in Figure 3.[4]

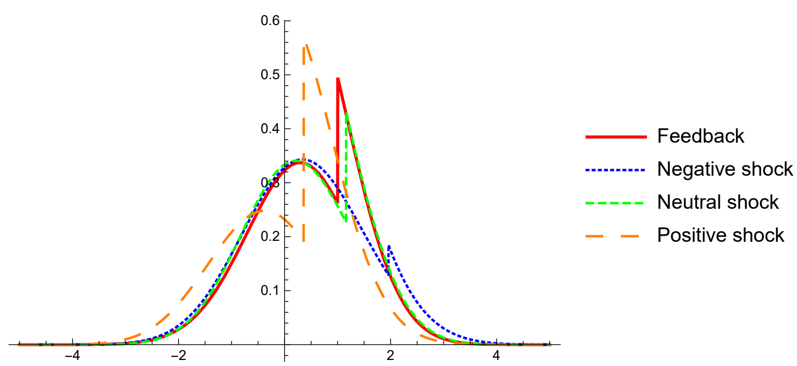


Figure 3: The resulting distributions with $s = 2$ and $\sigma = 0.4$. For the positive shock, $b_1 = 0.8$, for the neutral shock, $b_1 = 0$ and for the negative shock, $b_1 = -0.8$

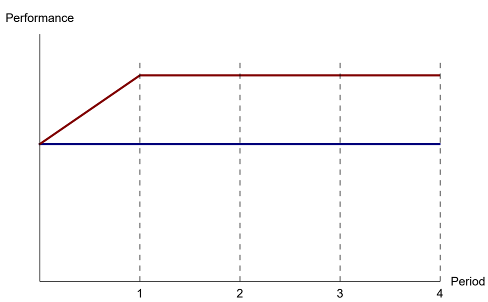
[4]To calculate the distributions, we need the convolution of a truncated normal distribution (for the (**Keep, Experiment**) strategy), which can be found for instance in Turban (2010).

2.3 Comparison of the predictions from the two channels

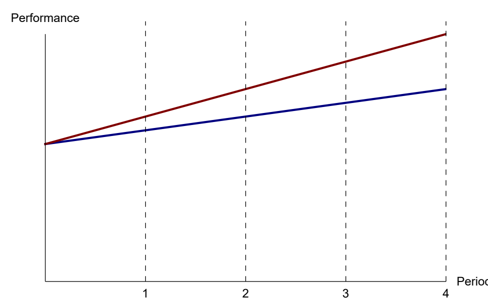
After having described the two ways in which RPF may affect performance, we can compare the two models' theoretical predictions. Under the assumptions made in the two models, both can theoretically explain why performance improves on average when RPF is provided. There are, however, two key differences between the two models' predictions: the ability to rationalise all types of data and the evolution over time.

The first difference between the effort and technology models regards their ability to rationalise data in a single period. As noted in the description of the effort channel, the term (the g function) added to the utility function affects the marginal utility of effort and so performance. By having very few restrictions on the g function, we can rationalise any distribution of a positive RPF effect.[5] The technology channel, however, does not allow for such easy rationalisation.

The second difference relates to their implications in multiple-round settings. Figure 4 displays the evolution of the control (blue curve) and RPF (red curve) groups over time in both models. In the effort model (Figure 4a), the term added to the utility function is only added once in the first period. The distribution of performances shifts in the first period (depending on the shape of the g function) and the two distributions subsequently remain stable over time. On the other hand, in the technology model (Figure 4b) both groups improve across the periods. Providing RPF increases the speed of the increase for the RPF group as compared to the control group.



(a) Effort model



(b) Technology model

Figure 4: The evolution of the average in the RPF (red curve) and control (blue curve) groups in the two models. The lines here are purely illustrative and may not correspond to realistic parameter values.

3 Empirics

We now describe a field experiment in which the task to be performed, math tests at one-week intervals, leaves large room for improvement. As a consequence, we expect to confirm the predictions of the

[5] As noted above, relaxing the monotonicity assumption allows *any* distribution to be rationalised by the effort channel.

technology channel.

3.1 Experimental Design

The experiment took place in 2011 in Iran.[6] The subjects were 14 year-old girls (Eighth grade) in single-sex schools. They took four math exams, which were framed as part of the curriculum. Students were not aware that they were participating in an experiment. They were only told that they would take four exams, and that the two top-ranked students in each class would receive a prize. There were around 30 students in each class.

The experiment was run with tournament incentives: in each class (there were three classes per school), the two best performers received a prize of 1.000.000 Iranian Rials.[7] The final rankings were calculated as the sum of the scores over the last three exams. The first test can thus be considered as a trial.

In the control group, students were only informed about their individual grade; in the treatment group, students were also (privately) told their rank in the class. The total number of students in the no-feedback treatment (control group) is 91, with 83 in the RPF treatment (treatment group). Students in the same group belong to the same school. To ensure comparability, we ran pilot studies to find two schools that were sufficiently similar. As the schools were quite far away from each other, there was little chance of the students realising that a similar (but different) tournament was being organised in another school.

All of the exams consisted of 40 multiple-choice questions. The difficulty of the questions were evaluated using pilot studies, so that the grades across tests are comparable. Improvements across rounds can therefore only be due to improvements in performance (i.e. better math skills rather than easier tests). For each question, four options were presented including one correct answer. Each correct answer obtained four points and each incorrect answer was penalised by one point. Unanswered questions were not penalised.

To ensure that the students would have time to react to the RPF information, the exams were one week apart, and the results (either individual score alone or individual score and ranking) were given to students the day after they had taken the exam.

3.2 Results

The histograms of the scores of students in the treatment and control groups appear in Figure 5. These show that (1) there is a strong positive average effect of RPF on exam scores and (2) students who receive RPF continually improve, as the treatment effect appears to increase over time.

[6] The reason for choosing Iran is that, surprisingly enough, it was much easier to obtain permission to run experiments in schools there than in France.

[7] At the time roughly equivalent to 50€, about one week's wages for a low-skilled worker.

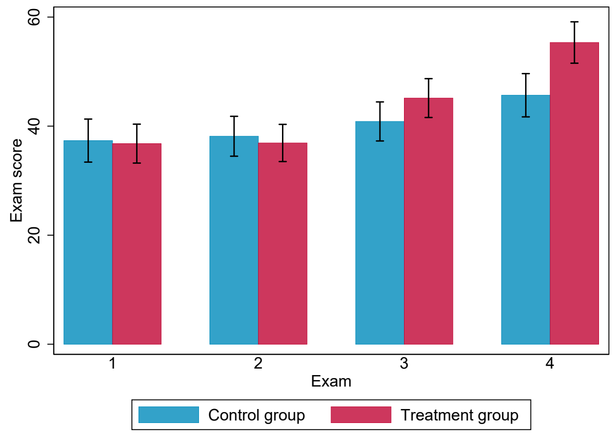


Figure 5: The evolution of average exam scores over time for the control and treatment groups.

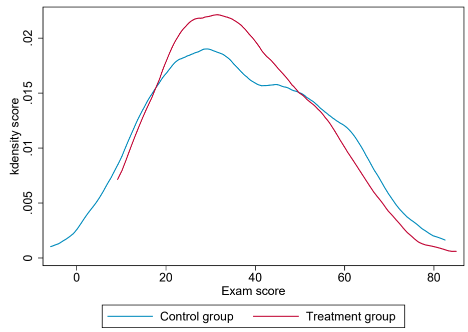
Moreover, as indicated in Figure 6, the RPF appears to shift the entire distribution upwards.

We further test these results using a standard difference in differences framework (Equation 12: the coefficient of interest is β_3). The results in Table 2 further indicate that the size of the RPF effect is large (at about 10 points for a maximum score of 160 points) and robust to the introduction of the controls available (parents' education and number of siblings).

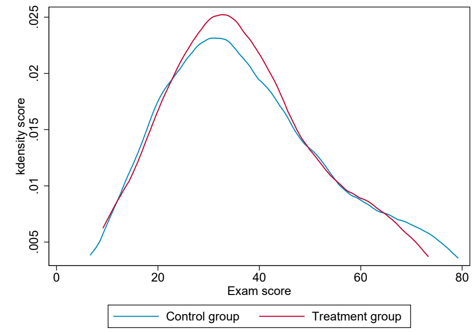
$$Performance_{it} = \beta_0 + \beta_1 Treatment_i + \beta_2 Period_4 + \beta_3 (Treatment_i * Period_4) + \gamma X_i + \varepsilon_{it} \quad (12)$$

The case of low-ability students deserves particular attention. It is often suspected that RPF may harm the weakest students. Our design allows us to identify the sub-group of students who, after round 3, no longer have any chance of winning a prize, even if they obtain the highest possible score. Even so, the average scores in this group do rise between Periods 3 and 4 (see Figure 7). On a more general note, we also see in Figure 7 that RPF has a more homogeneous effect.

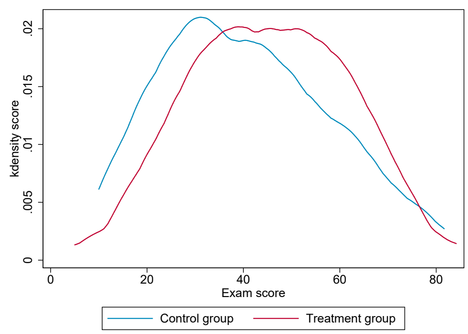
Overall, our empirical findings fit nicely with the predictions of the technology channel: there is evolution in the treatment effect across periods, with an increasing effect in the last periods. This feature is hard to explain using the effort channel. The fact that the treatment effect is more or less constant all along the performance distribution is also predicted by the technology channel. Since there obviously is room for improvement here, we conjecture that students became more efficient by gradually improving their learning technology.



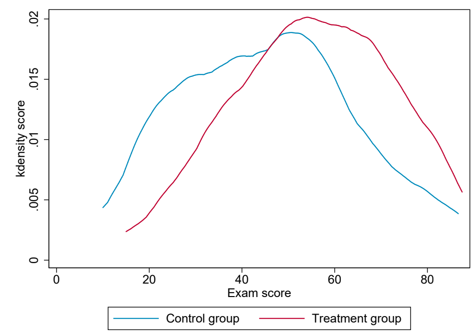
(a) Exam 1



(b) Exam 2

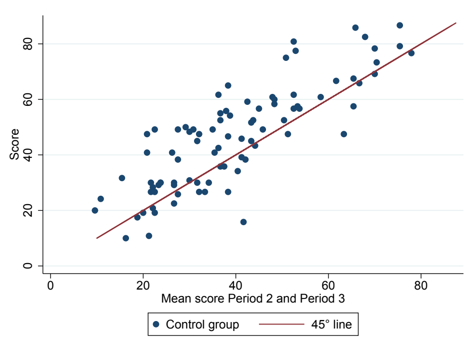


(c) Exam 3

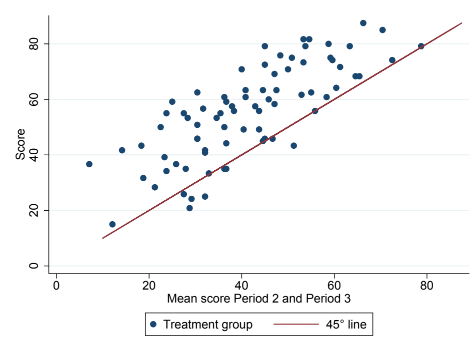


(d) Exam 4

Figure 6: The change in the densities of exam scores over time for the control (green) and treatment (orange) groups



(a) Control group



(b) Treatment

Figure 7: The comparison of performance in the final Period to the average in Periods 2 and 3. Each dot represents an individual. The X-coordinate is the average score in Periods 2 and 3; the Y-coordinate is the score in Period 4.

Table 2: The results of difference-in-differences estimations of providing feedback on exam scores

	(1)	(2)	(3)	(4)
Treatment	-.555	-2.964	-2.729	-3.443
	(2.675)	(2.630)	(2.609)	(2.668)
Period 4	8.318***	8.301***	8.304***	8.167***
	(2.818)	(2.801)	(2.804)	(2.822)
Period4 * Treatment	10.226***	9.538**	9.542**	9.736**
	(3.845)	(3.756)	(3.732)	(3.767)
Mother's education		2.541***	1.262	.686
		(.703)	(.863)	(.902)
Father's education			1.797**	2.002**
			(.823)	(.851)
Number of siblings				-2.509**
				(1.120)
Constant	37.348***	29.236***	26.374***	31.990***
	(1.987)	(2.993)	(3.227)	(4.161)
R^2	.150	.169	.179	.189
No. obs	336	321	321	315

* p<0.10, ** p<0.05, *** p<0.01. The dependent variable is the exam score. The initial period in the estimations is Period 1. Robust standard errors appear in parentheses.

4 Conclusion

RPF is often assumed to affect performance by changing optimal effort. We here propose an alternative channel, namely technology. We identify "room for improvement" as a key distinguishing feature of the technology versus the effort channels in the effect of RPF on performance. A first test of our approach is to make sense of the conflicting results found in previous work, in the light of whether there is large room for improvement. Why exactly RPF may have a positive or negative effect was a source of some scepticism among researchers. The technology channel is certainly not the definitive answer, but it at least suffices to organise the existing empirical evidence in a natural way. The second test of the technology channel comes from a (natural) field experiment. The results here are in line with the theoretical predictions of the technology channel.

This paper therefore proposes a wider view of RPF. Traditionally, probably the best-known use of RPF is in tournaments, for instance among workers in a firm who compete for promotion. Tournaments have long been recognised as creating powerful incentives, especially for top performers (see Sheremeta (2016) for an in-depth and stimulating discussion of the pros and cons of tournaments). RPF is used in these cases to let top performers know that the reward is within reach. The interpretation of the subsequent improvement of the top performers is then assumed to take place through the effort channel: top performers will work harder. Although very common, tournaments are nonetheless a specific use of RPF, resulting in RPF being closely associated with greater competition. Furthermore, tournaments (and RPF) may discourage low performers, conveying an image of promoting winners and stigmatising poor performers.

On the contrary, the technology channel proposes a different (perhaps simultaneous) explanation of how RPF affects performance that has little to do with competition. To illustrate, consider a poor performer in a situation in which performers are not subject to any kind of competition (their payoffs do not depend on others' performance) and RPF is not publicly provided (there is no social-image concern, like shaming). By (privately) knowing that many others are doing much better, she can learn how good her attempts to improve her technology have been. As a result, RPF provides her with crucial information on improving her performance. RPF can therefore have a positive effect on performance without being associated with incentives, nor appealing to anything like a taste for ranking.

The findings from this paper are related to two branches of the literature. They could first be of particular interest in the field of education. RPF is traditionally seen as detrimental in education, as it is argued to negatively affect poor performers. Note that RPF can be publicly- or privately-provided. RPF is generally associated with public rankings, which indeed may have an adverse effect on poor performers (via for instance poorer self-image). As a consequence, teachers are often reluctant to provide rankings to young pupils. However, the effect of RPF on performance via the technology channel does not require RPF to be public. As shown in a number of experiments, private RPF can be beneficial even for poor performers, and may therefore be a cheap and efficient way of improving performance.

This paper is also related to the literature on social norms. In the dominant interpretation, social norms, once made salient to individuals, influence behaviour as individuals want to comply with the norm. Our work here suggests that there also exists an alternative interpretation. Social norms can indeed be considered as RPF as they may convey information about the technology used. One good example is electricity consumption. Electricity consumers have been found to reduce their consumption when provided with information about their consumption relative to that of similar households in their neighbourhoods, without any additional incentives (see Goldstein et al. (2008) and Schultz et al. (2007)). In these situations, individuals may learn thanks to RPF how to use electricity in a more efficient way, and change their behaviour accordingly. Individuals may thus comply with the norm, not (only) because they wish to comply *per se*, but because they can then learn about better technologies.

References

Anderson, A. and L. M. Cabral (2007). Go for broke or play it safe? dynamic competition with choice of variance. *The RAND Journal of Economics 38*(3), 593–609.

Aoyagi, M. (2010). Information feedback in a dynamic tournament. *Games and Economic Behavior 70*(2), 242–260.

Azmat, G., M. Bagues, A. Cabrales, and N. Iriberri (2018). What you don't know... can't hurt you? A field experiment on relative performance feedback in higher education. *Management Science (in press).*

Azmat, G. and N. Iriberri (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.

Azmat, G. and N. Iriberri (2016). The provision of relative performance feedback: An analysis of performance and satisfaction. *Journal of Economics & Management Strategy 25*(1), 77–110.

Bandiera, O., V. Larcinese, and I. Rasul (2015). Blissful ignorance? A natural experiment on the effect of feedback on students' performance. *Labour Economics 34*, 13–25.

Barankay, I. (2011). Gender differences in productivity responses to performance rankings: Evidence from a randomized workplace experiment. *Unpublished working paper, The Wharton School*.

Blanes i Vidal, J. and M. Nossol (2011). Tournaments without prizes: Evidence from personnel records. *Management Science 57*(10), 1721–1736.

Che, Y.-K. and J. Horner (2019). Optimal design for social learning. *Quarterly Journal of Economics*, forthcoming.

Ederer, F. (2010). Feedback and motivation in dynamic tournaments. *Journal of Economics & Management Strategy 19*(3), 733–769.

Eriksson, T., A. Poulsen, and M. C. Villeval (2009). Feedback and incentives: Experimental evidence. *Labour Economics 16*(6), 679–688.

Farrokhi-Kashani, M. (2012). *Improving students' performance via relative performance feedback: evidence from three field experiments*. Ph. D. thesis, Paris 1.

Fershtman, C. and U. Gneezy (2011). The tradeoff between performance and quitting in high power tournaments. *Journal of the European Economic Association 9*(2), 318–336.

Fischer, M. and V. Wagner (2018). Effects of timing and reference frame of feedback: Evidence from a field experiment.

Gershkov, A. and M. Perry (2009). Tournaments with midterm reviews. *Games and Economic Behavior 66*(1), 162–190.

Gill, D., Z. Kissová, J. Lee, and V. Prowse (2018). First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Management Science 65*, 494–507.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: Gender differences. *Quarterly Journal of Economics 118*(3), 1049–1074.

Goldstein, N. J., R. B. Cialdini, and V. Griskevicius (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research 35*(3), 472–482.

Goltsman, M. and A. Mukherjee (2011). Interim performance feedback in multistage tournaments: The optimality of partial disclosure. *Journal of Labor Economics 29*(2), 229–265.

Halac, M., N. Kartik, and Q. Liu (2017). Contests for experimentation. *Journal of Political Economy 125*(5), 1523–1569.

Hannan, R. L., R. Krishnan, and A. H. Newman (2008). The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. *The Accounting Review 83*(4), 893–913.

Hvide, H. K. (2002). Tournament rewards and risk taking. *Journal of Labor Economics 20*(4), 877–898.

Kremer, I., Y. Mansour, and M. Perry (2014). Implementing the wisdom of the crowd. *Journal of Political Economy 122*(5), 988–1012.

Kuhnen, C. M. and A. Tymula (2012). Feedback, self-esteem, and performance in organizations. *Management Science 58*(1), 94–113.

Kuziemko, I., R. W. Buell, T. Reich, and M. I. Norton (2014). last-place aversion: Evidence and redistributive implications. *The Quarterly Journal of Economics 129*(1), 105–149.

Renault, J., E. Solan, and N. Vieille (2017). Optimal dynamic information provision. *Games and Economic Behavior 104*, 329–349.

Schultz, P. W., J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science 18*(5), 429–434.

Seel, C. and P. Strack (2013). Gambling in contests. *Journal of Economic Theory 148*(5), 2033–2048.

Sheremeta, R. M. (2016). The pros and cons of workplace tournaments. *IZA World of Labor*.

Tran, A. and R. Zeckhauser (2012). Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics 96*(9-10), 645–650.

Turban, S. (2010). Convolution of a truncated normal and a centered normal variable. Technical Report 17, University of Columbia.

Wirtz, J. (2016). Feedback and learning in tournaments.