An Experimental Study on Responsibility as a Motivating Factor

by

Emre Aktürk

A Dissertation Submitted to the
Graduate School of Social Sciences and Humanities
in Partial Fulfillment of the Requirements for
the Degree of

Master of Arts

in

Economics

June 28, 2025

An Experimental Study on Responsibility as a Motivating Factor

Koç University

Graduate School of Social Sciences and Humanities

This is to certify that I have examined this copy of a master's thesis by

Emre Aktürk

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Prof. Seda Ertaç Güler (Advisor)

_____

Assoc. Prof. Emre Ekinci

_____

Asst. Prof. Mert Gümren

Date: _____

To My Family

ABSTRACT

An Experimental Study on Responsibility as a Motivating Factor
Emre Aktürk
Master of Arts in Economics
June 28, 2025

This study investigates the motivational effects of responsibility, as a prosocial incentive to put effort, compared to only material incentives. In a field experiment on elementary school students, participants were randomly assigned to one of three incentive schemes while completing a simple cognitive task: a material incentive group earning one point per correct answer, a higher-reward material incentive group earning two points per correct answer, and a responsibility group earning one point for themselves and one for an unknown peer per correct answer. The study examines performance differences across treatments, focusing on gender and ability as factors of heterogeneity. Results show that elementary school students' performance is on average insensitive to both material and prosocial incentives. However, girls and boys' effort response to responsibility is different, with responsibility motivating boys more positively than girls. We also analyze the links between altruism and the response to treatments.

ÖZETÇE

Sorumluğunun Motivasyona Etkisi Üzerine Deneysel Bir Çalışma
Emre Aktürk
Ekonomi, Yüksek Lisans
28 Haziran 2025

Bu çalışma, sorumluluğun çaba göstermeye yönelik prososyal bir teşvik olarak motivasyonel etkilerini maddi teşviklerle karşılaştırarak incelemektedir. Bir saha deneyinde, ilkokul öğrencileri basit bir bilişsel görevi tamamlarken rastgele olarak üç farklı teşvik grubundan birine atanmıştır: doğru cevap başına bir puan kazanan maddi teşvik grubu, doğru cevap başına iki puan kazanan yüksek ödüllü maddi teşvik grubu ve doğru cevap başına hem kendisi hem de tanımadığı bir akranı için birer puan kazanan sorumluluk grubu. Çalışma, cinsiyet ve beceri farklılıklarını heterojenlik unsurları olarak ele alarak, uygulanan teşvik türlerine göre performans farklılıklarını incelemektedir. Sonuçlar, ilkokul öğrencilerinin ortalama olarak hem maddi hem de sorumluluk teşviklerine genel olarak duyarsız olduğunu göstermektedir. Ancak, kız ve erkek öğrencilerin sorumluluğa verdikleri çaba tepkisi farklılık göstermekte; sorumluluk erkek öğrencileri, kızlara kıyasla daha olumlu şekilde motive etmektedir.

ACKNOWLEDGMENTS

I am grateful to my thesis advisor, Prof. Dr. Seda Ertaç, whose expertise in experimental economics, along with her unwavering guidance and support, played a crucial role in the development of this thesis. I am also thankful to Assoc. Prof. Emre Ekinci and Asst. Prof. Mert Gümren for joining my thesis committee and for their insightful feedback.

TABLE OF CONTENTS

List of Tables viii

List of Figures ix

Chapter 1: Introduction 1

Chapter 2: Literature Review 4

Chapter 3: Experimental Design 9

Chapter 4: Data 12

Chapter 5: Results 13

Chapter 6: Conclusion 27

Bibliography 29

Appendix A: 32

LIST OF TABLES

4.1 Summary Statistics 12

5.1 OLS Regression Results on Scores 16
5.2 OLS Regression Results on Scores by Gender 18
5.3 Interaction Effects of Treatment and Gender on Scores 20
5.4 OLS Regression Results on Scores by Cognitive Ability Level 22
5.5 OLS Regression Results on Scores by Altruism Level 25

A.1 OLS Regression Results on Scores (For Understanding = 1) 32
A.2 OLS Regression Results on Scores by Gender (For Understanding = 1) 33
A.3 Interaction Effects of Treatment and Gender on Scores (For Understanding = 1) 33
A.4 OLS Regression Results on Scores by Cognitive Ability Level (For Understanding = 1) 34
A.5 OLS Regression Results on Scores with Interaction Effects of Treatment and Cognitive Ability 34
A.6 OLS Regression Results on Scores with Interaction Effects of Treatment and Cognitive Ability (Understanding = 1) 35
A.7 OLS Regression Results on Scores by Altruism Level (For Understanding = 1) 36
A.8 OLS Regression Results on Scores with Interaction Effects of Treatment and Altruism 36
A.9 OLS Regression Results on Scores with Interaction Effects of Treatment and Altruism (Understanding = 1) 37

LIST OF FIGURES

3.1 Example Question in Number Finding Test 10

5.1 Distribution of Scores - Material x1 13
5.2 Distribution of Scores - Material x2 14
5.3 Distribution of Scores - Responsibility 14
5.4 Average Score in Number Finding Test over Treatment 15
5.5 Average Score in Number Finding Test over Treatment and Gender . 18
5.6 Average Score in Number Finding Test over Treatment and Cognitive Ability 22
5.7 Average Score in Number Finding Test over Treatment and Altruism 24

A.1 Example Question in IQ (Raven's Progressive Matrices) Test 32
A.2 Number of Donations 35

Chapter 1

INTRODUCTION

Understanding what drives individuals to exert effort is a central topic in economics. Standard economic models portray individuals as rational agents who act selfishly to maximize their total earnings (Hochman and Ariely, 2015). Consequently, a large body of research has focused on how material incentives, in the form of monetary and non-monetary motivators, affect performance (Bellemare and Shearer, 2009; Jeffrey, 2009; Sittenthaler and Mohnen, 2020; Condly et al., 2003). While these studies find that material incentives have a positive effect on motivation, the relative impact of monetary and non-monetary incentives tend to differ between studies.

On the other hand, growing evidence from behavioral economics and psychology suggests that material incentives are not the only or even necessarily the most effective type of incentive. As such, non-material incentives such as praise, feedback, goal setting, performance monitoring and recognition have gained attention as alternative methods of enhancing motivation (Bareket-Bojmel and Hochman, 2014; Lee et al., 2018; Grant and Gino, 2010; Nagin et al., 2002; Stajkovic and Luthans, 2003).

Moreover, incentives can influence individuals in different ways. In this regard, a substantial body of experimental literature finds that men and women tend to differ in terms of their risk aversion, competitiveness, and social preferences (Croson and Gneezy, 2009). These differences between men and women are likely to lead to differences in the impact incentives have over their motivation. Therefore, researchers have also examined the effect of incentives separately on males and females (Sittenthaler and Mohnen, 2020; Niederle et al., 2003; Migheli, 2010). Additionally, as

children grow and develop, their reaction towards incentives might change. Thus, there is also a growing literature on the effect incentives have on children (Bettinger, 2012; Fryer, 2010; Levitt et al., 2012; Jalava et al., 2015).

Furthermore, the development of models that consider social preferences, which recognize that individuals derive utility not only from their own outcomes but also from the outcomes of others, have broadened our view towards incentive schemes. This understanding has led to increased interest towards prosocial incentives and researchers have examined how these can influence motivation (Anik et al., 2013; Gosnell et al., 2019). However, this is a relatively new area of research, and the results are not uniform. Moreover, there is also a growing literature focusing on leadership (Chakravarty et al., 2011; Ertac and Gurdal, 2012), but these studies are focusing on the effect of leadership on risk taking.

While prosocial incentives and leadership have been receiving attention, the impact of responsibility, defined as being accountable not only for one's own outcome but also for anothers, on motivation remains understudied, especially in the context of children. However, considering humans altruistic tendencies (Fehr and Fischbacher, 2003; Artinger et al., 2010; Engel, 2010; Andreoni et al., 2008), we would expect responsibility to act as a motivator.

Therefore, this thesis investigates the effect of responsibility as a prosocial incentive by a field experiment on elementary school students, who were asked to solve a simple test of finding numbers. Before solving the test, participants were randomly assigned to one of three conditions; an individual material incentive group where each correct answer earned one point for the student, a high-reward individual material incentive group where each correct answer earned two points for the student, and a responsibility group where each correct answer earned one point for the student and one for an anonymous peer who could not participate in the tests. By comparing performance across these groups, the study isolates the motivational effect of responsibility and individual material incentives, and examines how it differs by gender, ability and altruism.

The findings reveal that neither material incentives nor responsibility have a significant effect on the motivation of elementary school students. Moreover, re-

sponsibility is found to have a positive effect on the motivation of boys while negatively effecting the motivation of girls. These effects were not statistically significant when assessed individually, however, the difference between them was statistically significant.

The remainder of this paper is organized as follows: Section 2 presents related literature, Section 3 explains the experimental design, Section 4 describes the data, Section 5 presents the results and finally Section 5 provides a conclusion.

Chapter 2

LITERATURE REVIEW

As mentioned previously, there exists ample research focusing on the effect of monetary and non-monetary (assets that have monetary value but are not cash) incentives on individual performance. While there is no universal consensus on which type is more effective, most studies find that both monetary and non-monetary incentives, which together can be grouped as material incentives, tend to have a positive effect on performance. In this regard, Bellemare and Shearer (2009) conducted a field experiment in a Canadian tree planting firm and found that a surprise monetary bonus led to higher daily output. Jeffrey (2009) showed, in a laboratory setting, that both cash and equivalent-value non-cash incentives increased performance, with non-cash rewards generating greater effects. Sittenthaler and Mohnen (2020) further confirm that monetary, non-monetary, and mixed incentives all positively impact performance. However, their study found no significant difference across the effects of the three types of incentive schemes when considering aggregate results. A meta-analysis by Condly et al. (2003) supports these findings, showing that monetary and non-monetary incentives improve performance, with monetary rewards having a greater impact than non-monetary equivalents such as gifts or meals.

Beyond material incentives, behavioral economists and psychologists have also focused on how non-material motivators influence performance. In a factory setting, Bareket-Bojmel and Hochman (2014) compared monetary rewards, a family pizza voucher (equivalent in value to the monetary reward), and verbal praise and found that all incentive schemes improved performance, with no significant differences between them. Lee et al. (2018), focused on feedback as a non-material incentive and found that feedback based on relative performance was more effective than absolute feedback, and that criticism outperformed praise in motivating computer program-

mers in a laboratory experiment. Grant and Gino (2010) showed that expressions of gratitude led to an increase in prosocial behavior. In a field experiment Nagin et al. (2002) found that employees respond to a reduction in monitoring by quickly and sharply increasing the rate at which they engage in malfeasance. In a meta-analysis, Stajkovic and Luthans (2003) found that money, feedback, and social recognition each had a significant impact on performance. However, when these three were used in combination, they produced the strongest effect.

While the overall data presents mixed results regarding the impact of material and non-material incentives, it is also important to separately examine males and females because studies find men and women to be different in terms of their risk preferences, social preferences, and competitive preferences. An extensive survey covering experimental studies in this regard has been done by Croson and Gneezy (2009). Therefore, such differences between men and women are likely to lead to variations in the impact that incentives have on motivation and thus researchers have also examined the effect of incentives separately on males and females.

As previously mentioned, Sittenthaler and Mohnen (2020) found that monetary, non-monetary, and mixed incentives all have a significant positive impact on performance and there are overall no significant differences between the three types of motivators. However, when examining the results separately for men and women, the authors find that men's performance is highest in the monetary treatment, whereas women's performance is highest in the non-monetary treatment. In a laboratory experiment Niederle et al. (2003) found that as the competitiveness of the environment increases, a significant increase in performance of men occur which is not observed within women. That is, under a piece-rate scheme where participants was rewarded for every maze they solved, the authors found no significant gender differences in performance. However, under a competitive pay scheme where only the participant who solved the most mazes was rewarded, they found a significant difference in the performance of men and women and the performance of men was found to be significantly higher. In contrast, Migheli (2010) observed no gender difference in performance under competitive conditions, but found that women performed better under piece-rate schemes, and were less influenced by the incentive structures

overall.

While the papers explained so far focused on adults, there is also a growing literature on the effect that incentives have on children. The effect of different incentive schemes on children is particularly important given that countries have been implementing incentive programs to enhance the education and health of children all around the world. For example, by the Progresa program in Mexico parents and children are eligible to receive cash rewards for school attendance and regular check-ups, by the Education Maintenance Allowance (EMA) program in United Kingdom[1] students received weekly payments for school attendance and by the Coshocton's program established by the Coshocton's foundation, primary school students in the third, fourth, fifth, and sixth grade was able to receive cash payments for the successful completion of their standardized testing (Bettinger, 2012). The rationale behind such payments is that while academic achievement in early grades can greatly improve students long-run success, these long-run benefits are intangible to many children at young ages (Bettinger, 2012).

Although there have been an increasing number of studies focusing on the effect incentives have on children, there is no consensus regarding the results. Bettinger (2012) focused on measuring the effects of Coshocton's program on student achievement. The finding shows that while there is a positive and significant effect on math test scores, there are no significant effects on reading, social science, or science. Moreover, also no significant difference between boys and girls were found. On the other hand, in series of experiments conducted over 200 schools across three cities in the United States of America, Fryer (2010) finds no statistically significant effect of financial incentives on students. However, in a field experiment including over 6,000 elementary and high school students in three school districts in and around Chicago, Levitt et al. (2012) finds that both financial and non-financial incentives have a significant impact on performance. In fact, they find that children who are at a younger age tend to be insensitive to the size of the monetary rewards suggesting

[1]The EMA program continues to function in Northern Ireland, Scotland and Wales, however it is no longer functioning in England. Students in England who are between the ages of 16 and 19 can instead apply to 16 to 19 Bursary Fund to get support with their education-related costs (UK Government, 2025).

limited understanding of these returns. Moreover, younger children are also found to be more responsive to non-financial rewards. In terms of gender, the authors find that boys are more responsive to incentives than girls. In another field experiment conducted on 1,045 sixth grade students by Jalava et al. (2015), the study finds that boys only increase their performance when facing rank-based grading whereas girls respond as strongly to rank-based grading as boys, but also respond strongly to being rewarded a certificate.

On the other hand, the recent development of models that consider social preferences, which recognize that individuals derive utility not only from their own outcomes but also from the outcomes of others, has led to increased interest towards focusing on prosocial incentives as a mean to enhance motivation. In this regard, Anik et al. (2013) examined the impact of prosocial bonuses on employees, where employees were rewarded with money to spend on others. They found that prosocial bonuses in the form of donations to charity led to happier and more satisfied employees. Moreover, they also showed that prosocial bonuses in the form of expenditures on teammates led to better performance in teams. However, Gosnell et al. (2019), in a field experiment with airline captains, found that while goal-setting and performance monitoring had an effect on performance, prosocial incentives, in the form of donations to charity for reaching targets, had no effect.

Supporting the rise of social preference models, there has also been studies focusing on the altruistic tendencies of humans. In this regard, the Dictators Game has been an used as a mean to show and measure human altruism. In this game, the "dictator" is given an amount of money and decides how much of it to keep and how much of it to give to another player. Although the "dictator" has all the power, empirical evidence shows most dictators tend to give part of the money. For example, in a laboratory experiment, Artinger et al. (2010) found the mean offer in the Dictators Game to be 25%, indicating significant altruism. In fact, in a meta study covering 129 papers studying the Dictators Game, Engel (2010) finds that on average dictators give 28%. Moreover, the study finds that women tend to give significantly more than men and that the amount given increases with age. Other games have also been used to measure human preferences and altruism. In a study

Andreoni et al. (2008) examines the role of altruism over the findings of experimental studies covering not only the Dictator Game but also Prisoners Dilemma, public goods games and trust games.

There is also a literature focusing on leadership and making decisions that effect others. However, researchers have approached this topic from a risk-taking perspective rather than a motivational one. For example, in a controlled laboratory experiment, Chakravarty et al. (2011) find that individuals tend to be less risk-averse when making decisions for an anonymous stranger. On the other hand, in an experimental study, Ertac and Gurdal (2012) find that individuals tend to take less risk when deciding for a group compared to when deciding only for themselves. Their study also reveals gender differences such that compared to men, women are found to be less willing to make decisions on behalf of the group. Moreover, women who choose to lead are not significantly different from those who do not in terms of the level of risk they take for themselves or for the group. However, considering men, those who are willing to lead tend to take on more risk when deciding for the group.

Despite extensive research on the altruistic tendencies of individuals, a growing body of literature on prosocial incentives and experimental studies focusing on leadership, the literature has not yet analyzed the impact of responsibility as a prosocial incentive. Therefore, this study contributes to the literature by examining the impact of responsibility as a motivational factor among children, with particular attention to gender, ability and altruism differences.

Chapter 3

EXPERIMENTAL DESIGN

To assess the impact of responsibility and material incentives on children's performance, we conducted an experiment on 937 elementary school students. The experiments were conducted in June 2023 in ten public schools across six districts in Istanbul. They were in the form of tasks that were completed by the students in their own classrooms using electronic tablets. [1]

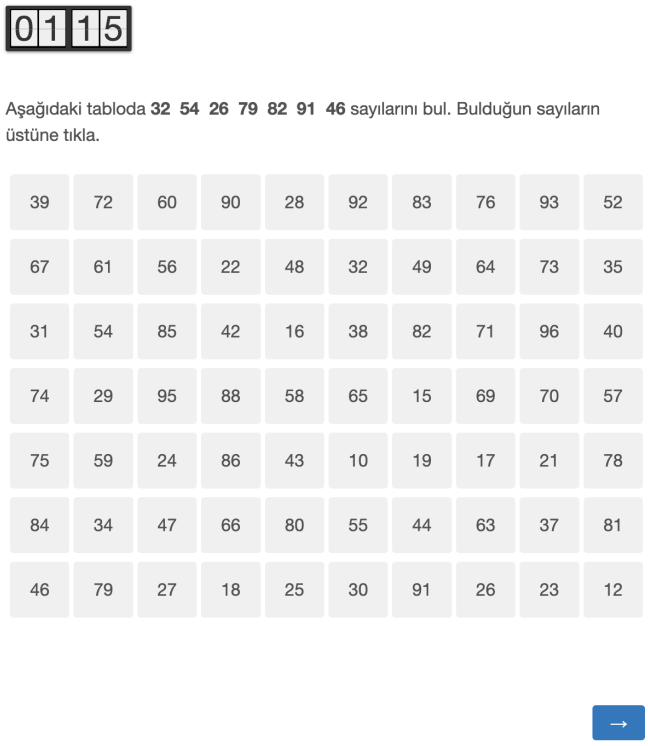
High performance in the experiment was rewarded, as students collected gift points which enabled them to earn presents, such as pencils, balls, and stickers at the end of the day. Before the tablets were distributed and the experiment began, it was explained to the students that they were not allowed to talk or share their answers with each other. To maintain privacy, the school bags of the students were placed between students who were sitting next to each other.

The focus of this thesis is the Number Finding Test, which was implemented to assess the impact of responsibility on motivation, compared to material incentives. In the Number Finding Test, the students were shown certain numbers and were asked to find these among a larger set of numbers in a limited time of 1 minute and 15 seconds. Before the task started, students were explained how the task worked and then a practice question appeared where 3 numbers were given and the participants were asked to find and select these among 20 numbers in 15 seconds. After the practice question was shown and the students were familiar with the test, the participants were randomly assigned to three groups. One third of the students were assigned to the group in which for each correct answer they earned 1 gift point, one third were assigned to the group in which for each correct answer they earned 2 gift points and one third were assigned to the group in which for each

[1]Experiments were conducted with parental consent and IRB approval. Data were collected anonymously.

correct answer the student earned 1 gift point for himself/herself and 1 point for another student from another school who could not participate in this activity. In the thesis, these groups are respectively referred to as "Material x1", "Material x2" and "Responsibility". Students read the rules of these treatments on their tablet, as the randomization was done at the individual level. After the students were assigned to their groups, a question was asked to them to check whether they understood which group they are in and how many points they will earn for themselves and others for each correct answer. If the student gave a wrong answer to this question the correct group of the student and how many points will be earned was explained once again in big, bold letters. Then, the test began and seven two-digit numbers were shown to the students who were asked to find these among seventy two-digit numbers in a limited time of 1 minute and 15 seconds. An example question is shown in Figure 3.1. However, even though there was a time limit of 1 minute and 15 seconds, the students were able to end the question earlier if they wanted.

Figure 3.1: Example Question in Number Finding Test



We also constructed a measure of cognitive ability. In this regard, we used an IQ test (Raven's Progressive Matrices) in which the participants were asked to identify the correct missing piece that completes a matrix. An example questions from the IQ test is provided in the Appendix.

At the end of the experiment, a question was asked to the students to measure their level of altruism. It was explained to the students that there were some schools that could not be visited and the students in such schools were not able to earn gift points and get presents. In this regard, in addition to the gift points that the students won, four more points were given, and it was asked that from these four points how many of it they would want to give away to the students in schools who could not participate. The response to this question was used as a measure to provide information on how altruistic the participant was.

Chapter 4

DATA

The data consists of a total of 937 elementary school students. It was collected from 10 different schools across six districts in Istanbul. Of the 937 observations, 485 (51.8%) belong to females and 452 (48.2%) belong to males. In terms of the treatment groups, 291 (31.1%) students belong to Material x1, 328 (35.0%) belong to Material x2 and 318 (33.9%) belong to Responsibility treatment. Therefore, both in terms of gender and treatment groups there is a balanced distribution within the data. In terms of understanding which group they belong to, 621 (66.3%) participants correctly identified their group. 63.1% of the participants are classified to having a high level of altruism while 47.5% are classified of having an above-median level of cognitive ability. Table 4.1 presents the summary statistics of the data.

Table 4.1: Summary Statistics

Total Participants	937
Female Participants	485
Percentage of Female Participants	51.8%
Participants in Material x1	291
Percentage of Participants in Material x1	31.1%
Participants in Material x2	328
Percentage of Participants in Material x2	35.0%
Participants in Responsibility	318
Percentage of Participants in Responsibility	33.9%
Participants Correctly Identifying Group	621
Percentage of Participants Correctly Identifying Group	66.3%
Participants with Non-Zero Donations	591
Percentage of Participants with Non-Zero Donation	63.1%
Participants with Above Median Cognitive Ability	445
Percentage of Participants with Above Median Cognitive Ability	47.5%

Chapter 5

RESULTS

Figures 5.1, 5.2 and 5.3 present the histograms for the distribution of scores in the treatment groups Material x1, Material x2 and Responsibility respectively. The figures reveal that the distribution of scores is similar across the three treatment groups, resembling a normal distribution. Moreover, it should also be noted that most students did not find all numbers within the given time, which is reassuring because it shows that the task is not too easy and there is room for treatment effects to operate.

Figure 5.1: Distribution of Scores - Material x1

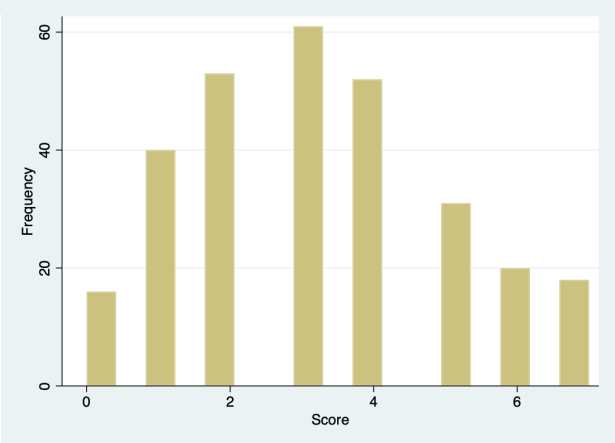


Figure 5.2: Distribution of Scores - Material x2

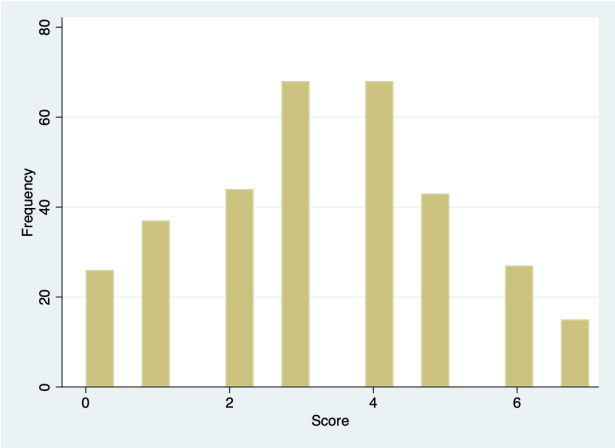


Figure 5.3: Distribution of Scores - Responsibility

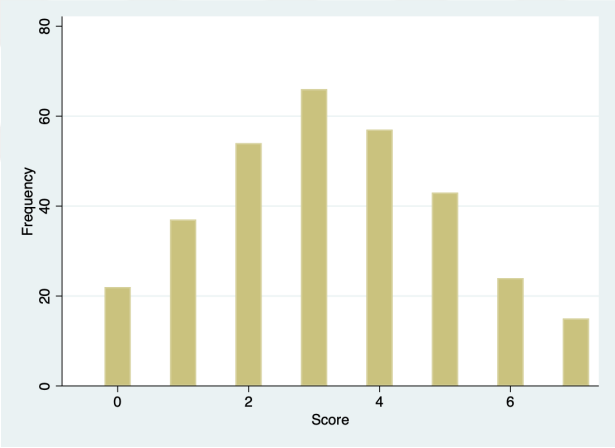
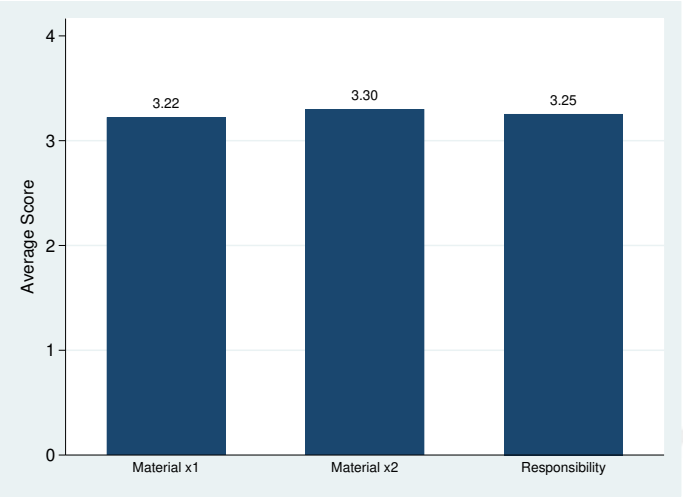


Figure 5.4 presents the average scores in the Number Finding Test over the three treatment groups. The figure reveals that there is no notable difference between the average scores of the three groups.

Figure 5.4: Average Score in Number Finding Test over Treatment



To more rigorously assess the effects, we conduct Ordinary Least Squares (OLS) regression analyses. In this regard, throughout the study, we estimate the following OLS regression model to evaluate the effects of different incentive treatments on motivation considering sub-groups within the data.

$$\text{Score}_i = \beta_0 + \beta_1 \text{Material x2}_i + \beta_2 \text{Responsibility}_i + \gamma X_i + \varepsilon_i \tag{5.1}$$

In this specification, the dependent variable Score_i represents the total number of correct answers provided by participant i. The variable Material x2_i is a dummy variable equal to 1 if the participant was assigned to the high-reward material incentive group and 0 otherwise. Similarly, Responsibility_i is a dummy variable equal to 1 if the participant was assigned to the responsibility treatment group and 0 otherwise. The omitted baseline category is Material x1. The vector X_i includes control variables such as gender, experimenter (accounts for the research assistants explaining the tests and supervising the experiments), cognitive ability, treatment comprehension (based on whether the student correctly identified the treatment group to which they were assigned to, in the quiz question), and school. It must also be noted that, the standard error are clustered at the class level.

This specification enables us to identify the average treatment effects of different incentive schemes on performance, controlling for individual and contextual

characteristics. Additionally, to ensure robustness, all regression models have been replicated using only the subset of students who correctly identified their treatment group (Understanding = 1). These regressions are presented in the Appendix.

Result 1: *Elementary school students are not sensitive to either of the incentive treatments.*

Table 5.1 presents the OLS regression results on scores of the participants. The results indicate that neither material incentives nor responsibility have a statistically significant effect on motivation. Moreover, when we conduct Wald tests to determine whether the coefficients of Material x2 and Responsibility are statistically significantly different within the models, we find that in none of the models there is a statistically significant difference between the two coefficients, as the p-values range between 0.7412 and 0.9980. Furthermore, when we use ordered logit instead of OLS, we similarly find no significant effect of material incentives or responsibility on motivation.

Table 5.1: OLS Regression Results on Scores

	(I)	(II)	(III)	(IV)	(V)	(VI)
Material x2	0.078	0.090	0.093	0.093	0.054	0.056
	(0.144)	(0.146)	(0.146)	(0.142)	(0.147)	(0.149)
Responsibility	0.031	0.056	0.083	0.066	0.048	0.056
	(0.145)	(0.143)	(0.136)	(0.134)	(0.133)	(0.139)
Constant	3.223***	3.349***	2.754***	2.175***	2.117***	3.966***
	(0.111)	(0.127)	(0.141)	(0.204)	(0.219)	(0.701)
Controls(Gender)	No	Yes	Yes	Yes	Yes	Yes
Controls (Experimenter)	No	No	Yes	Yes	Yes	Yes
Controls (Cognitive Ability)	No	No	No	Yes	Yes	Yes
Controls (Understanding)	No	No	No	No	Yes	Yes
Controls (School)	No	No	No	No	No	Yes
Observations	937	937	937	937	937	935
R^2	0.000	0.006	0.037	0.057	0.058	0.065

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Incentive x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.

These findings suggest that elementary school students are not sensitive to material or prosocial incentives. A possible explanation could be that young children

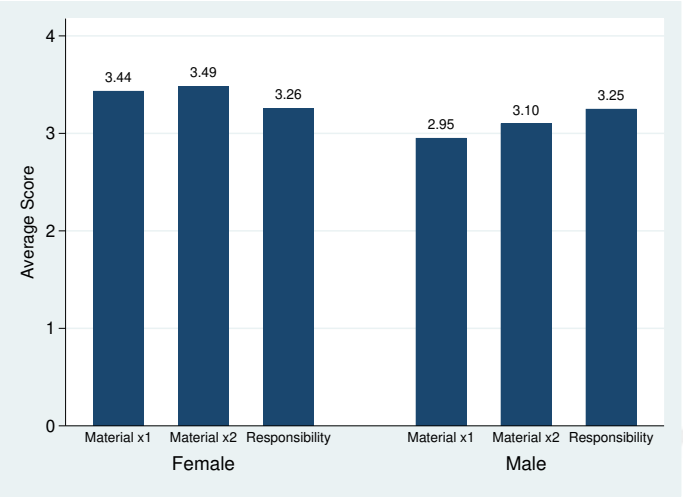
tend to exert effort and try to do their best intrinsically regardless of the external reward structure. At this developmental stage, students might be less influenced by external motivators and more driven by a natural desire to complete tasks in the best way possible.

Result 2: *Responsibility has a positive effect on boys and a negative effect on girls. While these effects are not individually significant, the difference between them is statistically significant.*

The results may be different for girls and boys as research have shown that girls and boys differ in terms of their risk preferences, social preferences, and competitive preference (Croson and Gneezy, 2009). Thus, we analyze whether the impact of the treatments differs between boys and girls.

Firstly, we conduct a Kolmogorov–Smirnov (K-S) test considering boys and girls. The test yields a combined K–S statistic of 0.0782 with a p-value of 0.114 indicating that the distribution of scores between girls and boys are not significantly different. In Figure 5.5 we analyze also over treatment as the figure presents the average scores in the Number Finding Test over treatment and gender. The figure reveals that within boys, those receiving responsibility treatment have the highest average score of 3.25, which is 0.30 points (4.3%) higher than the average of the benchmark group Material x1. On the other hand, considering girls, those receiving responsibility treatment have the lowest average score of 3.26, which is 0.18 points (2.6%) lower than the average of the benchmark group Material x1. Therefore, the effect of responsibility seems to effect boys and girls differently. On the other hand, when we consider the average scores in Material x2 we see that it is very close to Material x1 for both boys and girls indicating that higher rewards may not have a large impact on their motivation.

Figure 5.5: Average Score in Number Finding Test over Treatment and Gender



Overall, Figure 5.5 suggests that the impact of responsibility may differ by gender. To examine these effects more rigorously, we also conduct OLS regression analyses. The results are presented in Table 5.2. Columns I-V are the results for male students whereas columns VI-X are the results for female students.

Table 5.2: OLS Regression Results on Scores by Gender

	Males					Females				
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
Material x2	0.149	0.167	0.178	0.136	0.118	0.050	0.057	0.048	0.015	0.031
	(0.217)	(0.209)	(0.209)	(0.203)	(0.207)	(0.167)	(0.176)	(0.168)	(0.193)	(0.196)
Responsibility	0.298	0.352	0.343	0.340	0.340	-0.177	-0.170	-0.181	-0.201	-0.197
	(0.218)	(0.217)	(0.222)	(0.223)	(0.225)	(0.213)	(0.200)	(0.191)	(0.196)	(0.201)
Constant	2.953***	2.594***	2.161***	2.044***	3.754***	3.436***	3.530***	2.497***	2.424**	2.462**
	(0.157)	(0.188)	(0.270)	(0.316)	(1.271)	(0.144)	(0.857)	(0.915)	(0.932)	(0.937)
Controls (Experimenter)	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Controls (Cognitive Ability)	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Controls (Understanding)	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Controls (School)	No	No	No	No	Yes	No	No	No	No	Yes
Observations	452	452	452	452	452	485	485	485	485	483
R^2	0.004	0.028	0.041	0.045	0.052	0.003	0.051	0.081	0.081	0.089

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Material x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.

Table 5.2 reveals that, for both boys and girls, neither the material incentives nor the responsibility have a statistically significant effect on performance. However, when we consider the average scores of girls in the three treatment groups, a power analysis reveals that detecting a statistically significant effect at 80% power would

require at least 265 participants per group whereas our sample includes approximately 160 girls per group. This indicates that the analysis may be underpowered to detect differences. Similarly, for boys, considering the average scores of the three groups, detecting the observed differences with 80% power would require to have at least 173 participants per group, whereas in our analysis we have around 150 boys per group.

On the other hand, when we conduct Wald tests, we find that in none of the regression models there is a statistically significant difference between the coefficients of Material x2 and Responsibility, as the p-values range between 0.3506 and 0.4826 for boys and 0.2515 and 0.2981 for girls. However, when we compare the effect of responsibility treatment across males and females using seemingly unrelated estimation (suest) followed by a Wald test considering the full models presented in columns V and X, we find a chi-squared statistic of 3.39 with a p-value of 0.0657. This result indicates that the effect of responsibility differs (weakly) significantly between boys and girls. On the other hand, considering the coefficients for Material x2 over the same models, no statistically significant difference is found (p-value = 0.7728).

Moreover, we also conducted regressions with interactions between treatment and gender. The regression results are presented in Table 5.3. The table reveals that the interaction between responsibility and male is positive and statistically significant at the 10% significance level when considering models IV and V (p-values are 0.094 and 0.096 respectively). Furthermore, The gender difference in the effect of the Responsibility treatment corresponds to 27.8% of the standard deviation of performance among girls, whereas 27.5% of the standard deviation of performance among boys. This result, consistent with the findings from Table 5.2 and following Wald tests, indicates that the effect of responsibility differs significantly for boys and girls.

Table 5.3: Interaction Effects of Treatment and Gender on Scores

	(I)	(II)	(III)	(IV)	(V)
Material x2	0.050	0.057	0.043	-0.013	0.004
	(0.167)	(0.173)	(0.166)	(0.182)	(0.186)
Responsibility	-0.177	-0.172	-0.190	-0.225	-0.218
	(0.213)	(0.201)	(0.194)	(0.194)	(0.199)
Male	-0.482**	-0.484**	-0.495**	-0.512**	-0.502**
	(0.208)	(0.199)	(0.207)	(0.206)	(0.208)
Material x2 X Male	0.099	0.087	0.117	0.144	0.112
	(0.246)	(0.242)	(0.242)	(0.250)	(0.255)
Responsibility X Male	0.476	0.517	0.522	0.554*	0.551*
	(0.318)	(0.317)	(0.324)	(0.325)	(0.326)
Constant	3.436***	3.565***	2.950***	2.883***	0.889***
	(0.144)	(0.277)	(0.303)	(0.291)	(0.262)
Controls (Experimenter)	No	Yes	Yes	Yes	Yes
Controls (Cognitive Ability)	No	No	Yes	Yes	Yes
Controls (Understanding)	No	No	No	Yes	Yes
Controls (School)	No	No	No	No	Yes
Observations	937	937	937	937	935
R^2	0.009	0.040	0.060	0.062	0.069

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Material x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.

Overall, these results suggest that responsibility has a positive effect on the performance of boys, whereas a negative effect on the performance of girls. Although these effects are not statistically significant when examined individually, the difference between them is found to be statistically significant, indicating that there is a difference in how responsibility effects the performance of boys and girls at elementary school.

For a robustness check of the OLS regressions, when we restrict the sample to the students who understood the treatment (results are presented in Table A.2 in the Appendix), we find that when the regressions are evaluated individually the results are very similar to the results obtained from the overall data. On the other hand, when examine the regression with the interaction effects restricted for only the students who understood the treatment (results are presented in Table A.3 in the Appendix) we find that the interactions are no longer significant. However the p-values are very close to being significant at the 10% level (p-values are between

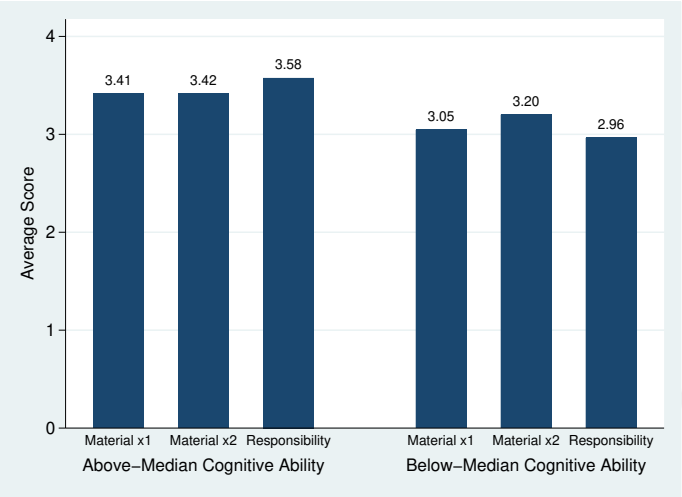
0.104 and 0.146)

Result 3: *Elementary school children are not sensitive to motivators, regardless of their level of cognitive ability.*

Children who have different levels of cognitive ability might react differently towards different types of incentives schemes. Therefore, we also examine the results separately for students who have a below-median and above-median cognitive ability.

Firstly, when we conduct a K-S test for students with below-median and above-median cognitive ability, we get a test statistic of 0.0935 with a p-value of 0.033. This indicates that the distribution of scores between students who have different levels of cognitive ability are significantly different at the 5% level. Moreover, Figure 5.6 considers also treatment as it presents the average scores over treatment and cognitive ability. The figure reveals that for students who have an above-median cognitive ability, the participants in the responsibility treatment obtained the highest average score whereas the average score of students in the treatment groups Material x1 and Material x2 are very similar to each other. On the other hand, considering those who have a below-median cognitive ability, the figure reveals that students who are in the responsibility treatment have the lowest average score whereas students who are in the Material x2 treatment have the highest average score. Furthermore, students with above-median cognitive ability score significantly higher (mean = 3.58) than students with below-median cognitive ability (mean = 2.96) within the responsibility treatment group and when we conduct a ttest we find that the difference is statistically significant (p=0.0028).

Figure 5.6: Average Score in Number Finding Test over Treatment and Cognitive Ability



To have a more rigorous analysis, we also conducted OLS regressions based on cognitive ability levels. Table 5.4 presents the results. Columns I-V show the results for the students who have an above-median cognitive ability whereas columns VI-X show the results for the students who have a below-median cognitive ability.

Table 5.4: OLS Regression Results on Scores by Cognitive Ability Level

	Above-Median Cognitive Ability					Below-Median Cognitive Ability				
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
Material x2	0.001	-0.000	0.050	0.061	0.031	0.155	0.167	0.134	0.011	0.025
	(0.222)	(0.219)	(0.223)	(0.214)	(0.215)	(0.199)	(0.201)	(0.192)	(0.204)	(0.205)
Responsibility	0.162	0.191	0.244	0.249	0.209	-0.082	-0.066	-0.043	-0.101	-0.072
	(0.243)	(0.238)	(0.240)	(0.237)	(0.239)	(0.169)	(0.168)	(0.165)	(0.172)	(0.172)
Constant	3.414***	3.618***	2.658***	2.685***	4.608***	3.046***	3.112***	2.858***	2.713***	0.565
	(0.165)	(0.183)	(0.185)	(0.237)	(0.144)	(0.152)	(0.173)	(0.289)	(0.305)	(1.083)
Controls (Gender)	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Controls (Experimenter)	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Controls (Understanding)	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Controls (School)	No	No	No	No	Yes	No	No	No	No	Yes
Observations	445	445	445	445	444	492	492	492	492	491
R^2	0.002	0.017	0.055	0.055	0.064	0.003	0.005	0.037	0.044	0.060

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Material x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.

The table reveals that both material incentives and responsibility do not have a statistically significant effect on the performance of elementary school students either with above-median cognitive ability or below-median cognitive ability. More-

over, when we conduct Wald tests, we find that within none of the regression models there is a statistically significant difference between the coefficients of Material x2 and Responsibility as the p-values range between 0.3126 and 0.4476 for students with above-median cognitive ability and between 0.1770 and 0.5981 for students with below-median cognitive ability. Furthermore, we compare the regression coefficients of Responsibility and Material x2 across students with above-median cognitive ability and below-median cognitive ability using suest followed by a Wald Test. Considering the full models of columns V and X, we find p-values of 0.3312 and 0.9829, respectively for Responsibility and Material x2. This indicates that there is no statistically significant difference in how material incentives or responsibility affect the motivation for those who have a above-median or below-median level of cognitive ability.

Moreover, we also examined the interaction between Responsibility and Material x2 with the students performance in the IQ test. The results, presented in Table A.5 in the Appendix, indicate that neither interaction is statistically significant.

Therefore, neither material incentives nor responsibility have a significant effect on motivation of elementary school students, regardless of the level of cognitive ability the student has. This result is consistent with and strengthens the previous finding that young children are not sensitive to external incentive schemes.

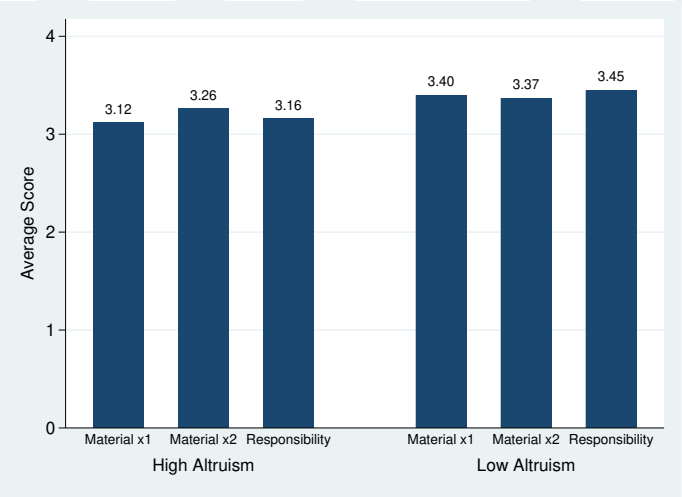
Result 4: *Elementary school children are not sensitive to motivators, regardless of their level of altruism.*

Altruism is recognized as a key driver of prosocial behavior (Benabou and Tirole, 2010). Based on this, we would expect the motivation of students who are more altruistic to be more positively affected by the responsibility treatment compared to those who are less altruistic. In order to test this hypothesis, we analyze the performance of students based on their level of altruism. As explained in the experimental design section, at the end of the experiments the students were given four extra points and were asked how many of these they would want to give away to students from other schools who were not able to participate in the experiments.

Students who chose to donate at least one extra point in response to this question were categorized as having a high level of altruism, whereas students who chose to keep all extra points were categorized as having a low level of low altruism. Based on this classification 63.1% of the participants have a high level of altruism. The distribution of the number of donations is presented in Figure A.2 in the Appendix.

Firstly, when we conduct a K-S test for students with high and low levels of altruism, we get a test statistic of 0.0629 with a p-value of 0.354 indicating that the distribution of scores between students who have a high and low level of altruism are not significantly different. Figure 5.7 considers also treatment as it presents the average scores over treatment and altruism level. The figure reveals that the average scores in the three treatments are similar to each other for students who have high level of altruism or low level of altruism.

Figure 5.7: Average Score in Number Finding Test over Treatment and Altruism



To have a more rigorous analysis, we also conducted OLS regressions based on altruism levels. Table 5.5 presents the results. Columns I-VI show the results for the students who have a high level of altruism whereas columns VII-XII show the results for the students who have a low level of altruism.

Table 5.5: OLS Regression Results on Scores by Altruism Level

	High Altrusim						Low Altruism					
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)	(XI)	(XII)
Material x2	0.139	0.172	0.150	0.135	0.069	0.071	-0.032	-0.031	0.018	0.050	0.079	0.051
	(0.183)	(0.185)	(0.182)	(0.184)	(0.197)	(0.199)	(0.211)	(0.210)	(0.210)	(0.200)	(0.196)	(0.209)
Responsibility	0.036	0.091	0.116	0.098	0.066	0.074	0.050	0.047	0.082	0.064	0.076	0.043
	(0.170)	(0.176)	(0.167)	(0.161)	(0.158)	(0.166)	(0.265)	(0.266)	(0.256)	(0.249)	(0.244)	(0.255)
Constant	3.120***	3.333***	2.609***	2.025***	1.952***	3.300***	3.398***	3.361***	3.550***	2.535***	3.000***	3.587***
	(0.147)	(0.159)	(0.144)	(0.229)	(0.238)	(0.201)	(0.194)	(0.235)	(0.817)	(0.362)	(0.938)	(0.899)
Controls (Gender)	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Controls (Experimenter)	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Controls (Cognitive Ability)	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Controls (Understanding)	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Controls (School)	No	No	No	No	No	Yes	No	No	No	No	No	Yes
Observations	591	591	591	591	591	589	346	346	346	346	346	346
R^2	0.001	0.019	0.066	0.087	0.091	0.098	0.000	0.001	0.020	0.038	0.038	0.055

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Material x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.

The table reveals that both material incentives and responsibility do not have a statistically significant effect on the performance of students either with high altruism or low levels of altruism. Moreover, when we conduct Wald tests, we find that within none of the regression models there is a statistically significant difference between the coefficients of Material x2 and Responsibility as the p-values range between 0.5683 and 0.9886 for students with high altruism and between 0.7391 and 0.9903 for students with low altruism. Furthermore, we compare the regression coefficients of Responsibility and Material x2 across students with high altruism and low altruism using suest followed by a Wald Test. Considering the full models of columns VI and XII, we find p-values of 0.9199 and 0.9491, respectively for Responsibility and Material x2. This indicates that in elementary school there is no statistically significant difference in how material incentives or responsibility affect the motivation for students who have a high or low level of altruism.

Moreover, we also examined the interaction between Responsibility and Material x2 with the students level of altruism (measured by the number of donations). The results, presented in Table A.8 in the Appendix, indicate that neither interaction is statistically significant.

Therefore, the regressions and the following tests indicate that in elementary school neither material incentives nor responsibility have a significant effect on motivation, regardless of the level of altruism the student has. Similar to the previous

result, this result is also consistent with and strengthens the previous finding that children who are at a young age are not sensitive to external motivators.

Chapter 6

CONCLUSION

This thesis investigated the effect of responsibility as a prosocial incentive on the motivation of elementary school students. In a field experiment involving a simple number finding test, participants were randomly assigned to one of three incentive schemes; a material incentive group, a high-reward material incentive group, or a responsibility group where correct answers benefited both the student and an anonymous peer. The aim was to assess how responsibility compares to material incentives in motivating effort, and investigate whether this effect varies by gender, level of cognitive ability, or level of altruism.

This study makes an important contribution to the literature by exploring the role of responsibility as a prosocial incentive as this topic has remained understudied, particularly in the context of children. While existing research has focused on material and non-material motivators, prosocial incentives such as responsibility have received little empirical attention. This gap is especially pronounced in the educational settings, where understanding what motivates children is crucial in order to design and implement effective interventions. By isolating the motivational effects of responsibility and comparing them to material incentives, this study provides novel insights into how children respond to prosocial motivators.

The findings reveal several key insights. First, elementary school students did not show significant performance differences across the three treatments, suggesting limited sensitivity to both material and prosocial incentives. Furthermore, when the results were analyzed separately based on students' levels of cognitive ability and altruism, no significant treatment effects emerged within any subgroup. This may reflect an intrinsic motivation to engage with tasks at younger developmental stages.

Another explanation could be that the task is not conducive to higher effort, that is, it may be difficult to increase performance with effort. However, the fact that

there is some positive effect on performance of boys with responsibility suggests that the task is not completely immune to effort-based performance changes. Still, further work should look at other tasks where effects of responsibility can be measured, to see if any of the effects documented here are task-specific.

Furthermore, responsibility appeared to have a gender-differentiated effect on motivation such that it positively affected boys while negatively affecting girls. Although these effects were not statistically significant when evaluated individually, the difference between them is found to be statistically significant.

Overall, this study suggests that external incentives have limited impact on the motivation of elementary school students. Moreover, the absence of significant differences across treatments, even when accounting for cognitive ability and altruism, suggests the potential dominance of intrinsic motivation at early developmental stages. These findings indicate that responsibility does not uniformly enhance the motivation of children at young ages and underscore the importance of considering age-specific motivational drivers when designing interventions aimed at improving student performance. Future research should further explore how different types of tasks and settings might interact with prosocial incentives, and whether similar findings hold for individuals who are older or challenges that are more complex.

BIBLIOGRAPHY

Andreoni, J., Harbaugh, W. T., and Vesterlund, L. (2008). Altruism in experiments. In *The new Palgrave dictionary of economics*, pages 1–7. Springer.

Anik, L., Aknin, L. B., Norton, M. I., Dunn, E. W., and Quoidbach, J. (2013). Prosocial bonuses increase employee satisfaction and team performance. *PloS one*, 8(9):e75509.

Artinger, F., Exadaktylos, F., Koppel, H., and Sääksvuori, L. (2010). Fairness in simple bargaining games: the role of empathy and theory of mind. *Max Planck Institute of Economics Working Paper*.

Bareket-Bojmel, L. and Hochman, G. (2014). It's (not) all about the jacksons: Testing different types of short-term bonuses in the field. *Journal of Management*, 9999.

Bellemare, C. and Shearer, B. (2009). Gift giving and worker productivity: Evidence from a firm-level experiment. *Games and Economic Behavior*, 67:233–244.

Benabou, R. and Tirole, J. (2010). Individual and corporate social responsibility. *Economica*, 77.

Bettinger, E. P. (2012). Paying to learn: The effect of financial incentives on elementary school test scores. *Review of Economics and Statistics*, 94(3):686–698.

Chakravarty, S., Harrison, G., Haruvy, E., and Rutström, E. (2011). Are you risk averse over other people's money? *Southern Economic Journal*, 77:901–913.

Condly, S., Clark, R., and Stolovitch, H. (2003). The effects of incentives on workplace performance: A meta-analytic review of research studies. *Performance Improvement Quarterly*, 16:46 – 63.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47:448–74.

Engel, C. (2010). Dictator games: a meta study. *Max Planck Institute for Research on Collective Goods, Working Paper Series of the Max Planck Institute for Research on Collective Goods*, 14.

Ertac, S. and Gurdal, M. (2012). Deciding to decide: Gender, leadership and risk-taking in groups. *Journal of Economic Behavior Organization - J ECON BEHAV ORGAN*, 83.

Fehr, E. and Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425:785–91.

Fryer, R. (2010). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics*, 126:1755–1798.

Gosnell, G., List, J., and Metcalfe, R. (2019). The impact of managerial practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy*, 128.

Grant, A. and Gino, F. (2010). A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior. *Journal of personality and social psychology*, 98:946–55.

Hochman, G. and Ariely, D. (2015). Behavioral economics. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, pages 1–14.

Jalava, N., Joensen, J. S., and Pellas, E. (2015). Grades and rank: Impacts of non-financial incentives on test performance. *Journal of Economic Behavior Organization*, 115:161–196. Behavioral Economics of Education.

Jeffrey, S. (2009). Justifiability and the motivational power of tangible noncash incentives. *Human Performance*, 22:143–155.

Lee, J., Keil, M., and Te'eni, P. (2018). The effects of relative and criticism-based performance appraisals on task-level escalation in an it project: a laboratory experiment. *European Journal of Information Systems*, 27:1–19.

Levitt, S., List, J., Neckermann, S., and Sadoff, S. (2012). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *SSRN Electronic Journal*.

Migheli, M. (2010). Gender at work: Productivity and incentives. *Department of Public Policy and Public Choice - POLIS, POLIS Working Papers*.

Nagin, D., Rebitzer, J., Sanders, S., and Taylor, L. (2002). Monitoring, motivation and management: The determinants of opportunistic behavior in a field experiment. *American Economic Review*, 92.

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

Sittenthaler, H. and Mohnen, A. (2020). Cash, non-cash, or mix? gender matters! the impact of monetary, non-monetary, and mixed incentives on performance. *Journal of Business Economics*, 90.

Stajkovic, A. and Luthans, F. (2003). Behavioral management and task performance in organizations: Conceptual background, meta-analysis, and test of alternative models. *Personnel Psychology*, 56:155 – 194.

UK Government (2025). Education maintenance allowance (ema). Available at: https://www.gov.uk/education-maintenance-allowance-ema.

Appendix A

Figure A.1: Example Question in IQ (Raven's Progressive Matrices) Test

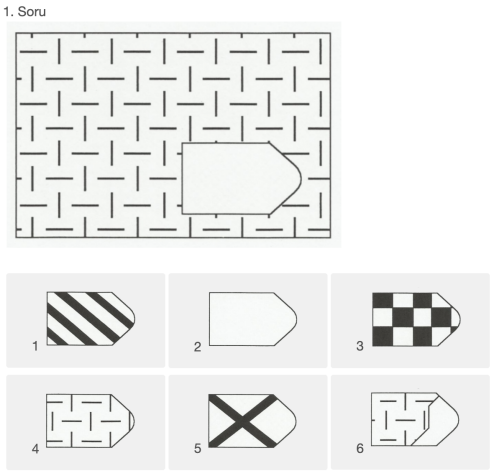


Table A.1: OLS Regression Results on Scores (For Understanding = 1)

	(I)	(II)	(III)	(IV)	(V)
Material x2	0.024	0.016	0.036	0.066	0.065
	(0.185)	(0.185)	(0.184)	(0.184)	(0.186)
Responsibility	0.076	0.075	0.106	0.108	0.112
	(0.184)	(0.182)	(0.178)	(0.176)	(0.181)
Constant	3.309***	3.421***	2.880***	2.430***	0.938***
	(0.142)	(0.157)	(0.308)	(0.375)	(0.228)
Controls(Gender)	No	Yes	Yes	Yes	Yes
Controls (Experimenter)	No	No	Yes	Yes	Yes
Controls (Cognitive Ability)	No	No	No	Yes	Yes
Controls (School)	No	No	No	No	Yes
Observations	621	621	621	621	620
R^2	0.000	0.004	0.028	0.038	0.051

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Incentive x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.

Table A.2: OLS Regression Results on Scores by Gender (For Understanding = 1)

	Males				Females			
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
Material x2	0.118	0.191	0.249	0.195	-0.085	-0.067	-0.086	-0.041
	(0.260)	(0.253)	(0.254)	(0.270)	(0.225)	(0.233)	(0.228)	(0.229)
Responsibility	0.290	0.379	0.408	0.394	-0.132	-0.163	-0.196	-0.160
	(0.238)	(0.256)	(0.260)	(0.270)	(0.253)	(0.235)	(0.229)	(0.235)
Constant	3.080***	2.256***	1.943***	4.650***	3.532***	3.378***	2.516***	2.782***
	(0.205)	(0.214)	(0.331)	(1.154)	(0.193)	(0.730)	(0.860)	(0.754)
Controls (Experimenter)	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Controls (Cognitive Ability)	No	No	Yes	Yes	No	No	Yes	Yes
Controls (School)	No	No	No	Yes	No	No	No	Yes
Observations	296	296	296	296	325	325	325	324
R^2	0.004	0.048	0.053	0.063	0.001	0.053	0.072	0.100

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Material x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.

Table A.3: Interaction Effects of Treatment and Gender on Scores (For Understanding = 1)

	(I)	(II)	(III)	(IV)
Material x2	-0.085	-0.061	-0.081	-0.053
	(0.225)	(0.232)	(0.227)	(0.227)
Responsibility	-0.132	-0.134	-0.168	-0.154
	(0.254)	(0.239)	(0.234)	(0.239)
Male	-0.452	-0.458	-0.523*	-0.525*
	(0.282)	(0.284)	(0.292)	(0.291)
Material x2 x Male	0.203	0.189	0.294	0.233
	(0.312)	(0.317)	(0.320)	(0.322)
Responsibility x Male	0.422	0.487	0.560	0.535
	(0.333)	(0.331)	(0.339)	(0.337)
Constant	3.532***	3.641***	3.199***	1.202***
	(0.193)	(0.476)	(0.479)	(0.307)
Controls (Experimenter)	No	Yes	Yes	Yes
Controls (Cognitive Ability)	No	No	Yes	Yes
Controls (School)	No	No	No	Yes
Observations	621	621	621	620
R^2	0.006	0.030	0.041	0.054

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Material x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.

Table A.4: OLS Regression Results on Scores by Cognitive Ability Level (For Understanding = 1)

	Above-Median Cognitive Ability				Below-Median Cognitive Ability			
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
Material x2	0.114	0.052	0.096	0.049	-0.043	-0.034	-0.017	-0.017
	(0.254)	(0.251)	(0.250)	(0.254)	(0.286)	(0.289)	(0.278)	(0.290)
Responsibility	0.226	0.187	0.232	0.191	-0.075	-0.066	-0.018	-0.008
	(0.295)	(0.293)	(0.293)	(0.304)	(0.281)	(0.286)	(0.293)	(0.300)
Constant	3.345***	3.558***	3.665***	5.428***	3.265***	3.294***	3.418**	1.809
	(0.186)	(0.224)	(0.722)	(0.900)	(0.238)	(0.243)	(1.465)	(1.471)
Controls (Gender)	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Controls (Experimenter)	No	No	Yes	Yes	No	No	Yes	Yes
Controls (School)	No	No	No	Yes	No	No	No	Yes
Observations	313	313	313	313	308	308	308	307
R^2	0.002	0.013	0.060	0.070	0.000	0.001	0.022	0.035

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Material x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.

Table A.5: OLS Regression Results on Scores with Interaction Effects of Treatment and Cognitive Ability

	(I)	(II)	(III)	(IV)	(V)
Material x2	0.439	0.473	0.444	0.365	0.375
	(0.430)	(0.431)	(0.430)	(0.458)	(0.465)
Responsibility	-0.063	-0.037	0.013	-0.018	0.014
	(0.430)	(0.432)	(0.434)	(0.443)	(0.437)
IQ Test Score	0.092***	0.092***	0.087***	0.084***	0.081***
	(0.025)	(0.025)	(0.026)	(0.027)	(0.028)
Material x2 x IQ Test Score	-0.033	-0.035	-0.032	-0.028	-0.029
	(0.036)	(0.035)	(0.035)	(0.037)	(0.038)
Responsibility x IQ Test Score	0.008	0.007	0.005	0.006	0.004
	(0.039)	(0.039)	(0.039)	(0.040)	(0.040)
Constant	2.204***	2.324***	2.498***	2.463***	0.336
	(0.313)	(0.317)	(0.383)	(0.366)	(0.309)
Controls (Gender)	No	Yes	Yes	Yes	Yes
Controls (Experimenter)	No	No	Yes	Yes	Yes
Controls (Understanding)	No	No	No	Yes	Yes
Controls (School)	No	No	No	No	Yes
Observations	937	937	937	937	935
R^2	0.031	0.036	0.062	0.063	0.070

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Material x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.

Table A.6: OLS Regression Results on Scores with Interaction Effects of Treatment and Cognitive Ability (Understanding = 1)

	(I)	(II)	(III)	(IV)
Material x2	-0.013	0.060	0.044	0.109
	(0.572)	(0.579)	(0.572)	(0.576)
Responsibility	-0.089	-0.033	-0.006	0.087
	(0.632)	(0.647)	(0.649)	(0.644)
IQ Test Score	0.054	0.059	0.054	0.052
	(0.036)	(0.037)	(0.037)	(0.037)
Material x2 x IQ Test Score	0.007	-0.000	0.002	-0.004
	(0.044)	(0.044)	(0.044)	(0.044)
Responsibility x IQ Test Score	0.015	0.010	0.010	0.002
	(0.052)	(0.053)	(0.053)	(0.053)
Constant	2.668***	2.725***	2.904***	0.727*
	(0.497)	(0.495)	(0.636)	(0.370)
Controls (Gender)	No	Yes	Yes	Yes
Controls (Experimenter)	No	No	Yes	Yes
Controls (School)	No	No	No	Yes
Observations	621	621	621	620
R^2	0.017	0.021	0.042	0.054

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Material x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.

Figure A.2: Number of Donations

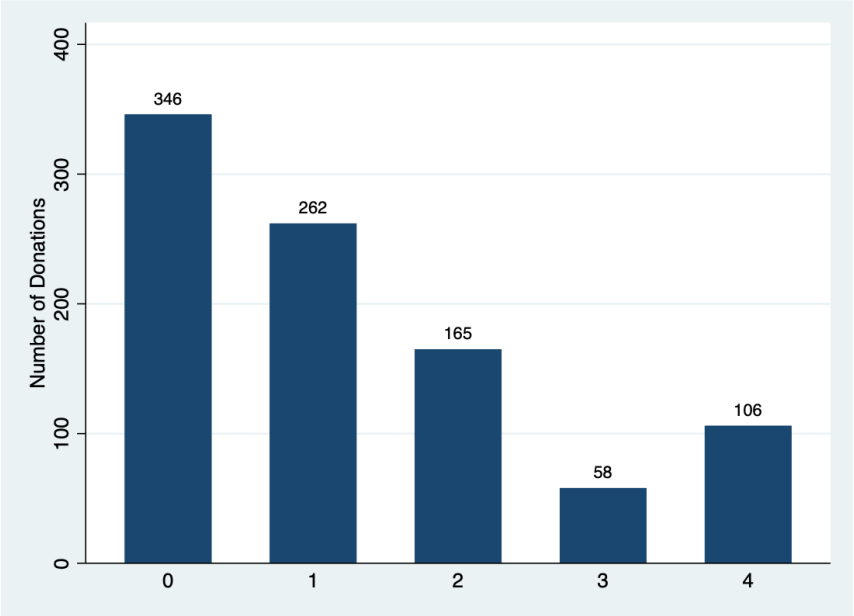


Table A.7: OLS Regression Results on Scores by Altruism Level (For Understanding = 1)

	High Altrusim					Low Altruism				
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
Material x2	0.065	0.068	0.092	0.127	0.112	-0.037	-0.038	-0.001	0.021	0.014
	(0.236)	(0.232)	(0.242)	(0.243)	(0.247)	(0.292)	(0.294)	(0.304)	(0.301)	(0.307)
Responsibility	0.084	0.084	0.116	0.124	0.121	0.072	0.072	0.055	0.046	-0.003
	(0.228)	(0.230)	(0.215)	(0.207)	(0.218)	(0.325)	(0.325)	(0.320)	(0.317)	(0.322)
Constant	3.275***	3.448***	2.730***	2.140***	0.894***	3.361***	3.368***	3.083***	2.897***	3.018***
	(0.174)	(0.182)	(0.384)	(0.501)	(0.301)	(0.277)	(0.313)	(0.919)	(0.449)	(0.953)
Controls (Gender)	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Controls (Experimenter)	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Controls (Cognitive Ability)	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Controls (School)	No	No	No	No	Yes	No	No	No	No	Yes
Observations	378	378	378	378	377	243	243	243	243	243
R^2	0.000	0.010	0.055	0.073	0.082	0.001	0.001	0.036	0.040	0.072

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Material x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.

Table A.8: OLS Regression Results on Scores with Interaction Effects of Treatment and Altruism

	(I)	(II)	(III)	(IV)	(V)	(VI)
Material x2	-0.104	-0.101	-0.071	-0.063	-0.101	-0.097
	(0.185)	(0.189)	(0.181)	(0.179)	(0.181)	(0.191)
Responsibility	-0.080	-0.073	-0.015	-0.041	-0.060	-0.061
	(0.207)	(0.208)	(0.200)	(0.198)	(0.197)	(0.204)
Donations	-0.218***	-0.218**	-0.198**	-0.177**	-0.176**	-0.169**
	(0.081)	(0.082)	(0.079)	(0.079)	(0.078)	(0.080)
Material x2 x Donations	0.146	0.152	0.130	0.124	0.124	0.122
	(0.097)	(0.096)	(0.093)	(0.095)	(0.095)	(0.099)
Responsibility x Donations	0.094	0.107	0.079	0.087	0.088	0.094
	(0.094)	(0.098)	(0.096)	(0.096)	(0.096)	(0.097)
Constant	3.495***	3.615***	3.701***	3.094***	3.029***	1.032***
	(0.152)	(0.176)	(0.287)	(0.321)	(0.315)	(0.316)
Controls (Gender)	No	Yes	Yes	Yes	Yes	Yes
Controls (Experimenter)	No	No	Yes	Yes	Yes	Yes
Controls (Cognitive Ability)	No	No	No	Yes	Yes	Yes
Controls (Understanding)	No	No	No	No	Yes	Yes
Controls (School)	No	No	No	No	No	Yes
Observations	937	937	937	937	937	935
R^2	0.011	0.016	0.045	0.063	0.064	0.070

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Material x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.

Table A.9: OLS Regression Results on Scores with Interaction Effects of Treatment and Altruism (Understanding = 1)

	(I)	(II)	(III)	(IV)	(V)
Material x2	-0.132	-0.147	-0.134	-0.101	-0.094
	(0.263)	(0.265)	(0.262)	(0.261)	(0.269)
Responsibility	0.011	0.007	0.048	0.040	0.025
	(0.261)	(0.261)	(0.265)	(0.262)	(0.271)
Donations	-0.151	-0.150	-0.159	-0.146	-0.136
	(0.118)	(0.117)	(0.111)	(0.109)	(0.109)
Material x2 x Donations	0.131	0.136	0.139	0.136	0.131
	(0.139)	(0.138)	(0.135)	(0.133)	(0.136)
Responsibility x Donations	0.057	0.060	0.048	0.057	0.072
	(0.145)	(0.147)	(0.141)	(0.140)	(0.144)
Constant	3.491***	3.598***	3.702***	3.253***	1.218***
	(0.218)	(0.240)	(0.485)	(0.500)	(0.335)
Controls (Gender)	No	Yes	Yes	Yes	Yes
Controls (Experimenter)	No	No	Yes	Yes	Yes
Controls (Cognitive Ability)	No	No	No	Yes	Yes
Controls (School)	No	No	No	No	Yes
Observations	621	621	621	621	620
R^2	0.004	0.008	0.032	0.042	0.053

Note: Standard errors in parentheses account for clustering at the class level. The omitted category is Material x1. Asterisks next to coefficients indicate significance at the 10/5/1 percent level.