European Commission / DG Justice

DIPARTIMENTO PER LE PARI OPPORTUNITÀ- PRESIDENZA DEL CONSIGLIO DEI MINISTRI
in partnership with
DONDENA RESEARCH CENTER ON SOCIAL DYNAMICS AND PUBLIC POLICY, UNIVERSITA' BOCCONI

WS2 – A pilot experiment of new working arrangements

1. Research report of the experiment on the Italian company

1.1 Design of the experiment

The objective of this workstream is to produce innovative policy research through a pilot experiment conducted on an Italian company with the scope of assessing the causal impact of policies related to flexible working arrangements on productivity of workers, wellbeing of individuals, work-life balance.

The analysis of the literature in WS1 highlighted the lack of rigorous analysis to understand the consequences of flexible work arrangements. Without a fully-fledged randomised and structured experiment, it is impossible to provide clear conclusions or to describe the mechanisms that play a role in achieving the results of flexible work on productivity, well-being and work-life balance.

ACEA Company, multi-utility of Rome, was selected for the experiment. Taking into consideration criteria linked to the representativeness compared to the population of ACEA employees and the measurability of the productivity of the persons involved in the trial, we identify a sample of 310 workers and their respective supervisors in the company.
The sample is randomly split into two sub-groups, respectively made up of 200 and 100 people: the former group is subject to the treatment (Treated group), while the latter continue to work according to the pre-existing arrangements, constituting the Control group. The treatment consists in the identified flexible work arrangement (see workstream 1): in accordance with the supervisor, the worker can work outside the company, in the place he/she prefers and at the time he/she prefers one day per week (not to be sub-divided). The treatment lasts for 9 months.

Table 1 summarizes the characteristics of the randomized groups.

Table 1: Randomised target population

Gender	Age bracket	Position	Level	Sample	Treated
Male	27/45	Employee	1	80	52
		Labourer	2	30	19
	46/63	Employee	3	51	33
		Labourer	4	13	8
Female	27/45	Employee	5	70	45
		Manager	6	6	4
	46/68	Employee	7	57	37
		Manager	8	3	2
Total				**310**	**200**

We use questionnaires to collect information on the workers in the final sample. All workers in the sample and their supervisors answer to a pre-policy questionnaire, aimed at gathering information not directly collected by the company, with the aim of having a background on the starting situation of the participants. The themed areas of investigation of the questionnaire are summarized in table 2.

Table 2: Themed areas

Worker Questionnaire	Supervisor Questionnaire
Family	Flexibility
Productivity	Underlying costs
Environmental impact	Productivity
Flexibility	Commitment to company
Wellness	
Work-life balance	
Commitment to company	

The experiment lasts 9 months. After that period, all workers in the sample are asked to fill in a post-policy questionnaire, containing the same questions asked before, and additional questions for the treated group related to the experience of flexible work.
The questionnaires are available upon request.

To analyse the results of the experiment based on the answers to the questionnaires, we use a ***"Difference in difference"*** model. The difference-in-difference estimate is a methodology used mainly in the socio-economic field to estimate the effect of a "treatment" (for example, the introduction of a policy measure) on a group of people ("treated"), as opposed to a second group of people not exposed to the treatment ("controls"). The two groups are observed in two periods, one before and one after the treatment. Figure 1 summarizes the working of the model, and figure 3 provides a graphical representation of the methodology adopted in our specific experiment.

Figure 1: *"Difference in difference"* Model

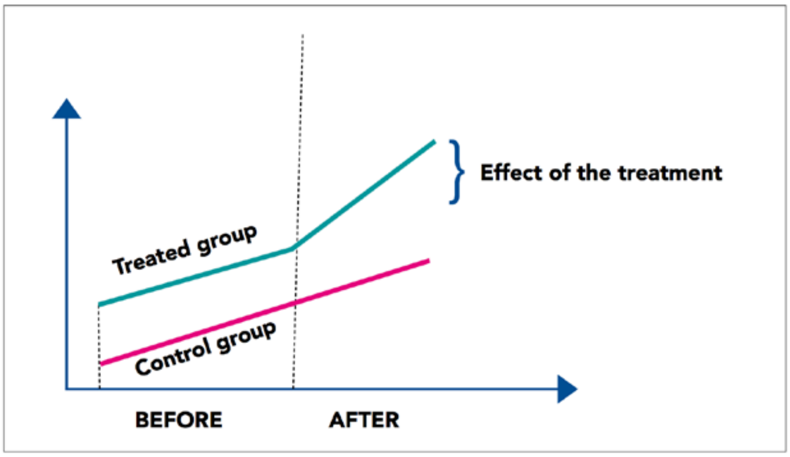
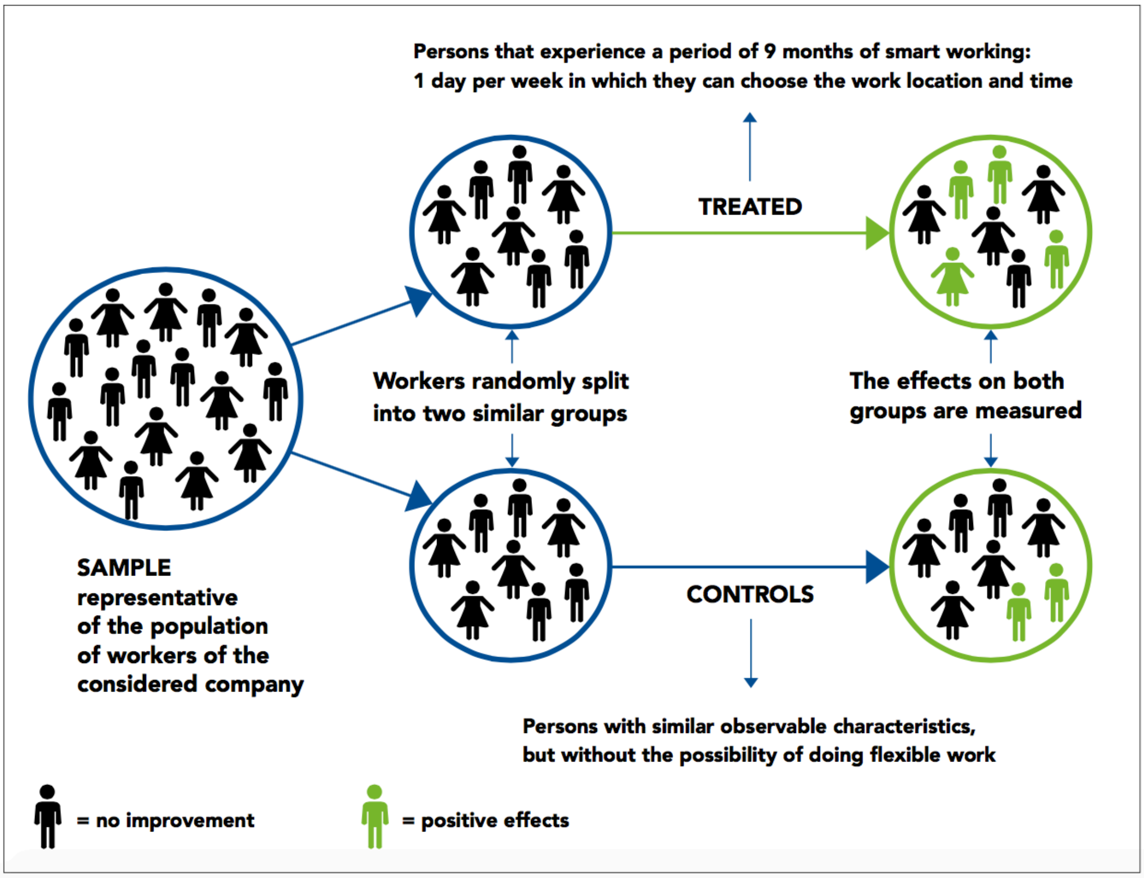


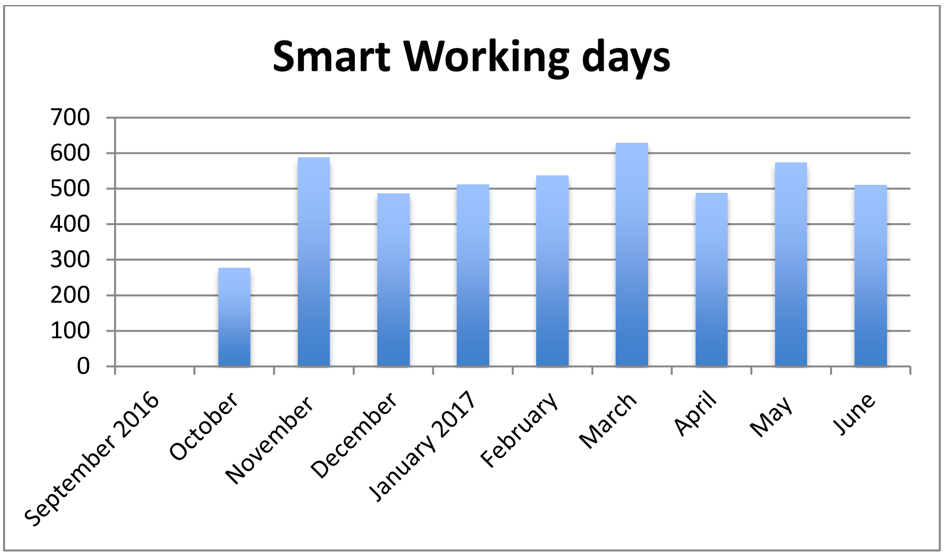
Figure 2: Graphical representation of trial methodology



1.2 Results

The smart working days used during the treatment period are summarized in figure 3.

Figure 3: Monthly use of smart working

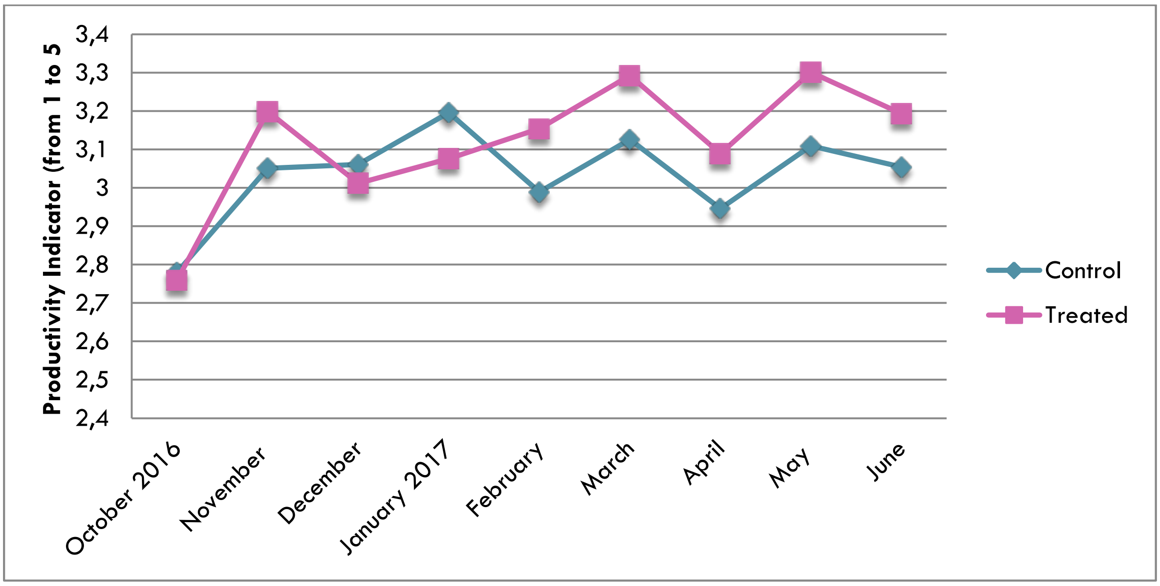


We assess the effect of the introduction of the smart-working policy on the following outcomes:

Starting from productivity, we measure it in three ways: productivity self-provided by the worker (from the worker's questionnaire), reported from the supervisor (from the supervisor's questionnaire) and objective data provided by the company. This last data is measured each month starting from the month before the trial (September 2016) to the last complete month of the trial (June 2017) by an indicator which assumes values from 1 (low productivity) to 5 (high productivity).

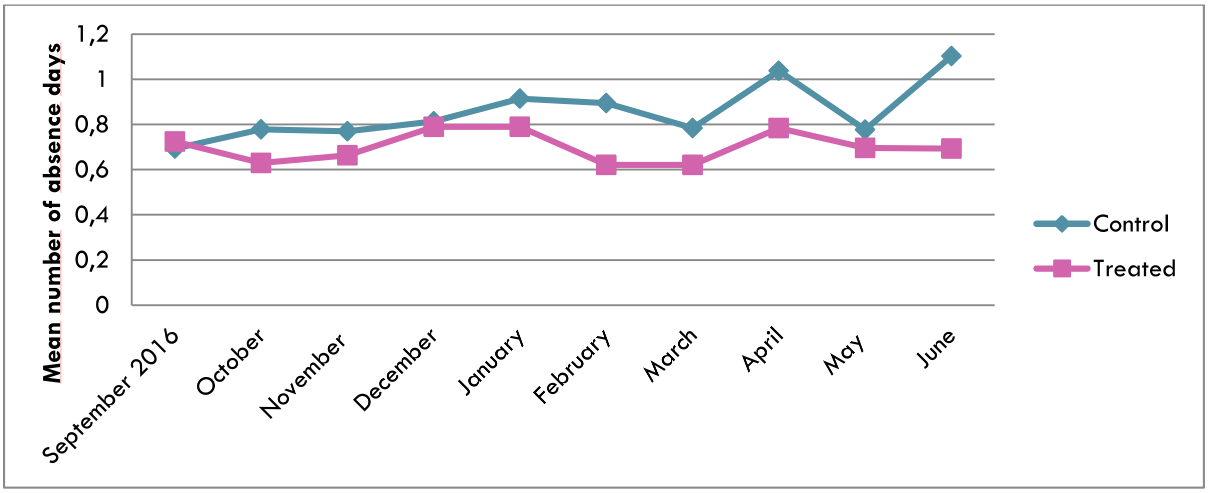
At the end of the trial period of smart working in the company, all these measures show an increase. In particular, in relation to objective measure, figure 4 shows that after four months, the productivity of the treatment group of workers is steadily higher than the productivity of the control group.

Figure 4: Objective productivity



Another objective data measured by the company relates to the days of absence from work due to sickness or special leave. Figure 5 shows that from the comparison between treated and control groups, the total of monthly absences for the treated group decreases during the treatment.

Figure 5: Sickness and special leave absences



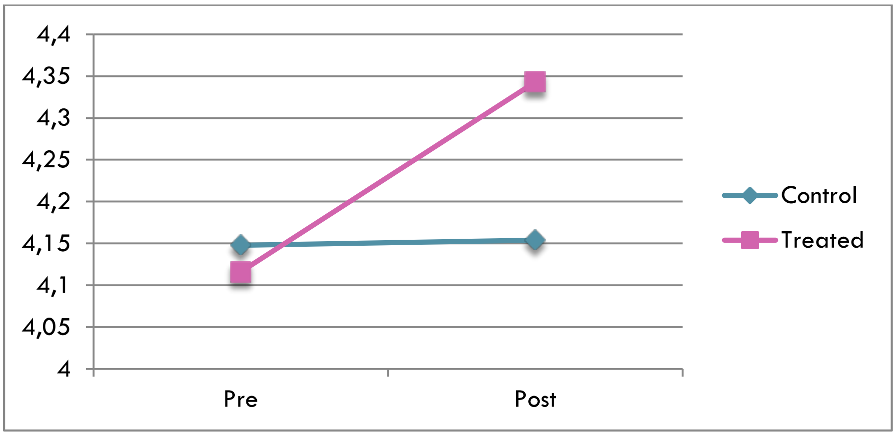
When we consider days of absences from work for holidays instead, there are no differences between the treated group and the control group (figure 6). Thus, there is no connection between smart working and holidays.

Figure 6: Days of paid leave

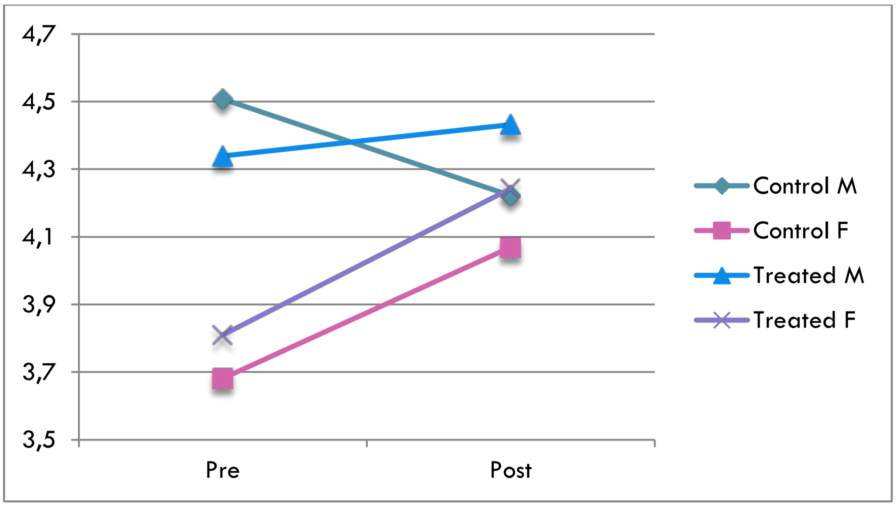
We then move to our second category of outcomes: wellbeing. The wellbeing of employees is measured on the basis of the statements of employees' satisfaction in relation to several dimensions of quality of life, which include both their working sphere (*Salary, Health, Work, Social Life, Free Time, Life in general*) and their emotive sphere (*Ability to concentrate on activities, Loss of sleep due to worry, Useful role in working life, Ability to make decisions, Ability to appreciate daily activities of a normal day, Feeling under pressure, Feeling of not being able to overcome difficulties*).
The worker is asked to express an opinion on his/her satisfaction on a scale ranging from 1 (not at all satisfied) to 7 (completely satisfied). Individuals responded to the same pre-post experiment questions and we can compare them.
The analyses conducted on the data collected through the questionnaires show a statistically significant difference between the treated and the control group in the various dimensions of quality of life before and after the treatment. The difference signals an increase of wellbeing for treated workers with respect to control ones. Figure 7 shows wellbeing indicators for the pre and post treated and control groups. We also report the graph divided between male and female workers, when the gender difference is significant.

Figure 7. Wellbeing indicators: Average Satisfaction

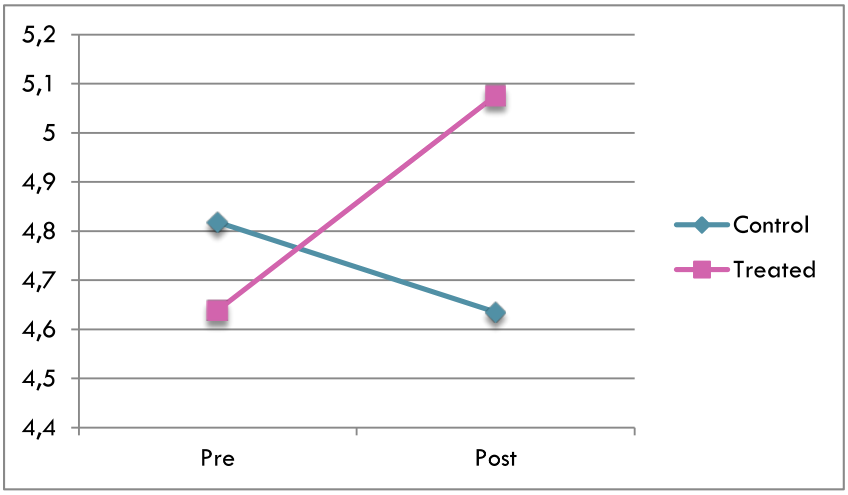
a) Average Satisfaction with Income



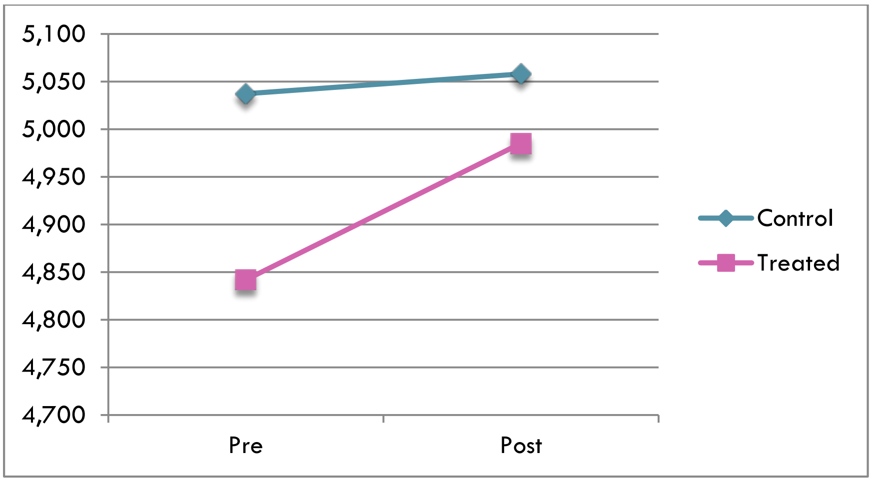
a1) Average Satisfaction with Income by gender



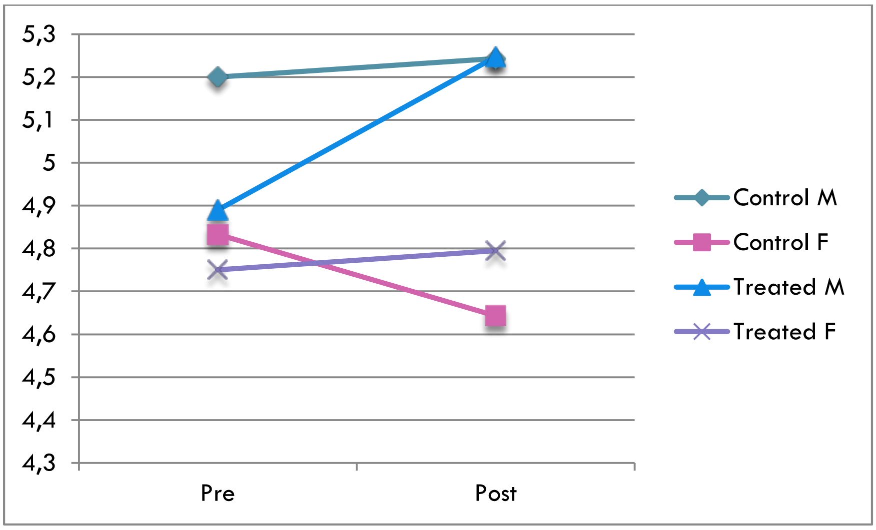
b) Average Satisfaction with Health



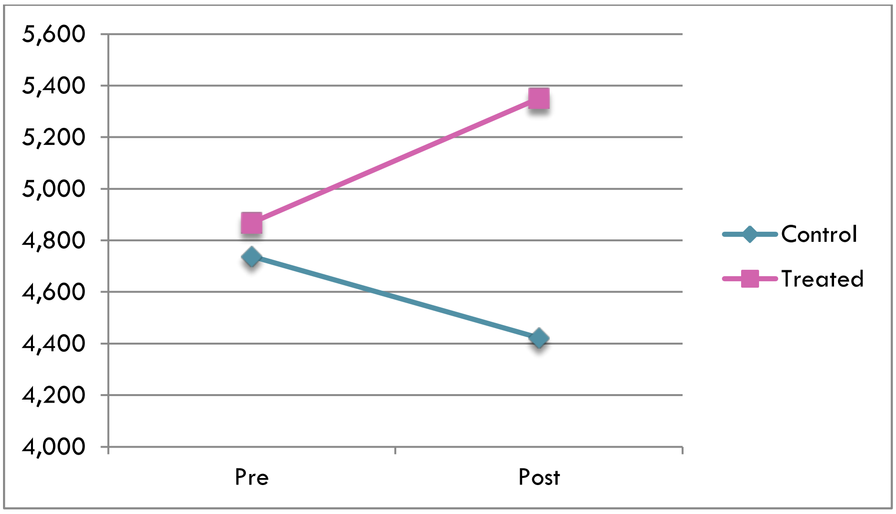
c) Average Satisfaction with Work



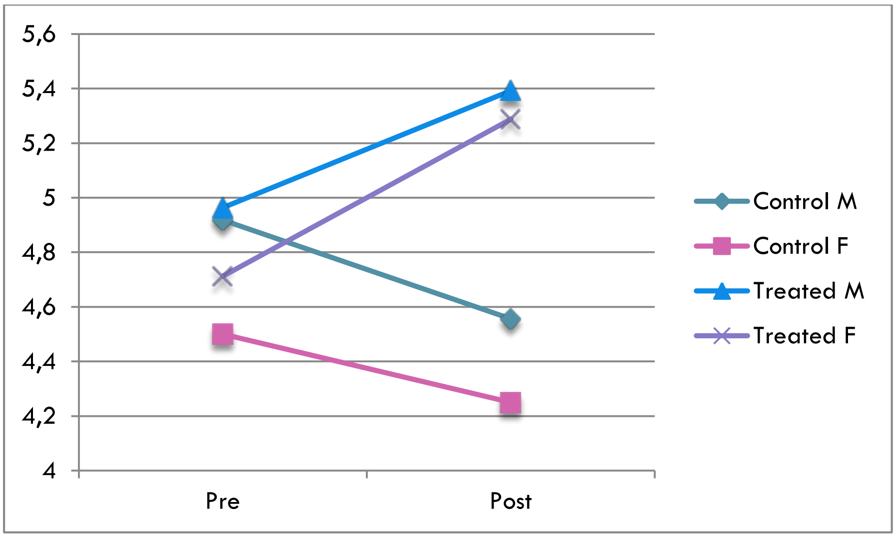
c1) Average Satisfaction with Work by gender



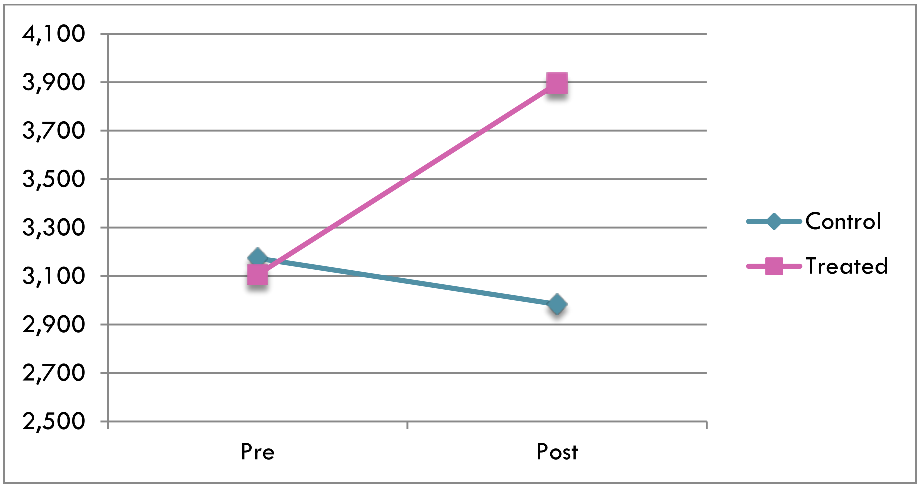
d) Average Satisfaction with Social Life



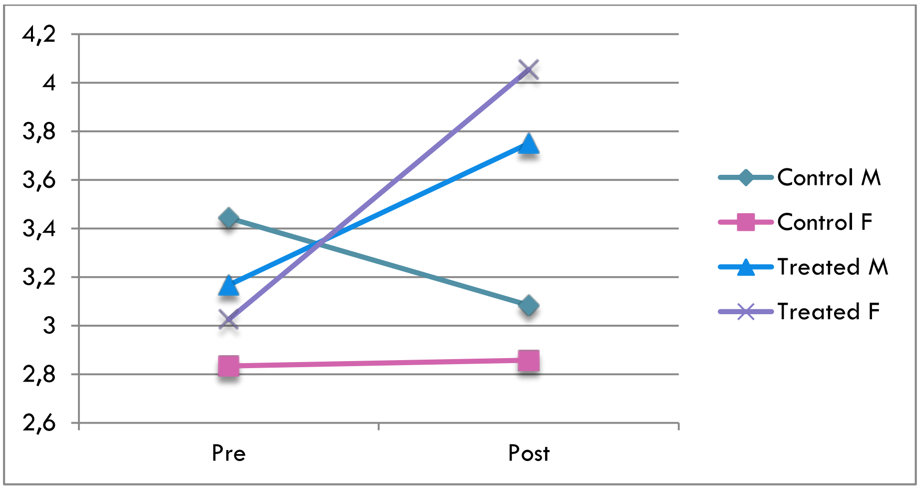
d1) Average Satisfaction with Social Life by gender



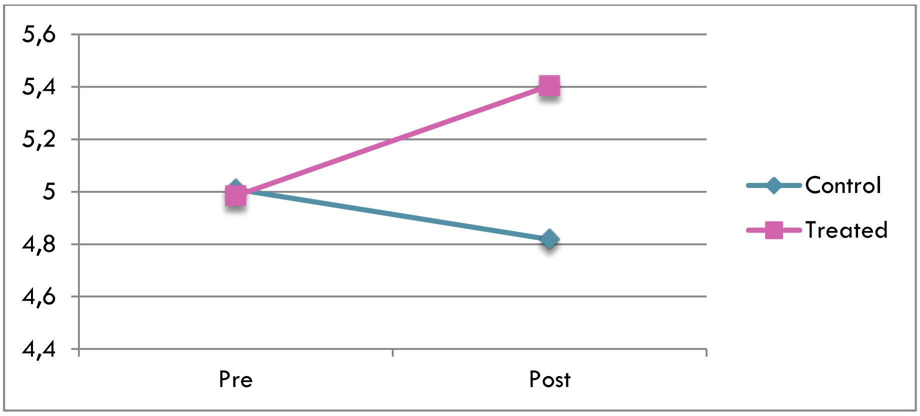
e) Average Satisfaction with Leisure Time



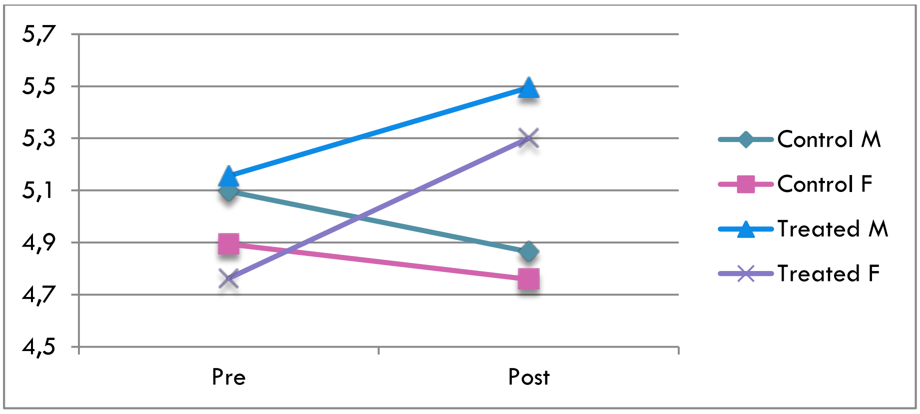
e1) Average Satisfaction with Leisure Time by gender



f) Average Satisfaction with Life in General



f1) Average Satisfaction with Life in General by gender

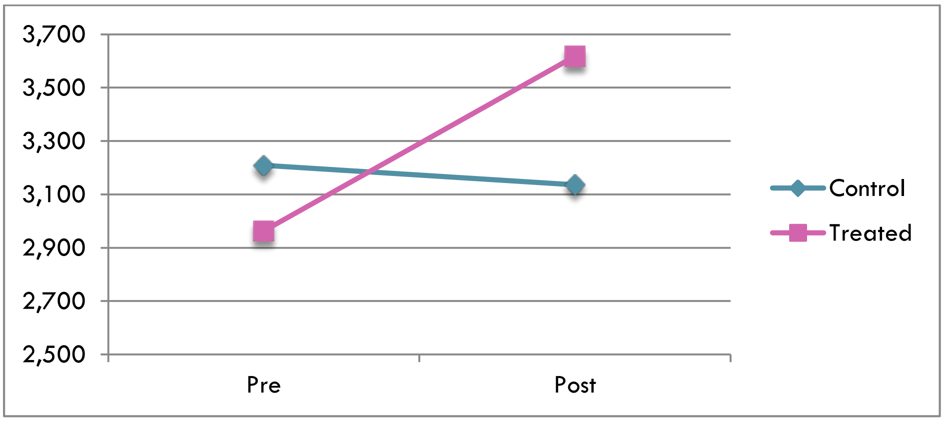


The graphs demonstrate a stark increase in satisfaction for the treated group and show that, for some of the indicators, the increase for women is higher than that for men.

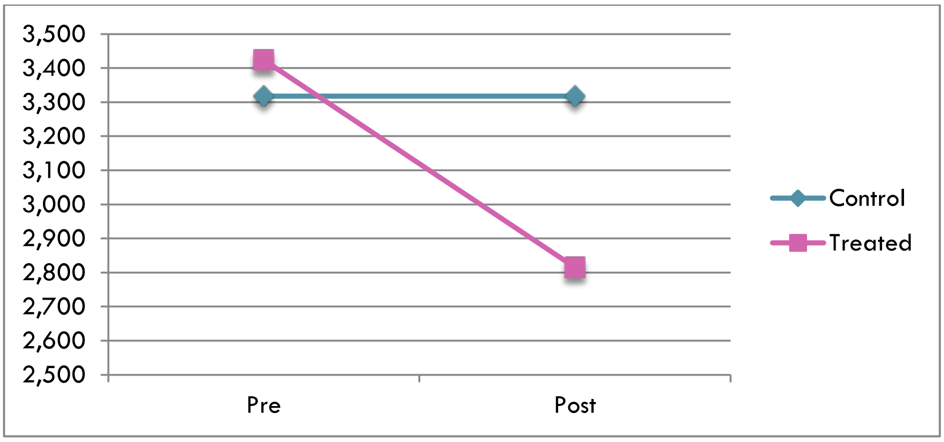
We also investigate some aspects of wellbeing related to the ability to do some activities perceived on a scale from 1 (Much less than usual) to 5 (much more than usual). Figure 8 compares the average values of treated and control groups, pre-post experiment.

Figure 8: Wellbeing indicators: Ability

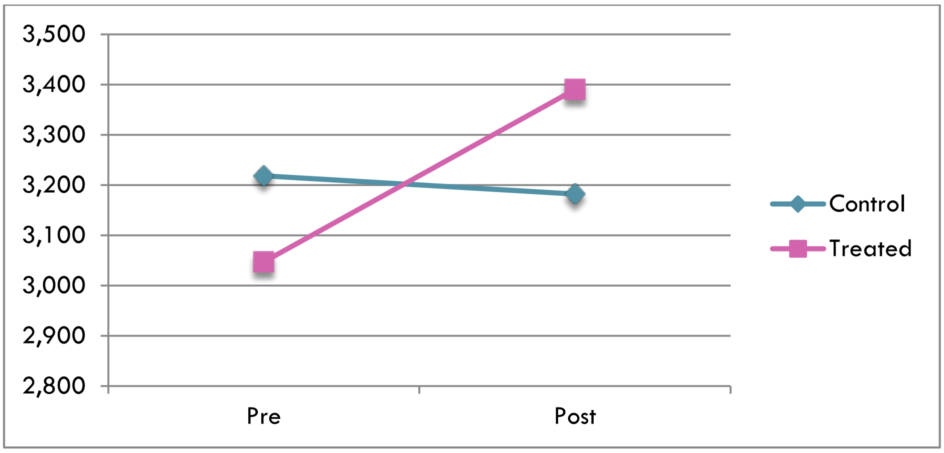
a) Ability to concentrate on their activities



b) Loss of sleep due to concerns[1]

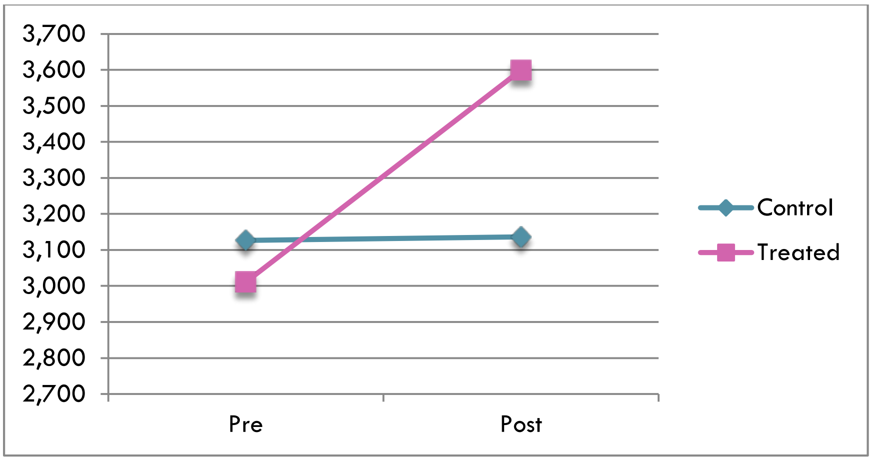


c) A useful role in working life

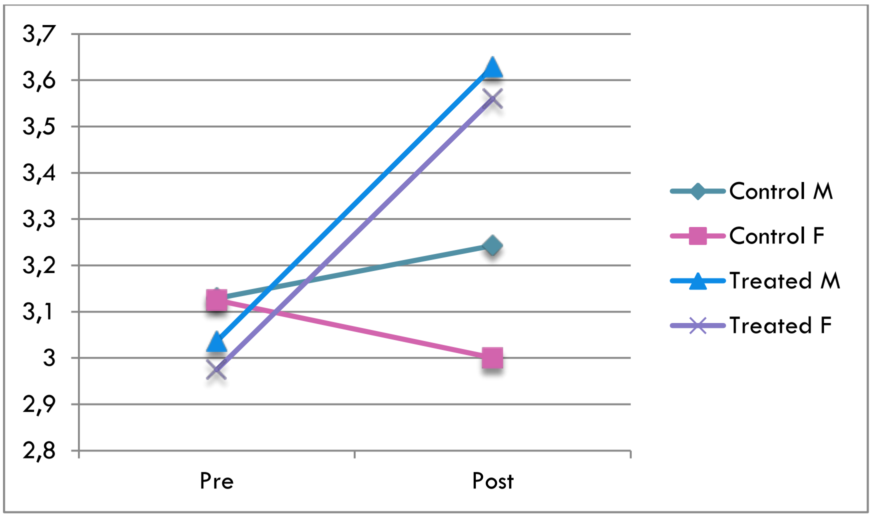


[1] In this case since the loss of sleep is a negative factor, the decrease of this indicator after the treatment for the treated groups signals an improvement of wellbeing.

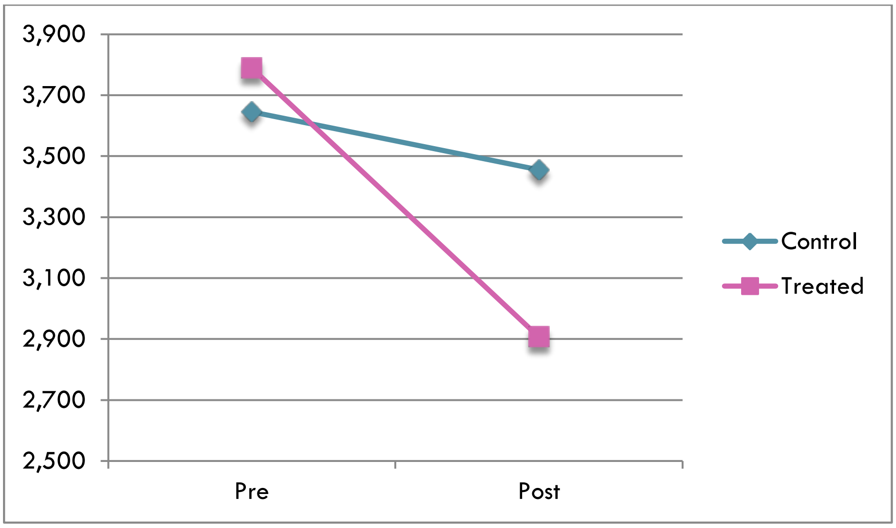
d) Ability to appreciate day-to-day activities of a normal day



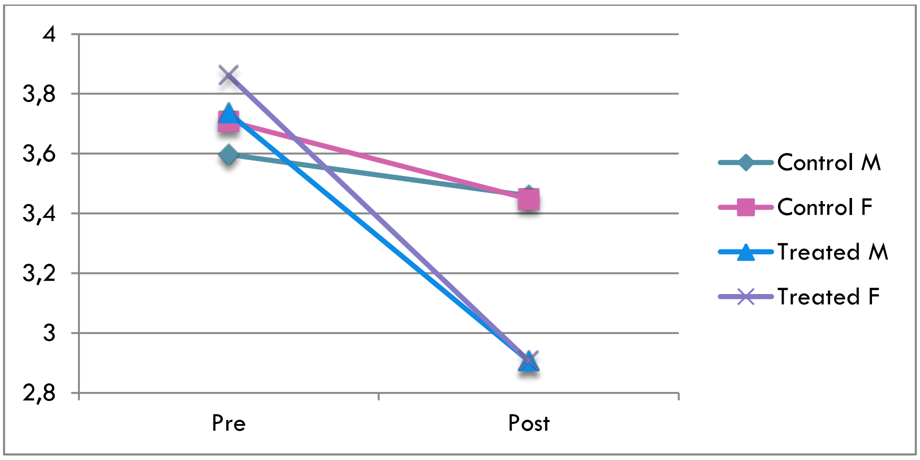
d1) Ability to appreciate day-to-day activities of a normal day by gender



e) Feel under stress



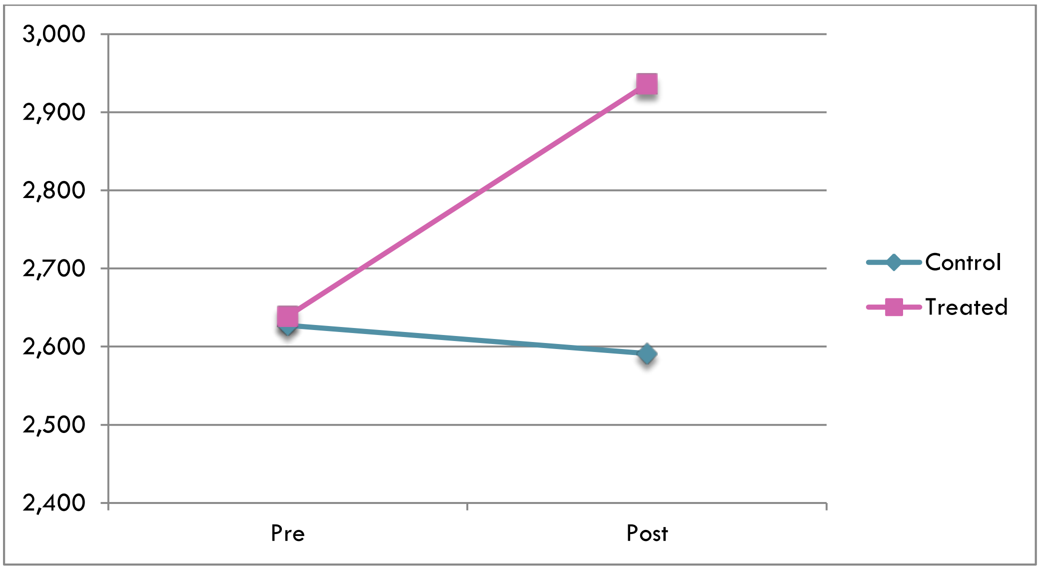
e1) Feel under stress by gender



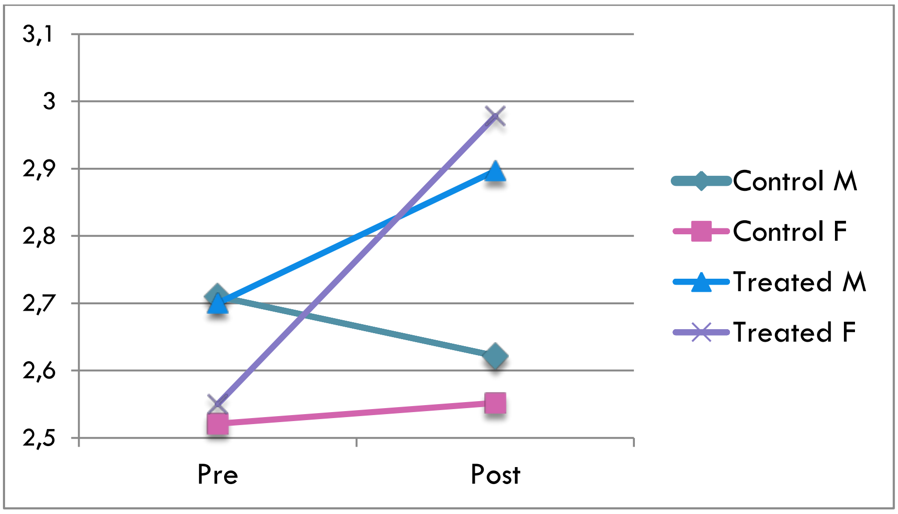
We then analyse indicators of <u>work-life balance</u>. Workers respond with a scale ranging from 1 (very unsatisfied) to 5 (very satisfied). Figure 9 shows that the satisfaction following the experiment significantly increases for the treated group with respect to the control group. The effect is larger for women (panel b).

Figure 9. Work-life balance

a) Satisfaction with adapting working hours to private life

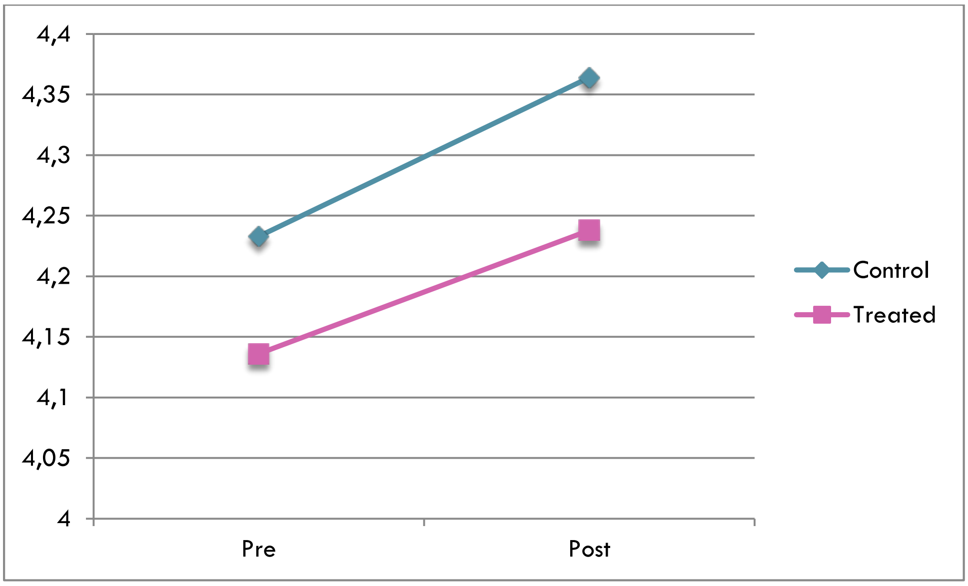


b) Satisfaction with adapting working hours to private life by gender



Finally, figure 10 shows that the commitment to the company increases during the policy period for both groups of workers, probably because both groups appreciate the involvement of the company with the new work practice.

Figure 10. Commitment



2. A lab-experiment on stress and productivity

What is driving the positive link between flexibility of work and productivity? How is this related to an increase in wellbeing? A possible channel of connection is the reduction of stress for workers in the presence of smart-working. Being able to organise their time more autonomously and to balance family and working life better thanks to the flexibility of time and space at work, workers are expected to be less stressed. Less stress may translate into greater productivity. This effect may differ by gender.

To test the validity of the link between stress, flexibility of time and productivity, and the gender differences, as part of the E.L.E.N.A. project, a lab experiment was performed which involved 311 university students. The trial subjects underwent three logical-analytical tests: two performed without time limits and one with a defined time limit. In this second case, a watch also reminds the subjects about the remaining time, thus generating anxiety.

The context we choose to evaluate the lik between stress, time constraints and productivity refers to a large literature which studies the performance of students in math tests – a task which is easy to reproduce in lab and which has been show to appropriately reproduce the behaviour of individuals subject to specific tasks in different contexts. This context is particularly interesting for our scope because the literature on students' math tests has investigated the existence of gender gaps (see Ellison and Swanson, 2010, among many others). Despite female students having an overall better performance at school than male students (Goldin, Katz, and Kuziemko, 2006), males have higher scores in standardized math tests than females. A growing literature has studied whether this gender gap is due to nature – that is, males and females having different brain structures, which lead to different math ability – or nurture – that is, differences in how male and female students are educated in, or primed toward, math at school and at home. A recent contribution (Niederle and Vesterlund, 2010) underlines the role of competition in explaining this gender difference. Males, unlike females, have been shown (Gneezy, Niederle, and Rustichini, 2003) to improve their performance in a more competitive environment. Since math tests are often used as entry test in college or else, and thus involve a high degree of competition, males exploit their competitive hedge and do better than females.

Yet math tests are also done under stringent time constraints. This aspect has been less investigated. We contribute in understanding the role of time constraints in the link between stress and productivity. The time constraint may lead in general to a lower performance, and, in particular to higher gender gaps, if males become less anxious, and thus have a better response than female, under time pressure.
To assess the role of time pressure on the performance of male and female students in math tests, we design an experiment in which each student performs a math test under no time pressure and under time pressure (either moderate or high). Our first research goal is to establish whether time pressure is associated with a decrease in productivity and whether this

differs by gender. Our second goal is to understand which mechanisms may help male and female students to cope with anxiety and time pressure. To this respect, we measure the students' working memory and their level of focus (or effort) in performing a simple task, which is our reference "benchmark".

We collect four types of scores: (1) Focus or effort (2) Working Memory, (3) Math questions solved under no time pressure and (4) Math questions solved under time pressure. A random extraction at the end of the experiment defines on which of these parts students' performance will be paid.

In the Effort task, students are given either 12 strings of 50 letters and then asked to count how many times a specific letter appears in each string. This "effort task" scores from 0 to 60 points. In the Working Memory task, after each string exercise is done, students are asked to memorize a random number different from the one they just counted. Once they finish counting all the strings, they need to write in a row the 12 numbers they had to memorize. This "memory task" varies from 0 to 120 points.

Example:

Task 1: Count the number of "F" that appears in the following string and enter it in the box below.

LHGANITBRLINIOOQPDHEGPOEDSGSBMQRADLIMDHCAFNSQSOKIR

Task 2: Memorize number 2, Task 1 and 2 repeated for 10 or 12 times.

Final Task: Type in the boxes below the numbers that appeared on the screen after each string, starting from the first in order of appearance.

Both untimed and timed Math tasks consist of 10 exercises SAT-like. In the "untimed" task, students need to solve the aforementioned exercises under their own pace, with no time pressure. This score varies from 0 to 200 points, i.e. 20 points per question. In the "timed" task, students must solve 10 exercises under certain pressure, which varies between 12, 10 and 7.5 minutes. In the first and second cases, students had a timer on the top of their screens showing how much time they had left to solve the whole task. In the latter scenario, participants had a timer of 45 seconds for each question and once time was up, the survey would auto-advance to the next question, regardless of having picked an answer or not. This score too varies from 0 to 200.

Example 1:

If $y = x^3 + 2x + 5$ *and* $z = x^2 + 7x + 1$, *what is* $2y + z$ *in terms of* x*?*

A. $3x^3 + 11x + 11$

B. $2x^3 + x^2 + 9x + 6$

C. $2x^3 + x^2 + 11x + 11$

D. $2x^3 + 2x^2 + 18x + 12$

Example 2:

A group of 202 people went on an overnight camping trip, taking 60 tents with them. Some of the tents held 2 people each, and the rest held 4 people each. Assuming all the tents were filled to capacity and every person got to sleep in a tent, exactly how many of the tents were 2-person tents?

A. 30

B. 20

C. 19

D. 18

Table I presents the results of the experiment divided by gender. We report the average score in the different tests, the standard deviation, the minimum and the maximum value. The 25th Percentile value gives us the value below which we can find the 25% of the sample population, and similarly the 75th percentile columns reports the value below which we can find the 75% of the sample population. Over this value we have the top 25% scores for the population

Table I: Main results

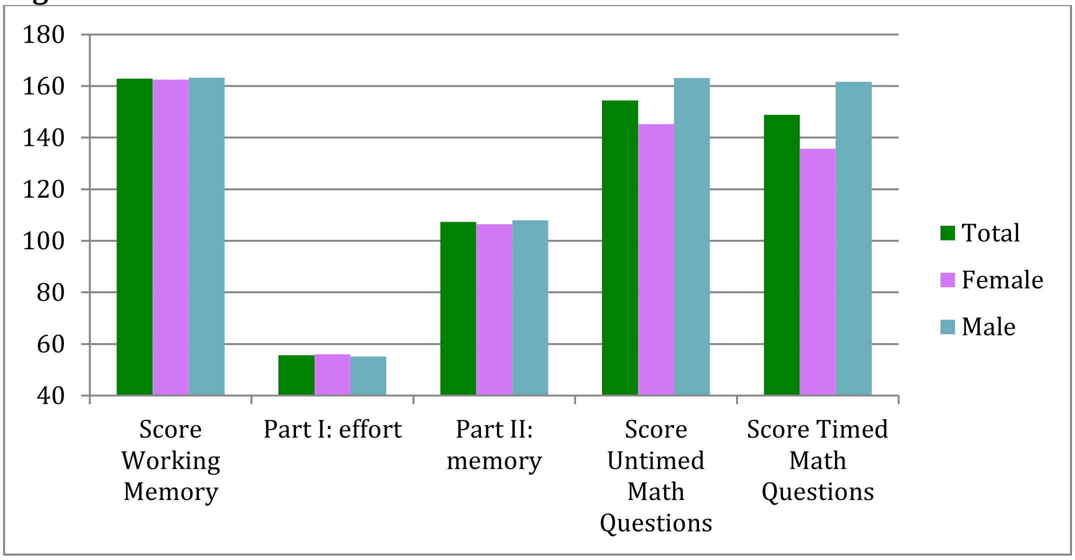
Total (N=311)	Mean	Standard Deviation	Min	25th Percentile	Median	75th Percentile	Max
Score Working Memory	162.8	21.6	70	155	170	180	180
Part I: effort	55.6	5.6	25	55	55	60	60
Part II: memory	107.3	19.9	24	100	120	120	120
Score Untimed Math Questions	154.4	33.9	60	140	160	180	200
Score Timed Math Questions	148.9	35.0	20	120	160	180	200

Female (N=152)	Mean	Standard Deviation	Min	25th Percentile	Median	75th Percentile	Max
Score Working Memory	162.5	21.5	90	152.5	170	180	180
Part I: effort	56.0	5.0	35	55	55	60	60
Part II: memory	106.5	19.8	40	100	120	120	120
Score Untimed Math Questions	145.3	34.8	60	120	150	170	200
Score Timed Math Questions	135.7	36.7	20	100	140	160	200

Male (N=159)	Mean	Standard Deviation	Min	25th Percentile	Median	75th Percentile	Max
Score Working Memory	163.2	21.7	70	155	170	180	180

Part I: effort	55.2	6.1	25	54	55	60	60
Part II: memory	108.0	20.0	24	100	120	120	120
Score Untimed Math Questions	163.1	30.6	60	140	160	180	200
Score Timed Math Questions	161.6	28.1	40	140	160	180	200

Figure 11: Mean of the score



Observing the average score, we note that male and female perform very similarly on working memory tests, our benchmark which captures the ability of subjects. However, the final score for women is higher for the effort than for the memory part, while for male is almost the same. When we consider the math questions, the entire sample performs better without time limits. However while for male the difference is very small, for women it is large: the average score in the untimed situation is 10 points higher than under the timed one.

The results of this lab experiment show that the existence of the time limit leads to a reduction in the performance of the subjects. Stress generated by inflexible work times is negatively related to productivity. This effect is very limited for male, while it is strong for female, thus suggesting that flexible work arrangements may be particularly useful for women's productivity.