Replication of

Screening, Competition and Job Design: Economic Origins of Good Jobs

by Bartling, B./Fehr, E./Schmidt, K.M. (2012)
in: The American Economic Review, 102(2), pp. 834–864.

Replication Authors:
Felix Holzmeister, Jürgen Huber, Michael Kirchler and Michael Razen

In an employer-employee setting, Bartling et al. test whether adding a screening opportunity (informing the employer about the effort of the work in the past three periods) for an employer that can offer full or limited discretion contracts increases efficiency or not compared to a base treatment without screening opportunity. They find that a screening opportunity increases efficiency (average per-period total surplus).

Hypothesis to bet on:

Adding a screening opportunity (informing the employer about the effort of the worker in the past three periods) for an employer that can offer full or limited discretion contracts increases efficiency (a comparison in average per-period total surplus between the base treatment and the screening treatment).

Power Analysis

The original p-value is 0.007 (Wilcoxon-Mann-Whitney test using matching group averages as the unit of observation; p. 854 under result 6): *"Overall, the average per-period total surplus is 58 percent higher in the screening treatment – a difference that is highly significant (rank-sum test on matching group averages, $p = 0.007$)."*

The original sample size is 216 participants (108 in the base treatment and 108 in the screening treatment). To achieve 90% power the required sample size is 312 participants.

Sample

The sample for replication consists of 360 students (5 sessions à 36 subject in each of the two treatments) at the University of Innsbruck in Austria. There are no exclusion criteria.

Materials

We use the material of the original experiment (programmed in z-Tree) along with the original German instructions, both available at the journal's webpage.

Procedure

We follow the procedure of the original article, with only slight but unavoidable deviations as outlined below. The following summary of the experimental procedure is therefore based on the section "I. Experimental Design and Procedures" (pp. 839–841) in the original study.

Subjects intending to participate have to register for two sessions and will be randomly assigned to one of the treatments once registration has closed. Upon arrival at the lab, half of the subjects are randomly and anonymously assigned the role of an employer, the other half the

role of an employee. The experiment is framed as an employment relationship.

The employer can offer an employment contract to the employee that specifies a fixed wage w and a requested, nonbinding, effort level $\hat{e}$. If the employee rejects the contract offer, no wage is paid, no effort is exerted, and both parties receive their reservation payoffs of 0. If the employee accepts, the employer must pay the offered wage, irrespective of the actual effort the employee chooses.

There are two types of contracts that the employer can offer: a *contract with full discretion* and a *contract with limited discretion* which differ in two dimensions:

(i) *Feasible effort levels*: In a contract with full discretion, the employee can choose any effort level between 1 and 10, whereas he must choose an effort level of at least 3 in a contract with limited discretion given he accepts the contract.

(ii) *Efficiency*: The effort efficiency of the relationship is characterized by an employee's efficiency parameter $b = 5$ in a contract with full discretion, while the efficiency parameter is only $b = 4$ in a contract with limited discretion.

All employers and employees know the payoff functions, the set of feasible contracts, and hence the efficiency implications of limited discretion.

In each period, an employer is randomly matched with one of the employees. In the base treatment, the employer does not receive any information about his current employee, while he receives an imperfect signal about his current employee's track record in the screening treatment: he is informed about his current employee's effort choices in the last three periods. Note that an employer neither observes the contract types, the wage offers, nor the requested effort levels that his current employee faced in the last three periods. Employees know that future employers will be able to observe their current effort choices. Apart from the information given to the employers in the screening treatment, the two treatments are identical.

As in the original experiment, sessions of 36 subjects for each treatment are run and two matching groups of 9 employers and 9 workers in each session are implemented. As each session involves 15 periods where employers are randomly matched with employees, some of the subjects interact with the same opponent more than once. However, subjects do not know that they are divided into two matching groups, nor do they know whether and, if so, with whom they interact more than once. We will randomly allocate subjects into the different treatments.

After all rounds have been played, subjects will be privately paid in cash based on the same incentives as in the original study (average earnings were CHF 43.00 per subject in the original study).

Analysis

The analysis will be performed exactly as in the original article. That is, Wilcoxon-Mann-Whitney tests on matching group averages will be conducted.

Differences from Original Study

The replication procedure is identical to that of the original study, with some unavoidable deviations. This replication will be performed at the University of Innsbruck in Innsbruck, Austria, in 2015, on students from the University of Innsbruck, while the original data was gathered at the Zurich University in Zurich, Switzerland, in 2007, on students from Zurich University and the Federal Institute of Technology in Zurich. The experiment will be in German as in the original study.

In a third treatment the original study also looks at labor market competition, but the focus of the replication is on the effect of the screening opportunity, hence we replicate only treatments base and screening. As no participants in the original experiment participated in more than one session and treatment this should make no difference.

Replication Results

In the replication experiment, the average per-period total surplus per matching group in the screening treatment is 80.28% higher than in the base treatment without screening opportunity (see Table 1) – a difference that is statistically significant with a p-value of 0.001 (Wilcoxon-Mann-Whitney test).

As a measure for the effect size we take the difference of average per-period total surplus between the two treatments, which equals 7.16 in the replication experiments. In comparison to the effect size of 5.92 in the original study, the relative effect size of the replication experiment equals 120.95% (7.16/5.92).

As highlighted in the table, the results of the replication experiments are even more pronounced compared to the original study. While the matching group averages of the screening treatment are similar, average incomes and total surplus in the base treatment are lower in the replication study. These differences result in an even more pronounced efficiency-enhancing effect in the replication study as all treatment differences are statistically significant at the 1%-level. Thus, the treatment difference for the employees turns out to be significant as well in the replication study.

Unplanned Protocol Deviations

Due to the unavailability of a room with 36 cubicles, the replication experiment was conducted in two separate (but basically similar) laboratories next to each other. That is, instead of treating both matching groups of 18 participants in one single room as in the original study, each matching group was seated in a separate room (importantly, all 36 subjects showed up simultaneously). Apart from that, the replication experiment has been conducted exactly the way as outlined above, without further deviations from the protocol.

Discussion

Given the criteria and procedure outlined above, the hypothesis of interest has been replicated at a significance level of $\alpha < 5\%$. The relative effect size equals 120.95% and the p-value of the hypothesis test is 0.001.

Table 1: Comparison of per-period matching group averages of employers, employees, and the total surplus in both treatments

	Original Study				
	Base	*Screening*	*Difference in %*	*z-value*	*N*
Employers	1.86	6.55	251.66	2.882***	12
Employees	8.39	9.62	14.77	1.281	12
Total Surplus	**10.25**	**16.17**	**57.80**	2.722***	12
	Replication Study				
	Base	*Screening*	*Difference in %*	*z-value*	*N*
Employers	2.54	6.46	154.63	3.628***	20
Employees	6.38	9.61	50.70	2.646***	20
Total Surplus	**8.91**	**16.07**	**80.28**	3.250***	20

*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level