PERUVIAN ECONOMIC ASSOCIATION

The Effect of Monetary Incentives on Cognitive Effort, Emotions and Test-Solving Performance

Juan F. Castro

Gustavo Yamada

Hans Contreras

Freddy Linares

Herwig Watson

Working Paper No. 131, December 2018

The Effect of Monetary Incentives on Cognitive Effort, Emotions and Test-Solving Performance[1]

Juan F. Castro /a **Gustavo Yamada** /a **Hans Contreras** /b **Freddy Linares** /c **Herwig Watson** /c

This version: November 2018

Abstract

The relation between monetary incentives, cognitive effort and task performance has been extensively studied. There is, however, scant experimental evidence about the concurrent effect of incentives on cognitive effort and emotions, and its implications for task performance. It is well documented that high-stakes tests correlate with students' anxiety and performance, but the available evidence is not causal. In this paper we estimate the effect of providing a monetary prize on the cognitive effort, emotions and efficacy exhibited by a group of university students when solving a set of four mathematics and logical reasoning questions. The prize was conditional on answering all questions correctly and was randomly assigned within a group of 126 participants. We find that the incentive produced more cognitive effort but this did not translate into increased test-solving efficacy. We provide evidence suggesting that the absence of increased efficacy despite the greater input of cognitive effort can be linked to the participants' emotional response to the prize.

JEL codes: D91, C91.
Keywords: Cognitive effort, emotions, monetary incentives, eye-tracking, facial expressions.

/a Department of Economics at Universidad del Pacifico.
/b Universidad Nacional Mayor de San Marcos.
/c Neurometrics.

[1] Authors would like to thank Hugo Fernández, Alexandra Heredia and Juan Marca for excellent research assistance. They are also grateful for comments received during the Research Seminar at Universidad del Pacifico.

1. Introduction and motivation

Cognitive effort, emotions and task-solving efficacy relate to each other in a complex manner. Cognitive effort can be understood as the degree of engagement with a demanding task. It involves difficulty, motivation, attention and cognitive control, although none of these elements is self-sufficient to define the concept (Westbrook and Braver, 2015). Emotions have several definitions but all of them have in common that they consist of three processes: physiological processes that regulate the body, subjective experience that regulates behavior, and expressive processes that regulate social coordination (Hannula, 2015).

Based on these definitions, it is reasonable to expect that greater cognitive effort will lead to increased task-solving efficacy. In fact, a conscious decision to deploy more cognitive resources will likely be motivated by the desire to improve performance in a task. The relation between emotions and task-solving efficacy is less clear as it will likely depend on the emotion analyzed and the characteristics of the task involved. Some emotions can facilitate the creative aspects of problem solving and direct attention and intuition, while others can bias cognitive processing (Fiedler, 2001; Pekrun and Stephens, 2010; Hannula, 2015). Many studies have explored this relation by manipulating the affective states of experiment participants and have found that a positive affect tends to improve performance (see, for example, Rader and Hughes, 2005; Isen et al., 1987; Spering et al., 2005; Bryan and Bryan, 1991; Thompson et al., 2001).

The relation between monetary incentives, cognitive effort and task performance has been extensively studied in the literature (Westbrook and Braver, 2015; Bonner, et al., 2002; Bonner, et al., 2000). A review of the available experimental evidence reveals mixed results as well as great heterogeneity in the type of task proposed and the method employed to measure cognitive effort (see Appendix 1). All the studies surveyed in Appendix 1 report a positive effect of the incentive on cognitive effort. The effect of the incentive on task performance, however, is less clear. For example, O'Neil et al. (2001) asked students to solve mathematics items and to self-report the level of effort they employed during the task. They found that the group that received the monetary incentive reported greater effort but did not outperform the group that did not receive the incentive. Heitz et al. (2008) asked participants to recall letters and answer a reading

comprehension question. They measured cognitive effort using pupillary dilation and found that the monetary reward increased effort as well as performance.

The relation between monetary incentives, emotions and task performance is less explored. Only two of the eight studies surveyed in Appendix 1 measured the effect of the incentive on some dimension of participants' emotion. Meloy et al. (2006) asked participants to make choices between three pairs of options and ranked performance according to participants' predecisional distortion of information[2]. They found that the monetary incentive increased effort, elevated mood and degraded performance. Konheim-Kalstein and van den Broek (2008) asked students to recall information from texts. They found that readers who received a monetary incentive spent more time reading and had a better performance but did not exhibit a different emotional response.

In the education literature, several studies have explored how the consequences that stem from the results of a test can affect students' emotions and performance. It is widely documented how high-stake tests correlate with anxiety which, in turn, correlates with performance (see, for example, von der Embse and Hasson, 2012; Keogh, et al., 2004; Mc Donald, 2001; Raffety, et al., 1997; Cassady and Johnson, 2002). None of these studies, however, have produced experimental evidence. They are based on measuring students' degree of anxiety before a high-stake test and estimating its correlation with test performance. For example, von der Embse and Hasson (2012) administrated a 23-item anxiety scale to 75 high school students one week before they took the Ohio Graduation Test. They found that the anxiety scale had a negative correlation with test performance.

In this paper, we estimate the effect of providing a monetary prize on the cognitive effort, emotions and efficacy exhibited by a group of university students when solving a set of four mathematics and logical reasoning questions. The incentive was conditional on answering all four questions correctly, and was randomly assigned within a group of 126 participants.

We measured participants' cognitive effort using their degree of visual interaction with the information required to solve each question and their pupillary dilation while solving the test. Visual interaction and pupillary dilation were measured using an eye-

[2] The predecisional distortion of information is the systematic evaluation of new information to favor the currently preferred alternative.

tracker. For each question, we defined a set of areas with relevant information to solve the problem (areas of interest). We measured the degree of visual interaction considering three different outcomes: (i) the number of times the participant had a fixation point within and area of interest (fixation count); (ii) the total time spent as fixation points within an area of interest (time spent); and (iii) the number of times a participant returned to an area of interest and left at least one fixation point (revisits).

Emotions were gauged based on participants' facial expressions. Facial expressions were linked to emotions using the FACET algorithm.[3] There are seven basic emotions identified by this algorithm: joy, anger, sadness, surprise, fear, contempt and disgust. Finally, the efficacy depended on whether the participant provided the correct answer or not to each question.

This paper seeks to contribute to the literature in two ways. First, it provides causal evidence on the concurrent effect of incentives on cognitive effort and emotions, and its implications for test-solving performance. As surveyed in Appendix 1, three studies have found that the additional cognitive effort produced by the incentive did not lead to increased task-performance (Pochon, et al., 2002; O'Neil, et al., 2005; Meloy, et al., 2006). These authors proposed that emotions could also be affected in a way that degraded performance but only one study effectively measured the effect of the incentive on participants' emotions and found significant results. Our paper provides more evidence on this particular mechanism that explains why is that monetary incentives will not necessarily lead to increased performance despite triggering a greater deployment of cognitive resources.

The second contribution is related to the specific task considered for our experiment. As already explained, the task consists in solving mathematics and logical reasoning questions of the type usually employed in standardized tests. In this regard, the paper also contributes to the education literature by providing experimental evidence on the effect of shifting the consequences of test results on students' emotional status and test performance. Previous studies have found that the imminence of a high-stakes test correlates with students' anxiety which, in turn, has a negative correlation with performance. In this study we produce an exogenous shift in the consequence of providing correct answers in a test and measure participants' emotions. This should

[3] See Littlewort (2011) for a revision of facial coding techniques.

enable us to advance the results produced thus far with causal evidence about the effect of high-stakes testing on students' emotions and test results.

Our main results can be summarized as follows. We found robust evidence that the incentive caused an increase in cognitive effort. In fact, the group that received the incentive exhibited a larger aggregate fixation count, time spent and number of revisits than the control group, in three out of the four questions. In one of the four questions we did not find statistically significant differences at the aggregate level. However, the group that received the incentive did show a significantly larger fixation count and time spent in three out of the six areas of interest of the question. Participants who received the incentive also showed a larger pupillary dilation during the experiment, with statistically significant results in two out of the four questions.

Our results also show evidence that, at best, the incentive had no effect on participants' efficacy in solving the test. Differences in test-solving efficacy were not statistically significant in any of the four questions considered. However, point estimates of the proportion of correct answers were smaller (between 8 and 10 percentage points) in three out of the four questions for the group that received the incentive.

Our results also reveal that the incentive increased the intensity of facial expressions related to surprise and fear, and reduced the presence of facial expressions related to feelings of disgust and sadness. Although statistically significant results are not present in all cases (probably due to the small sample size), this pattern is consistent in all four questions.

Overall, our results show that the incentive produced more cognitive effort but this did not translate into increased test-solving efficacy. Moreover, the evidence suggests the incentive can exert a negative effect on the probability of providing a correct answer. Our results also suggest that the absence of increased efficacy despite the greater input of cognitive effort can be linked to the participants' emotional response to the incentive.

The rest of the paper is organized in the following way. Section 2 explains the design of the experiment and describes the outcome variables. Section 3 presents and discusses our main results. Section 4 closes with some concluding remarks.

2. Experimental design

2.1 Participants and procedure

The experiment was carried out during the third quarter of 2017 at Universidad del Pacifico in Lima, Peru. We selected a random sample of 300 Economics students among first year and senior students, and they were invited to participate in the experiment. Out of these 300 potential participants, 126 showed up to perform the task.

The task consisted in solving four multiple-choice mathematics and logical reasoning questions with a time limit of 1.5 minutes per question. In Appendix 2 we present the four questions in the same format as they were presented to the participants. Participants were called to solve the task one by one. Before starting the test, all participants were informed about the nature of the task and informed consent was collected. In addition, participants were randomly assigned to the treated or control group. Those assigned to the treated group were informed that they could receive a prize of S/.20 (around US$ 6.00) if all four questions were answered correctly. Out of the 126 participants, 70 were assigned to treatment and 56 to the control group.

The experiment took place in a specially conditioned space, with sufficient light and noise free. Test questions were presented in a computer screen and participants' answers were recorded by the computer. The equipment was calibrated before the experiment for each individual in order to assure the best quality of eye-tracking and facial expression data.

Random assignment of participants to the treatment or control group should ensure comparability between these groups. To corroborate this, in Table 1 we present the differences between these two groups across nine student characteristics related to their academic performance, socioeconomic status, age and sex. No statistically significant differences were found in any of these characteristics.

Table 1
Balance in student characteristics

VARIABLES	(1) Cumulative GPA	(2) Number of attended semesters	(3) Age	(4) Older cohort (Older=1)	(5) Sex (Male=1)	(6) Tuition scale (First=1)	(7) Tuition scale (Second=1)	(8) Tuition scale (Third=1)	(9) Tuition scale (Fourth=1)
Incentive (Yes=1)	-0.0441	-0.764	-0.246	-0.0821	0.0643	0.0464	0.0107	-0.0571	0.0179
	(0.248)	(0.875)	(0.471)	(0.0886)	(0.0894)	(0.0818)	(0.0672)	(0.0791)	(0.0808)
Constant	13.61***	7.893***	21.16***	0.625***	0.536***	0.268***	0.161***	0.286***	0.268***
	(0.192)	(0.643)	(0.342)	(0.0652)	(0.0672)	(0.0597)	(0.0495)	(0.0609)	(0.0597)
Observations	126	126	126	126	126	126	126	126	126
R-squared	0.000	0.006	0.003	0.007	0.004	0.003	0.000	0.004	0.000

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
Results correspond to an OLS regression of the student characteristic on an intercept and an indicator that has the value of 1 if the participant received the incentive and the value of 0 otherwise.

2.2 Outcome variables

Eyetracking and facial analysis has many benefits. These include objective results about where and what each individual sees and their facial expressions in a particular time frame. These results are useful for uncovering information about variables that are not easy to quantify such as individuals' cognitive load and emotional arousal.

In this experiment we took advantage of the eyetracking and facial recognition technology to gauge participants' cognitive effort and emotions. Cognitive effort is measured on the basis of participants' degree of visual interaction with the stimulus provided and their degree of pupillary dilation while solving the test. To account for the degree of visual interaction, for each question we defined a set of areas containing information required to solve the task. We labeled these as "areas of interest" (AOI). In Appendix 2 we present the areas of interest related to each question.

Three different outcomes were produced on the basis of these areas. First, the *fixation count*, defined as the number of times a participant had a fixation point within an area of interest. Each fixation usually lasts between 100 and 300 milliseconds. Second, the *time spent*, defined as the total time spent as fixation points within an area of interest. Third, the number of *revisits*, defined as the number of times a participant returns to an area of interest and leaves, at least, one fixation point.

It is well documented that pupillary dilation is related to changes in cognitive load (see, for example, Graur and Siegle, 2013; van Rijn et al., 2012; Zylberberg, et al., 2012; Steidtmann, et al., 2010). For this study, we employed the Mean Pupil Diameter Change (MPDC) methodology. The eye-tracker enables us to measure participants' pupillary diameter (in millimeters) throughout the experiment. For each question, these measures were normalized with respect to the sample mean. In addition, for each question we also calculated an average pupil size for every participant.

Participants' facial expressions throughout the experiment were filmed and linked to seven possible emotions using the FACET algorithm. The seven emotions identified by this methodology are: joy, anger, sadness, surprise, fear, contempt and disgust. For each participant and each question, we calculated the proportion of time he/she displayed every emotion with respect to the total time spent in the question.

Finally, for every question we identified whether each participant provided a correct answer, an incorrect answer, or left the question unanswered.

3. Results and discussion

In this section we present and discuss the main results of the experiment. Tables 2 to 5 present the results for the degree of visual interaction with each of the four questions. The first row in each table presents the aggregate results for all the areas of interest of the question. The rows that follow present the results for each area of interest separately. Outcomes correspond to the average fixation count, time spent and number of revisits for the group that received the incentive and the group that did not receive the incentive. For each outcome, we also report the difference between these groups, its standard error and degree of significance.

Inspection of Tables 2 to 5 reveals that the incentive produced a positive effect on the degree of visual interaction with the areas of the stimulus containing relevant information to solve the task. If we aggregate across areas of interest, the group that received the incentive had a larger fixation count, time spent and number of revisits with the areas of interest in three out of the four questions of the test. There are no significant differences at the aggregate level for question 3. However, we do observe a positive and significant difference in favor of the group that received the incentive in terms of fixation count and time spent in three out of the six areas of interest defined for this question.

In Appendix 3 we present the effects of the incentive on the degree of visual interaction estimated using pre-treatment controls. The covariates included in these regressions are the same as those reported in Table 1. As expected because of the absence of significant differences between the control and treatment group, all our results are robust to the inclusion of these controls.

Table 2
Question 1: Degree of visual interaction with the stimulus

	Fixation Count (No.)			Time Spent (ms.)			Revisits (No.)		
	Incentive	No Incentive	Difference	Incentive	No Incentive	Difference	Incentive	No Incentive	Difference
AOI's Question 1	**56.04**	**42.25**	**13.79*** (4.564)**	**10,551.68**	**7,667.68**	**2,884*** (907.4)**	**22.94**	**19.53**	**3.409** (1.623)**
AOI 1	3.26	2.84	0.423 (0.530)	694.62	502.09	192.5* (113.3)	2.20	1.15	1.055*** (0.329)
AOI 2	0.56	0.84	-0.282 (0.281)	84.51	141.77	-57.25 (47.92)	0.13	0.27	-0.139 (0.139)
AOI 3	37.71	28.19	9.520*** (3.274)	6,966.00	4,895.00	2,071*** (627.4)	13.33	12.44	0.884 (1.133)
AOI 4	10.13	5.82	4.305*** (1.043)	1,715.63	1,192.47	523.2** (207.6)	3.84	2.87	0.970** (0.430)
AOI 5	5.46	4.56	0.899 (0.619)	1,190.26	935.99	254.3* (143.0)	3.44	2.80	0.639 (0.427)

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
For each area of interest (AOI), results correspond to an OLS regression of the outcome on an intercept and an indicator that has the value of 1 if the participant received the incentive and the value of 0 otherwise.

Table 3
Question 2: Degree of visual interaction with the stimulus

	Fixation Count (No.)			Time Spent (ms.)			Revisits (No.)		
	Incentive	No Incentive	Difference	Incentive	No Incentive	Difference	Incentive	No Incentive	Difference
AOI's Question 2	**106.99**	**71.72**	**35.27*** (8.279)**	**25,046.00**	**14,224.00**	**10,822*** (2,141)**	**40.14**	**30.56**	**9.589*** (3.048)**
AOI 1	3.63	1.98	1.646** (0.621)	459.66	266.20	193.5** (84.54)	2.03	1.46	0.565 (0.391)
AOI 2	4.55	4.56	-0.00895 (0.891)	756.87	696.08	60.79 (145.4)	3.70	2.38	1.323** (0.602)
AOI 3	85.81	55.43	30.39*** (7.350)	21,488.81	11,674.41	9,814*** (2,019)	26.90	21.50	5.397** (2.384)
AOI 4	5.73	3.98	1.746** (0.723)	942.46	668.67	273.8** (125.1)	2.27	1.54	0.736** (0.336)
AOI 5	7.27	5.77	1.502* (0.886)	1,425.35	965.98	459.4** (175.5)	5.25	3.68	1.567** (0.605)

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
For each area of interest (AOI), results correspond to an OLS regression of the outcome on an intercept and an indicator that has the value of 1 if the participant received the incentive and the value of 0 otherwise.

Table 4
Question 3: Degree of visual interaction with the stimulus

	Fixation Count (No.)			Time Spent (ms.)			Revisits (No.)		
	Incentive	No Incentive	Difference	Incentive	No Incentive	Difference	Incentive	No Incentive	Difference
AOI's Question 3	**84.94**	**78.91**	**6.031 (7.648)**	**16,836.61**	**15,442.08**	**1,395 (1,709)**	**32.68**	**33.37**	**-0.688 (2.557)**
AOI 1	7.19	5.19	2.002** (0.996)	1,270.87	888.75	382.1** (171.7)	3.93	3.58	0.345 (0.605)
AOI 2	2.05	2.69	-0.638 (0.582)	406.18	474.72	-68.54 (109.8)	1.45	1.61	-0.157 (0.395)
AOI 3	3.21	4.04	-0.822 (0.766)	549.35	635.53	-86.18 (130.8)	1.58	2.13	-0.550 (0.408)
AOI 4	61.51	58.87	2.640 (6.100)	12,442.53	11,829.09	613.4 (1,427)	20.76	21.23	-0.475 (1.817)
AOI 5	3.51	2.66	0.855* (0.455)	626.34	447.18	179.2** (81.74)	1.14	1.41	-0.268 (0.271)
AOI 6	7.46	5.46	1.994** (0.793)	1,541.34	1,166.82	374.5** (182.8)	3.81	3.40	0.416 (0.438)

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
For each area of interest (AOI), results correspond to an OLS regression of the outcome on an intercept and an indicator that has the value of 1 if the participant received the incentive and the value of 0 otherwise.

Table 5
Question 4: Degree of visual interaction with the stimulus

	Fixation Count (No.)			Time Spent (ms.)			Revisits (No.)		
	Incentive	No Incentive	Difference	Incentive	No Incentive	Difference	Incentive	No Incentive	Difference
AOI's Question 4	**111.10**	**67.58**	**43.53*** (9.868)**	**25,486.65**	**13,364.76**	**12,122*** (2,366)**	**61.20**	**41.36**	**19.84*** (5.980)**
AOI 1	1.96	1.29	0.662** (0.267)	303.06	236.81	66.25 (44.99)	0.87	0.60	0.270 (0.226)
AOI 2	6.46	4.55	1.915** (0.839)	1,237.56	889.27	348.3** (160.8)	3.97	3.55	0.426 (0.624)
AOI 3	25.94	17.51	8.429*** (2.866)	6,262.52	3,423.38	2,839*** (714.2)	14.19	10.10	4.095** (1.693)
AOI 4	35.64	20.36	15.28*** (3.683)	8,100.31	4,414.59	3,686*** (874.1)	19.64	12.40	7.247*** (2.175)
AOI 5	23.65	12.37	11.29*** (2.457)	5,598.46	2,160.18	3,438*** (567.3)	13.23	8.07	5.158*** (1.471)
AOI 6	17.45	11.50	5.949*** (2.010)	3,984.74	2,240.53	1 744*** (475.4)	9.29	6.65	2.642** (1.072)

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
For each area of interest (AOI), results correspond to an OLS regression of the outcome on an intercept and an indicator that has the value of 1 if the participant received the incentive and the value of 0 otherwise.

Figure 1 and Table 6 present the results obtained for pupillary dilation. Figure 1 shows the evolution of the average pupil size of participants in the control and treatment group while solving each question. We can observe that the treatment group exhibits a larger pupillary dilation during most of the time invested in solving questions 1 and 4. This is confirmed in Table 6 where the group that received the incentive presents a positive and statistically significant difference with respect to the control group in terms of its average pupillary dilation while solving questions 1 and 4. The size of the effects reported in Table 6 is robust to the inclusion of pre-treatment controls as shown in Table 3.5 in Appendix 3.

Table 6
Mean Pupil Diameter Change (in mm)

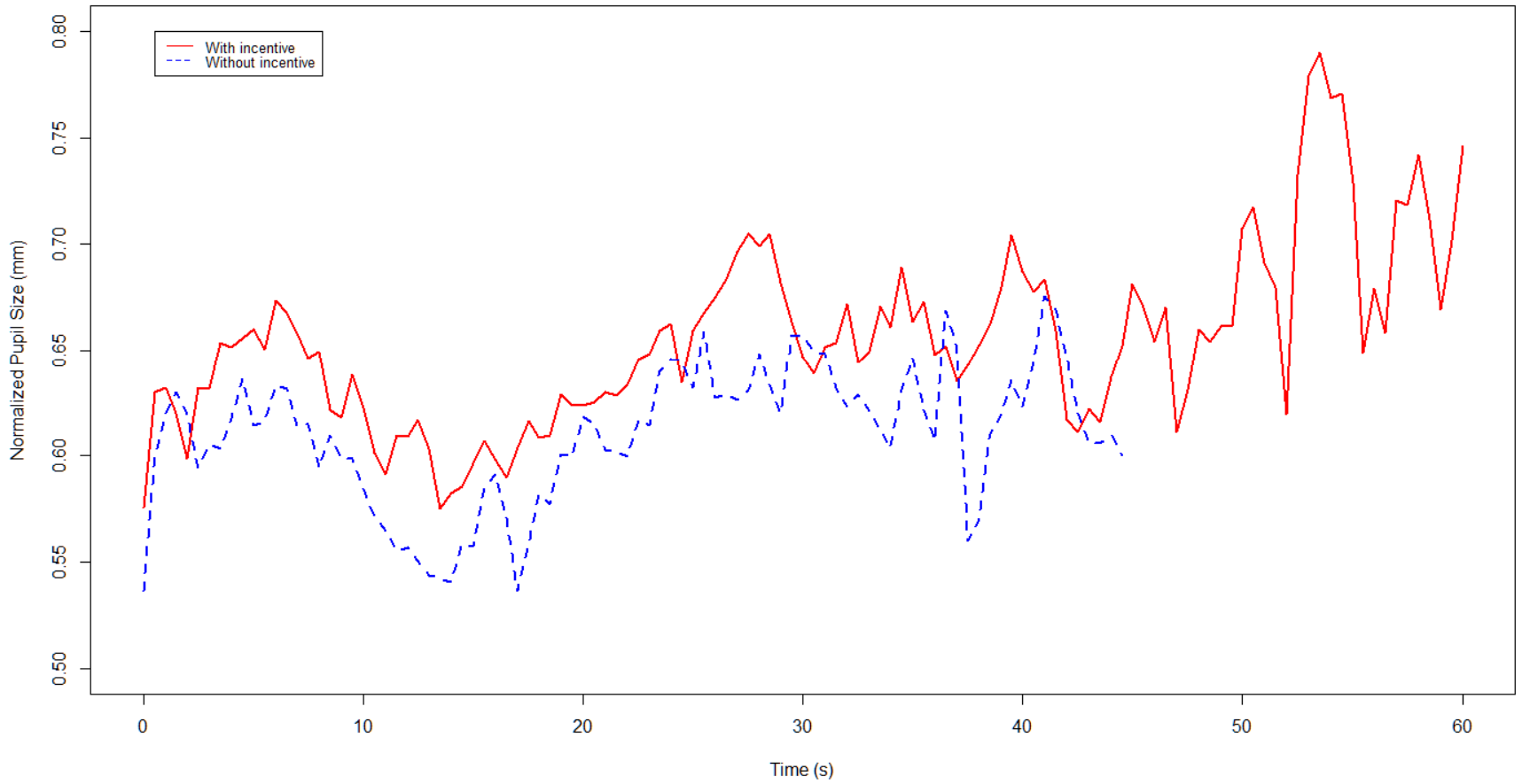
	Incentive	No incentive	Difference
Question 1	0.6272	0.5930	0.0342** (0.0169)
Question 2	0.6420	0.6220	0.0200 (0.0210)
Question 3	0.6141	0.6200	-0.00595 (0.0195)
Question 4	0.6707	0.6350	0.0357* (0.0211)

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
For each question, results correspond to an OLS regression of the outcome on an intercept and an indicator that has the value of 1 if the participant received the incentive and the value of 0 otherwise.

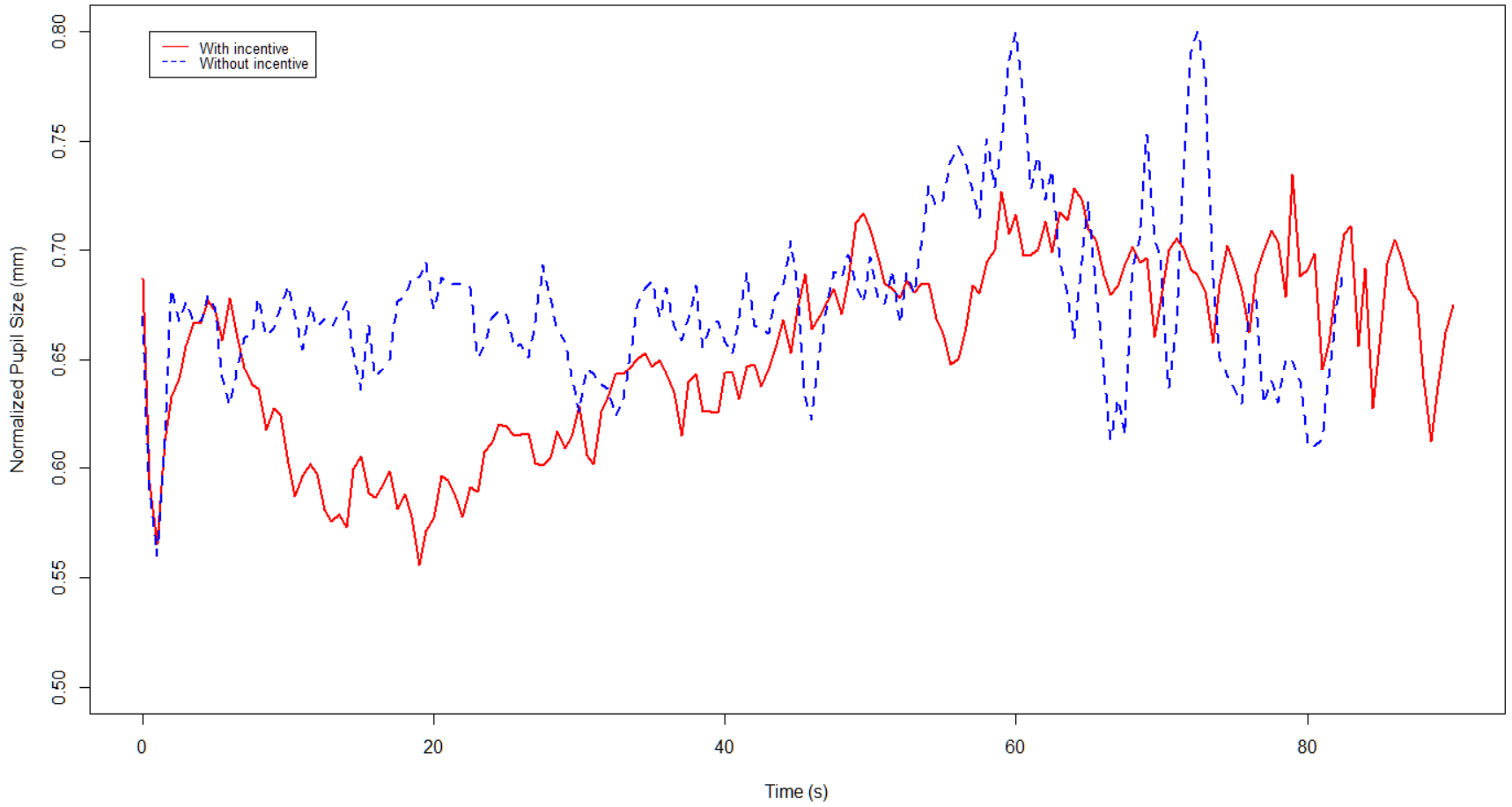
Table 7 summarizes the results for test-solving performance. We could not find any statistically significant difference between the treatment and control group in terms of the proportion of correct answers. If we consider the point estimates, however, the group that received the incentive exhibits a proportion of correct answers between 7 and 10 percentage points smaller in questions 2, 3 and 4. In addition, the treatment group presents a greater likelihood of failing to provide an answer in question 2. This is the only statistically significant result with a difference of almost 14 percentage points. As shown in Table 3.6 in Appendix 3, all these results are robust to the inclusion of pre-treatment controls.

Figure 1
Normalized pupil size while solving each question

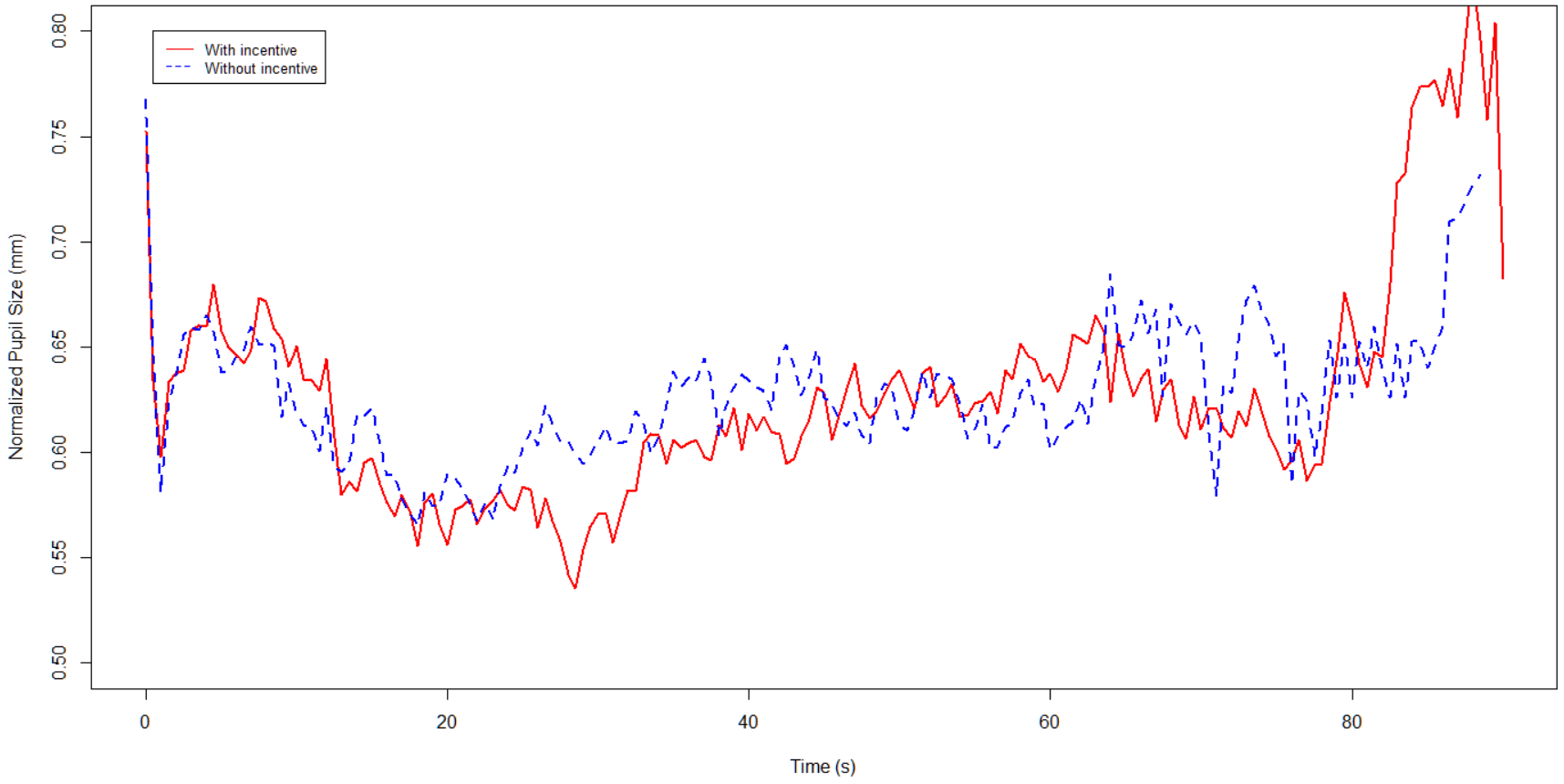
Question 1



Question 2


With incentive
Without incentive
Normalized Pupil Size (mm)
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0
20
40
60
80
Time (s)

Question 3


With incentive
Without incentive
Normalized Pupil Size (mm)
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0
20
40
60
80
Time (s)

Question 4

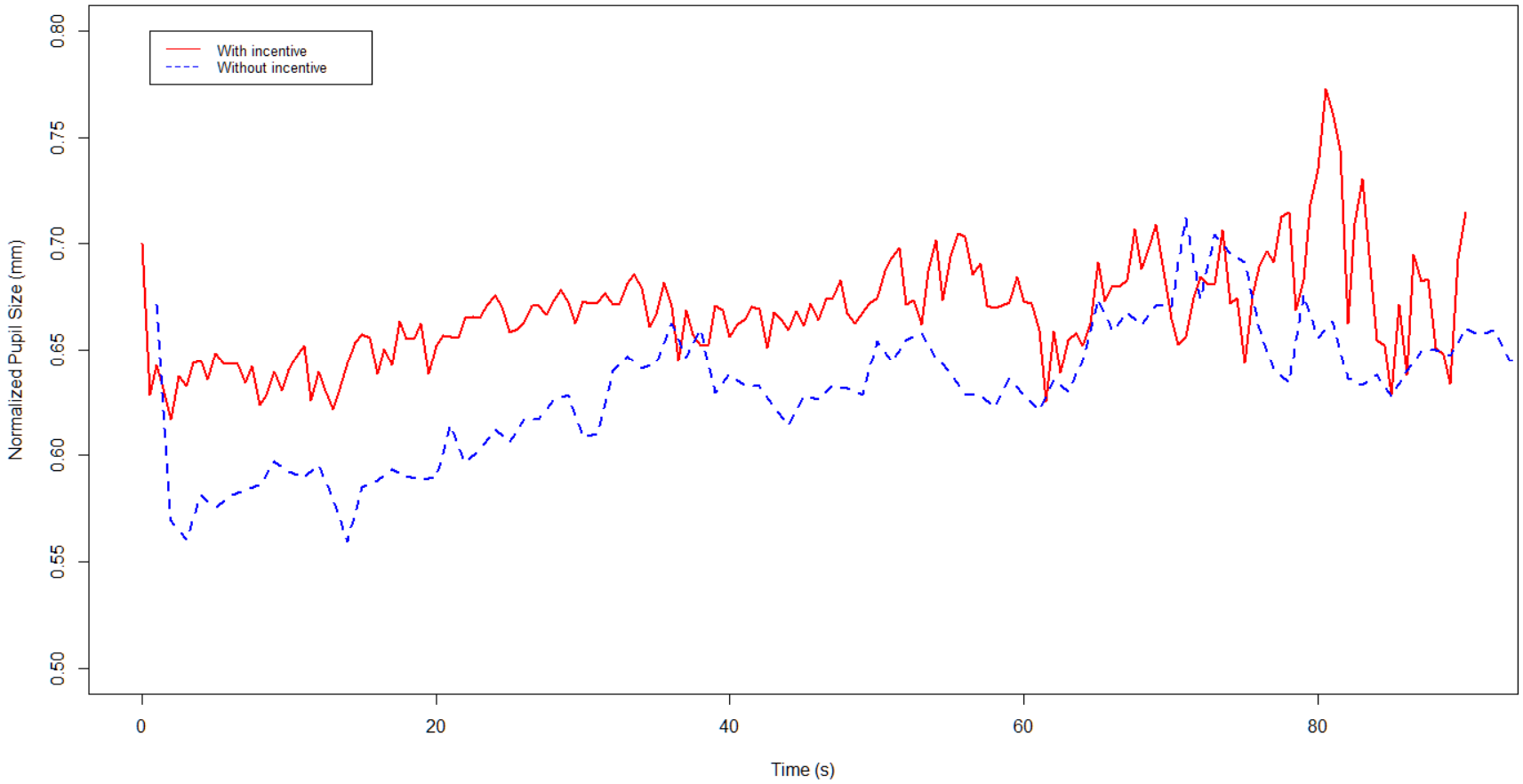

With incentive
Without incentive
Normalized Pupil Size (mm)
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0
20
40
60
80
Time (s)

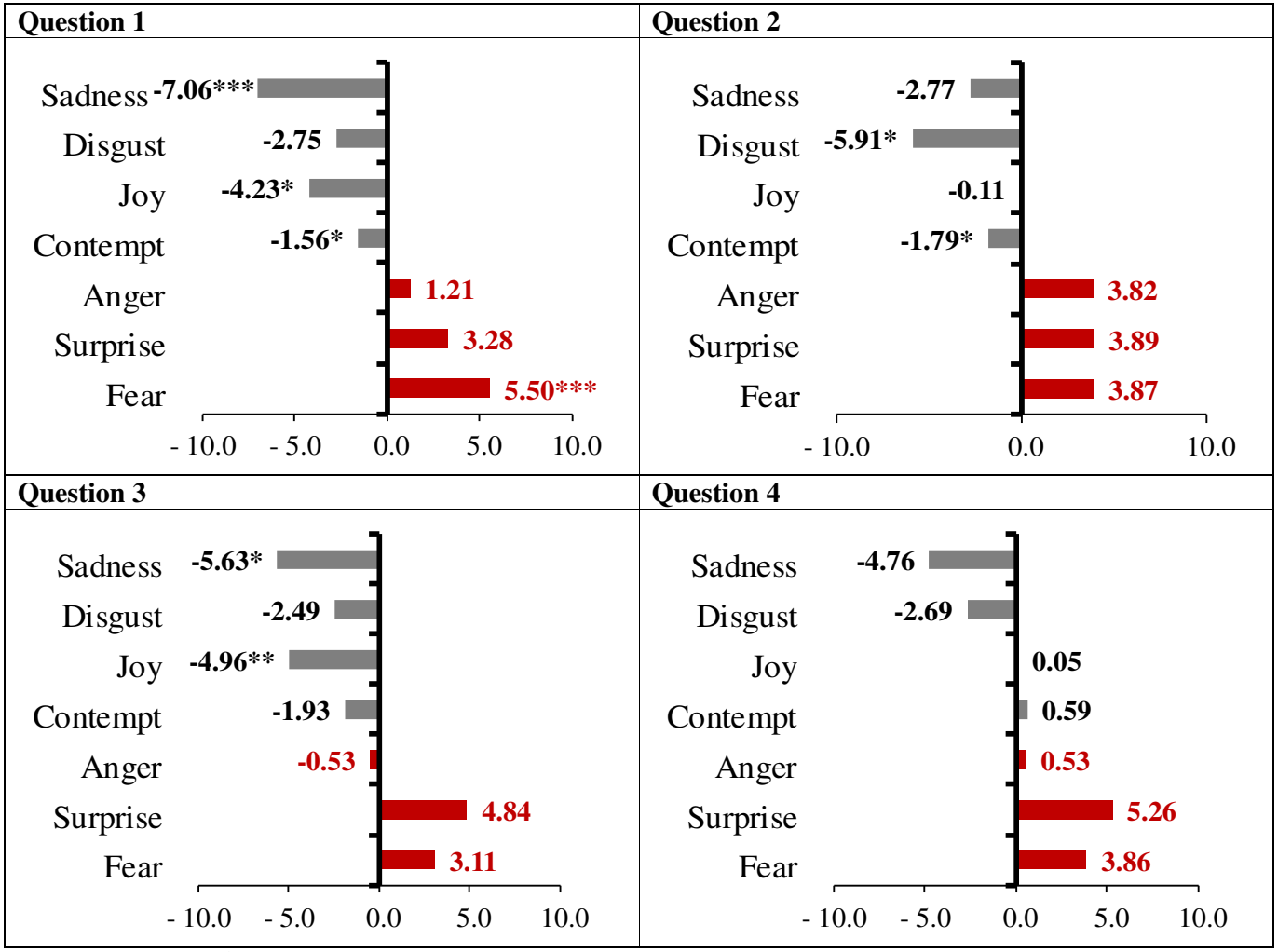
Table 7
Test-solving performance: proportion of correct, incorrect and no answers

	Outcome (%)	Incentive	No incentive	Difference
Question 1	Correct answer	0.97	0.95	0.025 (0.0364)
	Incorrect answer	0.03	0.05	-0.025 (0.0364)
	Did not answer	0.00	0.00	0.00
Question 2	Correct answer	0.27	0.38	-0.104 (0.0844)
	Incorrect answer	0.57	0.61	-0.0357 (0.0888)
	Did not answer	0.16	0.02	0.139*** (0.0473)
Question 3	Correct answer	0.44	0.52	-0.0750 (0.0901)
	Incorrect answer	0.50	0.46	0.0357 (0.0902)
	Did not answer	0.06	0.02	0.0393 (0.0331)
Question 4	Correct answer	0.50	0.61	-0.107 (0.0892)
	Incorrect answer	0.40	0.32	0.0786 (0.0863)
	Did not answer	0.10	0.07	0.0286 (0.0500)

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
For each question, results correspond to an OLS regression of the outcome (correct, incorrect, did not answer) on an intercept and an indicator that has the value of 1 if the participant received the incentive and the value of 0 otherwise.

Finally, in Figure 2 we present the results obtained for the proportion of time that participants revealed each of the seven emotions gauged through their facial expressions. Although statistically significant results are not present in all cases (probably due to the small sample size), these results show that participants receiving the incentive consistently expressed more surprise and fear and less disgust and sadness than those allocated to the control group. As with the rest of outcomes, these results are robust to the inclusion of pre-treatment controls (see Table 3.7 in Appendix 3). In fact, after the inclusion of controls, results consistent across all four questions indicate that the incentive produced more anger, surprise and fear and less disgust, sadness and joy.

Figure 2
Proportion of time that participants expressed each emotion (difference between the treatment and control group)



*** p<0.01, ** p<0.05, * p<0.1.
For each question and emotion, results correspond to an OLS regression of the outcome on an intercept and an indicator that has the value of 1 if the participant received the incentive and the value of 0 otherwise.

4. Concluding remarks

In this paper we sought to estimate the effect of providing a monetary incentive on the cognitive effort, emotions and efficacy shown by a group of university students when solving a test containing four mathematics and logical reason questions. We used an eye-tracker and a facial recognition software to gauge a reliable and objective measure of the degree of visual interaction, pupillary dilation and proportion of time that each participant expressed a certain emotion.

As shown in previous studies, we found that the incentive produced an increase in cognitive effort. Participants randomly assigned to receive a monetary prize if they answered all questions correctly, exhibited a greater visual interaction with the information required to solve the task in three out of four questions and an increased pupillary dilation while solving the task in two out of four questions. We found no evidence, however, that this greater cognitive effort lead to an improvement in performance. We found no significant differences between the group that received the incentive and the group assigned to receive no incentive in terms of correct answers. Moreover, point estimates suggest that the incentive had a detrimental effect on performance. This result is not entirely new as some previous experimental studies have found no effects on performance despite the incentive triggered greater effort.

The main contribution of this study is the provision of experimental evidence about the role of emotions in the mechanism linking incentives with performance. In particular, we found that the incentive triggered an emotional response that can be related to feelings of surprise and fear. One previous study has found that a monetary incentive can increase effort, elevate participants' mood and make them more prone to assess new information as favoring a currently preferred alternative, a phenomenon known as predecisional distortion (Meloy, et al., 2006). This paper adds more evidence about the effect of incentives on emotions and on the possibility that this emotional response can offset the positive effect of effort on performance.

Two elements make this evidence especially novel and relevant. First, the task employed to measure performance required solving a set of questions similar to those typically found in a mathematics and logical reasoning test. This type of evaluation is

common in educational systems across the world. Thus, our experiment provides causal evidence on how shifting the consequences of test results can affect students' emotional status and performance.

Second, the methodology employed to gauge participants' emotional response to the incentive is quite versatile and allowed us to identify particular emotions from a wide range of alternatives. This aids in narrowing down the mechanism linking incentives, emotions and performance. In fact, one of the emotions identified as being triggered by the incentive is known to compromise efficacy. Fear directs attention towards threatening information (e.g. the thought the one might lose a prize) which functions like a resource-demanding secondary task. The overload of working memory, in turn, can be detrimental for problem solving (Ashcraft and Krause, 2007; Hannula, 2015).

An additional assessment of the role of fear and other particular emotions as mediators in the relationship between incentives and performance can be a promising avenue for future research. More evidence can be produced by randomly allocating monetary prizes of different sizes and evaluating if a larger incentive triggers a more intense emotional response. Also, within the group receiving the prize, one could induce a particular emotion to a random subsample. If those who were induced to the emotion exhibit the same degree of effort but a poorer performance than those who were not, we will have additional causal evidence about the role of the emotion as mediator.

References

Ashcraft, M. H., & Krause, J. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review, 14*(2), 243-248.

Bonner, S., & Sprinkle, G. (2002). The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society, 27*(4-5), 303-345.

Bryan, T., & Bryan, J. (1991). Positive mood and math performance. *Journal of Learning Disabilities, 24*, 490-493.

Cassady, J., & Johnson, R. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology, 27*, 270-295.

Fiedler, K. (2001). Affective states trigger processes of assimilation and accomodation. In *Theories of mood and cognition: A user's guidebook* (pp. 85-98). In L. L. Martin & G. L. Clore.

Graur, S., & Siegle, G. (2013). Pupillary Motility: Bringing Neuroscience to the Psychiatry Clinic of the Future. *Current neurology and neuroscience reports, 13*(365).

Hannula, M. S. (2015). Emotions in problem solving. In M. S. Hannula, *Mobile Linear Algebra with Sage* (pp. 269-288). S. J.

Heitz, R., Schrock, J., Payne, T., & Engle, R. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric dat. *Psychophysiology, 45*(119-129).

Hübner, R., & Schlösser, J. (2010). Monetary reward increases attentional effort in the flanker task. *Psychonomic Bulletin & Review, 17*(6), 821-826.

Isen, A., Daubman, K., & Nowicki, G. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology, 52*, 1122-1131.

Keogh, E., Bond, F., French, C., Richards, A., & Davis, R. (2004). Test anxiety, susceptibility to distraction and examination performance. *Anxiety, stress and coping, 17*, 241-252.

Kleinginna, J., & Kleinginna, A. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion, 5*(4), 345-379.

Konheim-Kalkstein, Y., & van den Broek, P. (2008). The Effect of Incentives on Cognitive Processing of Text. *Discourse Processes, 45*(2), 180-194.

Kool, W., Mc Guire, J., & Rosen, Z. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology, 139*(4), 665-682.

Li, S., Cox, A., Or, C., & Blandford, A. (2018). Effects of monetary reward and punishment on information checking behaviour: An eye-tracking study. *Applied Ergonomics, 70*, 110-117.

Littlewort, G. W. (2011). Proceedings from 2011 I.E. International Conference onAutomatic Face & Gesture Recognition and Workshops (FG 2011). *https://doi.org/10.1109/FG.2011.5771414*, pp. 298–305. IEEE.

Mc Donald, A. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology, 21*(1), 89-101.

Meloy, M., Russo, E., & Gelfand Miller, E. (2006). Monetary incentives and mood. *Journal of Marketing Research, 43*(2), 267-275.

O'Neil, H., Abedi, J., Lee, C., Miyoshi, J., & Mastergeorge, A. (2005). Monetary incentives for low-stakes tests. *Educational Assessment, 10*(3), 185-208.

Pekrun, R., & Stephens, E. (2010). Achievement emotions: A control value approach. *Social and Personality Psychology Compass, 4*(4), 238-255.

Pochon, J., Levy, R., Fossati, P., Lehericy, S., Poline, J., Pillon, B., . . . Dubois, B. (2002). The neural system that bridges reward and cognition in humans: An fMri study. *Proceedings of the National Academy of Sciences of the United States of America, 99*(8), 5669-5674.

Rader, N., & Hughes, E. (2005). The influence of affecrive state on the performance of a block design task in 6-and-7-year-old children. *Cognition and Emotion, 19*(1).

Raffety, B., Smith, R., & Ptacek, J. (1997). Facilitating and debilitating trait anxiety, situational anxiety, and coping with an anticipated stressor: A process analysis. *Journal of Personality & Social Psychology, 72*, 892-906.

Spering, M., Wagener, D., & Funke, J. (2005). The role of emotions in complex problem solving. *Cognition and Emotion, 19*(8), 1252-1261.

Steidtmann, D., Ingram, R., & Siegle, G. (2010). Pupil response to negative emotional information in individuals at-risk for depression. *Cognition and Emotion., 24*(3), 480-496.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, mood and the Mozart effect. *Psychological Science, 12*, 248-251.

van Rijn, H. D. (2012). *Pupil Dilation Co-Varies with Memory Strength of Individual Traces in a Delayed Response Paired-Associate Task.* PLoS ONE, 7(12).

von der Embse, N., & Hasson, R. (2012). Test anxiety and high-stakes test performance between school settings: Implications for educators. *Preventing School Failure: Alternative Education for Children and Youth, 56*(3), 180-187.

Westbrook, A., & Braver, T. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective & Behavioral Neuroscience, 15*(2), 395-415.

Zylberberg, A., Oliva, M., & Sigman, M. (2012). Pupil Dilation: A Fingerprint of Temporal Selection During the "Attentional Blink.". *Frontiers in Psychology, 3*(316).

Appendix 1: Experimental evidence on the effect of monetary incentives on cognitive effort, emotions and task-solving efficacy

Study	Task	Incentive	Experimental group	Cognitive effort measure	Emotions measure	Effect on cognitive effort[/1]	Effect on emotion[/1]	Effect on task performance
Li, et al., 2018	Participants had to transcribe information from a left panel to a right panel in computer.	$0.10 based on the performance of each individual trial. The participants could earn up to $13.	60 undergraduate students.	Number and duration of eye fixations.	NA	No effect	NA	Positive
Hübner, et al., 2010	Participants had to suppress inappropriate responses in a particular context (flanker task).	Participants could earn up to €5, depending on their performance.	104 undergraduate students.	Accuracy in successive trials with different time deadlines.	NA	Positive	NA	Positive
Heitz, et al., 2008	Participants had to recall letters and answer a reading comprehension question.	Incremental payment procedure.	160 participants between 18 and 35 years old.	Pupillary dilation.	NA	Positive	NA	Positive

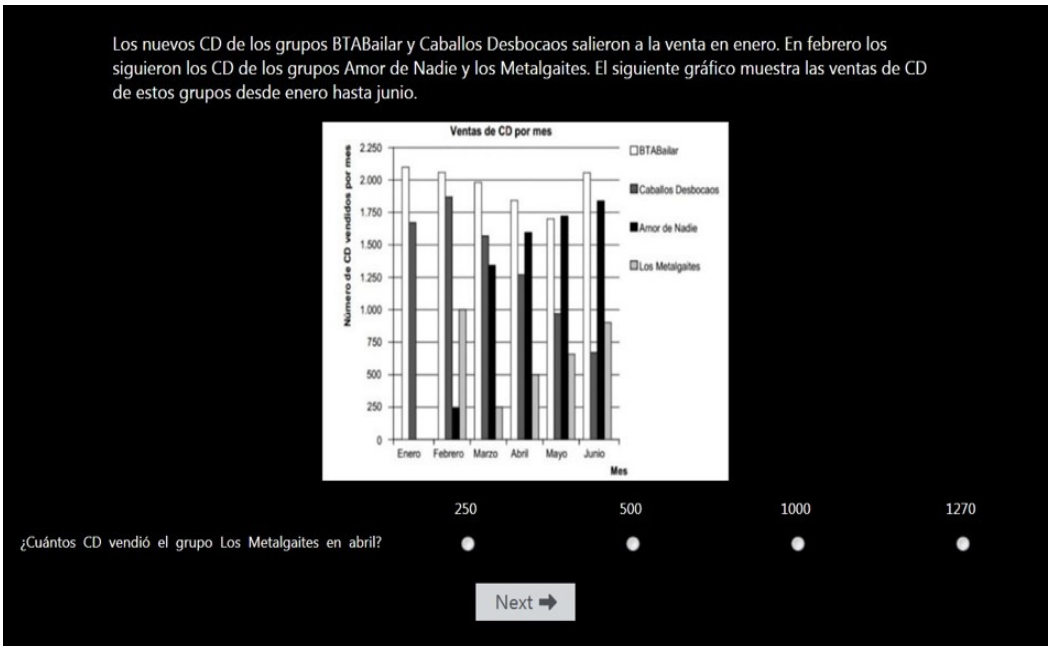
Study	Task	Incentive	Experimental group	Cognitive effort measure	Emotions measure	Effect on cognitive effort[/1]	Effect on emotion[/1]	Effect on task performance
Konheim-Kalstein and van den Broek, 2008	Participants had to recall information from six short narrative texts.	25 cents for each fact remembered. The participants could earn up to $12.	64 undergraduate psychology students.	Reading time per line of text.	PANAS questionnaire (identifies states of positive and negative affect).	Positive	No effect	Positive
Meloy, et al., 2006	Participants had to choose between a pair of fine-dining restaurants and resort destinations for a trip (Study 1) and scholarship applicants (Study 2).	Study 1: $2 for one correct choice and $12 for both choices correct. Study 2: $7 for being accurate in the task.	Study 1: 108 participants Study 2: 221 undergraduate students	Time spent on the task	Peterson and Sauber's (1983) Mood Short Form	Positive	Positive	Negative
Pochon, et al., 2002	Participants had to indicate whether a letter presented on the screen was similar or different from a letter previously presented.	Subjects were not informed about the exact amount of money. They could earn up to $285.	Six participants between 18 and 30 years old	Functional magnetic resonance imaging	NA	Positive	NA	None

Study	Task	Incentive	Experimental group	Cognitive effort measure	Emotions measure	Effect on cognitive effort[/1]	Effect on emotion[/1]	Effect on task performance
O'Neil, et al., 2001	Participants had to solve 20 mathematics items from the Third International Mathematics and Science Study (TIMSS).	$10 per correct item.	Two group studies of 12 grade students from nine schools: - Advanced Placement (AP) study: 160 students - Main study: 415 non-AP students	State Thinking Questionnaire: 6-item effort intensity scale where students indicate the level of effort they made during the test	NA	Positive	NA	None
Libby and Lipe, 1992	Participants had to recall information from a set of internal accounting controls. Then, they had to recognize which controls they have read before.	$0.10 per correct item.	121 college students	Time spent on the task	NA	Positive	NA	Positive

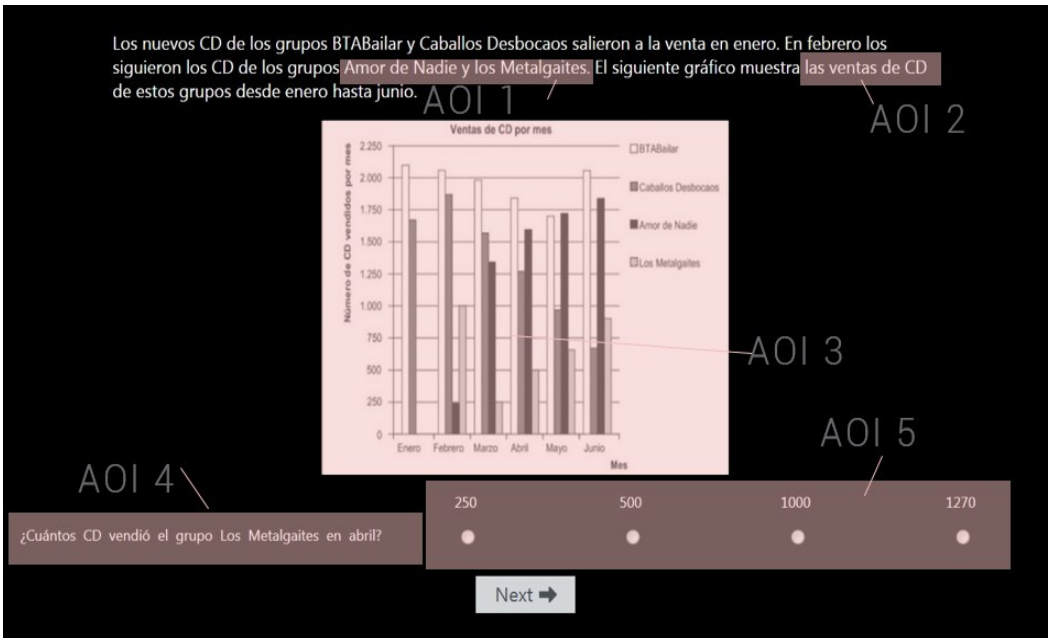
1/ No effect = no significant difference between the treated and control groups. Positive = the treated group had a better result than the control group. Negative = the control group had a better result than the treated group.

Appendix 3: Test questions and areas of interest

Question 1[4]

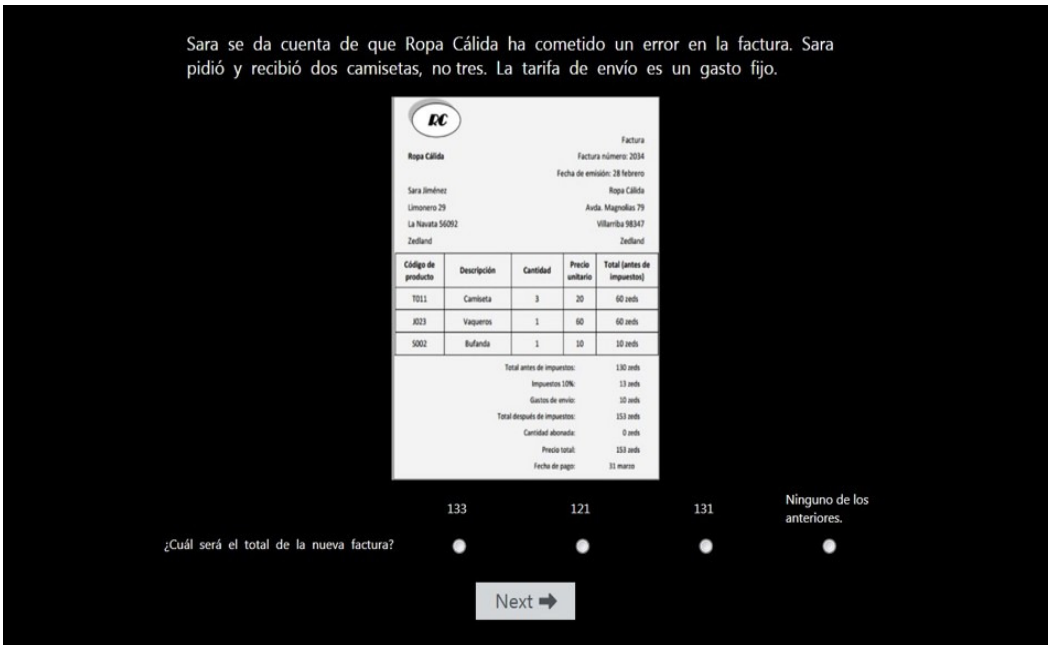


Question 1: Areas of interest

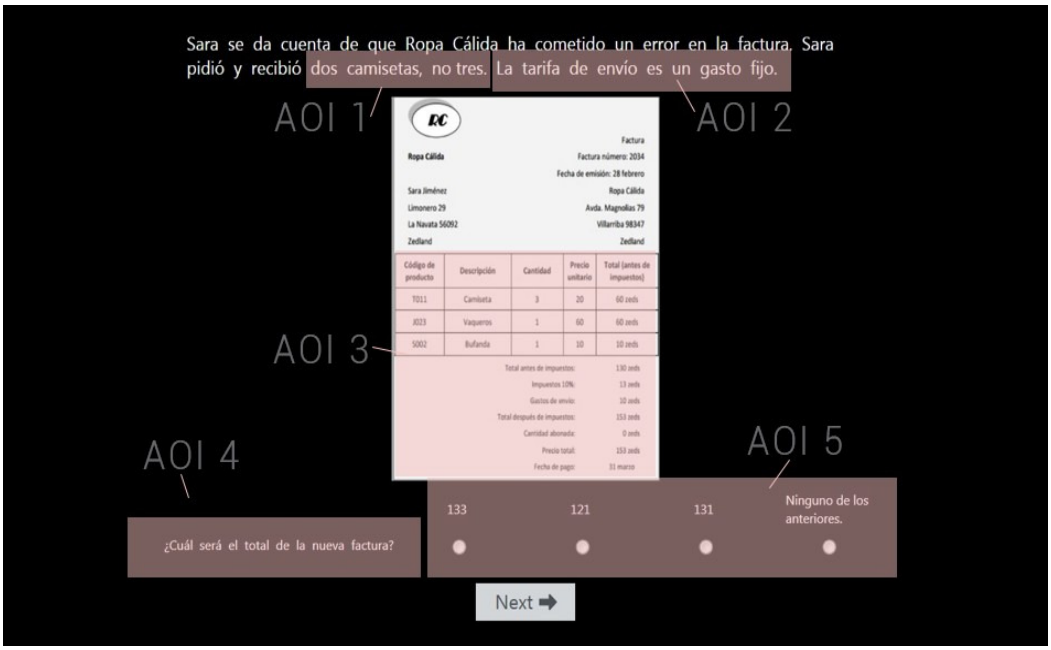


[4] Question 1: The new CDs of the groups *BTABailar* and *Caballos Desbocaos* were released for sale last January. The CDs of the groups *Amor de Nadie* and *Los Metalgaites* were released in February. The chart shows the sales of CDs of these groups from January to June. ¿How many CDs were sold by *Los Metalgaites* in April?

Question 2[5]

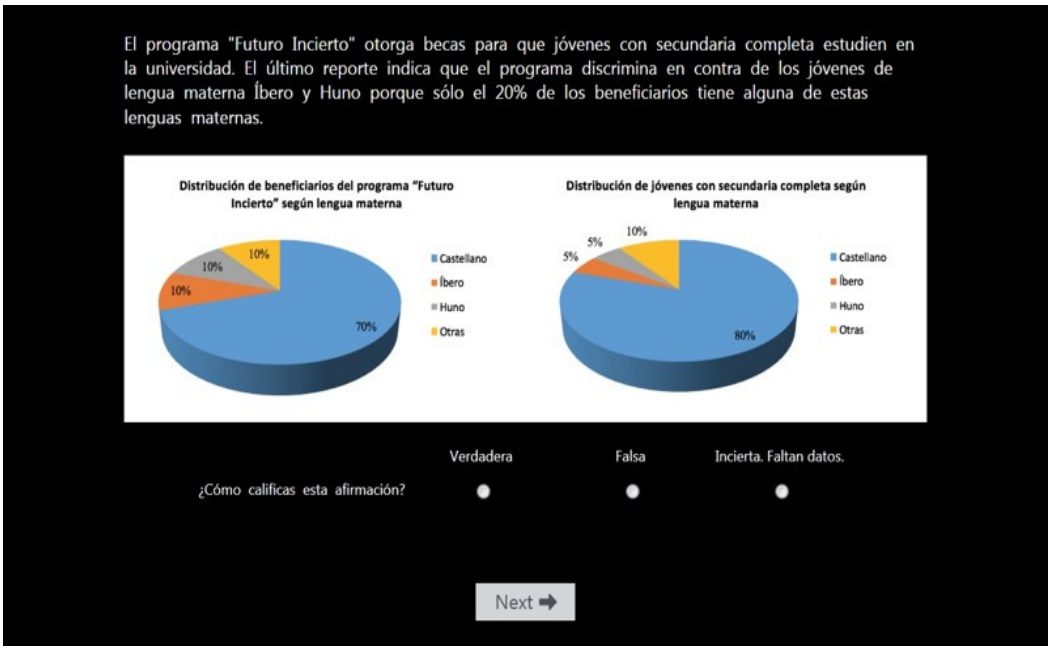


Question 2: Areas of interest

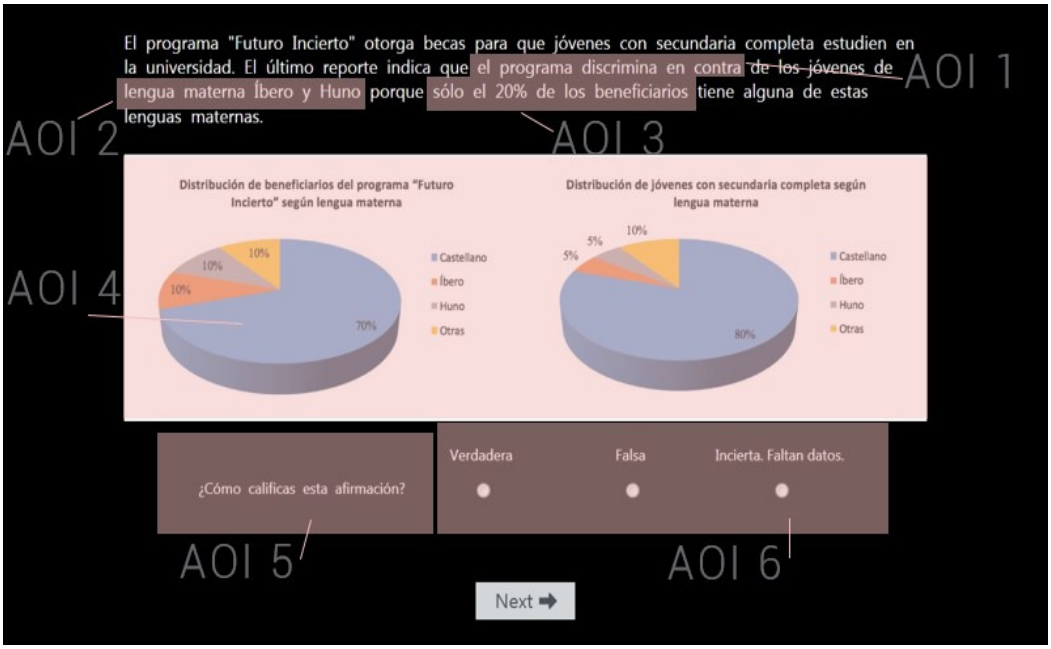


[5] Sara realizes that *Ropa Calida* made a mistake in her invoice. Sara bought and received two shirts, not three. The shipping fee is a fixed cost. ¿Which should be the total price of the invoice?

Question 3[6]

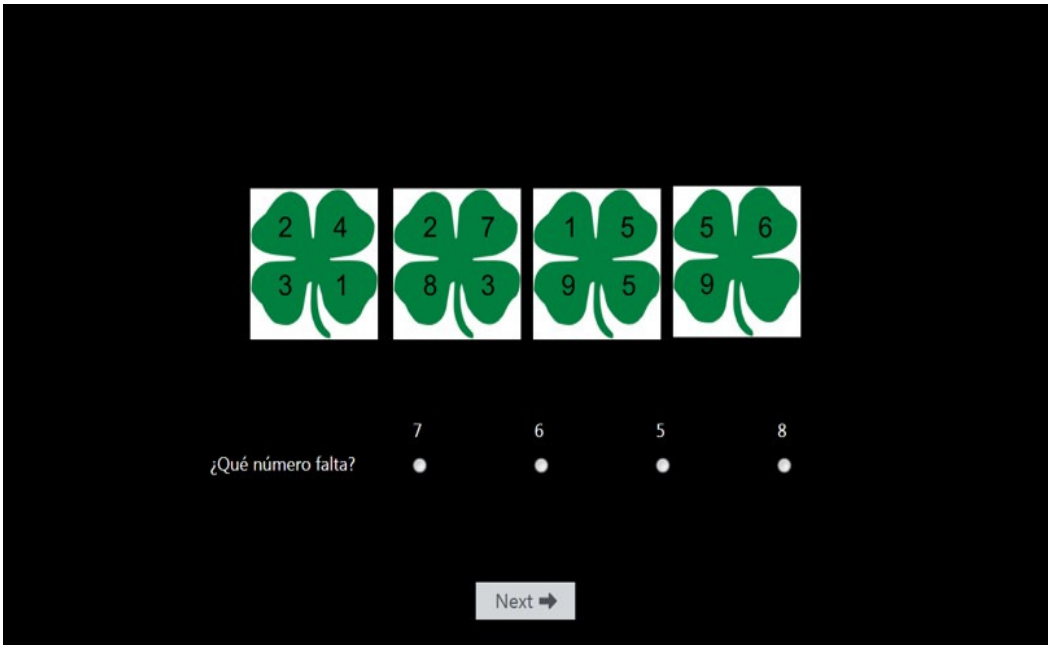


Question 3: Areas of interest

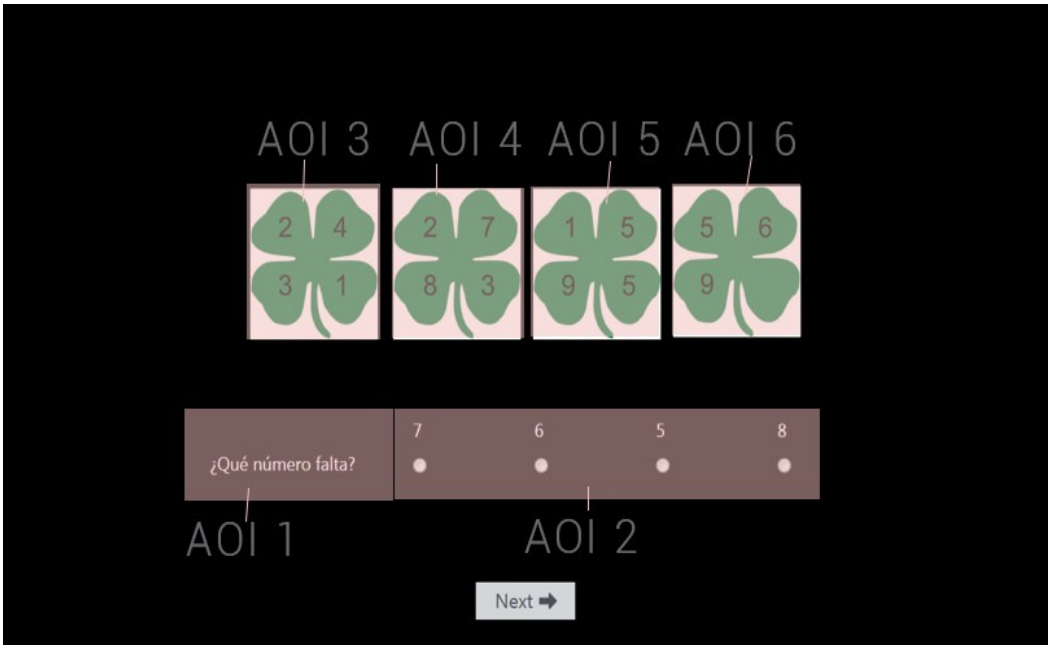


[6] The program *Futuro Incierto* awards college scholarships for students who have completed high school. The last report shows that the program discriminates against students with native languages *Íbero* and *Huno* because only 20% of the beneficiaries has one of these mother tongues. ¿Is this statement true, false or uncertain?

Question 4[7]



Question 4: Areas of interest



[7] ¿Which number is missing?

Appendix 3: Treatment effects including pretreatment controls

Table 3.1
Question 1: Degree of visual interaction with the stimulus

	Fixation Count (No.)	Time Spent (ms.)	Revisits (No.)
AOI's Question 1	**14.64*** (4.541)**	**3,016*** (905.2)**	**3.461** (1.646)**
AOI 1	0.430 (0.541)	217.8* (122.2)	1.080*** (0.340)
AOI 2	-0.213 (0.258)	-45.59 (43.44)	-0.108 (0.127)
AOI 3	10.04*** (3.249)	2,134*** (624.9)	0.788 (1.150)
AOI 4	4.417*** (1.072)	542.8*** (204.2)	1.027** (0.426)
AOI 5	0.942 (0.615)	278.4* (141.5)	0.674 (0.420)

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parenthesis. All regressions include participants' GPA, number of semesters attending the university, age, cohort, sex and tuition scale.

Table 3.2
Question 2: Degree of visual interaction with the stimulus

	Fixation Count (No.)	Time Spent (ms.)	Revisits (No.)
AOI´s Question 2	**36.04*** (8.467)**	**11,005*** (2,235)**	**9.473*** (2.985**
AOI 1	1.554** (0.615)	178.3** (86.14)	0.530 (0.402)
AOI 2	-0.0775 (0.939)	36.93 (147.9)	1.280** (0.608)
AOI 3	30.90*** (7.573)	9,987*** (2,110)	5.199** (2.308)
AOI 4	1.914** (0.739)	291.1** (133.1)	0.809** (0.342)
AOI 5	1.758** (0.880)	492.7*** (179.8)	1.655*** (0.617)

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parenthesis. All regressions include participants' GPA, number of semesters attending the university, age, cohort, sex and tuition scale.

Table 3.3
Question 3: Degree of visual interaction with the stimulus

	Fixation Count (No.)	Time Spent (ms.)	Revisits (No.)
AOI´s Question 3	**4.697** **(7.632)**	**1,098** **(1,734)**	**-1.272** **(2.514)**
AOI 1	1.890* (1.013)	367.3** (182.0)	0.174 (0.616)
AOI 2	-0.742 (0.585)	-92.55 (111.4)	-0.233 (0.394)
AOI 3	-0.874 (0.734)	-95.83 (126.6)	-0.582 (0.396)
AOI 4	1.395 (6.078)	334.1 (1,444)	-0.912 (1.768)
AOI 5	0.841* (0.460)	178.1** (82.69)	-0.238 (0.270)
AOI 6	2.186*** (0.793)	406.4** (190.4)	0.519 (0.438)

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parenthesis. All regressions include participants' GPA, number of semesters attending the university, age, cohort, sex and tuition scale.

Table 3.4
Question 4: Degree of visual interaction with the stimulus

	Fixation Count (No.)	Time Spent (ms.)	Revisits (No.)
AOI´s Question 4	**41.55***** **(9.830)**	**11,683***** **(2,388)**	**18.83***** **(6.106)**
AOI 1	0.615** (0.267)	66.17 (46.62)	0.230 (0.229)
AOI 2	1.939** (0.882)	348.8** (168.3)	0.453 (0.656
AOI 3	7.979*** (2.817)	2,738*** (724.3)	4.016** (1.756)
AOI 4	14.51*** (3.703)	3,456*** (868.2)	6.768*** (2.202)
AOI 5	10.50*** (2.430)	3,299*** (571.8)	4.776*** (1.451)
AOI 6	6.003*** (2.024)	1 774*** (495.3)	2.591** (1.084)

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parenthesis. All regressions include participants' GPA, number of semesters attending the university, age, cohort, sex and tuition scale.

Table 3.5
Normalized pupillary dilation

	Treatment effect
Question 1	0.0345** (0.0173)
Question 2	0.0127 (0.0224)
Question 3	-0.00381 (0.0192)
Question 4	0.0327 (0.0206)

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parenthesis.
All regressions include participants' GPA, number of semesters attending the university, age, cohort, sex and tuition scale.

Table 3.6
Test-solving performance

	Outcome	Treatment effect
Question 1	Correct answer (%)	0.027 (0.0382)
	Incorrect answer (%)	-0.027 (0.0382)
Question 2	Correct answer (%)	-0.103 (0.0856)
	Incorrect answer (%)	-0.0326 (0.0904)
	Not answer (%)	0.135*** (0.0505)
Question 3	Correct answer (%)	-0.0654 (0.0927)
	Incorrect answer (%)	0.0342 (0.0903)
	Not answer (%)	0.0312 (0.0279)
Question 4	Correct answer (%)	-0.0832 (0.0905)
	Incorrect answer (%)	0.0557 (0.0854)
	Not answer (%)	0.0276 (0.0519)

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parenthesis. All regressions include participants' GPA, number of semesters attending the university, age, cohort, sex and tuition scale.

Table 3.7
Proportion of time that participants expressed each emotion

	Emotion	Treatment effect
Question 1	Joy	-3.812 (2.328)
	Contempt	-1.330 (0.891)
	Disgust	-2.336 (3.135)
	Sadness	-7.126*** (1.674)
	Anger	1.116 (3.002)
	Surprise	3.887 (2.901)
	Fear	6.192*** (2.095)
Question 2	Joy	-0.0660 (3.082)
	Contempt	-1.702* (0.930)
	Disgust	-5.216 (3.470)
	Sadness	-3.172 (3.254)
	Anger	4.496 (3.533)
	Surprise	4.409 (3.258)
	Fear	4.611 (2.977)
Question 3	Joy	-4.735* (2.423)
	Contempt	-2.082* (1.228)
	Disgust	-2.189 (3.343)
	Sadness	-5.290 (3.196)
	Anger	0.0871 (3.732)
	Surprise	4.442 (3.125)
	Fear	3.792 (2.833)

	Emotion	Treatment effect
Question 4	Joy	-0.357 (3.502)
	Contempt	0.607 (1.175)
	Disgust	-2.728 (3.845)
	Sadness	-5.123 (3.697)
	Anger	1.476 (3.984)
	Surprise	5.942* (3.566)
	Fear	4.152 (3.120)

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parenthesis. All regressions include participants' GPA, number of semesters attending the university, age, cohort, sex and tuition scale.