Düsseldorf Institute
for Competition Economics

# DISCUSSION PAPER

No 203

## Peers or Parents? On Non-Monetary Incentives in Schools

Valentin Wagner,
Gerhard Riener

November 2015

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

d|u|p düsseldorf university press

IMPRINT

DICE DISCUSSION PAPER

Published by

düsseldorf university press (dup) on behalf of
Heinrich-Heine-Universität Düsseldorf, Faculty of Economics,
Düsseldorf Institute for Competition Economics (DICE), Universitätsstraße 1,
40225 Düsseldorf, Germany
www.dice.hhu.de

Editor:

Prof. Dr. Hans-Theo Normann
Düsseldorf Institute for Competition Economics (DICE)
Phone: +49(0) 211-81-15125, e-mail: normann@dice.hhu.de

DICE DISCUSSION PAPER



ISSN 2190-9938 (online) – ISBN 978-3-86304-202-8

# Peers or Parents? On Non-Monetary Incentives in Schools*

Valentin Wagner and Gerhard Riener

November 2015

**Abstract**

This paper presents the result of a field experiment on the effectiveness of non-monetary incentives on pupils' achievement on a mathematical multiple choice test. Our sample consists of 2113 pupils of deprived and high-achieving secondary schools in Germany. Based on a pre-study, we compare the effectiveness of (i) a medal (ii) a letter of praise to the parents and (iii) a delegation of choice over incentives. The effect of non-monetary incentives depends on pupils' socio-economic background. While they constitute a potentially cost-effective and easily implementable method of motivation in Non-High Schools, predetermined non-monetary incentives crowd out intrinsic motivation for pupils in High Schools. In contrast, the endogenous choice of the reward increases pupils' willingness to prepare for the test and mitigates the negative effect of predetermined external rewards in High Schools. Additionally, in the delegation treatment, we find that low-achieving pupils typically choose a reward with a higher signaling value to their parents, independent of the school type.

**Keywords:** Non-monetary incentives, field experiment, education, incentive choice, effort, socio-economic background

**JEL codes:** C93, I20, I21, J1

*Valentin Wagner: Düsseldorf Institute for Competition Economics, wagner@dice.hhu.de. Gerhard Riener: University of Mannheim and Düsseldorf Institute for Competition Economics, riener@dice.hhu.de. We would like to thank the teachers and pupils who participated in the experiment and the organizers of the Känguru-Wettbewerb for providing the test exercises. We are also grateful for comments and advice from Hans-Theo Normann, Damon Clark, Simon Wiederhold, Ludger Wössmann, Dirk Sliwka, Giovanni Peri, Claudia Möllers and participants in seminars at the University of Düsseldorf, the Ifo Center for the Economics of Education, participants in the Competition and Innovation Summerschool 2014, the International Workshop on Applied Economics of Education 2015, the Fourth SOLE/EALE World Conference 2015, the 10th Nordic Conference on Behavioral and Experimental Economics 2015 and the Workshop on Economics of Education 2015 in Mainz. Excellent research assistance was provided by David Heine. The usual disclaimer applies.

# 1 Introduction

Pupils often lack the motivation to study mathematics, although mathematical skills yield a large economic premium and are an important prerequisite for later employment possibilities and wages (Hanushek et al. 2015).[1] Pupils might under-invest in their own mathematical education because they are not aware of their own production function (Cunha and Heckman 2007), they may underestimate the return on education (Oreopoulos 2007; Gneezy, Meier, and Rey-Biel 2011)[2] or are afraid of not being accepted by their peers by performing in a manner that is not consistent with the group's expectations (Akerlof and Kranton 2005). Even if pupils recognize the individual importance of mathematical eduction, it has positive externalities, which may lead to sub-optimal investment. It is thus crucial for educational policy to understand how pupils are motivated to enhance their performance and improve their attitude toward mathematics.

An economist's natural recourse to increase performance is through financial incentives. However, implementing monetary incentives in an educational context entails at least three obstacles: i) it is potentially more cost-intensive than the status quo,[3] ii) there is low acceptance for "cash for grades" by teachers and parents who think that education has value and entails motivation in itself and iii) it raises ethical issues. Moreover, research on financial incentives in schools has revealed mixed results (Fryer 2011; Bettinger 2012; Levitt et al. 2012) as these incentives may crowd out internal curiosity and motivation to acquire new knowledge, which underlines the second concern.

In addition to financial incentives, non-monetary rewards that use public recognition of success may be effective. There is evidence that they have the power to keep up or increase workers' motivation (Kosfeld and Neckermann 2011; Kube, Marechal, and Puppea 2012) and might also work in schools because children are often higher motivated by short-run rewards than less tangible long-run rewards (Chelonis et al. 2004; Bettinger and Slonim 2007).[4] Furthermore, delegating the choice of the incentive to the recipient has been shown to have additional positive effects in experimental labor markets (Charness et al. 2012). Thus, incentives that aim at recognizing the achievement of a student within the class or that inform a student's parents

[1] Hanushek et al. (2015) show in a study across 22 countries that a one standard deviation increase in numeracy skills is associated with an average increase in hourly wages of 17.8% (see also Niederle and Vesterlund [2010]; Goodman [2012]).

[2] Cunha and Heckman (2007) develop a model of skill formation with multiple stages of childhood in which inputs at different stages are complements and self-productivity of investment is present. Due to dynamic complementarity—which means that the marginal productivity of investment depends on the level of skills produced by previous investments—it may be difficult for individuals to know their educational production function. Oreopoulos (2007) evaluates the impact of compulsory schooling on dropout rates. He finds that lifetime wealth increases by about 15% with an extra year of compulsory schooling. According to Oreopoulos (2007), dropouts likely forgo substantial gains to lifetime wealth because adolescents ignore or heavily discount future consequences when deciding to drop out of school. In the 1990 Eurobarometer Youth Survey, more than 50% of 16 to 25-year olds leaving school at the minimum age indicated that their reason for dropping out was lack of interest or that they saw no point in going on.

[3] For example Fryer (2011) distributed a total of $9.4 million (approx. $348.15 per pupil (treated and untreated)) and $650.000 (roughly $385 per treated student) were awarded by Angrist and Lavy (2009). Fryer (2011) tests the effectiveness of financial incentives in Dallas, New York and Chicago. He finds that the incentives offered for educational outputs (such as better grades) are less effective than incentives for educational inputs, such as attendance or reading books. Angrist and Lavy (2009) offered cash awards to students in Israel who passed their exams as part of an attempt to increase certification rates among low-achievers. These cash awards led to an increase in certification rates for girls but not for boys because girls devoted extra time to exam preparation.

[4] A recent literature review by Koch, Nafziger, and Nielsen (2014) offers an overview of other approaches in behavioral economics—such as self-control, willingness to compete, self-confidence and the influence of the environment—which can explain educational investment decisions and outcomes in education.

may provide a simple and cost effective way to circumvent the problems of financial incentives. These types of incentives are accepted and frequently used by teachers (Caffyn 1989)—who are, of course, important stakeholders in the implementation of the policy—and are politically feasible. Furthermore, pupils' empowerment by letting them participate in the learning environment is a positively valued feature.

Thus far, little research has focused on the effectiveness of social recognition of academic merit that may be a viable alternative in the educational sector. However, extrinsic non-monetary incentives do not come entirely free either as there is the danger of hidden costs. First—as with monetary incentives—there is the potential of crowding out intrinsic motivation if rewards are too low-powered or not properly designed (Gneezy and Rustichini 2000). Second, peer group effects may gain importance, as the performance or changes in performance will be made public (Bursztyn and Jensen 2015); thus, depending on the audience, recognition can have ambiguous effects (see also Austen-Smith and Fryer Jr [2005], on "Acting White").

The aim of this paper is to test the effects of a variety of recognitional incentives that differ with respect to the target audience—a medal awarded in front of a student's peers and a letter sent to a student's parents—that might potentially retain some of the power and simultaneously mitigate some of the problems of cash incentives. Therefore, we provide pupils in secondary schools with non-monetary incentives for individual improvement in a mathematical test. Rewards are exogenously determined in two of the treatments (Medal and Letter Treatment), whereas pupils can choose their reward in advance in the Choice Treatment where the choice of the incentive was delegated to the pupils. In particular, we are interested in how non-monetary incentives and delegation interact with gender and socio-economic background. Since reputational effects are grounded in social customs within the classroom and the family environment—and because we do not expect that the effect of recognitional incentives will change in the short or medium run—we focus on short-run effects.[5]

The German school system is characterized by early school tracking, and these tracking choices are highly correlated with pupils' socio-economic background (Dustmann 2004; Ditton 2007; Paulus and Blossfeld 2007).[6] We distinguish between High Schools and Non-High Schools; pupils attending High School belong to families of higher socio-economic status, on average. Our sample consists of younger pupils (fifth and sixth graders) because non-monetary incentives work better than financial rewards for this group, in particular (as has been argued by Levitt et al. [2012], since younger children who are less familiar with cash may be more responsive to non-financial rewards than older students who are more familiar with cash). Moreover, increasing educational inputs in younger ages is promising, as these inputs are likely to complement skill formation in later stages of education (Cunha and Heckman 2007). Our sample of selected schools matches important indicators of school success on the county level of North-Rhine Westphalia.

We present two sets of results. *First*, the overall result in which the effectiveness of non-monetary

[5]Altering the underlying attitudes towards educational achievements would require different type (and potentially more costly) interventions; we therefore will test the immediate applicability of an incentive scheme based on recognition.

[6]School types in Germany differ in their education of teachers because universities offer two different degree programs with different focus areas. Furthermore, students becoming High School teachers typically have a higher High School graduation score.

incentives is shown to depend on school type. Pooling over gender and age, we find that the effect on performance among High School pupils is negative in the Medal and Letter Treatment, while we observe no significant effect on performance for Non-High School pupils. The difference between the school types is significant for the Letter Treatment. Among High School pupils, this negative effect can be mitigated by providing the freedom to choose their incentives. *Second*, the choice of recognitional incentives depends on pupils' ability. We find that low-performing pupils in the Choice Treatment are more likely than high-performing pupils to choose the parents letter and that low performers in High Schools chose more often the letter to the parents than low performers in Non-High Schools. Furthermore, pupils in the Choice Treatment report significantly more often to have prepared for the test than pupils in the Control Treatment.

Although there is ample but mixed evidence on the effectiveness of incentivizing teachers using pupils' performance (Lavy 2002; Springer et al. 2011; Fryer 2013; Muralidharan and Sundararaman 2011),[7] few studies systematically evaluate the effects of incentivizing pupils. Furthermore, the different *level of achievement* regularly achieved at different schools has largely been neglected because previous experimental studies have mainly focused on deprived schools.[8] For developed countries, Fryer (2011) conducted a series of experiments in low-performing urban school districts in Chicago, Dallas and New York City and analyzes the effect of monetary incentives on learning input and testing output.[9] While input incentives (payments for reading a book) have a positive effect on performance, output incentives (payments for test performance) do not seem to be effective. Bettinger (2012) reports on a large scale policy experiment among students in grades three through six that took place in primary schools in Coshocton, Ohio and shows that monetary incentives may have different effects based on subject type, i.e., test scores increased in math but not in other subjects.

Notably, there are only a few experiments that use non-monetary incentives in schools, although it appears that incentives such as recognition are budget-neutral and are appreciated within the pedagogical community. To the best of our knowledge, the field experiment by Levitt et al. (2012) is the first study in schools in which the authors compared a recognitional non-monetary reward—an in-class trophy awarded for good

[7]While teacher incentives in developing countries have shown promising results (Lavy 2002; Muralidharan and Sundararaman 2011), experiments in the US suggest that teacher incentives are ineffective (Springer et al. 2011; Fryer 2013). For an overview of the effectiveness of performance-based pay systems on teachers see Neal (2011).

[8]A notable exception is Angrist and Lavy (2009). Cash awards were provided for low-achieving high school pupils in Israel. However, the sample consisted of 40 nonvocational high schools with the lowest Bargut (matriculation certificate) ratings in a national ranking.

[9]For developing countries, McEwan (2014) provides an overview of 76 experiments and randomized controlled trials (RCTs) conducted in schools from the mid-1970s to 2013 and reports that, on average, monetary grants and deworming had effects on learning outcomes that were statistically insignificant and close to zero. Nutritional treatments, treatments that provided information to parents or students and treatments that improved school management and supervision had small mean effect sizes (0.04–0.06 standard deviations) that were not always robust to controls for study moderators. The largest mean effect sizes included treatments with instructional materials (0.08), computers or instructional technology (0.15), teacher training (0.12), smaller classes, smaller learning groups within classes or ability grouping (0.12), student and teacher performance incentives (0.10) and contract or volunteer teachers (0.10).

Blimpo (2014) analyzes monetary team incentives in Benin and finds that an individual incentive treatment works best with middle-ability pupils and not as well for low- and high-ability pupils. For low-ability pupils, the bonus seems out of range, whereas high-ability pupils do not have to exercise much effort to achieve the bonus, which weakens the bonus payment. Teams that are in competition with other teams yielded the best results perhaps because in such a competitive setting, high-ability pupils are helping low-ability pupils in order to improve the team's average performance and thus the chance of winning the bonus.

Chand, Banerjee, and Bhargab (2008) find that uniforms, scholarships and textbooks are the three most important incentives in Indian schools. In their experiment, incentives are not provided for a specific test score but are based on performance in the previous year, attendance rate and socio-economic background.

performance on a test—to monetary rewards. Levitt et al. (2012) show that this recognition incentive has larger effects than monetary incentives for younger pupils. Furthermore, these authors find that incentives work better when the bonus is paid immediately instead of delayed by a month.[10] Jalava, ter Joensen, and Pellas (2014) is closest to our study; these authors analyze the effect of grading methods (rank-based grading vs. criterion-based grading) in Swedish schools and provide non-financial incentives—a certificate and a prize (refillable pencil). Their findings are comparable to ours as pupils are also in grade six and typically twelve years old. In the criterion-based grading treatment, pupils received grades on an A-F scale based on their performance, whereas in the rank-based grading treatment, the top three performing pupils within a class received a grade of A. Pupils in the "Certificate-Treatment" were promised a certificate if they exceeded the criterion-based score for A-B (18 points or more). An additional treatment rewarded pupils if they were among the top three performing pupils in their class. Jalava, ter Joensen, and Pellas (2014) find that the effectiveness of non-financial incentives differs across the test score distribution and with respect to gender. Boys and girls increase their performance equally in the rank-based grading treatment, but girls also respond strongly to the certificate reward. The non-financial incentives primarily work positively for pupils in the middle quartiles of the ability distribution and crowd out intrinsic motivation for low-ability pupils.

Our contribution to the literature is fourfold. First, we examine the effect of non-monetary incentives on high- and low-achieving school types, which are differentiated by the socio-economic status of their student bodies. Until now, the literature has focused primarily on incentives provided in deprived schools. However, it is of interest to policy makers, teachers and parents to learn how pupils of high-achieving schools react to these incentives.

Second, we contribute to the growing literature on the effects of empowerment on human performance by giving pupils the flexibility and freedom to choose their reward beforehand. Individuals who choose their wage payment exhibit higher performance in experimental labor markets (Charness et al. 2012). Individuals who choose an activity are likely to perform and cooperate better than those who are assigned an activity (Bo, Foster, and Putterman 2010; Sutter, Haigner, and Kocher 2010). Thus far, little is known of the effect on effort when people are free to determine their compensation scheme (see Mellizo, Carpenter, and Matthews [2014] for a recent study). To our knowledge, no study has yet applied this method to the educational sector.

Third, we extend the literature on non-monetary incentives in school by extending and comparing the set of non-monetary incentives. Thus far, a trophy (Levitt et al. 2012), a certificate and a refillable pencil (Jalava,

[10] In a university setting, Chevalier, Dolton, and Lührmann (2014) conducted a controlled field study among first-year undergraduate economics students that varied the incentives rewarding effort on a quiz. Incentives included, inter alia, additional educational material, a book voucher for the top performer or the quiz grade counted for 2.5% or 5% towards the final grade of the course. Chevalier, Dolton, and Lührmann (2014) find that assessment weighting is highly effective in improving quiz participation, which improves performance on the final exam. Each additional quiz improved grades by 0.15 of one standard deviation. Bigoni et al. (2015) tested the effects of non-monetary incentives (extra points on the next exam) on university students (depending on their performance in previous tests, students could earn bonus points for the final exam). They employ both a cooperative and a competitive treatment. In the cooperative treatment, a student's test score was increased by one extra point if her partner's score was sufficiently good. In the competitive treatment, a student's mark in a test was increased by two extra points if her score resulted was higher than her partner's. Although women did not respond to the incentive at all, men—and particularly low-ability men— performed better in the competitive treatment. Bigoni et al. (2015) find no difference between the control and cooperative treatment.

ter Joensen, and Pellas 2014) have been tested. This study is complementary to (Jalava, ter Joensen, and Pellas 2014), as we use an incentive scheme that attempts to avoid crowding out motivation for low-performing pupils in that pupils compete against their past scores and are awarded for self-improvements.

Fourth, we contribute to the literature on the signaling value of rewards. This signaling value is likely to change if the peer group differs in its socio-economic background (Bursztyn and Jensen 2015). Providing non-monetary incentives to pupils in both deprived and high-achieving schools (as differentiated by the socio-economic status of the student bodies), we can examine the correlation between signaling value and socio-economic status. Furthermore, by giving pupils the flexibility and freedom to choose an incentive, we can also analyze to whom—peers or parents—pupils want to signal their investment in education.

The paper is organized as follows. In Section 2 we give background information on the German school system and on the selection of test incentives. Section 3 explains the experimental design and Section 4 presents the data. In Section 5, we present our results, which are discussed in Section 6. Section 7 concludes.

## 2 Background and Selection of Test Incentives

The German school system offers a good setting in which to analyze the impact of non-monetary incentives on performance in different institutional environments because it children are segregated into high- and low-performing groups at the age of ten. We run our experiment on pupils in grades 5 and 6 (age groups ten and eleven). These grade levels serve to test, promote and monitor pupils and to decide in cooperation with parents on the suitability of pupils for the chosen type of school: suitability is assigned with successful promotion after the sixth grade. We provide a detailed description of the German school system in Appendix C.

Secondary school track choice has major effects on subsequent educational achievements and labor market outcomes (Dustmann 2004). It is important to understand the transition process from elementary school to secondary education to recognize how it translates into the social composition of pupils between school types. Peer composition in the classroom is determined by a tracking system, which begins after grade four of elementary school (at the age of 10).

Parental social status has a significant twofold influence on the choice of school type (Gresch, Baumert, and Maaz 2010). First, the social status of parents directly influences school performance in elementary school and hence the transition recommendation. Pupils from families with higher socio-economic status are more likely to be recommended to High School based on their better school performance. Second, parents from a privileged background put more emphasis in sending their children to academically advanced school types than parents with low socio-economic status (see Ditton [2007]; Paulus and Blossfeld [2007]). These parents are also more likely not to follow the recommendation and to enroll their child at a school type of their original choice if they do not receive the desired transition recommendation. For example, Dustmann (2004) shows that parental background is strongly related to children's secondary track choice. Furthermore, Jonkmann et al. (2010) provide a more recent and detailed overview about the dependency between parents' educational

background and children's tracking decision in Germany. They show that approximately 62% of pupils whose parents have the highest school graduation also attend High School. In comparison, approximately 35% of pupils whose parents have middle-level school graduation and only 14% of pupils whose parents have the lowest school graduation attend High School.

## 2.1 Selection of Schools and Multiple-Choice Test

**Schools** Using a list of schools that is publicly available from the Ministry of Education of North Rhine-Westphalia (NRW), we contacted 170 schools in the cities of Bonn, Cologne and Düsseldorf, which represent 9.5% of secondary schools in NRW.[11] Contact was first established via email and posted letter on November 19, 2013. As the average information transfer in school takes about two weeks (according to informal inquiries within schools), we contacted the schools again on December 9, 2013. About 33% of all schools responded, and 50% (28 schools) replied positively and agreed to a preparatory talk. In these talks, the experimental design was explained to at least one teacher per school and lasted about 30 minutes. Finally, 25 schools totaling 89 classes agreed to participate in the experiment.

**Multiple-Choice Test** We received permission to use questions from a mathematics competition test (Känguru) that is administered throughout Germany and in over 50 other countries. The mathematical test consisted of 14 multiple-choice pen-and-paper questions. Pupils were given 30 minutes to answer all the questions so that the test could be taken during a regularly scheduled teaching hour. The problems and the possible choices were presented on three question sheets and pupils received three, four or five points for correct answers, depending on the difficulty level of the questions.[12] There were five answering possibilities with only one correct answer per question, and pupils had to mark their answers on the same sheet. To prevent pupils from answering at random, one point was deducted for a wrong answer and zero points were given for no answer. To minimize cheating (see Jensen et al. [2002]; Behrman et al. [2011]; Armantier and Boly [2013]), we changed the order of questions for pupils within a class. The mathematical problems were a compilation of old questions of the Känguru competition test and differed among school types. We prepared one test for High Schools and another one for Non-High Schools. One test for all school types is not appropriate, as the questions would otherwise be to easy for High School pupils or too difficult for Non-High School pupils. We considerably reduced the length and complexity—particularly the verbal explanations—as many pupils in Non-High Schools have problems understanding lengthy text and lack abstraction capabilities (Retelsdorf and Möller 2008).

To fulfill privacy and data protection requirements, each test and questionnaire received a test identification number, so that pupils did not have to write down their names. This procedure is similar to that of

[11] In the 2012/13 school year, there were 2,018 secondary schools in North Rhine-Westphalia (NRW) with 37,451 school classes and a total of 1,295,741 pupils. Of these pupils, 12.26% attended Secondary General School, 23.07% attended Middle School, 18.95% Comprehensive School and 45.72% High School. The share of foreign pupils or pupils with migration background is as follows: *Secondary General School* 57.46%, *Middle School* 39.84%, *Comprehensive School* 46.02%, and *High School* 20.24%.

[12] There were five questions for three points, five questions for four points and four questions for five points.

evaluations of learning processes that are regularly carried out in various subjects.

## 2.2 Survey and Selection of Test Incentives

What type of non-monetary incentives could potentially work in the German school environment? To answer this question, we conducted a survey before implementing the field experiment in 11 classes in 4 schools with a total of 241 pupils of the same age group. This was a convenience sample gathered through personal contacts. The questionnaire consisted of two parts. On the first page, pupils were asked for three incentives that would motivate them to learn for a test. On the back of the sheet, pupils could mark their choices from a predefined selection. The number of answers was limited to three and pupils were asked to rank the answers (the questionnaire and a complete list of pre-selected incentives can be found in Appendix D). The selection of the reward options in the questionnaire was based on the concepts of Goal Theory,[13] the aspect of work avoidance and social recognition. We categorize these these incentives as follows: (i) *work avoidance* (ii) *mastery* (iii) *social appreciation*, which can be further distinguished in (iiia) *private* and (iiib) *public* (iv) *consumption* and (v) *curiosity*.

We find that pupils prefer *work-avoidance incentives* and *private social appreciation incentives* over *mastery incentives* and *public social appreciation incentives*. Overall, the most frequently chosen incentive was extra points for the next exam, which is consistent with the findings of Chevalier, Dolton, and Lührmann (2014), who show that participation in solving quizzes increased between 40% and 60% when the quiz grade counted (2.5% or 5%) toward the course's final grade. The least-favored incentive of our survey was to receive a certificate, which contrasts (at least in stated preferences) with the findings of Jalava, ter Joensen, and Pellas (2014)—that girls respond strongly to being rewarded with a certificate—indicating that there might be differences among pupils from different cultural backgrounds. Figure 5 in Appendix D presents the top answers.

Based on our survey results, we chose to assess the following incentives: (i) medal, (ii) letter of praise and (iii) a choice of incentives treatment where pupils could choose between (iii) "no-homework" voucher and (iv) surprise gift. The *homework voucher* could be used once during the semester and exempts pupils from homework in math. The *trophy* consisted of a small medal, worth about one Euro, that was awarded in front of the other pupils in the classroom. The *parents-letter* was a pre-formulated letter sent to parents and signed by the teacher praising pupils' performance (see Figures 6, 7 and 8). The *surprise* consisted of the medal plus the parental letter which was not revealed to the students beforehand.

We find small gender differences in the survey. Girls evaluate the *parents letter* slightly higher than boys,

[13] In addition, to set our work in the context of the pedagogical literature we refer to *Goal Theory* which is a widely used concept in pedagogy research. Goal Theory was developed to classify and explain motivation in school (see Ames [1992]) and therefore serves as one source for our survey incentives. The basic idea behind Goal Theory is that there are two main types of motivation: The *Ability Goal Orientation*, i.e., the motivation to be better than classmates and to earn good grades, and *Goal Mastery Orientation*, i.e., the motivation to expand knowledge in one subject and the joy of learning. Goal Theory has been extended by aspects such as *Work Avoidance* (Dowson and McInerney 2001) and *Social Goals* (Urdan and Maehr 1995). There are four types of Social Goals that can be a source of motivation: social recognition, social compliance, social solidarity and social care.

whereas boys evaluated the medal higher than girls. In our study, we did not include *Bonus Points* and *Mobile Phone* because teachers are often not allowed to give extra points for the following exam and the use of mobile phones is prohibited in almost all schools. However, *Bonus Points* might be promising to test in future research. For example, Chevalier, Dolton, and Lührmann (2014) have shown that assessment weighting is highly effective among university students.

***Finding from Survey:*** *Pupils prefer work-avoidance and private social appreciation incentives over mastery goal and public social appreciation incentives.*

# 3 Experimental Intervention

The study was conducted in 25 secondary schools with a total of 89 school classes in Bonn, Cologne and Düsseldorf, cities that are located in the federal state of North Rhine-Westphalia, Germany. During February and March 2014, 2,113 pupils in grades 5 and 6 participated; these students were 11 years old, on average, and 43.49% of the participants were female. There might be some selection on the school level, regarding which schools would participate; however, all the pupils of an included class participated. Therefore, we eliminate the potential sample selection bias that might arise with voluntary participation and self-selection of pupils, who are our main subject of interest.

**Treatments** We designed the following four treatments to analyze the effectiveness of non-monetary incentives and to evaluate the power of choice: the Control Treatment ( *Control*), the Letter Treatment (*Letter*), the Medal Treatment (*Medal*) and the Choice Treatment (*Choice*). The test was announced and the preparatory material was distributed one week in advance for all treatments. During the preparation week, teachers did not actively prepare pupils for the test. Teachers answered questions concerning the preparatory exercises only if pupils asked on their own initiative.

**Reward Conditions** The condition for receiving the reward in all incentivized treatments was an improvement in test grades compared with pupils' last midterm grade. Top-performing pupils who received the highest possible midterm grade received the reward if they did not perform worse. The rationale behind using a relative performance measure is to avoid demotivating low-performing pupils. A criterion-based incentive condition—one in which pupils must score above a predetermined benchmark—might demotivate low-performing pupils because these pupils may believe that the benchmark is not reachable. For example, Jalava, ter Joensen, and Pellas (2014) found that girls near and in the lowest decile were demotivated by a high threshold. The grading system of the test is designed such that the highest performing pupil in a class receives the highest possible grade and others are graded relative to the top performer. This grading scheme ensured that at least one reward was paid per class. Notably, we do *not* consider the difference between the

midterm grade and the test grade as a dependent variable in our later analysis. The test grade does not enter into our analysis, as we use it solely to determine who earns a reward.

**Control Treatment** Pupils in the control group were offered no reward for test performance. For this group, nothing changed from the usual test situation. The test scores of the control group serve as a baseline to estimate the effects of providing non-monetary incentives. The average treatment effect is the difference in the mean test score of each incentivized treatment and the control group.

**Fixed Treatments (Letter & Medal)** In the fixed treatment rewards, one week prior to the test, teachers explained to the pupils that they would earn a reward on the test if they could improve on their last midterm grade and also explained the reward condition and presented the class with a copy of the rewards. It was further explained that grading the test and thus receiving the reward would take one week at most. One week later, on the test day and shortly before the test, teachers reminded their class of the incentive and explained the reward conditions.

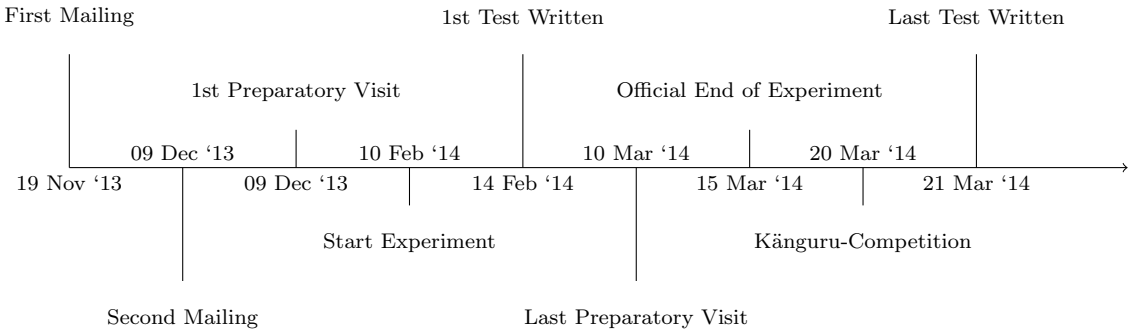
**Choice Treatment** In contrast to the fixed treatment rewards, the pupils taking those tests with the choice treatment rewards were given their choice of incentive beforehand. After announcing the test, each pupil could individually mark one incentive out of four (medal, parental letter, homework voucher and surprise) on a small card. This card was then collected by the teacher and the preparatory materials were distributed. The procedure on the test day remained the same as in the fixed treatments.[14] This treatment was inspired by recent results in real effort experiments, where Mellizo, Carpenter, and Matthews (2014) show that workers that voted to determine their compensation scheme exerted significantly more effort.

**Experimental Procedure** We visited the schools one time during the preliminary stage of the experiment. During this meeting, the exact schedule and expiration of the experiment was described and teachers' questions were answered. Each teacher received the instructions (again) in written form near the start of the experiment. In total, two envelopes at different points in time were sent to the teacher. The first envelope was distributed at the beginning of the experiment (February 10, 2014) and contained instructions regarding the announcement of the test, preparatory material for pupils and copies of the rewards to present in front of the class. The teachers communicated the test data to us via email. Two to three days in advance of the test date, teachers received the second envelope containing the actual tests, instructions for the test day and a list in which teachers entered the midterm grades and the corresponding test-id numbers. Sending the tests in a timely manner was important to reduce the risk that teachers—willingly or unwillingly—prepared pupils. Tests were corrected by the researchers and teachers were asked to answer a questionnaire at the close of the

[14] Since the number of participating schools was restricted, we did not test the "Voucher Treatment" and "Surprise Treatment". The rationale for choosing the "Medal Treatment" is the comparability to the study by Levitt et al. (2012). The "Letter Treatment" was chosen as this can be easily implemented by teachers and policy makers.

experiment as well as to collect all preparation material pupils received which were to be sent back to us.[15] Figure 1 shows a schematic timeline for the experiment.

Figure 1: Time-line




Our aim was to maintain a natural exam situation within the classroom. Therefore, the tests took place in regularly scheduled classes in which teachers were free to choose the test date during a predetermined period (February 10th – March 15th). In this manner, teachers could choose a suitable testing week in which no other class test was scheduled for which pupils had to study. Furthermore, we had to evaluate the trade-off between a potential loss of control and increased external validity for our results. We opted for the latter, and the experiment was conducted solely by the teachers as we did not want to change the natural class environment and thereby induce experimenter demand effects (see Zizzo [2010] for a discussion of experimenter demand effects). This would have seriously challenged the internal and external validity of our results. Thus, pupils were unaware that the test was part of an experiment.[16]

The test was announced one week in advance and teachers explained the bonus scheme in the event that the class had been assigned to an incentive treatment. In the same lesson, pupils received preparatory questions with attached solutions. Notably, this preparatory material did not prepare pupils with respect to the content of the curriculum but was instead intended to prepare pupils for the (multiple-choice) format of the test. In Section 6, we analyze the impact of preparation on pupils' achievement on the test in greater detail. We find that pupils who are significantly more likely to prepare for the test do not perform significantly better on the test. Thus, a difference in pupils' test achievement would not be the result of exerting more effort for test preparation.

The teachers clarified that pupils will be evaluated and graded and that test grades do not count for the school report. They did so in the framework of an evaluation of pupils' achievements that demonstrate their

[15] We intended to collect the preparation material to measure the impact of non-monetary incentives on the motivation to prepare for the test. However, a sensible analysis was not feasible because less than 50 percent of the preparation material was sent back.

[16] According to Zizzo (2010), experimenter demand effects are typically a problem only when they are positively correlated with the true experimental objectives' predictions. In our experiment, this would be the case if an unknown (external) person would have offered pupils a reward (or rewards) for good performance. However, if researchers were never present in the classroom, the pupils' natural environment would remain unchanged because teachers typically try to motivate pupils to increase their efforts in school. Thus we simply changed the way that teachers motivated pupils but not their objective, i.e., improvement in performance.

skills during a school year. Before the test started, teachers read the following text aloud in the classroom:

> *"The test contains a total of 14 tasks that must be solved within 30 minutes. For each task, there are 4 wrong and 1 correct answers. There are tasks that are worth 3 points for each correct answer, and others that are worth 4 or 5 points. If an incorrect answer is written, 1 point is deducted. If no answer is given, you receive 0 points. Calculators are not allowed, but "scratch paper" for sketches and small calculations are allowed, of course!"*

Pupils then had 30 minutes to answer all the test questions and a questionnaire that was attached to the end of the test. The tests were corrected centrally by the researchers, and the pupils received their rewards one week later.

**Randomization procedure** Randomization was performed using a classroom-based block randomization design (see Duflo, Glennerster, and Kremer [2007]; Bruhn and McKenzie [2009] regarding the rationale for the use of randomization). As there are at least three classes in almost every school, except for three Secondary General Schools that had only two classes. Our treatment assignment procedure ensures that the Control, Choice and at least one of the Fixed Treatments (Medal and/or Letter) was implemented in each school. The Medal and Letter Treatment was implemented simultaneously in schools with more than three classes. Table 12 in Appendix A.1 shows the randomization of treatments over all school types and reports average points by treatment group for the full sample and for boys and girls separately. Tables 8–11 in Appendix A.1 report the summary statistics by school type and treatment group for all variables we will use as controls in our analysis. On average, all of these variables do not differ from the control group at conventional levels of statistical significance, which indicates that the randomization procedure was successful. On average, subjects in our sample are 11.16 years old and have 0.92 older siblings. 43.49% of the subjects are female and 58.17% speak only German at home, while 37.59% speak another language and 4.24% speak two languages at home. The average midterm grade in mathematics is 2.86 on a scale from 1 to 6, where 1 is the highest and 6 is the lowest grade.

# 4 Data and Descriptive Statistics

Our primary variable of interest is the number of points scored on the test. Our identification of the average treatment effects on the test score relies on our block randomization strategy. We therefore compare test scores of pupils in the treatment groups to pupils in the control group.[17] The most important control variable is pupils' last midterm grade. The last midterm grades are reported by teachers and available for almost all pupils. Midterm grades in Germany combine the written and verbal performance of pupils wherein the written part has a larger influence on the final grade; thus, these grades are therefore a good measure of

[17] Remember, we do *not* consider the difference between the midterm grade and the test grade as a dependent variable. The test grade does not enter into our analysis, as we use it solely to determine who earns a reward.

math ability. Importantly, the midterm grades can be treated as exogenous in our analysis because they were given to pupils before teachers learned about our experiment.

Additional control variables on pupil level are gender, parents' education and a dummy for whether pupils are in grade 5 or 6. The latter variable controls for pupils'age and educational level at the same time. Parents' educational level is captured by the number of books at home (see Woessmann [2005]; Fuchs and Woessmann [2008] for an application in PISA studies).

We also include control variables at the classroom-level: teacher gender, teacher's working experience and the share of German-speaking pupils within a class. It seems that there is a common understanding in the literature that unobserved teacher characteristics may be more important than observed characteristics. However, among the observable teacher characteristics, many studies find a positive effect of teachers' experience on pupils achievement, (see Mueller [2013] for a literature review). The influence of teacher's gender on pupils (math) performance has been investigated by Carrell, Page, and West (2010), for example, who find that the professor's gender has little impact on male students but a powerful effect on female students' performance in math. As classes are closed entities with in-part strong peer effects, ethnic and gender composition might have an influence on pupils' performance (see, for example, Jensen and Rasmussen [2011]; Ohinata and Van Ours [2013]). Thus, to control for ethnic and gender composition effects we include the share of German-speaking pupils in the analysis.

At the school level, we collected data on the unemployment rate within the school district to control for school district specific effects. We also control for the fact that classes within a school took the test on different days. Therefore, the number of days between the test and the first test written in the respective school is controlled for.

Table 1 compares the descriptive statistics to the actual data in NRW. Although we cannot claim representativeness of our sample for the school population in NRW, our data are consistent with key school indicators from NRW. We included 2.067 observations in our analysis. 46 observations were dropped because of missing values.[18]

[18]Missing values were the result of incomplete pupil questionnaires. There are 23 missing values for the last midterm grade and 23 for pupils' gender.

Table 1: Comparison of important indicators: Experiment vs. North Rhine-Westphalia (in %)

| | *Experimental Data* | *North Rhine-Westphalia* |
|---|---|---|
| *A. non-High School* | | |
| Proportion Female | 44.63 | 47.39 |
| Proportion Pupil German | 57.45 | 50.11 |
| Class size | 26.08 | 25.37 |
| Proportion Teacher Female | 66.59 | 65.37 |
| *B. High School* | | |
| Proportion Female | 41.99 | 51.73 |
| Proportion Pupil German | 79.74 | 72.46 |
| Class size | 26.83 | 27.10 |
| Proportion Teacher Female | 55.15 | 59.16 |

*Note:* This table presents characteristics of the sample in the experiment by school type and compares it with the same indicators in North Rhine-Westphalia. The cell entries present the percentage shares of key school indicators.NRW school data are taken from the official statistical report of the ministry of education for the school year 2014/2015 (see `https://www.schulministerium.nrw.de/docs/bp/Ministerium/Service/Schulstatistik/Amtliche-Schuldaten/StatTelegramm2014.pdf`)

# 5 Results

## 5.1 Econometric Model

For purposes of estimating treatment effects on incentivized performance, we will apply negative binomial models. Negative binomial models are closely related to the Poisson models that are frequently used for count data, but negative binomial models do not require the restrictive assumption of unitary variance. As our data show a significant degree of overdispersion (Non-High School: $\ln \alpha = -2.004$, p-value $< 0.001$; High School: $\ln \alpha = -2.636$, p-value $= 0.001$), the negative binomial provides a basis for a more efficient estimation. We control for pupil, class and school variables described earlier and include school fixed effects. Furthermore, there has not been a change of teacher between the midterm grade and the test. Standard errors are clustered on classroom level—which is the level of randomization. We distinguish between High Schools and Non-High Schools as our main categorization of interest. The group of High Schools consists

of the German Gymnasium whereas Comprehensive, Middle and Secondary Schools belong to the group of Non-High Schools. We estimate the models separately for High Schools and Non-High Schools and allow for school fixed effects. Furthermore, our results are robust to multiple testing—linking equations by seemingly unrelated estimations[19] This leads us to the following negative binomial model:

$$\begin{aligned} E(points_i) &= m(\beta_0 + \beta_1\ Treat_i + \beta_2\ School\ Level_i + \beta_3\ Midterm_i \\ &\quad +\gamma P_i + \mu C_i + \alpha S_i + \delta School_i) \end{aligned} \tag{1}$$

$m(\cdot)$ is the mean function of the negative binomial model. $points_i$ is the number of points achieved on the test by pupil $i$, $Treat$ indicates the respective treatment, *School Level* indicates whether pupils are in grade 5 or grade 6, $Midterm$ is the grade in math on the last semester report, $P$ is a vector of pupil-level characteristics, $C$ a vector of class-level covariates, $S$ captures school-district-specific variables and $school_i$ controls for school fixed effects. As a robustness check, we estimated a linear model (OLS) using the same covariates, and the results do not change in either significance or size (see Appendix A.2).

## 5.2 Incentives and test performance

Table 2 presents estimates of the average treatment effects for High Schools and Non-High Schools. The dependent variable is the number of points received in the test. The average treatment effect is the difference in the mean test score of each treatment and the control group. First, we report on the average treatment effect of the Choice, Medal and Letter Treatments over all school levels. Subsequently, we report in more detail on the average treatment effects for boys and girls.

Table 2 reports the averages over treatments across school types. A special focus is on the Choice Treatment because this is the first study that evaluates the flexibility and freedom of choice on a set of permissible incentives in the educational sector. At first view, the results from Mellizo, Carpenter, and Matthews (2014) do not seem to extend to the educational sector. Although the coefficients are positive for pupils in the Choice Treatment in High Schools and Non-High Schools, they are small in High Schools (0.091, p = 0.920) and only slightly larger for those pupils in the Non-High Schools (1.109, p = 0.360) and nonetheless statistically insignificant.

Comparing school types, treatments work in opposite directions. In the Medal and Letter Treatment—pooling fifth and sixth graders—there are no significant differences in Non-High Schools but significant negative effects in High Schools (Medal: -2.006, p = 0.033; Letter: -2.586, p = 0.058).[20] We summarize these

[19] Seemingly unrelated estimation combines the parameter estimates, the variance and covariate variances of the separately estimated equations into a robust single parameter-vector and simultaneous variance covariance matrix. The advantage of seemingly unrelated estimations is the robustness to cross-equation correlation and between group heteroskedasticity; consequently, it can overcome the problem of multiple testing.

[20] Ordinary least square estimation on grade improvement—difference between midterm grade and grade in test—shows similar results. The Choice and Medal Treatments in Non-High Schools have (insignificant) positive coefficients—pupils received

findings in our first result:

**Result 1.** *Pooling over gender and grades, we find that pupils in High School significantly reduce their performance when incentives are exogenously predetermined (*Medal *and* Letter *treatments).*

Table 2: Treatment Effects

| | OLS | | | | Negative Binomial | | | |
|---|---|---|---|---|---|---|---|---|
| | *Non-High School* | | *High School* | | *Non-High School* | | *High School* | |
| *Treatments* | | | | | | | | |
| Choice | 1.109 | [1.128] | 0.087 | [0.869] | 1.109 | [1.211] | 0.091 | [0.907] |
| Medal | 0.597 | [0.991] | −1.709* | [0.918] | 0.362 | [1.143] | −2.006** | [0.943] |
| Letter | 0.941 | [1.192] | −2.262 | [1.523] | 0.867 | [1.219] | −2.586* | [1.364] |
| *Controls* | | | | | | | | |
| Pupil Covariates | Yes | | Yes | | Yes | | Yes | |
| Class/School Covariates | Yes | | Yes | | Yes | | Yes | |
| School FE | Yes | | Yes | | Yes | | Yes | |
| *N* | 1198 | | 869 | | 1198 | | 869 | |

*Note:* This table compares the result of a linear and negative binomial regression separately for High Schools and Non-High Schools including school fixed effects. Dependent variable: points in test. Covariates: last midterm grade, gender, number of books at home, academic year (grade 5 or 6), teacher' working experience (in years), teachers' gender, day differences between tests, unemployment rate of the school district and the proportion of German speaking pupils within the class. Standard errors are reported in parentheses and clustered on classroom-level. The number of clusters is 53 in Non-High Schools and 36 in High Schools. Robustness checks with multiple testing—seemingly unrelated regressions—show similar results. * p<0.10, ** p<0.05, *** p<0.01

As discussed in the Introduction, stereotype threats and non-conformity to role behavior may cause girls not to excel under incentives that emphasize personal achievement in mathematics. Table 3 reports average treatment effects for boys and girls by school type controlling for pupil, class and school covariates as well as school fixed effects. We find significant gender differences in the reaction to incentives for pupils in the Letter Treatment High Schools. Surprisingly, the test performance of boys significantly decreases in the Letter Treatment (-3.661, p = 0.005), whereas this decrease is not statistically significant for girls. In the Non-High School sample, the coefficients of non-monetary incentives have positive signs but are insignificant at conventional levels for boys and girls. We summarize this in our second result:

**Result 2.** *Non-monetary incentive have no heterogeneous gender effects in Non-High Schools. The letter of praise sent to parents is detrimental to the test performance of High School boys.*

better grades in the test compared to their midterm grade—and the Letter Treatment has a (insignificant) negative coefficient. Treatments in High School have negative and significant effects in the Medal and Letter Treatment. Overall, the percentage of pupils who improved their grade and hence received a reward in the incentivized treatments, is as follows: (i) Non-High School: Control 25.00%, Choice 25.44%, Medal 33.50%, Letter 17.00% (ii) High School: Control 29.88%, Choice 19.35%, Medal 16.96%, Letter 18.50%.

Table 3: Treatment effects by gender

| | *Non-High School* | | *High School* | |
|---|---|---|---|---|
| *Males* | | | | |
| Choice | 1.129 | [1.347] | −0.186 | [1.317] |
| Medal | 0.055 | [1.112] | −1.073 | [1.201] |
| Letter | 0.924 | [1.255] | −3.661*** | [1.305] |
| *N* | 665 | | 504 | |
| *Females* | | | | |
| Choice | 1.209 | [1.326] | 0.704 | [1.303] |
| Medal | 1.043 | [1.546] | −2.707 | [1.759] |
| Letter | 0.929 | [1.380] | −1.136 | [2.011] |
| *N* | 533 | | 365 | |
| *Controls* | | | | |
| Pupil Covariates | Yes | | Yes | |
| Class/School Covariates | Yes | | Yes | |
| School FE | Yes | | Yes | |

*Note:* This table reports the result of a negative binomial regression separately for boys and girls and for High Schools and Non-High Schools including school fixed effects. Dependent variable: points in test. Covariates: last midterm grade, number of books at home, academic year (grade 5 or 6), teacher' working experience (in years), teachers' gender, day differences between tests, unemployment rate of the school district and the proportion of German speaking pupils within the class. Standard errors are reported in parentheses and clustered on classroom-level. The number of clusters is 53 in Non-High Schools and 36 in High Schools. * p<0.10, ** p<0.05, *** p<0.01

**What might lead to the crowding out of motivation?** Notably, treatment effects are significantly negative in the Medal and Letter Treatment in High Schools. One reading of this result is that the incentives provided are too low-powered for pupils from higher socio-economic backgrounds. In the pupil questionnaire, we asked students how much the respective reward motivated them on a 1 (not at all) to 5 (very much) scale to verify whether the external rewards differ in their motivational power. We find that pupils in High Schools state a lower motivation than pupils in Non-High Schools in the Choice (3.587 v. 3.340, p = 0.002), Medal (3.473 v. 2.971, p < 0.001) and Letter Treatment (3.526 v. 2.714, p < 0.001). It is also conceivable that pupils in High School are more likely to be monetarily rewarded by parents for good grades. In Germany, parents from the highest tercile of the income distribution spent a monthly average of Euro 160, middle tercile: Euro 101 and bottom tercile: Euro 74 on the school education of their children.[21] However, this

[21]These figures are available at `http://www.vodafone-stiftung.de/ideen_foerdern_publikationen.html?&tx_newsjson_pi1[showUid]=30&cHash=e2270fb5104907e5c3be9121af72e237` (accessed November 12, 2015 )

cannot explain a *decrease* in performance in High Schools instead of just a smaller increase, as pupils in High Schools on average report that they are more than "not at all" motivated by the prizes.

Another explanation might be that reputational motivation differs with socio-economic background. Bénabou and Tirole (2006) assume that an agents' pro-social or antisocial behavior reflects an endogenous and unobservable mix of *intrinsic*, *extrinsic* and *reputational* motivations, whereas in our context prosocial behavior is the effort in school. The authors show that extrinsic rewards can result in a crowding out of the reputational value of good deeds because they create doubts about intentions, i.e., to what extent was performance increased for the incentive rather than for yourself?

There is evidence that the value for education and hence the reputation associated with academic achievements differs with socio-economic background. In a meta-study, Sirin (2005) shows that children from lower income families receive less parental attention in educational matters than their higher income classmates. According to Dwyer and Hecht (1992), one reason for low parental involvement in the education of their children might be a negative parental attitude. Parents who were never very successful in school or for whom school was a traumatic experience might not send a positive message to their children regarding the importance of education. Hence, an extrinsic non-monetary reward given by teachers could give pupils the recognition that they do not receive otherwise or signal to parents that they should praise their child. In terms of the Bénabou and Tirole (2006) model, (positive) extrinsic motivation exceeds the amount of (negative) reputational motivation. Furthermore, Avvisati et al. (2014) shows that motivating parents to become involved in their children's education can change pupils' behavior. Particularly with parental involvement, pupils developed more positive behavior and attitudes in school, notably in terms of truancy and disciplinary sanctions (see also Bergman [2014]).

### 5.3 Peers or Parents? On the Selection of Incentives

We now turn to the analysis of the Choice treatment. Classrooms are closed entities with in-part strong peer effects that may determine a child's behavior and attitude toward learning and effort (Carrell, Fullerton, and West 2009; Kremer, Duflo, and Dupas 2011; Sacerdote 2011). Giving pupils the possibility to choose their incentive beforehand can shed light on the question to whom pupils primarily want to reveal their educational performance and whether the freedom of choice translates into improved performance.

**Choice of incentives** Table 4 reports selected incentives by the midterm grade and the school type. Cell entries represent the share of pupils with the same midterm grade who have chosen the corresponding reward. Overall, pooling High School and Non-High School students, we observe that the surprise is very popular. Approximately one-third of all pupils chose the surprise as their desired reward. One explanation might be the high degree of curiosity among younger children (see Loewenstein [1994] for a psychological perspective of curiosity). Alternatively, pupils may simply dislike the remaining rewards. More importantly, the share of pupils who chose the medal or the voucher is decreasing in the midterm grade. In other words, high-achieving

pupils tend to be more likely to choose an incentive that has a signaling value to their peers. By contrast, the share of pupils who chose the parents letter is increasing in the midterm grade, which indicates that low-ability pupils chose a reward with a signaling value to parents. This is an indication that recognition by parents might be of great importance for those pupils.

Table 4: Chosen Incentives by Midterm Grade (in percent)

| Midterm Grade | *Medal* | *Letter* | *Voucher* | *Surprise* |
|---|---|---|---|---|
| *Panel A: Pooled (N=676)* | | | | |
| 1 | 32.00 | 8.00 | 26.00 | 34.00 |
| 2 | 19.81 | 19.32 | 24.64 | 36.23 |
| 3 | 20.48 | 28.11 | 19.28 | 32.13 |
| 4 | 15.15 | 37.88 | 18.18 | 28.79 |
| 5 | 10.35 | 24.11 | 15.79 | 31.58 |
| *Total* | 19.53 | 26.63 | 21.01 | 32.84 |
| *Panel B: Non-High School (N=399)* | | | | |
| 1 | 36.84 | 5.26 | 26.32 | 31.58 |
| 2 | 20.21 | 19.15 | 22.34 | 38.30 |
| 3 | 17.50 | 29.38 | 18.75 | 34.38 |
| 4 | 14.29 | 30.61 | 19.39 | 35.71 |
| 5 | 14.29 | 35.71 | 17.86 | 32.14 |
| *Total* | 18.05 | 26.57 | 20.05 | 35.34 |
| *Panel C: High School (N=277)* | | | | |
| 1 | 29.03 | 9.68 | 25.81 | 35.48 |
| 2 | 19.47 | 19.47 | 26.55 | 34.51 |
| 3 | 25.84 | 25.84 | 20.22 | 28.09 |
| 4 | 17.65 | 58.82 | 14.71 | 8.82 |
| 5 | 0.00 | 60.00 | 10.00 | 30.00 |
| *Total* | 21.66 | 26.71 | 22.38 | 29.24 |

*Note:* This table presents the choice of reward of pupils in the Choice Treatment. Cell entries present percentages. Panel A shows the result pooled over school types; panel B presents the choice of pupils in Non-High Schools and panel C the choice of pupils in High Schools. Within the German school system, 1 is the highest and 6 is the lowest possible. We do not report on the choice of pupils having the lowest midterm grade because we only have two observations in that group.

Estimating the multinomial logistic regression on the incentive selection distinguished by school types allows us to test whether choices indeed differ across high- and low-ability performers (Table 5). We find that low-ability pupils choose significantly more often the letter—the reward with a signaling value to their parents in High Schools (0.148, p = 0.002) as well as Non-High School (0.080, p = 0.014). The medal with its signaling value to peers seems to be chosen—although not significantly—more frequently by high

ability pupils in Non-High Schools (-0.076, p = 0.168) and by high-ability pupils in High Schools (-0.062, p = 0.280). Furthermore, low-ability pupils in High Schools are less likely to choose the homework voucher (-0.086, p = 0.021) and the surprise (-0.117, p = 0.036).

Comparing the choice of incentives by school type, we find small and insignificant differences between High Schools and Non-High Schools for high and middle performing pupils. However, pupils with midterm grade 4 in High School chose different incentives than their counterparts in Non-High Schools. While 58.82% of those pupils in High Schools chose the parents letter and only 8.82% chose the surprise, 30.61% of lower performing pupils in Non-High Schools chose the parents letter but 35.71% chose the surprise. Furthermore, as Table 19 in Appendix A.6 shows, we find no gender differences in the choice of incentives when we distinguish by school type and ability level.

**Chosen incentive and test performance** We can now examine the correlations between chosen incentives and test performance. Table 6 presents the estimates of the average effects on test performance for each reward in the Choice Treatment as well as the two Fixed Treatments. In Non-High Schools, we find a large positive and significant effect for those pupils who chose the Letter reward (2.640, p = 0.085). In High School, we find no statistically significant relationship for any of the chosen incentives. However, the difference between the positive coefficient of the "Chosen Medal" and the negative effect of the "Fixed Medal" is significant (chi2 = 5.50, p = 0.019).

**Result 3.** *Pupils who chose the letter significantly had a significantly higher test performance in Non-High Schools. When free to choose the medal in High School there crowding out vanishes.*

Table 5: Multinomial Logit Model of Chosen Incentives

| | *Pooled* | | *Non-High School* | | *High School* | |
|---|---|---|---|---|---|---|
| **A. Medal** | | | | | | |
| Midterm grade | −0.069* | [0.040] | −0.076 | [0.055] | −0.062 | [0.057] |
| Grade 6 | −0.230 | [0.508] | 0.410 | [0.719] | −2.343** | [1.026] |
| Female pupil | −0.186 | [0.222] | −0.213 | [0.287] | 0.084 | [0.316] |
| *Books at home* | | | | | | |
| (11-25) | −0.533 | [0.344] | −0.447 | [0.371] | −0.363 | [0.657] |
| (26-100) | −0.382 | [0.284] | −0.387 | [0.313] | −0.578 | [0.763] |
| (101-200) | −0.439 | [0.437] | −0.636 | [0.767] | −0.419 | [0.896] |
| (201-500) | −0.562 | [0.384] | 0.495 | [0.477] | −1.616 | [1.039] |
| (over 500) | 0.086 | [0.428] | 0.394 | [0.606] | −0.421 | [1.028] |
| (Not Reported) | −0.739* | [0.446] | −0.470 | [0.649] | −0.912 | [0.862] |
| Teacher experience (years) | 0.022 | [0.026] | 0.034 | [0.033] | −0.019 | [0.036] |
| Day difference | 0.023 | [0.030] | 0.007 | [0.038] | 0.190*** | [0.057] |
| Teacher female | 0.173 | [0.498] | 0.490 | [0.710] | −1.181 | [1.003] |
| Unemployment | 0.062 | [0.072] | 0.104 | [0.098] | −0.086 | [0.089] |
| Proportion German | −0.952 | [1.000] | −2.527 | [2.083] | −3.337* | [1.806] |
| Constant | −0.693 | [2.971] | −4.116 | [4.264] | 14.40** | [6.986] |
| **B. Letter** | | | | | | |
| Midterm grade | 0.093*** | [0.029] | 0.080** | [0.033] | 0.148*** | [0.047] |
| Grade 6 | 0.462 | [0.546] | 0.672 | [0.745] | −0.736 | [1.288] |
| Female pupil | 0.110 | [0.208] | 0.172 | [0.294] | 0.063 | [0.326] |
| *Books at home* | | | | | | |
| (11-25) | 0.145 | [0.289] | 0.154 | [0.336] | 0.037 | [0.357] |
| (26-100) | 0.255 | [0.335] | 0.026 | [0.341] | 0.368 | [0.461] |
| (101-200) | 0.273 | [0.428] | 0.305 | [0.532] | 0.181 | [0.599] |
| (201-500) | 0.623 | [0.552] | 0.196 | [0.628] | 0.478 | [0.848] |
| (over 500) | 0.299 | [0.523] | 0.821 | [0.681] | −0.066 | [0.731] |
| (Not Reported) | −0.012 | [0.488] | −0.228 | [0.581] | 0.386 | [0.732] |
| Teacher experience (years) | −0.012 | [0.025] | 0.002 | [0.032] | −0.038 | [0.047] |
| Day difference | 0.019 | [0.032] | 0.006 | [0.043] | 0.119* | [0.063] |
| Teacher female | 0.660 | [0.621] | 1.475* | [0.894] | −0.572 | [1.187] |
| Unemployment | 0.033 | [0.075] | 0.010 | [0.102] | −0.051 | [0.093] |
| Proportion German | −0.969 | [1.002] | −1.706 | [1.820] | −2.435 | [1.686] |
| Constant | −5.738 | [3.736] | −6.925 | [5.286] | 3.077 | [9.063] |

*Note:* This table presents the results of a multinomial logit model on the choice of incentive of pupils in the Choice Treatment (results of the multinomial logit model for the voucher and surprise are reported in Table 20 in Appendix A.6). The pupils which were not allocated to the Choice Treatment represent the baseline. Midterm grade is the variable of interest, a positive coefficient shows that low performing pupils are more likely to chose the reward as a high midterm grade resembles low performance in the German school system. A negative coefficient shows that high performers are more likely to chose the respective incentive. Covariates: last midterm grade, number of books at home, academic year (grade 5 or 6), gender, teacher' working experience (in years), teachers' gender, day differences between tests, unemployment rate of the school district and the proportion of German speaking pupils within the class.The number of observation is 2.067 for the pooled specification, 869 for High School and 1.098 for Non-High Schools. Standard errors are reported in parentheses and clustered on classroom-level. The number of clusters is 53 in Non-High Schools and 36 in High Schools. * p<0.10, ** p<0.05, *** p<0.01.

Table 6: Chosen incentives and points in test

| | *Non-High School* | | *High School* | |
|---|---|---|---|---|
| *Treatments* | | | | |
| Medal_Chosen | 1.745 | [1.887] | 0.545 | [0.913] |
| Letter_Chosen | 2.640* | [1.532] | −0.446 | [1.552] |
| Voucher_Chosen | −0.650 | [1.572] | −0.970 | [1.007] |
| Surprise_Chosen | 0.873 | [1.439] | 1.160 | [1.311] |
| Medal_Fixed | 0.419 | [1.123] | −1.932** | [0.943] |
| Letter_Fixed | 0.973 | [1.122] | −2.547* | [1.358] |
| *Controls* | | | | |
| Pupil Covariates | Yes | | Yes | |
| Class/School Covariates | Yes | | Yes | |
| School FE | Yes | | Yes | |
| *N* | 1198 | | 869 | |

*Note:* This table reports the result of a negative binomial regression for each incentive in the Choice Treatment and the incentives in the Fixed Treatments separately for High Schools and Non-High Schools including school fixed effects. Dependent variable: points in test. Covariates: last midterm grade, gender, number of books at home, academic year (grade 5 or 6), teacher' working experience (in years), teachers' gender, day differences between tests, unemployment rate of the school district and the proportion of German speaking pupils within the class. Standard errors are reported in parentheses and clustered on classroom-level. The number of clusters is 53 in Non-High Schools and 36 in High Schools. * p<0.10, ** p<0.05, *** p<0.01

Our Choice Treatment is closely related to a recent study by Bursztyn and Jensen (2015), who analyze peer effects in one natural and one field experiment when performance or investment in education is either observable or kept private. In the natural experiment, top students' performance declined by about 40 percentage points when revealed to the class, whereas lower performing students improved slightly.[22] In their field experiment, Bursztyn and Jensen (2015) find that investment in education depends on to whom the investment decision would be revealed. Students were offered complimentary access to an online SAT preparatory course and students' decision was either kept private or revealed to classmates. Students in honors classes were more likely to sign up for the preparatory course when the decision was made public rather than kept private, while students in non-honors classes were less likely to sign up if the decision was made public.[23]

As the medal in our study has a strong signaling value within the peer group, pupils' performances in the Medal Treatment can be compared to pupils' performance in the natural experiment of Bursztyn and Jensen

[22]On average, performance declined by 24 percent. The names of the top three scorers in the class were displayed on leaderboards.

[23]Students taking both honors and non-honors classes were 15 percentage points less likely in non-honors classes to sign up if the decision was public rather than private but were 8 percentage points more likely in honors classes to sign up if the decision was public.

(2015). They find that high-ability pupils' performance worsened, whereas low-ability pupils improved their performance. This indicates that high performers try to mimic low performance and low performers try to mimic high performers. We find contrary results for high and low achievers in our study (see Table 15 in Appendix A.3). Low performers in the Medal Treatment of all school types do not significantly increase their performance, whereas high performers do not significantly decrease their performance. These findings are in line with the theoretical model of Austen-Smith and Fryer Jr (2005) who show that in the single-audience case there is a separating equilibrium. Notably, we find that middle-ability pupils significantly decrease their performance—thus mimicking low-ability pupils instead of high-ability pupils.

The audience to whom performance is revealed was predetermined in Bursztyn and Jensen (2015). However, as there are at least three parties involved in pupils' educational achievement—peers, parents and teachers—it is of great interest not only to analyze the impact of awards on performance but also to whom pupils might want to signal good performance when they are free to do so? On the basis of the incentives in the Choice Treatment, we find that low-achieving pupils typically choose a reward with signaling value to their parents.

# 6 Further explanations and results

## Do treatment effects result due to increased test preparation or greater effort in the test?

It is difficult to disentangle whether improvements in educational outcomes are the result of increased efforts in studying the subject or the result of higher effort in solving test questions on the test day. The experimental design by Levitt et al. (2012)—incentives are announced immediately before the test with no advance notice—is one of the few that isolates the latter effect by not giving time to prepare for the test. These authors are therefore can attribute the incentive effects to greater short-run effort. Providing incentives without advance notice was not possible in our study because our aim was to analyze the effect of giving pupils the flexibility and freedom to choose an incentive. The choice of the preparatory material—old versions of the Känguru competition test—nevertheless allows us to isolate the effects of short-run effort. It is unlikely that pupils gained knowledge of the subject matter by solving the preparatory material. The material was designed *not* to prepare pupils with respect to the content of the curriculum but to familiarize them with the test format.

As we have previously shown, we find heterogeneous effects of the incentivized treatments on performance. In the pupil questionnaire, we asked pupils to state whether they prepared for the test using the provided material. Accordingly—if there is a learning effect—these incentivized pupils should also have prepared more often than pupils in the control treatment.

Table 17 in Appendix A.4 presents estimates of a logistic regression. The dependent dummy variable is whether pupils (self-reported) prepared for the test. We control for pupils' gender, school level, midterm grade, whether pupils like math (measured on a 1–5 scale) and school fixed effects.

We find that in all school types, pupils' willingness to prepare for the test in the incentivized treatments is higher than for pupils in the control group. The only exception are fifth graders of Non-High Schools in the Letter Treatment (-0.095, p = 0.301), which are less likely to prepare for the test than the control group. Overall, we find that pupils in the Choice Treatment, in particular, significantly increased the time they spent on preparation. The results are significant for pupils (grade 5 and 6) in High Schools in the Choice (grade 5: 0.136, p = 0.007; grade 6: 0.173, p = 0.003) and Letter Treatment (grade 5: 0.075, p = 0.045; grade 6: 0.166, p < 0.001) and for fifth graders of Non-High Schools in the Choice (0.208, p = 0.011) and Medal Treatment (0.186, p = 0.067).

We can now compare pupils' willingness to prepare with their actual test performance. Those pupils whose willingness to prepare for the test is positive do not gain significantly more points in the test. Furthermore, those who significantly improved or decreased performance—compared to the control group—have not prepared significantly more or less—compared to the control group—for the test. These results are an indicator that there is indeed no direct link between test preparation and test performance.

## Do external rewards change pupils risk preferences?

Answering multiple-choice questions without knowing the answer is a risky decision when points are deducted for a wrong answer. As we kept track of pupils' answering decision—wrong, omitted, correct—for each test question, we can analyze differences in risk taking between pupils from different socio-economic backgrounds (school types) and whether the provision of an external incentive leads to a shift in risk taking.

There is evidence that risk-aversion differs by socio-economic background. In a recent study, Deckers et al. (2015) analyzes how socio-economic status simultaneously shapes time preferences, risk preferences, social preferences, and IQ in one coherent framework. Children's risk preferences were measured with a coin flipping experiment in which children had to choose between a "sure coin" and a "lottery coin". Deckers et al. (2015) find that children from families with higher socio-economic status are less likely to be risk seeking.

In our study, pupils' risk aversion is measured by the share of total answers given and the share of correctly given answers. Pupils who answer more questions in total but also have a lower share of correctly given answers are considered more risk seeking. Table 7 reports the percentage share of correctly given answers as well as the share of answers given for incentivized and non-incentivized pupils distinguished by school types. We find significant differences between pupils of different socio-economic background and between incentivized and non-incentivized pupils in High Schools.

Pupils in High Schools answer fewer questions than pupils in Non-High Schools in the control group (79.22% vs. 93.91%, p < 0.001) as well as in the incentivized group (84.54% vs. 94.11%, p < 0.001). Nonetheless, pupils in High Schools give more correct answers than pupils in Non-High Schools (non-incentivized: 51.28% vs 34.16%, p < 0.001; incentivized: 46.44% vs. 36.66%, p < 0.001). This might be due to a sample selection effect as pupils in High School are in general more able students. Comparing incentivized pupils with the control group, we find no significant difference in pupils from Non-High Schools in the share of total

given answers but a significant increase in the share of correct answers (34.16% vs. 36.66%, p = 0.037). This is an indication that pupils in Non-High Schools become more risk averse and reduce their guessing behavior when incentivized. Conversely, pupils in High Schools answer significantly more questions when incentivized (84.54% vs. 79.22%, p < 0.001) but in contrast to Non-High School pupils, the accuracy of answering significantly declines (46.66% vs. 51.28%, p < 0.001). An increased number of questions with a simultaneously declining number of correct answers is a sign that pupils become more risk-seeking when incentivized, which is contrary to the findings of Deckers et al. (2015). As discussed above, the cause for this shift in risky behavior is a negative effect of extrinsic rewards. Pupils are not motivated to exert effort or even might increase their guessing behavior on purpose to avoid getting a reward and the accompanying signal of good performance to others.

Table 7: Share of correct and given answers

| | *Correct not-Incentivized* | *Correct Incentivized* | *Overall not-Incentivized* | *Overall Incentivized* |
|---|---|---|---|---|
| *Non-High School (N=1230)* | 34.16 | 36.66 | 93.91 | 94.11 |
| *High School (N=883)* | 51.28 | 46.44 | 79.22 | 84.54 |

*Note:* This table reports on the share of overall answered questions as well as the share of correctly answered questions for incentivized and not incentivized pupils separately for High Schools and Non-High Schools. Cell entries report percentages. "Correct not-Incentivized" is the share of correctly given answers of pupils in the Control Treatment on the number of given answers (correct + false). "Overall Incentivized" is the share of given answers of pupils in the (pooled) incentivized treatments (Choice, Medal and Letter).

## Further Results

A large part of the research in the economics of education involves the effects of educational inputs, such as teacher gender, teacher quality, and/or school resources on pupil achievement. However, no consensus has been reached regarding how these factors influence student performance (see Hanushek [1986]; Card and Krueger [1992]; Hoxby [2000]; Rivkin, Hanushek, and Kain [2005]). We now examine the correlations of some input factors and achievement to see how our sample compares with previous studies. Table 18 in Appendix A.5 shows the coefficients of teachers' working experience, parents' educational background, the gender of teachers and midterm grades on pupils achievement for the whole sample as distinguished by school types as well as by gender and pupils' ability.

**Socio Economic status and *Books at Home*** To further investigate the role of the socio-economic and foremost educational background of the parents in explaining the opposing effects for fifth graders in High Schools and sixth graders in Non-High Schools, we include *Books at home* as an explanatory variable. As expected, we find that the number of books is positively correlated with pupils' school achievement in both

High Schools and Non-High Schools. The effect seems to be strongest for pupils in High Schools whose parents have more than 200 books at home (201–500 books: 5.065, p = 0.002; over 500 books: 5.095, p = 0.003, Table 18 in Appendix A.5). Furthermore, there is a higher correlation between the education of the household and school achievement for boys than for girls. Analyzing the performance of German elementary pupils in the *Trends in International Mathematics and Science Study 2011* (TIMSS), Bos et al. (2012) find that fourth graders whose families have more than 100 books at home are one year ahead in mathematical skills in comparison with fourth graders who report that their families have fewer than 100 books at home. In our sample, for pupils in High Schools the modal response to the this question was 101–200 books at home, whereas the modal response for pupils in Non-High Schools was fewer than 100 books at home. Given the results of Bos et al. (2012), fifth graders in High Schools should have a similar level of mathematical skills than sixth graders in Non-High Schools, which holds true for our sample.

**Ability** Performance differences are driven not only by ability, but also by the amount of intrinsic motivation for the matter at issue. However, the previous literature has shown that extrinsic incentives tend to crowd out motivation for intrinsically motivated tasks (Frey 1994; Gneezy and Rustichini 2000; Frey and Jegen 2001). By asking pupils about their affinity for mathematics on a 1 (not at all) to 5 (very much) scale, we can approximate whether low- and high-performing pupils differ in their intrinsic motivation. We find that high performers have a significantly higher affinity toward mathematics (3.984) than low performers (3.150). Hence, providing extrinsic non-monetary incentives to pupils might lead to a poorer test performance for high-ability pupils if a potentially stronger internal motivation gets crowded out. Conversely, low-performing pupils—who lack internal motivation—might benefit by being extrinsically incentivized. Based on externally given midterm grades, we group pupils into *high*-, *middle*- and *low*-ability pupils. High-ability pupils refers to those with a midterm grade of 1 or 2; middle-ability pupils have a midterm grade of 3 and low-ability pupils are those with a midterm grade of 4, 5 or 6. The groups are of approximately equal size.

Table 15 in Appendix A.3 reports the average treatment effects by ability. We find differences between low- and high-ability pupils and differences between pupils at High Schools and Non-High Schools. Motivation is crowded out for low-ability pupils in High Schools in the Medal (-4.480, p = 0.013) and Letter Treatment (-5.672, p = 0.011). By contrast, high performers in High School do not seem to respond to the rewards in the Choice (0.595, p = 0.540), Medal (-0.240, p = 0.852) and Letter Treatment categories (-2.093, p = 0.262). In Non-High Schools, non-monetary incentives enhance test performance for both low-ability and high-ability pupils but decreases test performance for medium-ability performers. The effects are significant for high-ability pupils in the Letter Treatment (2.791, p = 0.011) and for middle-ability performing pupils in the Medal Treatment (-4.201, p = 0.009).

Our results can be compared to those of Leuven, Oosterbeek, and Klaauw (2010), who find that monetary incentives increase academic performance for the most able students but decrease performance for low-ability students. We find similar results for high performers in Non-High Schools and low performers in High Schools.

There are at least two mechanisms that can explain the results of Leuven, Oosterbeek, and Klaauw (2010): a pure "crowding out" effect and a "resignation" ("I won't make it in any case") effect. We find similar results although we use non-monetary rewards and different rewarding conditions. Students in the study of Leuven, Oosterbeek, and Klaauw (2010) had to pass all first-year requirements within one year according to a fixed (i.e., non-personalized) threshold. By using a relative rewarding scheme—pupils in our study had to improve relative to their past performance—we reduce or eliminate the "resignation effect". Overall, we find that average treatment effects are positive and highest for high-ability pupils in Non-High Schools.

We can further analyze whether there are also gender differences for low-, middle- and high-achieving pupils. Table 16 in Appendix A.3 reports NB estimates differentiated by ability and gender. We find pronounced and large gender differences for low-achieving pupils. Boys do not significantly respond to any type of incentive in High Schools and Non-High Schools. By contrast, intrinsic motivation is crowded out for girls in High Schools in the Choice (-4.727, p = 0.062) and Letter Treatments (-6.208, p = 0.062). In Non-High Schools, girls motivation is increased in the Medal Treatment (4.098, p = 0.029).

**Teachers' working experience**

Rivkin, Hanushek, and Kain (2005) show that mathematics teachers in their first year and—to a lesser extent—second- and third-year teachers perform significantly worse than more experienced teachers. There may be some additional gains to experience in the subsequent year or two, but the estimated benefits are small and not statistically significant in both mathematics and reading (see also Harris and Sass [2011]). In line with Rivkin, Hanushek, and Kain (2005), we find that teachers' experience is correlated with higher achievement in Non-High Schools (0.090, p = 0.015) and High Schools (0.058, p = 0.071). Boys in High Schools (0.077, p = 0.024) and girls in Non-High Schools (0.150, p = 0.002) achieve significantly higher test scores with a more experienced teacher, although there is no significant effect on girls in High Schools (0.019, p = 0.701) and boys in Non-High Schools (0.061, p = 0.106). Furthermore, low-ability pupils in Non-High Schools (0.181, $p < 0.001$), in particular, perform better than those with an inexperienced teacher. In addition, low-ability pupils in High Schools (0.165, p = 0.037) have better test performances with a teacher who has more experience. In all school types, there is no significant correlation for high-ability pupils. The most common channels in the literature that may explain these correlations are that i) experienced teachers are better able to use teaching strategies that respond to students' needs and learning styles, ii) experienced teachers focus more on low-ability pupils and iii) experienced teachers can better handle disturbances in class.

**Teachers' gender**

The results of the influence of teacher's gender are mixed so far. Carrell, Page, and West (2010) report that the gender of the professor has little impact on male students' performance in math but a powerful effect on female students' performance, whereas Antecol, Eren, and Ozbeklik (2015) find that in primary school, female students who were assigned to a female teacher suffered from lower math test scores at the end of the

academic year. Our study's findings support the findings of Antecol, Eren, and Ozbeklik (2015). We find that having a female teacher lowers (nonsignificant) test scores for girls in High Schools (-0.387, p = 0.736) and Non-High Schools (-2.909, p = 0.026). In contrast to Carrell, Page, and West (2010) and Antecol, Eren, and Ozbeklik (2015), we find a significant correlation for boys in High Schools (-3.109, p < 0.001), although the correlation for boys in Non-High Schools (0.429, p = 0.650) is insignificant.

# 7 Conclusion

In this paper, we analyze the effects of four non-monetary rewards on performances regarding an incentivized mathematical test in both deprived and high-achieving schools in Germany. Our experimental design allows to clearly analyze treatment effects, as pupils did not know that they were part of an experiment. We maintain a natural examination situation within the classroom by having the students take the test during a regularly scheduled math lesson and letting teachers conduct the experiment by themselves.[24]

The selection of rewards is motivated by a questionnaire conducted in the run-up to the experiment asking pupils about their preferences over seventeen pre-selected rewards. Based on the survey results, we test the effects of the following four incentives in the field: (i) medal, (ii) letter of praise, (iii) "no-homework" voucher and (iv) surprise gift.

We explore the short-run impact of non-monetary incentives on pupils' test scores by varying the degree of freedom with which pupils can chose their reward. In the Choice Treatment, pupils are free to choose one of the four incentives beforehand, whereas in the Medal and Letter Treatments, the respective reward is predetermined. Additionally, we compare the effects of non-monetary incentives on pupils with different socio-economic background.

Our general findings suggest that non-monetary incentives can be a potentially cost-effective way to increase achievement in deprived schools but that motivation is crowded-out with such rewards in high-achieving schools. We show that the counter effectiveness of incentives is driven by pupils socio-economic background. Moreover, we find that giving pupils the freedom to choose their incentive increases the willingness to learn for the test and mitigates the negative effects of external rewards in High Schools. Furthermore, low-performing pupils want to reveal their educational achievement to their parents. This might be interpreted as longing for appreciation and consideration by the parents.

A limitation of the experiment, as in most experiments, is that we can only learn the impact of treatments on the population studied, which is a broad—but not representative—sample of the population of pupils in Germany. However, we shed light on the effectiveness of non-monetary incentives in schools and show that they must be carefully designed and that pupils' socio-economic background must be taken into account.

The applicability of our incentives is confirmed by teachers. Overall, 44.31% of teachers are planning to use at least one incentive in the future. Again, there are differences between school types. While incentives

[24]Counter arguments for this kind of design might be a potential loss of control. However, we believe that teachers had no incentive to not follow our instructions.

are well received by teachers at Non-High Schools, only about 14% of High School teachers plan to use incentives in the future.[25]

It remains for future research to analyze the effectiveness of non-monetary incentives in the long term and to test the working of a *"Fixed Voucher"* and *"Fixed Surprise"* treatment, in addition to testing the remaining incentives suggested in our survey in Subsection 2.2. It would be also interesting to further investigate the impact of non-monetary incentives on pupils with different cultural backgrounds. Finally, more research is required on identifying potential non-monetary incentives for teachers.[26]

[25]The share of teachers planning to use future incentives by school types is roughly as follows: Secondary General School 66%; Middle School 56%; Comprehensive School 75%; High School 14%.

[26]Having informally asked teachers about their preferences, work avoidance (e.g., somebody else corrects tests) and public recognition incentives seem to be the most promising types of non-monetary incentives for teachers.

# References


Akerlof, George A., and Rachel E. Kranton, "Identity and the Economics of Organizations," *Journal of Economic Perspectives*, 19 (2005), 9–32.

Ames, Carole, "Classrooms: Goals, Structures, and Student Motivation," *Journal of Educational Psychology*, 84 (1992), 261.

Anders, Yvonne, Nele McElvany, and Jürgen Baumert, "Die Einschätzung lernrelevanter Schülermerkmale zum Zeitpunkt des Übergangs von der Grundschule auf die weiterführende Schule. Wie differenziert urteilen Lehrkräfte?" in *Der Übergang von der Grundschule in die weiterführende Schule. Leistungsgerechtigkeit und regionale, soziale und ethnisch–kulturelle Disparitäten,* Kai Maaz, Jürgen Baumert, Cornelia Gresch, and Nele McElvany, eds., Bildungsforschung. 34 (Bonn u.a.: Bundesministerium für Bildung und Forschung, Referat Bildungsforschung, 2010).

Angrist, Joshua, and Victor Lavy, "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *The American Economic Review*, 99 (2009), 1384–1414.

Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik, "The Effect of Teacher Gender on Student Achievement in Primary School," *Journal of Labor Economics*, 33 (2015), 63–89.

Armantier, Olivier, and Amadou Boly, "Comparing Corruption in the Laboratory and in the Field in Burkina Faso and in Canada," *The Economic Journal*, 123 (2013), 1168–1187.

Austen-Smith, David, and Roland G Fryer Jr, "An Economic Analysis of "Acting White"," *The Quarterly Journal of Economics*, 120 (2005), 551–583.

Avvisati, Francesco, Marc Gurgand, Nina Guyon, and Eric Maurin, "Getting Parents Involved: A Field Experiment in Deprived Schools," *The Review of Economic Studies*, 81 (2014), 57–83.

Behrman, Jere, Susan Parker, Petra Todd, and Kenneth I Wolpin, "Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools," (2011), available at: `http://economics.wustl.edu/files/economics/imce/todd_paper.pdf`.

Bergman, Peter, "Parent–Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment," Working Paper, 2014, available at: `http://www.columbia.edu/~psb2101/BergmanSubmission.pdf`.

Bettinger, Eric, "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores," *Review of Economics and Statistics*, 94 (2012), 686–698.

Bettinger, Eric, and Robert Slonim, "Patience Among Children," *Journal of Public Economics*, 91 (2007), 343–363.

Bigoni, Maria, Margherita Fort, Mattia Nardotto, and Tommaso G Reggiani, "Cooperation or Competition? A Field Experiment on Non-monetary Learning Incentives," *The BE Journal of Economic Analysis & Policy*, 15 (2015), 1753–1792.

Blimpo, Moussa P, "Team incentives for education in developing countries: A randomized field experiment in Benin," *American Economic Journal: Applied Economics*, 6 (2014), 90–109.

Bénabou, Roland, and Jean Tirole, "Incentives and Prosocial Behavior," *The American Economic Review*, 96 (2006), 1652–1678.

Bo, Pedro Dal, Andrew Foster, and Louis Putterman, "Institutions and Behavior: Experimental Evidence on the Effects of Democracy," *American Economic Review*, 100 (2010), 220–5–29.

Bos, Wilfried, Heike Wendt, Olaf Köller, and Christoph Selter, *TIMSS 2011 Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (New York, NY: Waxmann Verlag, 2012).

Bruhn, Miriam, and David McKenzie, "In Pursuit of Balance: Randomization in Practice in Development Field Experiments," *American Economic Journal: Applied Economics*, 1 (2009), 200–232, doi:10.1257/app.1.4.200.

Bursztyn, Leonardo, and Robert Jensen, "How Does Peer Pressure Affect Educational Investments?" *The Quarterly Journal of Economics*, 130 (2015), 1329–1367, doi:10.1093/qje/qjv021.

Caffyn, Rachel E., "Attitudes of British Secondary School Teachers and Pupils to Rewards and Punishments," *Educational Research*, 31 (1989), 210–220.

Card, David, and Alan B Krueger, "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, 100 (1992), 1–40.

Carrell, Scott, Richard L. Fullerton, and James West, "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, 27 (2009), 439–464.

Carrell, Scott E., Marianne E. Page, and James E. West, "Sex and Science: How Professor Gender Perpetuates the Gender Gap," *The Quarterly Journal of Economics*, 125 (2010), 1101–1144.

Chand, Vijaya Sherry, Tathagata Banerjee, and Chattopadhyay Bhargab, "An Exploratory Study of the Role of Educational Incentives in Primary Education in Gujarat," Indian Institute of Management Ahmedabad, Research and Publication Department, Working Paper WP2008–09–01, 2008, available at: `http://ideas.repec.org/p/iim/iimawp/wp02116.html`.

Charness, Gary, Ramón Cobo-Reyes, Natalia Jimenez, Juan A. Lacomba, and Francisco Lagos, "The Hidden Advantage of Delegation: Pareto Improvements in a Gift Exchange Game," *The American Economic Review*, 102 (2012), 2358–379, available at: `http://www.jstor.org/stable/41724625`.

Chelonis, John J., Rebecca A. Flake, Ronald L. Baldwin, Donna J. Blake, and Merle G. Paule, "Developmental Aspects of Timing Behavior in Children," *Neurotoxicology and Teratology*, 26 (2004), 461–476, doi:10.1016/j.ntt.2004.01.004.

Chevalier, Arnaud, Peter Dolton, and Melanie Lührmann, ""Making It Count": Evidence from a Field Study on Assessment Rules, Study Incentives and Student Performance," Institute for the Study of Labor (IZA) Discussion Paper No. 8582, 2014, available at: `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2517885`.

Cunha, Flavio, and James Heckman, "The Technology of Skill Formation," *American Economic Review*, 97 (2007), 31–47, doi:10.1257/aer.97.2.31, available at: `http://www.aeaweb.org/articles.php?doi=10.1257/aer.97.2.31`.

Deckers, Thomas, Armin Falk, Fabian Kosse, and Hannah Schildberg-Hörisch, "How Does Socio-Economic Status Shape a Child's Personality?" Institute for the Study of Labor (IZA), 2015.

Ditton, Hartmut, "Schulübertritte, Geschlecht und soziale Herkunft," in *Kompetenzaufbau und Laufbahnen im Schulsystem,* Hartmut Ditton, ed. (Münster u.a.: Waxmann, 2007).

Dowson, Martin, and Dennis M McInerney, "Psychological Parameters of Students' Social and Work Avoidance Goals: A Qualitative Investigation," *Journal of Educational Psychology*, 93 (2001), 35.

Duflo, Esther, Rachel Glennerster, and Michael Kremer, "Using Randomization in Development Economics Research: A Toolkit," *Handbook of Development Economics*, 4 (2007), 3895–3962.

Dustmann, Christian, "Parental background, secondary school track choice, and wages," *Oxford Economic Papers*, 56 (2004), 209–230, doi:10.1093/oep/gpf048, available at: `http://oep.oxfordjournals.org/content/56/2/209.abstract`.

Dwyer, David J, and Jeffrey B Hecht, "Minimal Parental Involvement," *School Community Journal*, 2 (1992), 53–66.

Frey, Bruno S., "How Intrinsic Motivation is Crowded out and in," *Rationality and Society*, 6 (1994), 334–352.

Frey, Bruno S, and Reto Jegen, "Motivation Crowding Theory," *Journal of Economic Surveys*, 15 (2001), 589–611.

Fryer, Roland, "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools," *Journal of Labor Economics*, 31 (2013), 373–427.

Fryer, Roland G, "Financial Incentives and Student Achievement: Evidence from Randomized Trials," *The Quarterly Journal of Economics*, 126 (2011), 1755–1798.

Fuchs, Thomas, and Ludger Woessmann, *What Accounts for International Differences in Student Performance? A Re–examination Using PISA Data* (Heidelberg: Springer, 2008).

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel, "When and Why Incentives (Don't) Work to Modify Behavior," *The Journal of Economic Perspectives*, 25 (2011), 191–209.

Gneezy, Uri, and Aldo Rustichini, "Pay Enough or Don't Pay at All," *The Quarterly Journal of Economics*, 115 (2000), 791–810.

Goodman, Joshua Samuel, "The Labor of Division: Returns to Compulsory Math Coursework," John F. Kennedy School of Government, Harvard University HKS Faculty Research Working Paper Series RWP12-032, 2012, available at: `http://dash.harvard.edu/handle/1/9403178`.

Gresch, Cornelia, Jürgen Baumert, and Kai Maaz, "Empfehlungsstatus, Übergangsempfehlung und der Wechsel in die Sekundarstufe I: Bildungsentscheidungen und soziale Ungleichheit," in *Bildungsentscheidungen,* Jürgen Baumert, Kai Maaz, and Ulrich Trautwein, eds. (VS Verlag für Sozialwissenschaften, 2010).

Hanushek, Eric A, "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature*, 24 (1986), 1141–1177.

Hanushek, Eric A, Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann, "Returns to skills around the world: Evidence from {PIAAC}," *European Economic Review*, 73 (2015), 103–130.

Harris, Douglas N, and Tim R Sass, "Teacher Training, Teacher Quality and Student Achievement," *Journal of Public Economics*, 95 (2011), 798–812.

Hoxby, Caroline M, "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics*, (2000), 1239–1285.

Jalava, Nina, Juanna Schr ter Joensen, and Elin Pellas, "Grades and Rank: Impacts of Non–Financial Incentives on Test Performance," *Journal of Economic Behavior & Organization*, (2014), doi:10.1016/j.jebo.2014.12.004.

Jensen, Lene Arnett, Jeffrey Jensen Arnett, S Shirley Feldman, and Elizabeth Cauffman, "It's Wrong, but Everybody Does It: Academic Dishonesty among High School and College Students," *Contemporary Educational Psychology*, 27 (2002), 209–228.

Jensen, Peter, and Astrid Würtz Rasmussen, "The Effect of Immigrant Concentration in Schools on Native and Immigrant Children's Reading and Math Skills," *Economics of Education Review*, 30 (2011), 1503–1515.

Jonkmann, Kathrin, Kai Maaz, Marko Neumann, and Cornelia Gresch, "Übergangsquoten und Zusammenhänge zu familiärem Hintergrund und schulischen Leistungen. Deskriptive Befunde," in *Der Übergang*

*von der Grundschule in die weiterführende Schule. Leistungsgerechtigkeit und regionale, soziale und ethnisch–kulturelle Disparitäten,* Kai Maaz, Jürgen Baumert, Cornelia Gresch, and Nele McElvany, eds., Bildungsforschung 34 (Bonn u.a.: Bundesministerium für Bildung und Forschung, Referat Bildungsforschung, 2010).

Koch, Alexander, Julia Nafziger, and Helena Skyt Nielsen, "Behavioral Economics of Education," *Journal of Economic Behavior & Organization*, (2014), doi:10.1016/j.jebo.2014.09.005.

Kosfeld, Michael, and Susanne Neckermann, "Getting More Work for Nothing? Symbolic Awards and Worker Performance," *American Economic Journal: Microeconomics*, 3 (2011), 86–99.

Kremer, Michael, Esther Duflo, and Pascaline Dupas, "Peer Effects, Teacher Incentives, and the Impact of Tracking," *American Economic Review*, 101 (2011), 1739–1774.

Kube, Sebastian, Michel Andre Marechal, and Clemens Puppea, "The Currency of Reciprocity: Gift Exchange in the Workplace," *The American Economic Review*, 102 (2012), 1644–1662.

Lavy, Victor, "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement," *Journal of Political Economy*, 110 (2002), 1286–1317.

Leuven, Edwin, Hessel Oosterbeek, and Bas Klaauw, "The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment," *Journal of the European Economic Association*, 8 (2010), 1243–1265.

Levitt, Steven D, John A List, Susanne Neckermann, and Sally Sadoff, "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance," National Bureau of Economic Research Working Paper No. 18165, 2012, available at: `http://www.nber.org/papers/w18165`.

Loewenstein, George, "The Psychology of Curiosity: A Review and Reinterpretation," *Psychological Bulletin*, 116 (1994), 75.

McEwan, Patrick J, "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments," *Review of Educational Research*, (2014), doi:10.3102/0034654314553127, available at: `http://rer.sagepub.com/content/early/2014/10/03/0034654314553127.abstract`.

Mellizo, Philip, Jeffrey Carpenter, and Peter Hans Matthews, "Workplace Democracy in the Lab," *Industrial Relations Journal*, (2014), doi:10.1111/irj.12054.

Mueller, Steffen, "Teacher Experience and the Class Size Effect – Experimental Evidence," *Journal of Public Economics*, 98 (2013), 44–52, doi:10.1016/j.jpubeco.2012.12.001.

Muralidharan, Karthik, and Venkatesh Sundararaman, "Teacher Performance Pay: Experimental Evidence from India," *The Journal of Political Economy*, 119 (2011), 39–77, doi:10.1086/659655.

Neal, Derek, "The Design of Performance Pay in Education," National Bureau of Economic Research Working Paper No. 16710, 2011, available at: http://www.nber.org/papers/w16710.

Niederle, Muriel, and Lise Vesterlund, "Explaining the Gender Gap in Math Test Scores: The Role of Competition," *The Journal of Economic Perspectives*, (2010), 129–144.

Ohinata, Asako, and Jan C Van Ours, "How Immigrant Children Affect the Academic Achievement of Native Dutch Children," *The Economic Journal*, 123 (2013), F308–F331.

Oreopoulos, Philip, "Do Dropouts Drop out Too Soon? Wealth, Health and Happiness from Compulsory Schooling," *Journal of Public Economics*, 91 (2007), 2213–2229.

Paulus, Wiebke, and Hans-Peter Blossfeld, "Schichtspezifische Präferenzen oder sozioökonomisches Entscheidungskalkül? Zur Rolle elterlicher Bildungsaspirationen im Entscheidungsprozess beim Übergang von der Grundschule in die Sekundarstufe," *Zeitschrift für Pädagogik*, 53 (2007), 491–508.

Retelsdorf, Jan, and Jens Möller, "Entwicklungen von Lesekompetenz und Lesemotivation Schereneffekte in der Sekundarstufe?" *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 40 (2008), 179–188.

Rivkin, Steven G, Eric A Hanushek, and John F Kain, "Teachers, Schools, and Academic Achievement," *Econometrica*, 73 (2005), 417–458.

Sacerdote, Bruce, "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" *Handbook of the Economics of Education*, 3 (2011), 249–277, available at: http://ideas.repec.org/h/eee/educhp/3--04.html.

Sirin, Selcuk R, "Socioeconomic Status and Academic Achievement: A Meta–analytic Review of Research," *Review of Educational Research*, 75 (2005), 417–453.

Springer, Matthew G, Dale Ballou, Laura Hamilton, Vi-Nhuan Le, JR Lockwood, Daniel F McCaffrey, Matthew Pepper, and Brian M Stecher, "Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT)." 2011.

Sutter, Matthias, Stefan Haigner, and Martin G Kocher, "Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations," *The Review of Economic Studies*, 77 (2010), 1540–1566.

Urdan, Timothy C, and Martin L Maehr, "Beyond a Two–Goal Theory of Motivation and Achievement: A Case for Social Goals," *Review of Educational Research*, 65 (1995), 213–243.

Woessmann, Ludger, "The Effect Heterogeneity of Central Examinations: Evidence from TIMSS, TIMSS–Repeat and PISA," *Education Economics*, 13 (2005), 143–169.

Zizzo, Daniel John, "Experimenter Demand Effects in Economic Experiments," *Experimental Economics*, 13 (2010), 75–98, doi:10.1007/s10683--009--9230--z.

# A Tables

## A.1 Randomization Tables

Table 8: Randomization Check: Secondary General School

| | (1) Control | (2) Choice | (3) Medal | (4) Letter | (5) Overall |
|---|---|---|---|---|---|
| Age | 11.485 | 12.419 | 12.000 | 11.308 | 12.005 |
| | (0.115) | (0.082) | (0.196) | (0.237) | (0.069) |
| Month of Birth | 6.686 | 6.768 | 6.406 | 5.143 | 6.592 |
| | (0.481) | (0.395) | (0.612) | (0.931) | (0.265) |
| Number of Older Siblings | 1.072 | 1.423 | 1.188 | 2.133 | 1.330 |
| | (0.123) | (0.143) | (0.231) | (0.601) | (0.095) |
| Female Pupil | 0.471 | 0.427 | 0.406 | 0.533 | 0.444 |
| | (0.060) | (0.046) | (0.088) | (0.133) | (0.033) |
| Language German | 0.414 | 0.410 | 0.250 | 0.267 | 0.380 |
| | (0.059) | (0.046) | (0.078) | (0.118) | (0.032) |
| *N* | 73 | 118 | 33 | 15 | 239 |

*Note:* This table reports the results of the randomization check in General Secondary School. Cell entries present means and standard errors are reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 9: Randomization Check: Middle School

| | (1) Control | (2) Choice | (3) Medal | (4) Letter | (5) Overall |
|---|---|---|---|---|---|
| Age | 11.032 | 11.104 | 11.491 | 11.173 | 11.161 |
| | (0.051) | (0.062) | (0.076) | (0.067) | (0.032) |
| Month of Birth | 6.858 | 6.796 | 6.770 | 7.544 | 6.984 |
| | (0.279) | (0.299) | (0.392) | (0.348) | (0.161) |
| Number of Older Siblings | 0.965 | 0.893 | 0.957 | 1.075 | 0.969 |
| | (0.080) | (0.079) | (0.115) | (0.089) | (0.044) |
| Female Pupil | 0.425 | 0.522 | 0.429 | 0.364 | 0.439 |
| | (0.033) | (0.035) | (0.046) | (0.038) | (0.019) |
| Language German | 0.438 | 0.525 | 0.555 | 0.439 | 0.482 |
| | (0.033) | (0.035) | (0.046) | (0.039) | (0.019) |
| *N* | 228 | 206 | 120 | 165 | 719 |

*Note:* This table reports the results of the randomization check in Middle School. Cell entries present means and standard errors are reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 10: Randomization Check: Comprehensive School

| | (1) Control | (2) Choice | (3) Medal | (4) Letter | (5) Overall |
|---|---|---|---|---|---|
| Age | 11.438 | 11.462 | 11.118 | 11.414 | 11.376 |
| | (0.102) | (0.091) | (0.114) | (0.103) | (0.051) |
| Month of Birth | 6.641 | 6.821 | 7.340 | 7.451 | 7.049 |
| | (0.526) | (0.482) | (0.587) | (0.466) | (0.255) |
| Number of Older Siblings | 1.297 | 1.077 | 0.755 | 1.403 | 1.154 |
| | (0.139) | (0.140) | (0.117) | (0.167) | (0.074) |
| Female Pupil | 0.469 | 0.450 | 0.472 | 0.472 | 0.465 |
| | (0.063) | (0.056) | (0.069) | (0.059) | (0.030) |
| Language German | 0.645 | 0.550 | 0.566 | 0.667 | 0.607 |
| | (0.061) | (0.056) | (0.069) | (0.056) | (0.030) |
| *N* | 65 | 80 | 54 | 73 | 272 |

*Note:* This table reports the results of the randomization check in Comprehensive School. Cell entries present means and standard errors are reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 11: Randomization Check: High School

| | (1) Control | (2) Choice | (3) Medal | (4) Letter | (5) Overall |
|---|---|---|---|---|---|
| Age | 11.051 | 10.848 | 10.638 | 10.926 | 10.876 |
| | (0.055) | (0.046) | (0.050) | (0.062) | (0.027) |
| Month of Birth | 7.076 | 7.389 | 7.602 | 7.375 | 7.345 |
| | (0.284) | (0.262) | (0.325) | (0.346) | (0.150) |
| Number of Older Siblings | 0.769 | 0.719 | 0.618 | 0.661 | 0.699 |
| | (0.057) | (0.066) | (0.059) | (0.066) | (0.031) |
| Female Pupil | 0.419 | 0.412 | 0.376 | 0.479 | 0.420 |
| | (0.032) | (0.030) | (0.036) | (0.039) | (0.017) |
| Language German | 0.749 | 0.689 | 0.730 | 0.673 | 0.711 |
| | (0.028) | (0.028) | (0.033) | (0.036) | (0.015) |
| *N* | 242 | 279 | 189 | 173 | 883 |

*Note:* This table reports the results of the randomization check in High School. Cell entries present means and standard errors are reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 12: Treatment Randomization: Average Test Scores

| | *Control* | *Choice* | *Medal* | *Letter* |
|---|---|---|---|---|
| *Non-High School* | | | | |
| *Full Sample* | | | | |
| N individuals | 366 | 404 | 207 | 253 |
| Test Score | 8.945 | 10.129 | 11.275 | 11.245 |
| | (10.650) | (11.719) | (11.823) | (11.580) |
| Standardized Test Score | -0.247 | -0.141 | -0.041 | -0.039 |
| | (0.953) | (1.049) | (1.058) | (1.036) |
| *Boys* | | | | |
| N individuals | 202 | 209 | 115 | 150 |
| Test Score | 10.203 | 10.861 | 11.461 | 12.560 |
| | (10.474) | (11.568) | (12.125) | (12.549) |
| Standardized Test Score | -0.135 | -0.076 | -0.022 | 0.076 |
| | (0.937) | (1.0354) | (1.085) | (1.123) |
| *Girls* | | | | |
| N individuals | 160 | 193 | 89 | 102 |
| Test Score | 7.519 | 9.368 | 10.865 | 9.157 |
| | (10.758) | (11.914) | (11.562) | (9.635) |
| Standardized Test Score | -0.375 | -0.209 | -0.075 | -0.228 |
| | (0.963) | (1.066) | (1.035) | (0.862) |
| *High School* | | | | |
| *Full Sample* | | | | |
| N individuals | 242 | 279 | 189 | 173 |
| Test Score | 14.888 | 14.147 | 12.206 | 13.486 |
| | (9.689) | (10.299) | (10.755) | (11.269) |
| Standardized Test Score | 0.285 | 0.218 | 0.045 | 0.159 |
| | (0.867) | (0.922) | (0.963) | (1.009) |
| *Boys* | | | | |
| N individuals | 137 | 164 | 116 | 88 |
| Test Score | 16.431 | 14.793 | 13.543 | 13.352 |
| | (10.566) | (10.645) | (11.098) | (10.187) |
| Standardized Test Score | 0.423 | 0.276 | 0.164 | 0.147 |
| | (0.946) | (0.953) | (0.993) | (0.912) |
| *Girls* | | | | |
| N individuals | 99 | 115 | 70 | 81 |
| Test Score | 12.929 | 13.226 | 9.971 | 13.580 |
| | (8.144) | (9.758) | (9.575) | (12.603) |
| Standardized Test Score | 0.109 | 0.136 | -0.155 | 0.168 |
| | (0.729) | (0.873) | (0.857) | (1.128) |

*Note:* The table displays the descriptive statistics of test scores and the number of students in each of the treatment groups and the control group. Both average points scored on the test and standardized test scores (with mean 0 and standard deviation 1). Standard errors are displayed in parentheses. In our final analysis, we included 2.067 observations. 46 observations were dropped because missing values. There are 23 missing values for the last midterm grade and 23 for pupils' gender.

## A.2 Robustness Checks

Table 13: Robustness Check—Treatment Effects without Covariates

| | *Non-High School* | | *High School* | |
|---|---|---|---|---|
| *Treatments* | | | | |
| Choice | 2.064** | [1.000] | −0.941 | [1.160] |
| Medal | 1.188 | [1.027] | −3.511** | [1.623] |
| Letter | 1.344 | [1.058] | −1.019 | [1.709] |
| *Controls* | | | | |
| Pupil Covariates | No | | No | |
| Class/School Covariates | No | | No | |
| School FE | Yes | | Yes | |
| *N* | 1230 | | 883 | |

*Note:* This table reports the result of a negative binomial regression without covariates separately for High Schools and Non-High School including school fixed effects. Standard errors are reported in parentheses and clustered on classroom-level. Dependent variable: points in test. The number of clusters is 36 in High Schools and 53 in Non-High Schools. Results are robust to multiple testing (seemingly unrelated estimation) * p<0.10, ** p<0.05, *** p<0.01

Table 14: Robustness Check—Treatment Effects by School Level

| **Panel A: Regression** | OLS | | | | Negative Binomial | | | |
|---|---|---|---|---|---|---|---|---|
| | *Non-High School* | | *High School* | | *Non-High School* | | *High School* | |
| *Treatments* | | | | | | | | |
| Choice | 0.367 | [1.818] | −0.417 | [1.218] | 0.336 | [1.881] | −0.479 | [1.183] |
| Medal | −0.579 | [1.301] | −2.668** | [1.150] | −0.827 | [1.582] | −2.591** | [1.222] |
| Letter | 0.839 | [1.544] | −4.833*** | [1.707] | 0.801 | [1.628] | −4.389*** | [1.518] |
| Grade 6 | 1.599 | [1.416] | 5.679*** | [1.564] | 1.684 | [1.389] | 5.995*** | [1.710] |
| Choice × Grade 6 | 2.000 | [2.123] | 0.540 | [1.773] | 2.163 | [2.222] | 0.842 | [1.874] |
| Medal × Grade 6 | 2.563 | [1.865] | 2.034 | [1.918] | 2.667 | [2.105] | 1.189 | [1.959] |
| Letter × Grade 6 | 0.623 | [2.124] | 5.420** | [2.755] | 0.538 | [2.146] | 4.293* | [2.773] |
| *Control* | | | | | | | | |
| Pupil Covariates | Yes | | Yes | | Yes | | Yes | |
| Class/School Covariates | Yes | | Yes | | Yes | | Yes | |
| School FE | Yes | | Yes | | Yes | | Yes | |
| *N* | 1198 | | 869 | | 1198 | | 869 | |
| **Panel B: Contrasts** *Treatment vs. No Treatment in Grade 6* | | | | | | | | |
| Choice | 2.368** | [1.127] | 0.123 | [1.323] | 2.500** | [1.190] | 0.363 | [1.556] |
| Medal | 1.983 | [1.295] | −0.634 | [1.413] | 1.840 | [1.840] | −1.401 | [1.558] |
| Letter | 1.462 | [1.587] | 0.587 | [2.024] | 1.339 | [1.499] | −0.096 | [2.146] |

*Note:* Panel A compares the results of a linear (OLS) and a negative binomial regression (marginal effects) for pupils in grade 5 separately for Non-High Schools and High Schools including school fixed effects. Panel B compares linear (OLS) and the negative binomial treatment effects for pupils in grade 6. Grade 6: 0=pupils in grade 5, 1=pupils in grade 6. Dependent variable: points in test. Covariates: last midterm grade, number of books at home, academic year (grade 5 or 6), teacher' working experience (in years), teachers' gender, day differences between tests, unemployment rate of the school district and the proportion of German speaking pupils within the class. Standard errors are reported in parentheses and clustered on classroom-level. The number of clusters is 53 in Non-High Schools and 36 in High Schools. * p<0.10, ** p<0.05, *** p<0.01.

## A.3 Treatment effects by ability (and gender)

Table 15: Treatment effects by pupils midterm grade

| | *Non-High School* | | *High School* | |
|---|---|---|---|---|
| *Low Performing Pupils* | | | | |
| Choice | 0.421 | [1.119] | −2.969 | [1.907] |
| Medal | 0.572 | [1.326] | −4.480** | [1.796] |
| Letter | 0.512 | [1.387] | −5.672** | [2.229] |
| *N* | 408 | | 144 | |
| *Middle Performing Pupils* | | | | |
| Choice | 0.033 | [1.911] | 0.831 | [1.794] |
| Medal | −4.201*** | [1.617] | −3.069* | [1.564] |
| Letter | −1.407 | [1.779] | −3.803** | [1.910] |
| *N* | 428 | | 288 | |
| *High Performing Pupils* | | | | |
| Choice | 1.579 | [1.267] | 0.595 | [0.970] |
| Medal | 0.979 | [1.208] | −0.240 | [1.284] |
| Letter | 2.791** | [1.103] | −2.093 | [1.868] |
| *N* | 362 | | 437 | |
| *Controls* | | | | |
| Pupil Covariates | Yes | | Yes | |
| Class/School Covariates | Yes | | Yes | |
| School FE | Yes | | Yes | |

*Note:* This table reports the result of a negative binomial regression separately for low-, middle- and high-ability pupils and separately for High Schools and Non-High School including school fixed effects. Dependent variable: points in test; Covariates: last midterm grade, gender, number of books at home, academic year (grade 5 or 6), teacher's working experience (in years), teacher's gender, day differences between tests, unemployment rate of the school district and the proportion of German speaking pupils within the class. High-ability pupils refers to those with a midterm grade of 1 or 2; middle-ability pupils have a midterm grade of 3 and low-ability pupils are those with a midterm grade of 4, 5 or 6. The groups are of approximately equal size. Standard errors are reported in parentheses and clustered on classroom-level. The number of clusters is 36 in High Schools and 53 in Non-High Schools. * p<0.10, ** p<0.05, *** p<0.01.

Table 16: Treatment effects by pupil gender and midterm grade

| | **Low Performers** | | | | **Middle Performers** | | | | **High Performers** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Panel A: Regression** | *Non-High School* | | *High School* | | *Non-High School* | | *High School* | | *Non-High School* | | *High School* | |
| *Treatments* | | | | | | | | | | | | |
| Choice | −0.322 | [1.538] | −0.856 | [2.926] | 0.752 | [2.170] | 0.783 | [2.195] | 2.156 | [1.956] | −0.392 | [1.311] |
| Medal | −2.444 | [1.599] | −3.946 | [3.265] | −3.469* | [1.940] | −2.275 | [1.836] | 2.558* | [1.446] | 0.107 | [1.502] |
| Letter | −0.570 | [1.915] | −4.672 | [2.955] | −0.544 | [1.892] | −4.912** | [2.286] | 2.590 | [1.853] | −2.877 | [1.922] |
| Female | −3.217** | [1.357] | −0.787 | [3.736] | −1.178 | [1.603] | 0.420 | [1.247] | 0.639 | [2.746] | −1.940 | [1.570] |
| Choice × Female | 1.074 | [2.222] | −3.872 | [3.806] | −1.468 | [2.436] | 0.020 | [3.126] | −1.362 | [3.274] | 2.275 | [2.200] |
| Medal × Female | 6.542*** | [2.265] | −0.649 | [4.798] | −1.584 | [2.626] | −2.193 | [2.893] | −3.693 | [3.019] | −1.118 | [2.661] |
| Letter × Female | 2.120 | [2.806] | −1.536 | [4.362] | −1.786 | [2.247] | 2.679 | [2.551] | 0.817 | [4.321] | 1.589 | [1.962] |
| *Controls* | | | | | | | | | | | | |
| Pupil Covariates | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| Class/School Covariates | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| School FE | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| **Panel B: Contrast** | *Treatment vs. No Treatment for Females* | | | | | | | | | | | |
| Choice | 0.751 | [1.599] | −4.727* | [2.533] | −0.716 | [2.396] | 0.803 | [2.559] | 0.794 | [2.169] | 1.883 | [1.624] |
| Medal | 4.098** | [1.847] | −4.595 | [2.927] | −5.053** | [2.250] | −4.468* | [2.521] | −1.134 | [2.400] | −1.010 | [2.256] |
| Letter | 1.550 | [2.022] | −6.208* | [3.324] | −2.330 | [2.340] | −2.234 | [2.359] | 3.407 | [3.018] | −1.288 | [2.288] |
| *N* | 408 | | 144 | | 428 | | 288 | | 362 | | 437 | |

*Note*: Panel A reports the result of a negative binomial regression separately for low-, middle- and high-ability boys and separately for High Schools and Non-High School and also reports interaction effects with gender including school fixed effects. Panel B reports treatment effect for girls. Dependent variable: points in test; Covariates: last midterm grade, number of books at home, academic year (grade 5 or 6), teacher's working experience (in years), teacher's gender, day differences between tests, unemployment rate of the school district and the proportion of German speaking pupils within the class. Female: 0 = boys; 1 = girls. High-ability pupils refers to those with a midterm grade of 1 or 2; middle-ability pupils have a midterm grade of 3 and low-ability pupils are those with a midterm grade of 4, 5 or 6. The groups are of approximately equal size. Standard errors are reported in parentheses and clustered on classroom-level. The number of clusters is 36 in High Schools and 53 in Non-High Schools. * p<0.10, ** p<0.05, *** p<0.01.

## A.4 Test preparation

Table 17: Test Preparation by School Type

| **Panel A: Regression** | *Non-High School* | | *High School* | |
|---|---|---|---|---|
| *Treatments* | | | | |
| Choice | 0.208** | [0.082] | 0.136*** | [0.051] |
| Medal | 0.186* | [0.102] | 0.012 | [0.045] |
| Letter | −0.095 | [0.092] | 0.075** | [0.038] |
| Grade 6 | 0.009 | [0.127] | −0.335*** | [0.052] |
| Choice × Grade 6 | −0.112 | [0.140] | 0.037 | [0.082] |
| Medal × Grade 6 | −0.147 | [0.158] | 0.044 | [0.177] |
| Letter × Grade 6 | 0.159 | [0.160] | 0.090 | [0.058] |
| *Controls* | | | | |
| Female pupil | 0.130*** | [0.028] | 0.080** | [0.038] |
| MidTerm grade | 0.007 | [0.005] | 0.027*** | [0.008] |
| Like Maths | 0.040*** | [0.013] | 0.052** | [0.021] |
| School FE | Yes | | Yes | |
| *N* | 1189 | | 866 | |

| **Panel B: Contrasts** *Treatment vs. No Treatment in Year 6* | | | | |
|---|---|---|---|---|
| Choice | 0.096 | [0.101] | 0.173*** | [0.059] |
| Medal | 0.039 | [0.117] | 0.057 | [0.172] |
| Letter | 0.064 | [0.115] | 0.166*** | [0.042] |

*Note:* Panel A reports results of logistic regression (marginal effects) for pupils in grade 5 and the interaction terms for treatment and school level separately for Non-High Schools and High Schools including school fixed effects. Panel B reports the logistic treatment effects for pupils in grade 6. Grade 6: 0=pupils in grade 5, 1=pupils in grade 6. Dependent variable: prepared for test (*Did you prepare for the test?* 0=No, 1=Yes) . Covariates: last midterm grade, math curiosity (measured on 1 to 5 scale). Standard errors are reported in parentheses and clustered on classroom-level. The number of clusters is 53 in Non-High Schools and 36 in High Schools. * p<0.10, ** p<0.05, *** p<0.01..

## A.5 Effects of covariates

Table 18: Ordinary Least Squares regression of covariates on test performance

| | *Overall* | | *Male* | | *Female* | | *High Performers* | | *Low Performers* | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Panel A: Non-High Schools* | | | | | | | | | | |
| BooksHome (11-25) | 1.200 | [0.848] | 2.461** | [1.020] | −0.100 | [1.231] | 2.536 | [1.574] | −0.435 | [1.805] |
| BooksHome (26-100) | 1.673* | [0.939] | 2.621** | [1.144] | 0.760 | [1.288] | 3.718* | [1.902] | −0.061 | [1.854] |
| BooksHome (101-200) | 2.734*** | [1.028] | 2.062 | [1.470] | 3.331** | [1.568] | 4.305* | [2.322] | 0.419 | [1.982] |
| BooksHome (201-500) | 3.375** | [1.465] | 4.389** | [2.136] | 2.500 | [2.347] | 4.195* | [2.455] | 2.391 | [2.626] |
| BooksHome (over 500) | 2.747* | [1.503] | 3.256 | [2.271] | 2.515 | [2.293] | 1.934 | [3.268] | 0.904 | [3.330] |
| BooksHome (Not Reported) | 3.310*** | [1.254] | 3.882** | [2.038] | 3.066** | [1.340] | 1.982 | [3.686] | 1.831 | [2.460] |
| GradeMidTerm | −1.288*** | [0.098] | −1.295*** | [0.141] | −1.248*** | [0.141] | −2.610*** | [0.621] | −0.697** | [0.346] |
| TeacherExperience | 0.090** | [0.037] | 0.061 | [0.059] | 0.150*** | [0.048] | −0.006 | [0.039] | 0.181*** | [0.044] |
| TeacherFemale | −1.099 | [0.978] | 0.429 | [0.945] | −2.909** | [1.309] | 1.812 | [1.480] | −2.449** | [1.134] |
| *N* | 1198 | | 665 | | 533 | | 362 | | 408 | |
| *Panel B: High Schools* | | | | | | | | | | |
| BooksHome (11-25) | 3.071* | [1.768] | 4.552** | [1.982] | 2.834 | [3.509] | 4.134 | [3.245] | 2.165 | [2.165] |
| BooksHome (26-100) | 3.953** | [1.627] | 3.695** | [1.848] | 5.846 | [3.745] | 3.282 | [2.710] | 4.976* | [2.919] |
| BooksHome (101-200) | 2.490 | [1.674] | 3.056 | [2.047] | 3.938 | [4.014] | 2.145 | [2.806] | 5.006 | [3.305] |
| BooksHome (201-500) | 5.065*** | [1.614] | 4.990*** | [1.788] | 7.099* | [4.049] | 4.970* | [2.537] | 6.797* | [3.622] |
| BooksHome (over 500) | 5.095*** | [1.720] | 5.254** | [2.079] | 7.680* | [4.131] | 5.173* | [2.831] | 6.027* | [3.201] |
| BooksHome (Not Reported) | 2.936 | [1.828] | 3.473* | [1.954] | 4.943 | [4.772] | 2.700 | []2.499 | 1.238 | [3.593] |
| GradeMidTerm | −1.441*** | [0.129] | −1.581*** | [0.163] | −1.214*** | [0.179] | −2.646*** | [0.319] | −0.741* | [0.417] |
| TeacherExperience | 0.058* | [0.032] | 0.077** | [0.034] | 0.019 | [0.048] | 0.085 | [0.063] | 0.165** | [0.079] |
| TeacherFemale | −1.912** | [0.809] | −3.109*** | [0.800] | −0.387 | [1.148] | −1.808 | [1.139] | −3.854** | [1.853] |
| *N* | 869 | | 504 | | 365 | | 437 | | 144 | |

*Note:* Panel A reports results of negative binomial regression (marginal effects) of the covariates for pupils in Non-High School including school fixed effects. Panel B results of negative binomial regression (marginal effects) of the covariates for pupils in High School including school fixed effects. Dependent variable of the initial regression is points in test. Baseline for books at home is the category 0-10. Standard errors are reported in parentheses and clustered on classroom-level. The number of clusters is 53 in Non-High Schools and 36 in High Schools. * p<0.10, ** p<0.05, *** p<0.01.

## A.6 Selection of incentives

Table 19: Chosen Incentive by Gender and School Type (in percent)

| | *Medal* | *Letter* | *Voucher* | *Surprise* |
|---|---|---|---|---|
| *Non-High Schools (Fisher's exact = 0.136)* | | | | |
| Male (N=208) | 21.63 | 27.88 | 19.23 | 31.25 |
| Female (N=193) | 14.51 | 24.87 | 20.21 | 40.41 |
| *High Schools (Fisher's exact = 0.744)* | | | | |
| Male (N= 163) | 23.31 | 24.54 | 22.70 | 29.45 |
| Female (N=114 ) | 19.30 | 29.82 | 21.93 | 28.95 |
| *High Performers (Fisher's exact = 0.212)* | | | | |
| Male (N= 144) | 26.39 | 17.36 | 25.00 | 31.25 |
| Female (N= 113) | 16.81 | 16.81 | 24.78 | 41.59 |
| *Middle Performers (Fisher's exact = 0.711)* | | | | |
| Male (N= 125) | 22.40 | 29.60 | 18.40 | 29.60 |
| Female (N= 122) | 18.03 | 27.05 | 19.67 | 35.25 |
| *Low Performers (Fisher's exact = 0.584)* | | | | |
| Male (N=102) | 16.67 | 35.29 | 17.65 | 30.39 |
| Female (N= 69) | 10.14 | 43.48 | 17.39 | 28.99 |
| *All School types (Fisher's exact = 0.185)* | | | | |
| Male (N=371) | 22.37 | 26.42 | 20.75 | 30.46 |
| Female (N=307) | 16.29 | 26.71 | 20.85 | 36.16 |

*Note:* This table reports the percentage share of pupils' choice by gender separately for school types and ability levels. Fisher' s exact test reports on the difference in the proportions between boys and girls.

Table 20: Multinomial Logit Model of Chosen Incentives Cont.

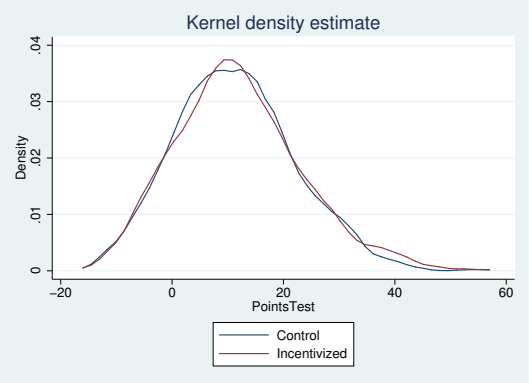
| | *Pooled* | | *Non-High School* | | *High School* | |
|---|---|---|---|---|---|---|
| **C. Voucher** | | | | | | |
| Midterm grade | −0.055** | [0.028] | −0.029 | [0.039] | −0.086** | [0.037] |
| Grade 6 | 0.078 | [0.586] | −0.346 | [0.759] | −0.264 | [1.077] |
| Female pupil | 0.199 | [0.232] | 0.287 | [0.247] | −0.086 | [0.397] |
| *Books at home* | | | | | | |
| (11-25) | −0.287 | [0.282] | −0.374 | [0.289] | 0.201 | [0.530] |
| (26-100) | −0.347 | [0.337] | −0.580* | [0.336] | 0.424 | [0.781] |
| (101-200) | −0.436 | [0.367] | −0.617 | [0.379] | 0.484 | [0.724] |
| (201-500) | −0.516 | [0.495] | −1.405 | [0.918] | 0.281 | [0.810] |
| (over 500) | −0.380 | [0.451] | 0.409 | [0.590] | 0.083 | [0.594] |
| (Not Reported) | −0.340 | [0.499] | −1.299* | [0.690] | 1.001 | [0.942] |
| Teacher experience (years) | 0.018 | [0.028] | 0.072* | [0.038] | −0.039 | [0.040] |
| Day difference | 0.038 | [0.029] | 0.005 | [0.042] | 0.125* | [0.069] |
| Teacher female | −0.317 | [0.608] | −0.218 | [0.864] | −0.778 | [0.869] |
| Unemployment | −0.030 | [0.071] | −0.035 | [0.113] | −0.123 | [0.111] |
| Proportion German | 0.390 | [1.134] | 0.260 | [1.896] | −1.241 | [2.319] |
| Constant | −2.483 | [3.632] | −0.794 | [4.698] | 1.768 | [7.382] |
| **D. Surprise** | | | | | | |
| Midterm grade | −0.048 | [0.031] | −0.031 | [0.031] | −0.117** | [0.056] |
| Grade 6 | 0.536 | [0.513] | 1.470** | [0.744] | −1.274 | [1.253] |
| Female pupil | 0.313* | [0.177] | 0.590*** | [0.200] | −0.005 | [0.280] |
| *Books at home* | | | | | | |
| (11-25) | −0.225 | [0.272] | −0.127 | [0.294] | 0.321 | [0.508] |
| (26-100) | −0.231 | [0.347] | −0.040 | [0.326] | −0.103 | [0.859] |
| (101-200) | 0.008 | [0.365] | 0.047 | [0.328] | 0.437 | [0.917] |
| (201-500) | 0.294 | [0.418] | 0.515 | [0.450] | 0.262 | [0.950] |
| (over 500) | −0.274 | [0.538] | 0.694 | [0.680] | −0.559 | [0.997] |
| (Not Reported) | −0.774* | [0.415] | −0.182 | [0.422] | −13.92*** | [0.904] |
| Teacher experience (years) | 0.015 | [0.024] | 0.025 | [0.033] | −0.009 | [0.038] |
| Day difference | 0.016 | [0.030] | 0.021 | [0.047] | 0.176** | [0.073] |
| Teacher female | 0.821 | [0.546] | 1.087 | [0.842] | −0.327 | [0.940] |
| Unemployment | 0.072 | [0.072] | 0.194 | [0.125] | −0.132 | [0.108] |
| Proportion Germa | −1.144 | [1.108] | −2.480 | [1.747] | −3.760* | [2.272] |
| Constant | −5.258 | [3.304] | −11.77** | [5.179] | 8.948 | [8.798] |
| *N* | 2067 | | 1198 | | 869 | |

*Note:* This table presents the results of a multinomial logit model on the choice of incentive of pupils in the Choice Treatment. The pupils which were not allocated to the Choice Treatment represent the baseline. Midterm grade is the variable of interest, a positive coefficient shows that low performing pupils are more likely to chose the reward as a high midterm grade resembles low performance in the German school system. A negative coefficient shows that high performers are more likely to chose the respective incentive. Covariates: last midterm grade, number of books at home, academic year (grade 5 or 6), gender, teacher' working experience (in years), teachers' gender, day differences between tests, unemployment rate of the school district and the proportion of German speaking pupils within the class.The number of observation is 2.067 for the pooled specification, 869 for High School and 1.098 for Non-High Schools. Standard errors are reported in parentheses and clustered on classroom-level. The number of clusters is 53 in Non-High Schools and 36 in High Schools. * p<0.10, ** p<0.05, *** p<0.01.

# B Kernel density plots by Treatment

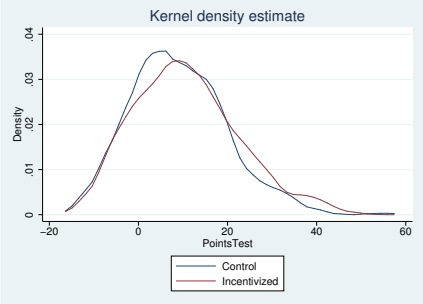
Figure 2: Kernel Density Estimation Control vs. Incentivized

(a) All School Types

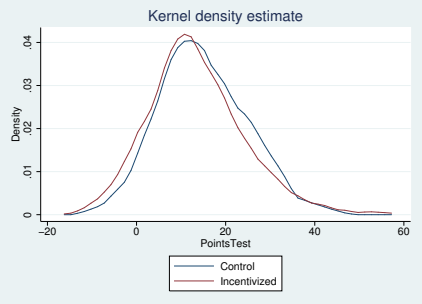



*Note:* This Figure presents Kernel density estimates for the test performance for incentivized and not incentivized pupils pooled over school types.

(b) Non-High School




(c) High School




*Note:* Figure (b) presents Kernel density estimates for the test performance for incentivized and not incentivized pupils in Non-High Schools. Figure (c) presents Kernel density estimates for the test performance for incentivized and not incentivized pupils in High Schools.

# C German School System

In the German school system, children are segregated into high and low performers at an early age. Elementary school in Germany runs from grade one at the age of 6 to grade four at the age of 9 or 10. School-aged children must attend the school in their school district.[27] With the semester report in grade four, parents receive a transition recommendation to which school type to send their child. This recommendation is given by the elementary school and is based on talent and performance (i.e., grades), social skills and social behavior and motivation and learning virtues (Anders, McElvany, and Baumert 2010). However, parents in NRW can decide to which type of secondary school they want to send their children, regardless of the recommendation. Nevertheless, depending on their capacity, secondary schools can decline applications.[28]

The German secondary school system consists mainly of four school types (approximate US equivalents in parenthesis): *Hauptschule* (Secondary General School), *Realschule* (Middle School), *Gesamtschule* (Comprehensive School) and *Gymnasium* (High School). In the following, we use the US equivalents. The average class size consists of 21–28 pupils and a typical week for fifth and sixth graders consists of approximately 37 school hours.[29] Typically, pupils remain in the same class from grade 5 until grade 10, at which time they turn into a course system. Therefore, classes are closed units in which most of the social interaction in pupils' school life takes place. The German grade system ranges from 1 to 6, in which 1 is the highest possible grade and 5 is the threshold for failing—the US equivalents are A+ to F.

The *Secondary General School* (grades five to nine or ten) provides pupils with a basic general education that prepares them, in particular, for a vocational job and finishes with a *Hauptschulabschluss* after grade nine or ten. Depending on performance, pupils can qualify to attend the advanced level of High School.

The *Middle School* (grades five to ten) encourages practical skills as well as interest in theoretical context. Pupils acquire an advanced education and career guidance skills. Furthermore, in grade six, pupils learn a second foreign language. After completion of the tenth grade—and depending on past performance and interest—pupils can change to a vocational training course or attend the advanced level of the High School if his/her grades are good enough. The minimum grade for continuing to High School is 3 on average in all subjects.[30]

Children attending the *Comprehensive School* (grades five to ten or twelve) have a longer period of common learning. Classes consist of children of all skill levels and career decisions are left open as long as possible. The majority of *Comprehensive Schools* are all-day schools in which all degrees of secondary

[27] In 2008, the forced allocation of pupils to the elementary school in their specific district, determined by address, was abolished in the federal state of North Rhine-Westphalia—where we conducted our experiment. This means that parents of the cohorts in our study were free to decide to which elementary school they sent their children.

[28] Criteria that may be used by the school principal for admission decisions are the number of siblings already attending the school, balanced ratios of girls and boys, distance to school and/or lottery procedure (see `http://www.schulministerium.nrw.de/docs/Recht/Schulrecht/APOen/HS-RS-GE-GY-SekI/APO_SI-Stand_-1_07_2013.pdf`).

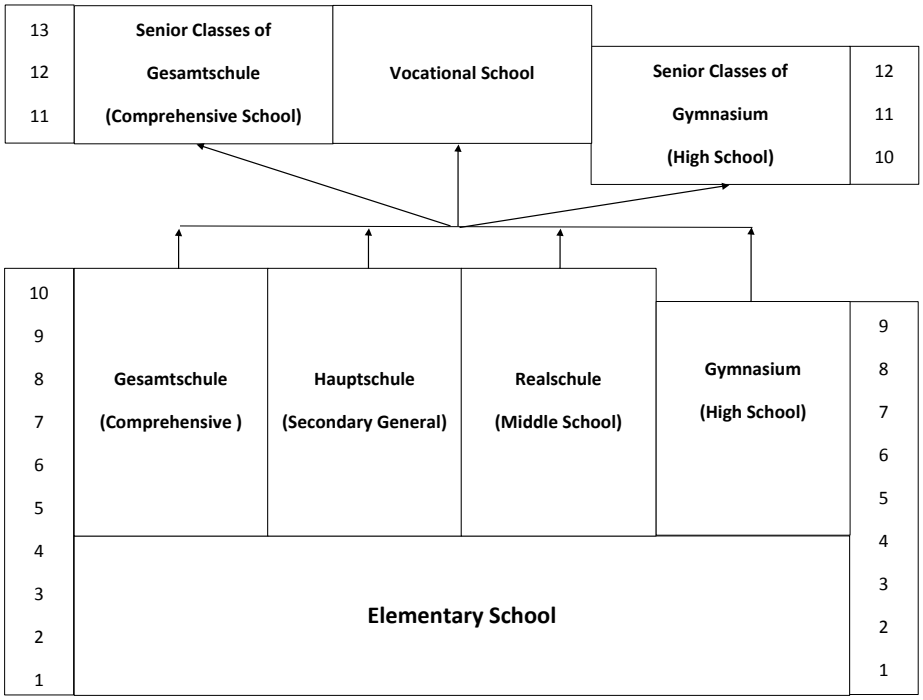
[29] This information is taken from the Ministry of Education and Further Education of NRW. For further information see `http://www.schulministerium.nrw.de/docs/bp/Ministerium/Service/Schulstatistik/Amtliche-Schuldaten/StatTelegramm2012.pdf`.

[30] A sufficient performance (grade 4) in a major subject—German, Mathematics, English—can be compensated by a good performance (grade 2) in another major subject. A maximum of three sufficient performances in a minor subject or two sufficient and one poor performance (grade 5) can be compensated for by an equal number of good performance in other minor subjects.

education can be achieved that are awarded at Secondary General School, Middle School and High School. As with the *Middle School*, pupils can qualify for the advanced level of the High School and obtain the Abitur (A-Level).

The *High School* (grades five to twelve) is the most academic school type. The final examination—the Abitur—entitles students to apply to University.[31] The aim of the High School is to give an in-depth general education, which is necessary for both higher education and for vocational training. The lessons should guide the analysis of complex problems and lead to abstraction, analytical and critical thinking capabilities.

Figure 3: Simplified illustration of the German School System




[31] All other education degrees from Secondary General School and Middle School can also be acquired at the High School.

# D Survey

Figure 4: Survey front and back side

## Student Questionnaire

Alter:_____ Class:_____ School:_____

Gender: _____Male _____Female

Teaching subject:_____

*Before you start, please remember to write down your age, class, school and gender. Completing the questionnaire should not take more than 10 minutes. Please remember to fill out the back side.*

You can get a reward for a good score in a test. Please think about 3 rewards that would motivate you to study for this test. Enter your ideas in the boxes below. You are not allowed to enter money and candies as a reward.

| |
|---|
| 1. |

| |
|---|
| 2. |

| |
|---|
| 3. |

**Please turn the page**

In the table below we listed rewards that you could receive for successful performance in a test. Please read over the table carefully. Then think about the rewards that would motivate you the most. Enter a 1 for the reward you like the most in the box to the left of it, a 2 for the reward you like second best and a 3 for the one you like third best.

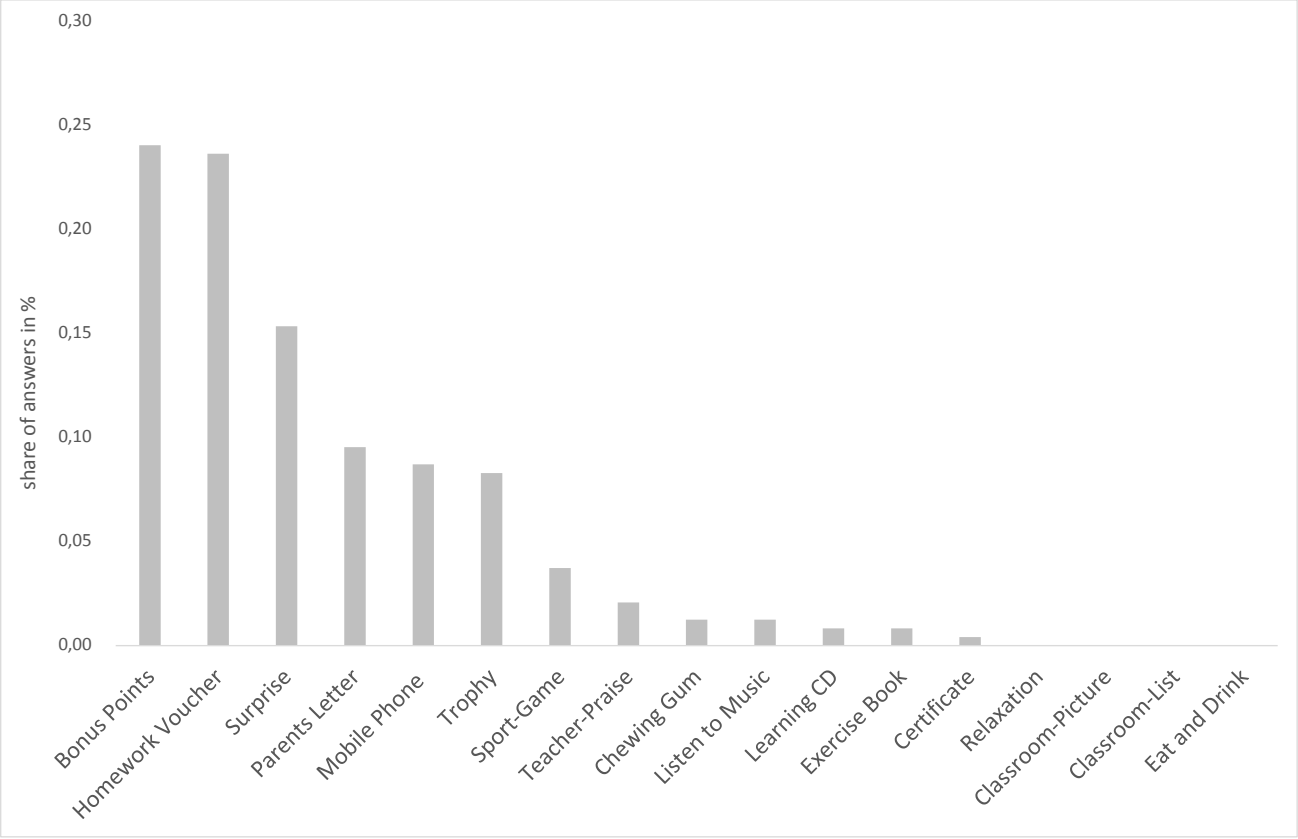
| | | | |
|---|---|---|---|
| | You get a “homework free voucher” in math. The voucher can be used once during the semester | | You are allowed to determine one game in sports teaching hour |
| | You receive a certificate | | You receive bonus points for the next written exam |
| | You get a small trophy | | You are allowed to listen to music in the last 5 min of one lesson |
| | You are allowed to eat a chewing gum during one hour of your choice. | | You are allowed to eat and drink during one hour of your choice. |
| | A picture of everybody who could improve its test score is hung up in the classroom | | You get a small surprise |
| | You get a learning-CD with exciting exercises | | You are allowed to relax in the last 5 min of one lesson |
| | The teacher praises you in front of the class | | A list of everybody who could improve its test score is hung up in the classroom |
| | Your teacher sends a letter to your parents in which he is praising your performance | | You are allowed to use your mobile phone for 5 min in one lesson |
| | You receive a booklet with exciting exercises | | |

**Thank you for your cooperation**

Table 21: Predetermined incentives given in survey

| Work Avoidance | Mastery Goal | Social Appreciation | | Consumption | Curiosity |
|---|---|---|---|---|---|
| | | Public | Private | | |
| *Homework voucher:* No homework in math. Voucher can be used once until the end of semester. | *Exercise Book*: Receiving a booklet with mathematical exercises | *Teacher-Praise*: Being Praised in front of the class | *Parents-Letter*: Teacher sends letter to parents, praising the pupil's performance | *Chewing Gum:* Being allowed to eat a chewing gum in one lesson | *Surprise*: Getting a small surprise reward |
| *Relaxation*: Relaxing 5 minutes of one lesson | *Learning CD*: Receiving a CD with mathematical games | *Classroom-Picture*: Picture of pupil who could improve its test score is hung up in classroom | *Trophy*: Getting a small Medal | *Listen to Music*: Being allowed to listen to music for 5 min | |
| *Bonus points*: Receiving extra points for next written exam | | *Classroom-List*: List of all who could improve their test score is hung up in classroom | *Certificate:* Receiving a certificate stating that test score could be improved | *Eat and Drink:* Being allowed to eat and drink during class | |
| | | | | *Mobile Phone*: Being allowed to play 5 min with mobile in one lesson | |
| | | | | *Sport-Game*: Determing one game in sports teaching hour | |

Figure 5: Survey Answers



# E Facsimile of Incentives

Figure 6: Example of No-Homework-Voucher

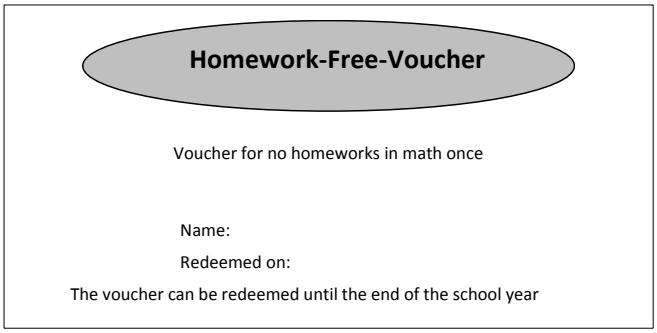


Figure 7: Example of Parents-Letter

*Dear Ms / Mr _____,*

*We are pleased to inform you that*

*Your son _____ / daughter _____, Class ___,*

*participated particularly engaged and motivated on a test in mathematics. We are pleased that _____ could improve compared to the current report mark and we hope that _____ continues his work as exemplary.*

*Sincerely yours*

_____

*Name of Teacher*

*Date: _____*

Figure 8: Picture of Medal

# PREVIOUS DISCUSSION PAPERS

203 Wagner, Valentin and Riener, Gerhard, Peers or Parents? On Non-Monetary Incentives in Schools, November 2015.

202 Gaudin, Germain, Pass-Through, Vertical Contracts, and Bargains, November 2015.
Forthcoming in: Economics Letters.

201 Demeulemeester, Sarah and Hottenrott, Hanna, R&D Subsidies and Firms' Cost of Debt, November 2015.

200 Kreickemeier, Udo and Wrona, Jens, Two-Way Migration Between Similar Countries, October 2015.
Forthcoming in: World Economy.

199 Haucap, Justus and Stühmeier, Torben, Competition and Antitrust in Internet Markets, October 2015.
Forthcoming in: Bauer, J. and M. Latzer (Eds.), Handbook on the Economics of the Internet, Edward Elgar: Cheltenham 2016.

198 Alipranti, Maria, Milliou, Chrysovalantou and Petrakis, Emmanuel, On Vertical Relations and the Timing of Technology, October 2015.

197 Kellner, Christian, Reinstein, David and Riener, Gerhard, Stochastic Income and Conditional Generosity, October 2015.

196 Chlaß, Nadine and Riener, Gerhard, Lying, Spying, Sabotaging: Procedures and Consequences, September 2015.

195 Gaudin, Germain, Vertical Bargaining and Retail Competition: What Drives Countervailing Power?, September 2015.

194 Baumann, Florian and Friehe, Tim, Learning-by-Doing in Torts: Liability and Information About Accident Technology, September 2015.

193 Defever, Fabrice, Fischer, Christian and Suedekum, Jens, Relational Contracts and Supplier Turnover in the Global Economy, August 2015.

192 Gu, Yiquan and Wenzel, Tobias, Putting on a Tight Leash and Levelling Playing Field: An Experiment in Strategic Obfuscation and Consumer Protection, July 2015.
Published in: International Journal of Industrial Organization, 42 (2015), pp. 120-128.

191 Ciani, Andrea and Bartoli, Francesca, Export Quality Upgrading under Credit Constraints, July 2015.

190 Hasnas, Irina and Wey, Christian, Full Versus Partial Collusion among Brands and Private Label Producers, July 2015.

189 Dertwinkel-Kalt, Markus and Köster, Mats, Violations of First-Order Stochastic Dominance as Salience Effects, June 2015.
Forthcoming in. Journal of Behavioral and Experimental Economics.

188 Kholodilin, Konstantin, Kolmer, Christian, Thomas, Tobias and Ulbricht, Dirk, Asymmetric Perceptions of the Economy: Media, Firms, Consumers, and Experts, June 2015.

187 Dertwinkel-Kalt, Markus and Wey, Christian, Merger Remedies in Oligopoly under a Consumer Welfare Standard, June 2015
Forthcoming in: Journal of Law, Economics, & Organization.

186 Dertwinkel-Kalt, Markus, Salience and Health Campaigns, May 2015
Forthcoming in: Forum for Health Economics & Policy

185 Wrona, Jens, Border Effects without Borders: What Divides Japan's Internal Trade?, May 2015.

184 Amess, Kevin, Stiebale, Joel and Wright, Mike, The Impact of Private Equity on Firms' Innovation Activity, April 2015.
Forthcoming in: European Economic Review.

183 Ibañez, Marcela, Rai, Ashok and Riener, Gerhard, Sorting Through Affirmative Action: Three Field Experiments in Colombia, April 2015.

182 Baumann, Florian, Friehe, Tim and Rasch, Alexander, The Influence of Product Liability on Vertical Product Differentiation, April 2015.

181 Baumann, Florian and Friehe, Tim, Proof beyond a Reasonable Doubt: Laboratory Evidence, March 2015.

180 Rasch, Alexander and Waibel, Christian, What Drives Fraud in a Credence Goods Market? – Evidence from a Field Study, March 2015.

179 Jeitschko, Thomas D., Incongruities of Real and Intellectual Property: Economic Concerns in Patent Policy and Practice, February 2015.
Forthcoming in: Michigan State Law Review.

178 Buchwald, Achim and Hottenrott, Hanna, Women on the Board and Executive Duration – Evidence for European Listed Firms, February 2015.

177 Heblich, Stephan, Lameli, Alfred and Riener, Gerhard, Regional Accents on Individual Economic Behavior: A Lab Experiment on Linguistic Performance, Cognitive Ratings and Economic Decisions, February 2015
Published in: PLoS ONE, 10 (2015), e0113475.

176 Herr, Annika, Nguyen, Thu-Van and Schmitz, Hendrik, Does Quality Disclosure Improve Quality? Responses to the Introduction of Nursing Home Report Cards in Germany, February 2015.

175 Herr, Annika and Normann, Hans-Theo, Organ Donation in the Lab: Preferences and Votes on the Priority Rule, February 2015.
Forthcoming in: Journal of Economic Behavior and Organization.

174 Buchwald, Achim, Competition, Outside Directors and Executive Turnover: Implications for Corporate Governance in the EU, February 2015.

173 Buchwald, Achim and Thorwarth, Susanne, Outside Directors on the Board, Competition and Innovation, February 2015.

172 Dewenter, Ralf and Giessing, Leonie, The Effects of Elite Sports Participation on Later Job Success, February 2015.

171 Haucap, Justus, Heimeshoff, Ulrich and Siekmann, Manuel, Price Dispersion and Station Heterogeneity on German Retail Gasoline Markets, January 2015.

170 Schweinberger, Albert G. and Suedekum, Jens, De-Industrialisation and Entrepreneurship under Monopolistic Competition, January 2015
Published in: Oxford Economic Papers, 67 (2015), pp. 1174-1185.

169 Nowak, Verena, Organizational Decisions in Multistage Production Processes, December 2014.

168 Benndorf, Volker, Kübler, Dorothea and Normann, Hans-Theo, Privacy Concerns, Voluntary Disclosure of Information, and Unraveling: An Experiment, November 2014.
Published in: European Economic Review, 75 (2015), pp. 43-59.

167 Rasch, Alexander and Wenzel, Tobias, The Impact of Piracy on Prominent and Non-prominent Software Developers, November 2014.
Published in: Telecommunications Policy, 39 (2015), pp. 735-744.

166 Jeitschko, Thomas D. and Tremblay, Mark J., Homogeneous Platform Competition with Endogenous Homing, November 2014.

165 Gu, Yiquan, Rasch, Alexander and Wenzel, Tobias, Price-sensitive Demand and Market Entry, November 2014
Forthcoming in: Papers in Regional Science.

164 Caprice, Stéphane, von Schlippenbach, Vanessa and Wey, Christian, Supplier Fixed Costs and Retail Market Monopolization, October 2014.

163 Klein, Gordon J. and Wendel, Julia, The Impact of Local Loop and Retail Unbundling Revisited, October 2014.

162 Dertwinkel-Kalt, Markus, Haucap, Justus and Wey, Christian, Raising Rivals' Costs through Buyer Power, October 2014.
Published in: Economics Letters, 126 (2015), pp.181-184.

161 Dertwinkel-Kalt, Markus and Köhler, Katrin, Exchange Asymmetries for Bads? Experimental Evidence, October 2014.

160 Behrens, Kristian, Mion, Giordano, Murata, Yasusada and Suedekum, Jens, Spatial Frictions, September 2014.

159 Fonseca, Miguel A. and Normann, Hans-Theo, Endogenous Cartel Formation: Experimental Evidence, August 2014.
Published in: Economics Letters, 125 (2014), pp. 223-225.

158 Stiebale, Joel, Cross-Border M&As and Innovative Activity of Acquiring and Target Firms, August 2014.

157 Haucap, Justus and Heimeshoff, Ulrich, The Happiness of Economists: Estimating the Causal Effect of Studying Economics on Subjective Well-Being, August 2014.
Published in: International Review of Economics Education, 17 (2014), pp. 85-97.

156 Haucap, Justus, Heimeshoff, Ulrich and Lange, Mirjam R. J., The Impact of Tariff Diversity on Broadband Diffusion – An Empirical Analysis, August 2014.
Forthcoming in: Telecommunications Policy.

155 Baumann, Florian and Friehe, Tim, On Discovery, Restricting Lawyers, and the Settlement Rate, August 2014.

154 Hottenrott, Hanna and Lopes-Bento, Cindy, R&D Partnerships and Innovation Performance: Can There be too Much of a Good Thing?, July 2014.

153 Hottenrott, Hanna and Lawson, Cornelia, Flying the Nest: How the Home Department Shapes Researchers' Career Paths, July 2015 (First Version July 2014).
Forthcoming in: Studies in Higher Education.

152 Hottenrott, Hanna, Lopes-Bento, Cindy and Veugelers, Reinhilde, Direct and Cross-Scheme Effects in a Research and Development Subsidy Program, July 2014.

151 Dewenter, Ralf and Heimeshoff, Ulrich, Do Expert Reviews Really Drive Demand? Evidence from a German Car Magazine, July 2014.
Published in: Applied Economics Letters, 22 (2015), pp. 1150-1153.

150 Bataille, Marc, Steinmetz, Alexander and Thorwarth, Susanne, Screening Instruments for Monitoring Market Power in Wholesale Electricity Markets – Lessons from Applications in Germany, July 2014.

149 Kholodilin, Konstantin A., Thomas, Tobias and Ulbricht, Dirk, Do Media Data Help to Predict German Industrial Production?, July 2014.

148 Hogrefe, Jan and Wrona, Jens, Trade, Tasks, and Trading: The Effect of Offshoring on Individual Skill Upgrading, June 2014.
Forthcoming in: Canadian Journal of Economics.

147 Gaudin, Germain and White, Alexander, On the Antitrust Economics of the Electronic Books Industry, September 2014 (Previous Version May 2014).

146 Alipranti, Maria, Milliou, Chrysovalantou and Petrakis, Emmanuel, Price vs. Quantity Competition in a Vertically Related Market, May 2014.
Published in: Economics Letters, 124 (2014), pp. 122-126.

145 Blanco, Mariana, Engelmann, Dirk, Koch, Alexander K. and Normann, Hans-Theo, Preferences and Beliefs in a Sequential Social Dilemma: A Within-Subjects Analysis, May 2014.
Published in: Games and Economic Behavior, 87 (2014), pp. 122-135.

144 Jeitschko, Thomas D., Jung, Yeonjei and Kim, Jaesoo, Bundling and Joint Marketing by Rival Firms, May 2014.

143 Benndorf, Volker and Normann, Hans-Theo, The Willingness to Sell Personal Data, April 2014.

142 Dauth, Wolfgang and Suedekum, Jens, Globalization and Local Profiles of Economic Growth and Industrial Change, April 2014.

141 Nowak, Verena, Schwarz, Christian and Suedekum, Jens, Asymmetric Spiders: Supplier Heterogeneity and the Organization of Firms, April 2014.

140 Hasnas, Irina, A Note on Consumer Flexibility, Data Quality and Collusion, April 2014.

139 Baye, Irina and Hasnas, Irina, Consumer Flexibility, Data Quality and Location Choice, April 2014.

138 Aghadadashli, Hamid and Wey, Christian, Multi-Union Bargaining: Tariff Plurality and Tariff Competition, April 2014.
Published in: Journal of Institutional and Theoretical Economics (JITE), 171 (2015), pp. 666-695.

137 Duso, Tomaso, Herr, Annika and Suppliet, Moritz, The Welfare Impact of Parallel Imports: A Structural Approach Applied to the German Market for Oral Anti-diabetics, April 2014.
Published in: Health Economics, 23 (2014), pp. 1036-1057.

136 Haucap, Justus and Müller, Andrea, Why are Economists so Different? Nature, Nurture and Gender Effects in a Simple Trust Game, March 2014.

135 Normann, Hans-Theo and Rau, Holger A., Simultaneous and Sequential Contributions to Step-Level Public Goods: One vs. Two Provision Levels, March 2014.
Published in: Journal of Conflict Resolution, 59 (2015), pp.1273-1300.

134 Bucher, Monika, Hauck, Achim and Neyer, Ulrike, Frictions in the Interbank Market and Uncertain Liquidity Needs: Implications for Monetary Policy Implementation, July 2014 (First Version March 2014).

133 Czarnitzki, Dirk, Hall, Bronwyn, H. and Hottenrott, Hanna, Patents as Quality Signals? The Implications for Financing Constraints on R&D?, February 2014.

132 Dewenter, Ralf and Heimeshoff, Ulrich, Media Bias and Advertising: Evidence from a German Car Magazine, February 2014.
Published in: Review of Economics, 65 (2014), pp. 77-94.

131 Baye, Irina and Sapi, Geza, Targeted Pricing, Consumer Myopia and Investment in Customer-Tracking Technology, February 2014.

130 Clemens, Georg and Rau, Holger A., Do Leniency Policies Facilitate Collusion? Experimental Evidence, January 2014.

129 Hottenrott, Hanna and Lawson, Cornelia, Fishing for Complementarities: Competitive Research Funding and Research Productivity, December 2013.

128 Hottenrott, Hanna and Rexhäuser, Sascha, Policy-Induced Environmental Technology and Inventive Efforts: Is There a Crowding Out?, December 2013.
Published in: Industry and Innovation, 22 (2015), pp.375-401.

127 Dauth, Wolfgang, Findeisen, Sebastian and Suedekum, Jens, The Rise of the East and the Far East: German Labor Markets and Trade Integration, December 2013.
Published in: Journal of the European Economic Association, 12 (2014), pp. 1643-1675.

126 Wenzel, Tobias, Consumer Myopia, Competition and the Incentives to Unshroud Add-on Information, December 2013.
Published in: Journal of Economic Behavior and Organization, 98 (2014), pp. 89-96.

125 Schwarz, Christian and Suedekum, Jens, Global Sourcing of Complex Production Processes, December 2013.
Published in: Journal of International Economics, 93 (2014), pp. 123-139.

124 Defever, Fabrice and Suedekum, Jens, Financial Liberalization and the Relationship-Specificity of Exports, December 2013.
Published in: Economics Letters, 122 (2014), pp. 375-379.

123 Bauernschuster, Stefan, Falck, Oliver, Heblich, Stephan and Suedekum, Jens, Why are Educated and Risk-Loving Persons More Mobile Across Regions?, December 2013.
Published in: Journal of Economic Behavior and Organization, 98 (2014), pp. 56-69.

122 Hottenrott, Hanna and Lopes-Bento, Cindy, Quantity or Quality? Knowledge Alliances and their Effects on Patenting, December 2013.
Published in: Industrial and Corporate Change, 24 (2015), pp. 981-1011.

121 Hottenrott, Hanna and Lopes-Bento, Cindy, (International) R&D Collaboration and SMEs: The Effectiveness of Targeted Public R&D Support Schemes, December 2013.
Published in: Research Policy, 43 (2014), pp.1055-1066.

120 Giesen, Kristian and Suedekum, Jens, City Age and City Size, November 2013.
Published in: European Economic Review, 71 (2014), pp. 193-208.

119 Trax, Michaela, Brunow, Stephan and Suedekum, Jens, Cultural Diversity and Plant-Level Productivity, November 2013.

118 Manasakis, Constantine and Vlassis, Minas, Downstream Mode of Competition with Upstream Market Power, November 2013.
Published in: Research in Economics, 68 (2014), pp. 84-93.

117 Sapi, Geza and Suleymanova, Irina, Consumer Flexibility, Data Quality and Targeted Pricing, November 2013.

116 Hinloopen, Jeroen, Müller, Wieland and Normann, Hans-Theo, Output Commitment through Product Bundling: Experimental Evidence, November 2013.
Published in: European Economic Review, 65 (2014), pp. 164-180.

115 Baumann, Florian, Denter, Philipp and Friehe Tim, Hide or Show? Endogenous Observability of Private Precautions against Crime When Property Value is Private Information, November 2013.

114 Fan, Ying, Kühn, Kai-Uwe and Lafontaine, Francine, Financial Constraints and Moral Hazard: The Case of Franchising, November 2013.

113 Aguzzoni, Luca, Argentesi, Elena, Buccirossi, Paolo, Ciari, Lorenzo, Duso, Tomaso, Tognoni, Massimo and Vitale, Cristiana, They Played the Merger Game: A Retrospective Analysis in the UK Videogames Market, October 2013.
Published under the title: “A Retrospective Merger Analysis in the UK Videogame Market” in: Journal of Competition Law and Economics, 10 (2014), pp. 933-958.

112 Myrseth, Kristian Ove R., Riener, Gerhard and Wollbrant, Conny, Tangible Temptation in the Social Dilemma: Cash, Cooperation, and Self-Control, October 2013.

111 Hasnas, Irina, Lambertini, Luca and Palestini, Arsen, Open Innovation in a Dynamic Cournot Duopoly, October 2013.
Published in: Economic Modelling, 36 (2014), pp. 79-87.

110 Baumann, Florian and Friehe, Tim, Competitive Pressure and Corporate Crime, September 2013.

109 Böckers, Veit, Haucap, Justus and Heimeshoff, Ulrich, Benefits of an Integrated European Electricity Market, September 2013.

108 Normann, Hans-Theo and Tan, Elaine S., Effects of Different Cartel Policies: Evidence from the German Power-Cable Industry, September 2013.
Published in: Industrial and Corporate Change, 23 (2014), pp. 1037-1057.

107 Haucap, Justus, Heimeshoff, Ulrich, Klein, Gordon J., Rickert, Dennis and Wey, Christian, Bargaining Power in Manufacturer-Retailer Relationships, September 2013.

106 Baumann, Florian and Friehe, Tim, Design Standards and Technology Adoption: Welfare Effects of Increasing Environmental Fines when the Number of Firms is Endogenous, September 2013.

105 Jeitschko, Thomas D., NYSE Changing Hands: Antitrust and Attempted Acquisitions of an Erstwhile Monopoly, August 2013.
Published in: Journal of Stock and Forex Trading, 2 (2) (2013), pp. 1-6.

104 Böckers, Veit, Giessing, Leonie and Rösch, Jürgen, The Green Game Changer: An Empirical Assessment of the Effects of Wind and Solar Power on the Merit Order, August 2013.

103 Haucap, Justus and Muck, Johannes, What Drives the Relevance and Reputation of Economics Journals? An Update from a Survey among Economists, August 2013.
Published in: Scientometrics, 103 (2015), pp. 849-877.

102 Jovanovic, Dragan and Wey, Christian, Passive Partial Ownership, Sneaky Takeovers, and Merger Control, August 2013.
Published in: Economics Letters, 125 (2014), pp. 32-35.

101 Haucap, Justus, Heimeshoff, Ulrich, Klein, Gordon J., Rickert, Dennis and Wey, Christian, Inter-Format Competition among Retailers – The Role of Private Label Products in Market Delineation, August 2013.

100 Normann, Hans-Theo, Requate, Till and Waichman, Israel, Do Short-Term Laboratory Experiments Provide Valid Descriptions of Long-Term Economic Interactions? A Study of Cournot Markets, July 2013.
Published in: Experimental Economics, 17 (2014), pp. 371-390.

99 Dertwinkel-Kalt, Markus, Haucap, Justus and Wey, Christian, Input Price Discrimination (Bans), Entry and Welfare, June 2013.
Forthcoming under the title “Procompetitive Dual Pricing” in: European Journal of Law and Economics.

98 Aguzzoni, Luca, Argentesi, Elena, Ciari, Lorenzo, Duso, Tomaso and Tognoni, Massimo, Ex-post Merger Evaluation in the UK Retail Market for Books, June 2013.
Forthcoming in: Journal of Industrial Economics.

97 Caprice, Stéphane and von Schlippenbach, Vanessa, One-Stop Shopping as a Cause of Slotting Fees: A Rent-Shifting Mechanism, May 2012.
Published in: Journal of Economics and Management Strategy, 22 (2013), pp. 468-487.

96 Wenzel, Tobias, Independent Service Operators in ATM Markets, June 2013.
Published in: Scottish Journal of Political Economy, 61 (2014), pp. 26-47.

95 Coublucq, Daniel, Econometric Analysis of Productivity with Measurement Error: Empirical Application to the US Railroad Industry, June 2013.

94 Coublucq, Daniel, Demand Estimation with Selection Bias: A Dynamic Game Approach with an Application to the US Railroad Industry, June 2013.

93 Baumann, Florian and Friehe, Tim, Status Concerns as a Motive for Crime?, April 2013.
Published in: International Review of Law and Economics, 43 (2015), pp. 46-55.

92 Jeitschko, Thomas D. and Zhang, Nanyun, Adverse Effects of Patent Pooling on Product Development and Commercialization, April 2013.
Published in: The B. E. Journal of Theoretical Economics, 14 (1) (2014), Art. No. 2013-0038.

91 Baumann, Florian and Friehe, Tim, Private Protection Against Crime when Property Value is Private Information, April 2013.
Published in: International Review of Law and Economics, 35 (2013), pp. 73-79.

90 Baumann, Florian and Friehe, Tim, Cheap Talk About the Detection Probability, April 2013.
Published in: International Game Theory Review, 15 (2013), Art. No. 1350003.

89 Pagel, Beatrice and Wey, Christian, How to Counter Union Power? Equilibrium Mergers in International Oligopoly, April 2013.

88 Jovanovic, Dragan, Mergers, Managerial Incentives, and Efficiencies, April 2014 (First Version April 2013).

87 Heimeshoff, Ulrich and Klein, Gordon J., Bargaining Power and Local Heroes, March 2013.

86 Bertschek, Irene, Cerquera, Daniel and Klein, Gordon J., More Bits – More Bucks? Measuring the Impact of Broadband Internet on Firm Performance, February 2013.
Published in: Information Economics and Policy, 25 (2013), pp. 190-203.

85 Rasch, Alexander and Wenzel, Tobias, Piracy in a Two-Sided Software Market, February 2013.
Published in: Journal of Economic Behavior & Organization, 88 (2013), pp. 78-89.

84 Bataille, Marc and Steinmetz, Alexander, Intermodal Competition on Some Routes in Transportation Networks: The Case of Inter Urban Buses and Railways, January 2013.

83 Haucap, Justus and Heimeshoff, Ulrich, Google, Facebook, Amazon, eBay: Is the Internet Driving Competition or Market Monopolization?, January 2013.
Published in: International Economics and Economic Policy, 11 (2014), pp. 49-61.

Older discussion papers can be found online at:
http://ideas.repec.org/s/zbw/dicedp.html

**Heinrich-Heine-University of Düsseldorf**

**Düsseldorf Institute for Competition Economics (DICE)**

Universitätsstraße 1 _ 40225 Düsseldorf
www.dice.hhu.de

ISSN 2190-9938 (online)
ISBN 978-3-86304-202-8