Journal of Research on Educational Effectiveness
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uree20

Monetary and Non-Monetary Student Incentives for Tutoring Services: A Randomized Controlled Trial

Matthew G. Springer Professor[a], Brooks Rosenquist[a] & Walker A. Swain[a]
[a] National Center on Performance Incentives, Peabody College of Vanderbilt University
Accepted author version posted online: 11 Apr 2015.

Monetary and Non-Monetary Student Incentives for Tutoring Services: A Randomized Controlled Trial

Matthew G. Springer, Professor[1,*] Brooks Rosenquist[1], and Walker A. Swain[1]

[1]National Center on Performance Incentives, Peabody College of Vanderbilt University

[*]Corresponding Author Email: matthew.g.springer@vanderbilt.edu, Vanderbilt University, Nashville, 37204 United States

Abstract

In recent years, the largely punitive accountability measures imposed by the 2001 No Child Left Behind Act have given way to an emphasis on financial incentives. While most policy interventions have focused primarily on linking teacher compensation to student test scores, several recent studies have examined the prospects for the use of financial incentives for students to improve student outcomes. This study builds on the existing literature by comparing student responses to monetary and non-monetary (certificates of recognition) incentives to increase participation in federally funded supplemental education services (SEdS). Approximately 300 eligible middle grade students who had signed up for SEdS were randomly selected and then assigned to either a control condition that received no incentive, a monetary reward group that could earn up to $100 for regular attendance, or a non-monetary group that could receive a certificate of recognition signed by the district superintendent. Although the benefits of the monetary incentives were negligible, the students in the certificate group attended 42.5 percent more of their allotted tutoring hours than those assigned to control. The effect of the certificate was particularly strong for female students, who attended 26 percent more of their allocated

tutoring hours compared to males who were also offered certificates. These results suggest the need for further research into the role of non-monetary incentives in motivating student behaviors. Also, the findings could be useful to policymakers at the state or district level seeking cost effective mechanisms to increase uptake of underutilized student supports.

Keywords

financial incentives, experiment, accountability

1. Introduction

In recent years, the largely punitive accountability measures imposed by the 2001 federal No Child Left Behind Act have given way to an emphasis on utilization of financial incentives in the Obama administrationôs requirements for Race to the Top competitive grants and NCLB waivers. The incentives pushed by the Department of Education have primarily focused on linking teacher compensation to student test score data. However, education policy researchers have also sought to examine the prospects for incentivizing another important group of participants in the education production functionð the students. This study seeks to build on a small but growing body of research on student incentives by comparing studentsôresponses to monetary and non-monetary participation incentives in a randomized controlled trial. Specifically, we examine whether different types of incentives can improve attendance of the underutilized, federally funded supplemental education services (SEdS).

The 2001 reauthorization of the Elementary and Secondary Education Act, commonly referred to as NCLB, required districts to make available free after-school tutoring for low-income students attending a Title I school that had failed to make adequate yearly progress towards its accountability goals (Springer, Pepper, Gardner, & Bower, 2009). Evidence of the effectiveness of these programs has ranged from positive to mixed and negligible.[1] However, while findings of impacts on achievement are markedly divergent, evaluations consistently find

[1] Studies finding positive impacts in mathematics and reading include: Rickles and Barnhart, 2007; Springer, Pepper, and Gosh-Dastidar, 2014; Zimmer et al, 2006; and Zimmer et al, 2007. Studies with findings of mixed impacts include: Heistad, 2007; and Rickles and White, 2006. Potter et al, 2007; Heinrich, Meyer, and Whitten, 2010; and Deke et al, 2012 had null findings.

that studentsô utilization of free SEdS programs is strikingly low. Analyzing SEdS in five large school districts, Berger and colleagues (2011) found that on average, only 18 percent of students eligible to participate registered for SEdS. Of those eligible students who did register, 28 percent never attended one tutoring session. Because participation in this kind of after-school tutoring is voluntary for students, it often competes with other extracurricular activities, and attendance typically declines as the school year progresses (GAO, 2006). Heinrich and colleaguesô 2010 analysis cautions against drawing causal conclusions but notes that student attendance at Milwaukee SEdS dropped dramatically (from 64% to 34%) in the year following restrictions that limited the use of incentives to encourage attendance to those ñdeemed educational (e.g., books, educational software, magazines, museum field trips, etc.)ò and explicitly prohibited vendors from offering ñmore popular incentives such as iPods, mall gift cards, movie passes, and pizza. (p. 26)ò

This lack of persistence in attendance appears to be problematic. In a 2012 review of studies of SEdS effectiveness in raising student test scores, Heinrich and Burch estimate that attendance of approximately 40 hours of tutoring may represent a ñcritical threshold,ò below which student test score gains are not typically realized. Research using data from the particularly large SEdS program in Chicago Public Schools, also found a significant dosage effect for each additional hour students attended at site-based tutoring (Heinrich & Nisar, 2013). While it is worth noting that research has documented dramatic variation in the quality and effectiveness of SEdS provision (e.g. Heinrich et al. 2014), this study does not aim to assess the impact of SEdS. Rather, we use the federally funded tutoring as an example of a poorly attended service for the purpose of evaluating the effects of different forms of student incentives.

Recent experimental evaluations of student incentives in the United States suggest that incentives are more likely to increase student achievement when targeted at inputs to the education production function, like attendance, rather than rewarding particular outcomes (Fryer, 2011). For example, monetary incentives for reading books or doing math practice problems (similar to what a student might do at tutoring sessions) improve achievement, where cash rewards for higher test scores or better grades produce no change. Here, we attempt to assess the effectiveness of incentives for attendance at SEdS.

In the 2010-11 school year, we collaborated with a large, Southern urban school district to conduct a randomized controlled trial evaluating the effectiveness of two types of incentives for student attendance at SEdS. Approximately 300 students who had signed up for SEdS were randomly selected and then assigned to one of three groups. Students assigned to the monetary treatment group could earn up to $100 for consistent attendance, those in the non-monetary group were eligible for certificates of recognition signed by the superintendent, and control students received no experimental incentives. The study focuses primarily on two research questions: What is the impact of monetary and non-monetary incentives on student attendance? And does the response to the incentives vary by gender? We also conduct exploratory analyses of the association between the incentive programs and studentsô intrinsic motivation using post-treatment survey results.

2. Review of Relevant Literature

2.1 SEdS and Student Achievement

Findings from evaluations of the effectiveness of SEdS have been decidedly mixed. Though they sometimes find positive effects for SEdS on student achievement (particularly in math), they

uniformly find low levels of registration, attendance, and persistence among students eligible to participate. Furthermore, early evaluations indicate the availability of quality SEdS providers is highly inconsistent (Heinrich, Meyer, & Whitten, 2010).

Site-specific studies of SEdS in Tennessee, Illinois, and Pennsylvania, as well as one national study of programs in large urban school districts show evidence of academic benefits for students who attend. Springer et. al. (2014) examined the effect of SEdS on student test score gains in Nashville, TN using data from 2003 to 2008. They consistently found statistically significant positive effects on test score gains in mathematics. The effects on test score gains in reading were not statistically significant. Two evaluations of SEdS in Chicago Public Schools, IL found larger gains in math and reading on the Iowa Test of Basic Skills for participating students who attended regularly (at least 30-40 hours) compared to eligible classmates in the same schools who did not receive SEdS (CPS, 2007). Similarly, a RAND study using data from 7 large urban school districts, found positive, statistically significant effects for SEdS registration in mathematics and reading in 5 out of the 7 districts they assessed with student fixed effects models (Zimmer et al. 2007). An evaluation of SEdS based tutoring in Pittsburg, PA conducted by the same research group found larger effects for math (0.15-0.20 depending on the model specifications) and no effects for reading (Zimmer et al. 2009).

Alternatively, evaluations of SEdS in Wisconsin, California, and Minnesota found no evidence of effects on participants. A study using similar methods in Milwaukee Public Schools, WI had null findings, and no consistent statistical relationship between a studentôs level of attendance and achievement (Heinrich et al., 2010). Studies in Minnesota (Burch 2007) and six other school districts (Deke et al, 2012) reported no evidence of effects on participants.

Perhaps most pertinent to this study, however, Heinrich and Nisar (2013), using rich longitudinal data from the large Chicago Public Schools SEdS system, found large positive effects for some providers, particularly at school-based programs where attendance was high, and estimated a consistent dosage response to additional hours of tutoring. While the authors employ a variety of sophisticated techniques to approximate causal estimates, they acknowledge that the differential rates of attendance pose a significant selection bias threat.

In sum, the literature on SEdS effectiveness, though mixed, is sufficient to classify the free services as potentially beneficial, and to perceive low rates of take-up and attendance as a policy problem. If students or parents are undervaluing the potential long-term benefits of attending such programs, one potential mechanism for shifting their calculation is the introduction of additional incentives, such as those examined in this study.

2.2 Student Incentives

That juvenile learners ï and especially adolescents ï may be under-motivated to achieve in school and apply suboptimal attention and effort is a theme that reoccurs frequently in educational research (Coleman, 1961; Finn, 1993; Fredricks, Blumenfeld, & Paris, 2004; National Research Council, 2003). The problems that accompany a lack of academic engagement are not unique to the U.S. setting, but have been identified as problematic across diverse international contexts (PISA, 2006; Wilms, 2003). A growing body of recent research examines policies that offer incentives for specific student activities or achievements to increase academic engagement, attendance, and effort in the US and around the world (e.g., Allan & Fryer, 2011; Fryer, 2011, 2012; Bettinger, 2010; Cha & Patel, 2010; Angrist & Lavy, 2009; Behrman et al 2011; Janvry et al, 2004; Janvry, 2006).

The primary finding from a series of randomized cash incentive programs in US cities was that incentives increase student achievement only when targeted at inputs to the education production function (Fryer, 2011). In Washington, D.C. and Dallas, TX, where the programs incentivized specific behaviors, the program effects were statistically significant and positive on both reading comprehension and reading exams; effects in vocabulary were positive but not statistically significant. In both cases however, treatment students performed no better on state exams. In New York City and Chicago, however, where the programs incentivized student outcomes, Fryer found no statistically significant impact on student achievement as the result of the treatment.[2]

In a later study that sought to align parental, student and teacher cash incentives for math achievement, students in the treatment school mastered a full standard deviation more math objectives on a computer exam, and parents attended nearly twice as many parent-teacher conferences (Fryer, 2012). Unfortunately the students scored worse on non-incentivized subjects than the students in the control schools. This suggests that the cash incentives promoted substitution of efforts away from other academic tasks.

One of the few studies to find effects for cash incentives for student test scores was Bettingerôs (2010) experimental evaluation of the Coshocton Incentive Program for elementary students in a poor Appalachian community in Ohio. Bettinger found that students in the incentivized treatment group scored 0.15 standard deviations higher in math than those in the control group. Bettinger also examined the impact of the incentives on studentôs intrinsic

[2] In Chicago the incentives had a positive, statistically significant impact on grade point averages and student attendance.

motivation and found no evidence that students in the treatment group had reduced intrinsic motivation.

2.3 Systematic Gender Differences in Response to Incentives

Prior experimental incentive studies have generally found female students to be more responsive (e.g., Angrist, Lang, & Oreopoulos, 2007; Angrist & Lavy, 2009) with the notable exception of Fryerôs 2010 experiments, which found in some cases boys benefited more from the incentives for reading and attendance. Reviewing previous literature, Levitt and colleagues (2012) note a general pattern that females tend to be more responsive to longer-term incentives (Angrist, Lang, & Oreopoulos, 2009; Angrist & Lavy, 2009) and boys perhaps more responsive to more immediate short-term incentives, especially when incentives are framed in the context of a competition (Gneezy, Niederle & Rustichini, 2003; Gneezy & Rustichini, 2004).

2.4. Extrinsic Rewards and Intrinsic Motivation

Researchers and practitioners have identified several reasons why non-monetary incentives could prove more effective motivators. Frey (2007), for example, acknowledges that, compared to monetary compensation, awards have the advantage of being less likely to crowd out recipientsô intrinsic motivation than monetary compensation. At the same time, Frey also notes that non-monetary awards also have the advantage of being more likely to reinforce bonds of loyalty and other positive relationship attributes and generally incur relatively low material costs for the presenter, especially relative to recipient valuation. Frey also notes that these kind of non-monetary incentives serve a strong signaling function: the presenter signals the kind of behavior that is desired and valued, and the recipient is able to signal to others the ability to display these kinds of behaviors. For a complete overview of the debate see, for example,

analyses by Cameron and Pierce (1994), a competing analysis by Deci, Koestner, and Ryan (1999), and a response by Cameron (2001).

3. Experimental Design, Interventions, and Sample

In the 2010-2011 school year, researchers working with a large, Southern urban school district identified 1,128 students in grades 5 through 8 who were eligible for and registered to receive SEdS. These students were enrolled in 14 different schools and were registered with 16 different SEdS providers. In the section below, we describe the experimental design, interventions, and characteristics of the sample.

3.1 Experimental Design

As displayed in Figure 1, a total of 309 students were randomly sampled from the 1,128 SEdS-eligible students.[3] Three students did not meet inclusion criteria, and another four students opted out of participating in the study after being notified of the project in October 2010. The 302 study-eligible students were then randomly assigned to one of three experimental conditions: a control group, a group which would receive monetary incentives for attendance, and a group which would receive symbolic, non-monetary recognition for their attendance. In total, 103 students were allocated to the control condition, 102 students to the non-monetary treatment condition, and 97 to the monetary treatment condition. All students and their parents were

[3] Assuming a mean difference of .125 percentage points with a standard deviation of .315, 304 participants were needed to provide power of .801. See Lipsey (1990) for a detailed discussion on design sensitivity.

notified of their experimental assignment in early-November 2010 and were eligible to begin attending SEdS later that month.[4]

3.2 Interventions

The assigned interventions are straightforward. Every student enrolled in SEdS had a learning plan or contract with their SEdS provider that identifies the subject(s) in which they were to be tutored and the total number of tutoring sessions they were to attend during the current academic year. Students are eligible for SEdS if they attend a failing school and qualify for free- or reduced-price lunch services. Students assigned to the non-monetary recognition incentive condition and their parents were told prior to attending tutoring that signed certificates from the superintendent of schools would be mailed to their homes upon completion of 25 percent and 75 percent of their allotted tutoring hours.[5] Similarly, students assigned to the monetary incentive condition were told prior to attending tutoring that $25 worth of points would be posted to an online account upon completion of 25 percent and 75 percent of their allotted

[4] We randomized eligible participants at the individual student level blocked by timing for their signing up for tutoring services. We used a simple randomization procedure blocked by enrollment date as Bruhn and McKenzie (2009) demonstrate that different randomization methods (e.g., pair-wise matching, stratification) perform similarly in populations of 300 or more.

[5] Students are allotted different hours of tutoring because providers of SEdS can charge different hourly rates. Tutoring providers invoice the school district for the number of hours students attend, up to a maximum per-student, per-year dollar allocation.

tutoring hours and that an additional $50 worth of points would be posted after they completed 100 percent of their allotted tutoring hours.

As displayed in Figure 2, if a student is allocated to the monetary treatment condition and their learning plan recommends 40 hours of tutoring and each session is 2 hours in length, the student will receive $25 in points after the 5th and 15th tutoring session and an additional $50 in points after the 20th session. The research team monitored attendance and distributed certificates and awards on a weekly basis. They were mailed each Friday from mid-November through May of the following calendar year.

During the design phase of the intervention, it was decided that we could not award students with cash. Monetary awards had to be made to students through an online awards platform that the students could then access at their home, school, tutoring provider, housing complex, or wherever else they had access to the internet. The online platform was designed and managed by a private firm that offers dynamic, customized award services to a number of education-related organizations, including Kaplan, Scientific Learning, National Education Association, National Science Teachers Association, Harvardôs Educational Innovation Laboratory, and Connections Academy. The format of the awards platform is very similar to Amazon.com and other online consumer retail sites, although the structure and content is tailored specifically to middle school students. In addition to a rewards catalog that was distributed to all participants that were randomly assigned to the monetary condition, students were offered 1,000s of reward choices online, including opportunities to make charitable donations or purchase electronics, sports equipment, educational materials, and gift cards to brand name retailers.

The actual payment methods are similar to those employed in other student incentive research projects (e.g., Fryer, 2010). The school district administrator received weekly invoices from financial officers with individual tutoring providers. The invoices identified the students and number of tutoring hours they attended for that week. The research team then used this information to calculate student attendance rates and process award information. For the monetary incentive condition, the research team processed individual student awards by notifying the online platform manager of student attendance rates and award amounts. The platform manager then credited each individual studentôs online account upon reaching each predetermined attendance threshold. Additionally, students received notification via mail (both electronic and US postal service) that points had been added to their online account. For the certificate condition, the research team produced a customized certificate of recognition, which was immediately mailed to the studentôs home address upon reaching the specified attendance thresholds.

Although the district administrators, SEdS providers, and site managers went to considerable lengths to ensure student access to computers and bonus award-related information, it is important to note that students did not receive actual cash in-hand. Additionally, there was approximately a one week delay between the time a student met a specific performance threshold and their receipt of the notice that points had been credited to their account, which could affect the strength of the incentive. For example, as noted in Levitt et al (2012) and elsewhere, all motivating power of the incentive vanishes in elementary and primary school student incentive experiments when rewards are handed out with a delay. This lag on incentive delivery may be less of a concern for the certificate of recognition condition as the certificate is mailed to a

studentôs parent(s) or guardian(s), who are less likely to exhibit similar levels of hyperbolic discounting. The concluding section discusses this dynamic further.

3.3 Sample

In expectation, the randomization of individuals to treatment and control conditions will ensure that all observable and unobservable characteristics of students and schools are balanced across the three groups. However, it is possible that our blocked randomization broke down and resulted in imbalances between the treatment and control conditions. To determine whether there were baseline imbalances between students participating in the treatment- and control-conditions, we tested for differences on observable student characteristics using a number of different tests. In addition to simple mean comparisons using a Studentôs t-test, we used Wilcoxonôs signed-rank test and a Kruskal-Wallis one-way analysis of variance when the population was not normally distributed (Kruskal and Wallis, 1952). We also implemented Hotellingôs t-test, which is the analog to a t-test when multiple variables are considered simultaneously. Finally, we ran a series of OLS and logit regressions with indicators for the monetary and certificate treatment conditions. Across all comparisons and statistical tests, we reject the hypothesis that the means of the treatment and the control conditions are different and that the means between the two treatment conditions are different.

As displayed in Table 1, the student sample is limited to middle-school students in grades 5 through 8. The lower grades are over-represented, with 37.21 percent of the sample in grade 5, which decreases monotonically to 18.27 percent in grade 8. Because SEdS target low income students in Title I schools, it is no surprise that 96.01 percent of students in this sample received free- or reduced-price lunch. Slightly more than half of the sample is categorized as African-

American with roughly 27 percent identified as Hispanic, 19 percent as White, and less than 2 percent as Asian. Approximately one out of every five students is labeled as special education and/or English language learner, with a difference of no more than 5 percentage points between conditions.[6]

Finally, it is important to note that the blocked randomization means students in the same school, grade, and classroom could be randomized to either a treatment or control condition. This creates the potential for spillover effects or resentful demoralization (Shadish, Cook, and Campbell, 2001). However, we do not believe this is a major concern given the relatively small number of students in overlapping schools and grades (recall we selected 309 students of 1,128 eligible students at 14 different schools in four different grade levels).

Table 2 displays summary statistics on test scores and select behavioral characteristics of students from the prior school year. We find that test score performance on the mathematics, reading, science, and social studies examinations was similar across groups. We also find that the average grade assigned to students across the three conditions was an 84.91 with average grade ranging between 71.5 and 94.56. Students also attended, on average, between 158 and 160 school days with attendance ranging between less than 100 to 172 days. Our sample had an

[6] We acknowledge that the assumption that students who are identified as needing special education services (or ELL services) would react the same to incentives is a broad one and most likely dependent on a wide variety of factors, which are outside the scope of this study. However, not surprisingly given the small numbers in each treatment group, we are unable to detect any significant differential responses to the various incentives for these students. We also estimated models without special education and ELL populations and results are qualitative similar.

average of 1.22 disciplinary events per child. When delineated by grade, all comparisons are similar to those reported in Table 2 and, once again, we detected no imbalances between treatment conditions or treatment and control conditions.

Table 3 displays summary statistics on the subjects in which students received tutoring and the total hours of tutoring specified in the studentôs individual learning plan with their SEdS provider. While the subject(s) in which a student is tutored remains unknown for nearly one-third of the sample, among students for whom this information is known, the majority are tutored in reading (37.85 percent) or both math and reading (21.58 percent). Table 3 further reports that, on average, in the treatment conditions and the control group, studentsô individual learning plans specify that they receive 31 hours of tutoring.

4. Data and Analytical Methods

We cleaned and merged relevant student, school, and provider information from multiple data sources to create a single data file for the 2009-10 and 2010-11 school years. Data were drawn from management information systems maintained by the school district, including test score files, enrollment history files, and federal program files. The enrollment history file contains student demographic information such as a unique student identifier, race, gender, date of birth, grade, free lunch status, and reduced lunch status. The file also provides a transactional enrollment history, which records dates of school enrollment and transfer for each student. The enrollment history file was supplemented with daily student attendance records to create an in-school attendance variable for each student.

The federal program file tracked the involvement of each student in SEdS on several dimensions, including student enrollment date, total hours scheduled, total hours attended, the

name of the tutoring provider, and the content area of tutoring (i.e., mathematics, reading, or both). Under mandate by the state department of education, this data is recorded and maintained by a designated SEdS coordinator at the district. SEdS attendance information is tracked through invoices submitted by providers. School-level SEdS coordinators confirm the accuracy of records in the federal program file at regular intervals throughout the school year.

To supplement the administrative data, the school district administered a student survey in April to May 2011. Surveys were mailed to student homes and follow-ups facilitated by SEdS providers and studentôs homeroom teacher. For the complete sample, the response rates were 69.61 percent for the control group, 66.67 percent for the monetary incentive group, and 69.90 percent for the non-monetary incentive group. Of the students that attended at least one tutoring session, response rates were 59.40 percent for the control group, 73.84 percent for the monetary incentive group, and 67.50 percent for the non-monetary incentive group.[7] For the analysis of survey results, we restrict the sample to those students that attended at least one tutoring session. The nature of the survey questions required student knowledge of the tutor, content of tutoring, and tutoring practices. Student who attended zero sessions would be unable to provide meaningful responses.

The survey contained measures of intrinsic and extrinsic motivation toward school developed by Harter (1981) and a series of items to gauge student perceptions of various aspects of tutoring services. For the purposes of this evaluation, we report on student perceptions of

[7] The supplementary online materials contain a comparison of respondents and non-respondents on observable characteristics.

SEdS providers and the intrinsic/extrinsic motivation items and scales. More information on the instrument and constructs can be found in the supplementary online materials.

4.1 Analytic Strategy

To judge the overall impact of the interventions as implemented, we estimate variants of the following OLS regression model, which we can interpret as the causal relationship between conditions and outcomes of interest:

$$Y_{ip} = \delta_0 + \delta_1 monetary_{ip} + \delta_2 certificate_{ip} + \varepsilon_{ip} \quad (1)$$

where, Y_{ip} represent the percentage of tutoring hours attended for student *i* in provider *p*; *monetary* is an indicator variable that equals one if a student was randomly assigned to the monetary treatment condition and zero if not; *certificate* is an indicator variable that equals one if a student was randomly assigned to the certificate treatment condition and zero if not.

Here, we are most interested in the estimates of δ_0, which indicates the average percentage of hours attended for students in the control group; $\delta_0 + \delta_1$, which indicates the average percentage of tutoring hours attended for students in the monetary incentive condition; $\delta_0 + \delta_2$, which indicates the average percentage of tutoring hours attended for students in the certificate incentive condition. The coefficient δ_2 differentiates the average percentage of tutoring hours attended for students in the monetary and control conditions, and δ_3 differentiates the average percentage of tutoring hours attended for students in the certificate incentive and control conditions.

An alternative specification of equation (1) can be expressed as:

$$Y_{ip} = \delta_0 + \delta_1 monetary_{ip} + \delta_2 certificate_{ip} + student_{ip}\delta_3 + \varphi_p + \varepsilon_{ip} \quad (1b)$$

where, *student* is a vector of baseline observable student-level characteristics, including binary variables for gender, free lunch status, ELL status, SPED status, race/ethnicity and a series of grade-level dummies and φ_p is a provider fixed effect, eliminating across provider variation from the estimates.

We also report estimates from a second model which can be expressed as:

$$Y_{ip} = \delta_0 + c\delta_1 monetary_{ip} + \delta_2 certificate_{ip} + \delta_3 (monetary_{ip} * female_{ip}) + \delta_4 (certificate_{ip} * female_{ip}) + \delta_5 female \quad (2)$$

where all variables are as previously defined in equation (1) and *female* is an indicator variable that equals one if a student is female and zero if a student is male. We also estimate this model with student and provider controls.

Here, we are most interested in the estimates on δ_0, which indicates the average percentage of hours attended for male students in the control group; $\delta_0 + \delta_5$, which indicates the average percentage of hours attended for female students in the control group; $\delta_0 + \delta_1$, which indicates the average percentage of tutoring hours attended for male students in the monetary incentive condition; $\delta_0 + \delta_1 + \delta_3 + \delta_5$, which indicates the average percentage of tutoring hours attended for female students in the monetary incentive condition; $\delta_0 + \delta_2$, which indicates the average percentage of tutoring hours attended for male students in the certificate condition; $\delta_0 + \delta_2 + \delta_4 + \delta_5$, which indicates the average percentage of tutoring hours attended for female students in the certificate condition. We are also very interested in δ_1, which differentiates the average percentage of tutoring hours attended for male students in the monetary and control conditions; δ_2, which differentiates the average percentage of tutoring hours attended for male

students in the certificate and control conditions; δ_3, which differentiates the average percentage of tutoring hours attended for female and male students in the monetary condition; δ_4, which differentiates the average percentage of tutoring hours attended for female and male students in the certificate condition. We also report estimates for specifications containing student and provider controls.

In this study, our primary outcome of interest is the expected percentage of tutoring hours attended. However, we also investigate take-up rates, as measured by a studentôs attendance at a minimum of one tutoring session, among students registered for SEdS. Take-up rates are of interest given widespread reports of the lack of initial attendance once individuals sign-up for tutoring services (e.g., Springer et al, 2014; GAO, 2006). We estimate take-up using both a linear probability model and a logistic regression model. The linear probability model is a special case of a binomial regression model where the relationship between whether or not a student attended a single tutoring session and her treatment classification is fitted by simple linear regression. The logistic regression framework measures the relationship between whether or not a student attended a single tutoring session and her treatment classification by using probability scores as the predicted values of the dependent variable specified by the following model:

$$\text{Probability(Student attends at least one hour of tutoring} \mid \text{X = x)} = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$$

where x is a vector including treatment-, student-, and provider level variables.

An ITT effect assumes that the results of an experiment are based on the initial treatment assignment and not on the treatment actually received, even though 30.5 percent of eligible students in our sample did not attend a single tutoring session. We believe the ITT estimates are

most relevant because, by all accounts, if the interventions were implemented in future years, it is likely that imperfect treatment implementation would continue to occur.

At the end of the results section, we draw on a number of items from the district administered student survey instrument. For all survey-related analyses, we limit the sample to students that attended at least one tutoring session. We limit the sample in this way because questions of interest on the survey assume that a student attended at least one session.[8] It is important to note that, due to this sample restriction, these are non-experimental estimates of the association between treatment condition students and their responses to items on the survey instrument.

5. Results

Our primary research questions of interest included: (1) what are the impacts of monetary and non-monetary incentives on student attendance at their tutoring provider? and (2) does the response to the incentives vary by gender? We then investigate the relationship between a studentôs experimental condition and responses to various survey items, mainly student perceptions of their service provider and individual responses to intrinsic/extrinsic motivation items.

5.1 Impact on Attendance

As is displayed by the intercept for model 1 in the first column of Table 4, students in the control group attended an average of 16.77 percent of their allotted tutoring hours. Students in the monetary incentive group attended an average of 6.45 percentage points more than those in

[8] As noted earlier, the supplementary online materials contains a comparison of the full and restricted survey samples.

the control group, but this difference was not significant at conventional levels. When we add controls for student and provider characteristics, the magnitude of the value on the monetary coefficient increases and this estimate becomes marginally significant such that students in the monetary condition attends 8.32 percent more of their allocated tutoring sessions when compared to the control group.

By contrast, there was a large positive effect on the average percentage of tutoring hours completed by the certificate group students. In this treatment group, the average percentage of sessions completed was 43.2 percentage points higher than that of the control group. These results are robust to controlling for student and provider characteristics.

Of potential concern is that the treatment reward structure has multiple performance thresholds that could potentially have effects on subsequent attendance. It is plausible that students may reach the first performance threshold (i.e., attend 25 percent of allocated hours) and become less likely to attend additional SEdS sessions because of the amount of work required to hit the next performance threshold (i.e., attend 75 percent of allocated hours).

To investigate this potential threat, we examine the proportion of students in each condition completing percentage of allocated hours over time. Figure 3 illustrates differences in take-up and persistence between the monetary, certificate, and control conditions. While there is an initial, significant difference in take-up between the certificate treatment and control conditions, we also find that the gap in rates of participation widens dramatically between the certificate and control groups over time. More specifically, we find that approximately 35 percent of students in the control group completed at least 10 percent of their allocated hours of tutoring (or about 3.1 hours), while 72 percent of students in the certificate group completed at

least 10 percent of their allocated hours. There is no clear evidence that the multiple performance reward structure of the treatments affected student attendance.

5.2 Differential Response by Gender

As displayed in Table 5, female students were more responsive to the certificate of recognition than their male counterparts. On average, females in the certificate condition attended 25 percentage points more of their allocated tutoring sessions than males in the same condition. Males who were eligible for certificate incentives also attended significantly more than those assigned to the control condition though the difference was nearly half the magnitude of that for females. There was no evidence of significant differential effects by gender for the monetary incentive condition. Females in the control group attended slightly smaller percentages of allocated tutoring sessions, although this difference was not statistically significant at conventional levels. Findings are robust to the inclusion of controls for student and provider characteristics.

In addition to looking at percentage of allocated hours completed, we also examined tutoring *take up* as an outcome, inspired by analyses of college going which look both at persistence and enrollment. Here, we operationalize tutoring take-up as attending at least one session of tutoring. We found that he effect of the non-monetary treatment on tutoring take-up varied greatly by gender. Only 54 percent of registered females in the control group attended at least one tutoring session, compared to 86 percent of registered females in the non-monetary group ($p<0.01$). In contrast, males in the non-monetary group actually had a lower take up rate than males in the control group (67 percent versus 73 percent), although the difference was not

significant at conventional levels. All estimates are reported in the supplementary online materials.

5.3 *Non-Experimental Evidence from Student Survey*

Table 6 displays summary statistics on survey items related to studentsô perceptions of their experience in after school tutoring. Estimates indicate the percentage of student respondents by treatment condition that reported a statement was ñPretty much trueò or ñVery much trueò. While there are several interesting differences in responses, none of the differences are statistically different, which is not unexpected given the imprecision of the estimates.

Table 7 displays estimates when the dependent variable comes from select subscales of intrinsic and extrinsic motivation scale as developed by Harter (1981). Estimates reported in Panel A relate to the academic curiosity or interest of the student. That is, whether a student works to satisfy his/her own interest and curiosity in learning or does the student do schoolwork in order to satisfy the teacher, obtain good marks and grades. None of the differences are statistically different from zero.

Estimates reported in Panel B relate a studentôs notion of academic success/failure. That is, does the student know when he/she succeeded or failed on school assignments or tests, or is the student dependent upon external sources of evaluation such as teacher feedback, grades, and marks. Estimates indicate that students assigned to the certificate condition are significantly more likely to report that they rely on external sources, which is likely a result of a studentôs awareness about the certificates of recognition received when they met specific performance thresholds associated with the intervention.

6. Discussion

The design of this randomized controlled trial allows us to both estimate the causal effect of a monetary incentive on student attendance of SEdS and directly compare its effectiveness with a low-cost alternative non-monetary incentive in the form of a certificate of recognition. The null findings for the monetary incentive, and large significant benefits of the certificate incentive have implications for both future research and current policy.

As indicated in the literature review, a number of randomized controlled trials testing financial rewards for students have not observed increased rates of the incentivized behavior relative to the control group. In some cases, it might be that the size of the monetary incentive was insufficient to motivate the behavior. While the size of the rewards offered here may seem large relative to the hourly-wage a student might be able to earn in the labor market, it is also important to remember that s/he was expected to invest 25 percent of allocated hours (or on average about 7 to 8 hours), stretched over the period of two or more weeks, before receiving the monetary incentive.

Adolescents, like those in our study, frequently exhibit decision-making in which short-term gains are weighed more heavily than long-term gains, described as hyperbolic discounting (Green, Fry, & Myerson, 1994; Whelan & McHugh, 2010). Levitt and colleagues (2012), for example, found that the promise of immediate financial rewards did result in increased test performance, with an effect size estimated at 0.1 to 0.2 standard deviations. However, in the group where students were told they would not receive their rewards for one month, test performance was no different from that of the control group. While this well-documented preference for immediate rewards among children could explain the ineffectiveness of the

relatively delayed monetary incentives, the role of benefit discounting for equally or more delayed non-monetary incentives is less obvious.

Empirical literature typically conceptualizes non-monetary incentives as having a signaling value. We conjecture that in the context of this experiment, the much larger positive effects of the non-monetary incentive may have been realized because of the signal's most proximal audience: because the certificate was mailed home, the signal of this award was observed not by peers, but by the studentôs parents and family. If the parent-adolescent relationship is one that can be characterized as exhibiting significant information asymmetry regarding the studentôs proacademic values, motivation and behavior, than this award in particular might be viewed as a signal of a studentôs proacademic values and behavior to studentsô parents. For some adolescent students, this kind of targeted signal would be particularly worthwhile, in that it excludes any similar but undesirable signaling to the studentôs peers.[9] Policymakers and practitioners seeking to increase attendance of underutilized student supports should note the cost-effectiveness of the certificate intervention. The costs of each intervention are relatively straightforward. For the certificate intervention it cost approximately $3 per

[9] Prior research suggests that students would prefer that different audiences receive different signals. Relative to peers, adolescents' parents place greater value on time spent on homework (Fordham, 1996), are less tolerant of misbehavior in class (Berndt, Miller, & Park, 1989), and place different relative values on reputation, popularity, and academic success (Coleman, 1961). It is likely then that students know or perceive their parents to value academic achievement and behavior – including tutoring attendance – more than students' peer groups. For this reason, incentives that aim to generate recognition or pride from a student's parents might be more effective than those awarded before students' peers.

certificate, which included a gold stamp sticker and postage, or approximately $9 per student completing 100 percent of allocated hours. For the monetary intervention it cost $25 for 25 percent and 75 percent threshold and $50 for the 100 percent threshold, or $100 per student completing 100 percent of the allocated hours (not accounting for postage, online delivery system, etc.). If we compute the cost required to obtain a single unit of effectiveness, which is calculated by taking the cost of a given alternative and dividing it by its effectiveness, we find the certificate intervention is more than 6,000 times more cost-effective than the monetary incentive. While these rough estimates paint a relatively extreme picture, they illustrate one of the important elements of our primary finding: non-monetary incentives can be both effective and cheap. If cost is one of the frequent barriers to translating research to practice, it should not be one here.

At the same time, the results of this experiment should not be interpreted as offering a single, comprehensive answer to an ongoing problem. Offering certificates of recognition to motivate all desirable student behavior would not necessarily have the same dramatic effects if implemented in a widespread and sustained manner. Economic theory suggests that the value ascribed to a certificate of recognition (be it signal value, symbolic value, or any variety of value) would likely have diminishing marginal utility. We would imagine that even in this study, the second certificate of recognition received by a student would not have the same impact as the first. Relatedly, a certificate of recognition might not have the same value and motivating power for a student who consistently brings home high-GPA report cards, compared to a student who brings home low-GPA report cards. Ultimately, implementation of non-monetary incentives would be

most effective if it takes into consideration the likelihood that these kinds of incentives may have diminishing marginal returns.

The findings presented in this study also offer insights about the important role of gender in studentsô response to incentives. As noted above, recent incentive studies have generally found female students to be more responsive (e.g., Angrist, Lang, & Oreopoulos, 2007; Angrist and Lavy, 2009), with the exception of Fryerôs 2012 experiments, which found in some cases boys benefited more from the incentives for reading and attendance. All of these studies, however, focus entirely on financial incentives and thus could only generate theories about differential gender responsiveness to financial rewards. With the introduction of non-monetary incentives as alternative treatments, we can begin to formulate a more general theory about gender differences in response to manipulations of incentives.

We find that female students benefited almost twice as much as their male counterparts from the non-monetary incentive. The patterns of heterogeneous response to this application of nonmonetary incentives suggest implications for the design and implementation of student incentive programs more generally. In this study, male students offered certificates of recognition were no more likely to take-up tutoring services than those in the control group, although the certificate did seem to be an effective incentive for encouraging persistence among those male students who attended at least one tutoring session. Put another way, males like those in this study may need additional support or incentives to take-up tutoring services and ñget over the hurdleò of attending tutoring session for the first time. Specifically, by addressing the barriers and incentives to male take-up of tutoring service, the district might expect to leverage the observed persistence-effect of the certificate incentive to an even greater degree.

More broadly, this analysis contributes to a body of evidence that suggests the need for purposeful testing for heterogeneous treatment effects. Researchers have for some time asserted that analysis of systematic differences in group average treatment effects are underutilized in program evaluation, and that results of these analyses can be particularly helpful for informing policy design, policy targeting, and selecting a given program or approach from a series of policy options (Djebbari & Smith, 2008; Heckman, Smith, & Clements, 1997). A number of authors have recently reiterated the importance of this kind of analysis (Imai & Ratkovic, 2013; Weiss, Bloom, & Brock, 2014). In particular, the potential for results of this kind of analysis to improve the effectiveness and efficiency of social programs is intuitive: tailoring programs to distinct needs of student subgroups and directing limited resources to where they might do the most good (Schuck & Zeckhauser, 2006).

Among a recent series of disappointing student incentive experiments, the results of this study should inspire measured optimism. In essence, we find that incentives can effectively increase participation in voluntary SEdS, and their costs can be low. The surprising dominance of the non-monetary incentive over relatively substantial monetary rewards as a student attendance motivator highlights the need for more nuanced research as to the differential effectiveness of the various types of incentives we can offer. Furthermore, the strong causal estimates of effectiveness for non-monetary incentives coupled with their low implementation costs suggest policymakers and practitioners looking to motivate participation in underutilized, potentially beneficial programs should consider immediate applications.

Acknowledgement

The authors are grateful to Dale Ballou, Chris Hulleman, Matthew Pepper, Morgan Polikoff, and Ron Zimmer for helpful comments and insight in developing this work. They also wish to acknowledge seminar participants at the University of Colorado ï Colorado Springs, Vanderbilt Universityôs Peabody College, and the annual meetings of the American Educational Research Association, Association for Public Policy Analysis and Management, Association for Education Finance and Policy, and U.S. Department of Educationôs Institute for Education Sciences. The usual disclaimers apply.

References

Allan, B. M., & Fryer, R. G. (2011). The Power and Pitfalls of Education Incentives. *The Hamilton Project*, (September). Retrieved from http://www.brookings-tsinghua.cn/~/media/Files/rc/papers/2011/09_incentives_fryer_allen/092011_incentives_fryer_allen_appendix.pdf

Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and Services for College Achievement: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics, 1*(1), 136-163.

Angrist, J., & Lavy, V. (2009). The effects of high stakes high school achievement awards: evidence from a randomized trial. *The American Economic Review, 99*(4), 1384-1414.

Arroyo, C.G., and Zigler, E. (1995). Racial identity, academic achievement, and the psychological well-being of economically disadvantaged adolescents. *Journal of Personality and Social Psychology, 69*, 903-914.

Berger, A., deSousa, J-M, Hoshen, G., Lampron, S., Le Floch, K.C., Petroccia, M., Shkolnik, J. (2011). *Supplemental Educational Services and Student Achievement in Five Waiver Districts.* Washington, DC: US Department of Education, Office of Planning, Evaluation, and Policy Development.

Behrman, J.R., Parker, S.W., and Todd, P.E. (2011). Do conditional cash transfers for schooling generate lasting benefits? A five-year follow-up on PROGRESA/Oportunidades. *Journal of Human Resources*, 46(1), 93-122.

Berndt, T.J., Miller, K.E., and Park, K. (1989). Adolescentsô perceptions of friendsô and parentsô influence on aspects of their school adjustment. *The Journal of Early Adolescence, 9*, 419-435.

Bruhn, M. and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics, 1, 200-232.*

Burch, P., Steinberg, M., and Donovan, J. (2007). Supplemental Educational Services Under NCLB: Policy Assumptions, Market Practices, Emerging Policy Issues. *Educational Evaluation and Policy Analysis*.

Burch, P. and Good, A. (2009). Getting to the Core: The role of Instructional Setting in Federally Mandated Tutoring. Working paper presented at the 2009 American Educational Research Association Conference, San Diego, CA.

Cha, P., & Patel, R. Rewarding progress, reducing debt. MDRC. October 2010, 1-61.

Chicago Public Schools: Office of Research, Evaluation, and Accountability (2005). *SEdS Tutoring Programs: An Evaluation of Year 3 in the Chicago Public Schools*. Chicago: Chicago Public Schools.

Chicago Public Schools: Office of Research, Evaluation, and Accountability (2005). *SEdS Tutoring Programs: An Evaluation of the Second Year*. Chicago: Chicago Public Schools.

Coleman, J.S. (1961). *The adolescent society.* New York: Free Press of Glencoe.

Deci, E., Koestner, R., Ryan, R, and Cameron, J. (2001). Extrinsic rewards and intrinsic

motivation in education: Reconsidered once again: Comment/Reply. Review of Educational Research, 71(1), 1-51.

Deke, J., Dragoset, L, Bogen, K, and Gill, B. (2012). *Impacts of Title I Supplemental Educational Services on Student Achievement*. Washington, DC: Department of Education, Institute of Education Sciences.

Djebbari, H., & Smith, J. (2008). Heterogeneous impacts in PROGRESA. *Journal of Econometrics, 145*(1), 64-80.

Downey, D.B., & Vogt Yuan, A.S. (2005). Sex differences in school performance during high school: puzzling patterns and possible explanations. *Sociology Quarterly 46*(2), 299ï 321.

Dweck, C. S., & Licht, B. G. (1980). Learned helplessness and intellectual achievement. In J. Garber & M. E. P. Seligman (Eds.), *Human helplessness: Theory and applications*. New York: Academic Press.

Eccles. J. S., Wigfield, A., Harold, R .. & Blumenfeld, P. B. (1993). Age and gender differences in childrenôs self- and task perceptions during elementary school. *Child Development, 64,* 830-847.

Fordham, S. (1996). *Blacked out: Dilemmas of race, identity, and success at Capital High.* Chicago: University of Chicago Press.

Fredericks, J., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male sex-typed domains. *Developmental Psychology 38*, 519-533.

Frey, B. S. (2007). Awards as compensation. *European Management Review, 4*(1), 6-14.

Fryer, R.G. & Torelli, P. (2010). An empirical analysis of ‘acting white’. *Journal of Public Economics 94*(5-6), 380-396.

Fryer, R.G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *Quarterly Journal of Economics*, 126(4), 1755-1798.

Fryer, R. G. (2012). Aligning Student, Parent, and Teacher Incentives: Evidence from Houston Public Schools. *NBER Working Paper, No. 17752*.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics, 118*(3), 1049-1074.

Gneezy, U., & Rustichini, A. (2004). Gender and competition at a young age. *American Economic Review*, 377-381.

Government Accountability Office (GAO). (2006). *No Child Left Behind Act: Education Actions Needed to Improve Local Implementation and State Evaluation of Supplemental Educational Services. Report 06-758* Washington, DC: Author.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology, 93,* 3-13.

Graham, S., Taylor, A. Z., & Hudley, C. (1998). Exploring achievement values among ethnic minority early adolescents. *Journal of Educational Psychology, 90*, 606-620.

Green, L., Fry, A. F., & Myerson, J. (1994). Discounting of delayed rewards: A life-span comparison.

Psychological Science, 5, 33ï 36.

Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology, 17*, 300-312.

Heckman, J. J., Smith, J., & Clements, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies, 64*(4), 487-535.

Heinrich, C.J. & Burch, P. (2012). *The Implementation and Effectiveness of Supplemental Education Services: A Review and Recommendation for Program Improvement*. Washington, DC: American Enterprise Institute for Policy Research.

Heinrich, C. J., Burch, P., Good, A., Acosta, R., Cheng, H., Dillender, M., ... & Stewart, M. (2014). Improving the Implementation and Effectiveness of Out-of-School-Time Tutoring. *Journal of Policy Analysis and Management*, *33*(2), 471-494.

Heinrich, C. J., & Nisar, H. (2013). The efficacy of private sector providers in improving public educational outcomes. *American Educational Research Journal*, 0002831213486334.

Heinrich, C.J., Meyer, R.H., & Whitten, G. (2010). Supplemental education services under No Child Left Behind: Who signs up, and what do they gain? *Educational Evaluation and Policy Analysis* 32, 273-298.

Heistad, D. (2005). *Analysis of 2005 supplemental education services in Minneapolis public schools: An application of matched sample statistical design.* Minneapolis: Minneapolis Public Schools.

Imai, K., & Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics, 7*(1), 443-470.

Jacob, B.A. (2002). Where the boys arenâ: noncognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review 21*, 589ï 98

Jacobs, J., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Ontogeny of children's self-beliefs: Gender and domain differences across grades l through 12. *Child Development 73*, 509-527.

Janvry, A, finana, F., Sadoulet, E., Vakis, R. (2004). Can conditional cash transfers serve as safety nets to keep children at school and out of the labor market?

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47*(260), 583-621.

Lipsey, M. (1990). Design sensitivity: Statistical power for experimental research. Newbury Park, CA: Sage.

Levitt, S.D., List, J.A ., Neckermann, S., Sadoff, S. (2012, June). *The Behaviorist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance.* (NBER Working Paper No. 18165). Retrieved from the NBER website: http://www.nber.org/papers/w18165.

Long, J.S. & Freese, J. (2006). *Regression Models for Categorical Dependent Variables Using Stata.* College Station, TX: Stata Press

McCall, R. B., Evahn, C., & Kratzer, L. (1992). *High school underachievers: What do they achieve as adults?* Newbury Park, CA: Sage.

National Research Council (2003). *Engaging Schools: Fostering High School Students' Motivation to Learn.* Washington, DC: National Academies Press

Potter, A., Ross, S.M., Paek, J., McKay, D., Ashton, J., and Sanders, W.L. (2007). *Supplemental educational services in the State of Tennessee: 2005 - 2006.* Center for Research in Educational Policy. Memphis, TN.

Riccio, J.A., Dechausay, N., Greenberg, D.M., Miller, C., Rucks, Z., and Verma, N. (2011). Toward Reduced Poverty Across Generations: Early Findings from New York Cityôs Conditional Cash Transfer Program. MDRC Working Paper.

Rickles, J.H. & Barnhart, M.K. (2007). *The Impact of Supplemental Educational Services Participation on Student Achievement: 2005-06.* Los Angeles: Los Angeles Unified School District Program Evaluation and Research Branch, Planning, Assessment and Research Division.

Rickles, J.H. & White, J. (2006). *The Impact of Supplemental Education Services Participation on Student Achievement* (Publication No. 295). Los Angeles, CA: Los Angeles Unified School District, Program Evaluation and Research Branch, Planning, Assessment, and Research Division.

Schuck, P. H., & Zeckhauser, R. (2007). *Targeting in social programs: Avoiding bad bets, removing bad apples.* Washington, DC: Brookings Institution Press.

Shadish, W., T. Cook and D.T. Campbell (2001). *Experimental and quasi-experimental designs for generalized causal inference.* New York: Houghton-Mifflin

Spencer, S. J., Steele, M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of' Experimental Social Psychology, 35*, 4-28.

Springer, M.G., Pepper, M.J., Gardner, C., Bower, C.B. (2009). Supplemental Educational Services Under No Child Left Behind. In M. Berends, M.G. Springer, D. Ballou, H.J. Walberg (Eds., 2009). *Handbook of Research on School Choice.* New York: Routledge.

Springer, M.G., Pepper, M.J. &. Ghosh-Dastidar, B. (2014). *Supplemental educational services and student test score gains: Evidence from a large, urban school district*. Journal of Education Finance.

Weiss, M. J., Bloom, H. S., & Brock, T. (2014). A conceptual framework for studying the sources of variation in program effects. *Journal of Policy Analysis and Management 33*(3), 778-808

Whelan, R., & McHugh, L. A. (2010). Temporal discounting of hypothetical monetary rewards by adolescents, adults, and older adults. *The Psychological record, 59*(2), 6.

Yonezawa, S., Wells, A., and Serna, I. (2002). Choosing Tracks: ñFreedom of choiceò in detracking schools. *American Educational Research Journal, 39*, 37-67

Zimmer, R., Christina, R., Hamilton, L.S., and Prine, D.W. (2006). *Evaluation of Two Out-of-School Program in Pittsburgh Public Schools: No Child Left Behind's Supplemental Educational Services and State of Pennsylvania's Educational Assistance Program.* RAND Working Paper. Santa Monica, CA: RAND Corporation.

Zimmer, R., Gill, B., Razquin, P., Booker, K., Lockwood, J.R., Vernez, G., Birman, B.F., Garet, M.S., and O'Day, J. (2007). *State and Local Implementation of the No Child Left Behind Act: Volume I - Title I School Choice, Supplemental Educational Services, and Student Achievement.* Washington, DC: U.S. Department of Education. Available at http://www.rand.org/pubs/reprints/2007/RAND_RP1265.pdf.

Table 1. Summary Statistics on Student Characteristics

	Full Sample	Certificate	Monetary	Control	Cert-Cont	Monet - Cont	Cert-Monet
Percent Female	50.50	55.34	47.92	48.04	7.3	-0.12	7.42
Percent Black	52.82	58.25	51.04	49.02	9.23	2.02	7.21
Percent White	19.27	18.45	16.67	22.55	-4.1	-5.88	1.78
Percent Hispanic	26.58	22.33	32.29	25.49	-3.16	6.8	-9.96
Percent Asian	1.33	0.97	0	2.94	-1.97	-2.94	0.97
Percent Free Lunch	89.70	91.26	90.63	87.25	4.01	3.38	0.63
Percent Reduced Price Lunch	6.31	5.83	4.17	8.82	-2.99	-4.65	1.66
Percent SPED	17.94	15.53	17.71	20.59	-5.06	-2.88	-2.18
Percent ELL	22.92	20.39	22.92	25.49	-5.1	-2.57	-2.53
Percent Grade 5	37.21	36.89	36.46	38.24	-1.35	-1.78	0.43
Percent Grade 6	26.58	27.18	29.17	23.53	3.65	5.64	-1.99
Percent Grade 7	17.94	23.3	14.58	15.69	7.61	-1.11	8.72
Percent Grade 8	18.27	12.62	19.79	22.55	-9.93	-2.76	-7.17
n	302	103	97	102	205	199	200

Table 2. Summary Statistics on Test Scores and Behavioral Characteristics (t-1)

	Full Sample		Certificate	Monetary	Control		Cert-Cont	Monet - Cont	Cert-Monet
Math (raw score)	483.78		482.26	489.04	480.36		1.9	8.68	-6.78
Reading (raw score)	484.32		483.71	486.94	482.43		1.28	4.51	-3.23
Science (raw score)	185.72		184.56	187.47	185.22		-0.66	2.25	-2.91
Social Studies (raw score)	190.39		187.97	193.67	189.75		-1.78	3.92	-5.70
Avg. Grade	84.91		85.05	84.75	84.92		0.13	-0.17	0.30
School Attendance (days)	158.64		160.27	159.92	158.08		2.19	1.84	0.35
Discipline	1.22		1.3	1.49	0.88		0.42	0.61	-0.19
n	181 - 298		65 - 102	90-95	56-101				

Notes: * significant at the 10% level; ** 5% level; *** 1% level

Table 3. Summary Statistics on Tutoring Subjects and Total Hours of Tutoring Assigned

	Full Sample		Certificate	Monetary	Control		Cert-Cont	Monet - Cont	Cert-Monet
Subjects									
Percent Math Only	9.98		6.8	12.5	10.78		-3.98	1.72	-5.7
Percent Reading Only	37.85		44.66	32.29	36.27		8.39	-3.98	12.37
Percent Both Math and Reading	21.58		25.24	17.71	21.57		3.67	-3.86	7.53
Percent Unknown	30.59		23.3	37.5	31.37		-8.07	6.13	-14.2
Hrs. Allocated	30.99		31.13	30.83	31		0.13	-0.17	0.3
n	302		103	97	102		205	199	200

Notes: * significant at the 10% level; ** 5% level; *** 1% level

Table 4. Impact of Interventions on Percentage of Allocated Tutoring Hours Attended

	Model 1	Model 2	Model 3
Non-Monetary	0.432	0.4198	0.435
	(0.0478)**	(0.0481)**	(0.0475)**
Monetary	0.0645	0.0616	0.0832
	(0.0486)	(0.0488)	(0.0488)*
Intercept	0.1677	0.1396	0.155
	(0.0338)	(0.0756)	(0.0239)
Student controls		√	√
Provider control			√
N	302	302	302
Adj. R-squared	0.236	0.243	0.3235

Notes: Robust standard error in parentheses. * significant at the 10% level; ** 5% level; *** 1% level

Table 5. Impact of Interventions on Percentage of Allocated Tutoring Hours Attended by Gender

	Model 1	Model 2	Model 3
Non-Monetary	0.2905	0.2875	0.3041
	(0.0683)**	(0.0688)**	(0.0673)**
*Non-Monetary*female*	0.2681	0.2667	0.2496
	(0.0950)**	(0.0956)**	(0.0922)**
Monetary	0.0133	0.0068	0.035
	(0.0668)	(0.0671)	(0.0662)
*Monetary*female*	0.1066	0.1133	0.0971
	(0.0965)	(0.0968)	(0.0931)
Female	-0.0951	-0.1019	-0.0845
	(0.0669)	(0.0681)	(0.0652)
Intercept	0.2126	0.1491	-0.0602
	(0.0466)	(0.0758)	(0.2420)
Student controls		√	√
Provider controls			√
n	302	302	302
Adj. R-squared	0.2505	0.2592	0.3365

Notes: Robust standard errors in parentheses. * significant at the 10% level; ** 5% level; *** 1% level

Table 6. Summary Statistics on Select Survey Items

	Full Sample	Certificate	Monetary	Control	Cert-Cont	Monet - Cont	Cert-Monet
Tutor knew a lot about reading and/or math	86.20	88.89	81.25	88.64	0.25	-7.39	7.64
Attending tutoring made me do better at reading and/or math	84.55	87.31	76.25	91.12	-3.81	-14.87	11.06
I really worked hard at tutoring	84.15	87.5	84.38	79.1	8.4	5.28	3.12
I talk with teacher about tutoring	79.93	79.36	83.34	76.42	2.94	6.92	-3.98
More likely to attend tutoring if I get a prize	45.48	45.16	45.83	…	…	…	-0.67
Attended tutoring because I had to go	45.63	36.06	58.33	43.18	-7.12	15.15	-22.27
n	140	54	48	38	92	86	102

Notes: * significant at the 10% level; ** 5% level; *** 1% level

Table 7. Relationship between Extrinsic/Intrinsic Motivation Subscales and Treatment Condition

	Panel A:			Panel B:		
	Curiosity/Interest vs. Please Teacher/Good Grades			Success/Failure vs. External Criteria		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Non-Monetary	0.2805	0.3685	0.5016	0.6221	0.6677	0.8139
	(0.2840)	(0.3024)	(0.3466)	(0.2823)**	(0.2946)**	(0.3121)**
Monetary	0.4086	0.3935	0.2868	0.2575	0.2230	0.4200
	(0.3039)	(0.3116)	(0.3789)	(0.2940)	(0.2991)	(0.3473)
Student controls		√	√		√	√
Provider fixed effect			√			√
N	82	82	82	82	82	82
Adj. R-squared	0.0014	0.0225	0.0413	0.0383	0.315	0.15

Notes: Robust standard error in parentheses. * significant at the 10% level; ** 5% level; *** 1% level

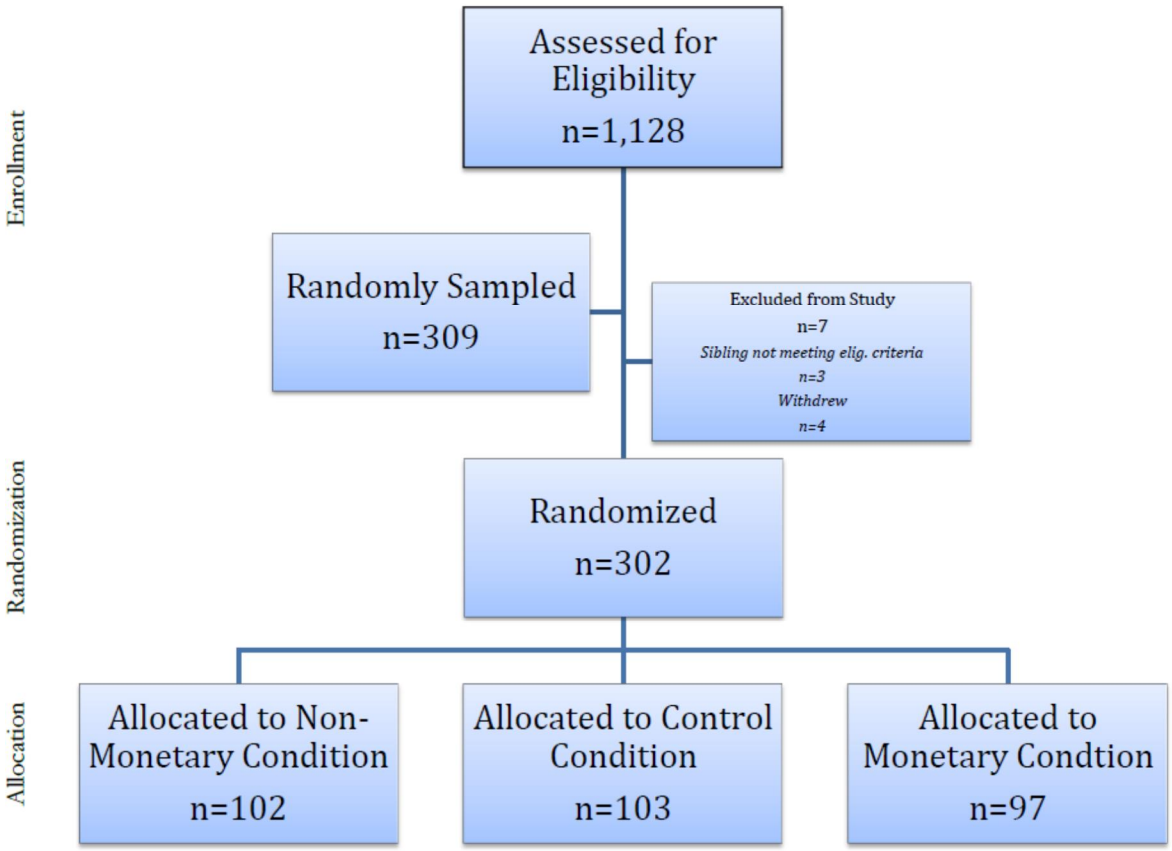

Figure 1. Consort Diagram

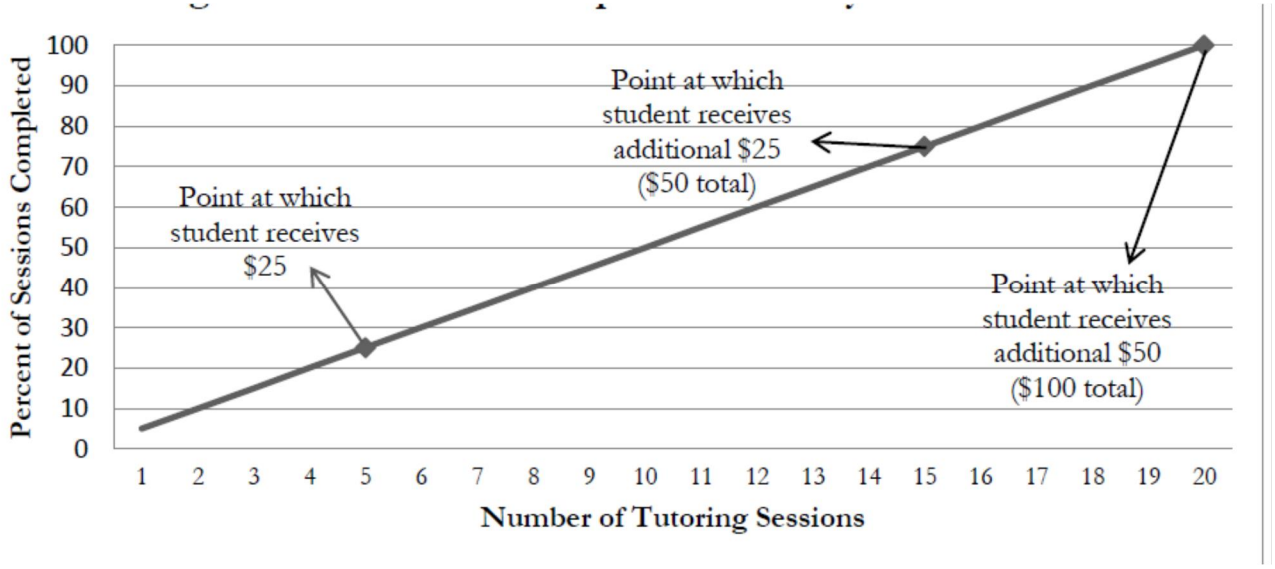

Figure 2. Illustrative Example of Monetary Intervention

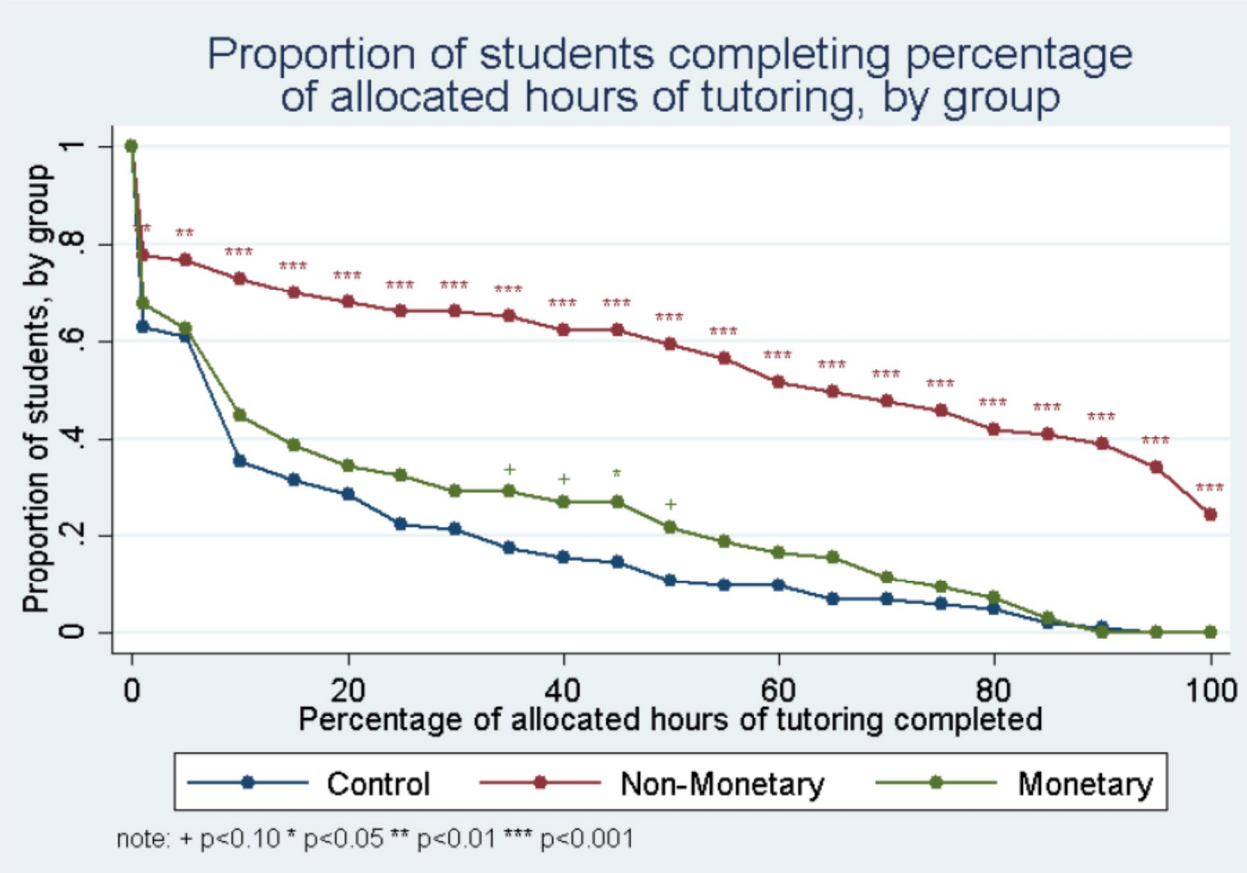

Figure 3. Proportion of students completing percentage of allocated hours of tutoring, by group