Paying for A's:
An Early Exploration of
Student Reward and Incentive Programs in Charter Schools

Margaret Raymond
CREDO
Stanford University

April 2008

Introduction

Recent public attention has spotlighted the use of reward programs for K-12 public school students. In districts such as Baltimore, Fulton County, Georgia, and New York City, the efforts of educators to structure reward systems are being met with support, skepticism and indignation. The programs use a variety of incentives such as cash or mp3 players to shape individual student behavior including, among other things, high school attendance, continuing enrollment or academic performance. Districts and schools also use rewards such as social events or concerts to reward school wide attainment of desirable outcomes, such as improved school-wide safety or attendance. The natural laboratory of charter schools offers a unique chance to investigate the performance implications of such programs.

Student incentive programs are an innovation in education that already has widespread practice in charter schools across the country. Where charter schools are permitted, the state authorizing legislation gives charter schools latitude to adopt new approaches to K-12 education, and many have elected to incorporate incentive schemes into their education programs. Until now, the outlines of those schemes and the degree to which they improve student academic achievement have not been examined.

Growing interest in rewards or incentives for students occurs against the policy backdrop of school accountability policies, which themselves employ incentives and sanctions (negative rewards). Underperforming schools and districts face shortening timelines for improvement. Reward systems for students are one among many innovations being tried across the country to jumpstart an increase in student motivation and results.

The research aim of this study was to determine if the presence of such incentives and rewards in charter schools significantly affects the academic outcomes of students. This study examines a non-random sample of charter schools and their decisions to use or forego an incentive program in their school to see if the systems enhance academic achievement gains. The findings show that the program designs exist in all grade levels but cover much of the same behavior and outcomes. They do vary in the intensity with which students are scrutinized with longer periods of review in higher grade levels.

The analysis shows consistent impact of the programs across grades and designs in terms of student achievement gains on state achievement tests in reading. At the same time, there is no evident impact in math.

In the next section, I develop a taxonomy for incentive programs to classify the reward programs used today in charter schools. The third section of this paper describes the schemes as they operate today, based on self-reported survey results. In the fourth section, the analytic methods used to study the impact of these systems are reviewed. The results are presented in the fifth. A discussion of further directions for this work concludes the paper.

A Taxonomy of Student Incentives

The programs studied in this analysis build on more traditional programs of incentives such as college scholarships and "bridge" programs. For the typical types of scholarship incentives there is a robust literature of their effectiveness focused on comparative college attendance rates. [1] In these studies, scholarship incentives create larger demand for higher education, but absent specific targeting do little to address the underlying structural inequalities across the economic spectrum.[2] At the secondary level, policies that reward students or their teachers for passing rigorous examinations such as Advanced Placement exams or national exams in other countries create positive results, though the results are shown to weaken when teachers are excluded or are not supportive of the idea.[3] In both types of incentive programs, the payoff (i.e., receipt of a scholarship) is realized well after the behavior that is targeted for reinforcement (i.e., high school completion or specific levels of achievement), and often awarded by individuals with no relationship to the recipients. On a stand-alone basis, then, these approaches are weak in terms of their direct influence on behavior. They also assume students possess long time horizons, appropriately value the reward, and calculate the inter-temporal tradeoffs in some predictable way -- clearly assumptions that require investigation.[4]

Both to create incentives for college attendance and to address some of the cultural and socio-economic barriers previously referenced, a number of "college-bound" programs modify the college scholarship concept in two ways. First, they reduce the payoff time and increase the likelihood of reward by making investments in college savings accounts based on academic results each grading period. Second, they incorporate out-of-school social and education supports while students are still in high school both to shape behavior in real time and to motivate them to pursue longer term education goals. In some cases continued program enrollment is the incentive for full participation in tutoring, mentoring, community service or career exploration activities. In other cases, small rewards are tied to these non-academic aspects of the program. A few studies have examined the effect of a promise or probability of future scholarships on secondary behavior and achievement, but this work is plagued by small, localized samples and few control observations.[5]

The incentive and reward programs in this study represent further evolution of the "college bound" design. These programs are operating in a wider range of grade levels than the programs previously discussed. The behavior is more granular and short-term; in addition to grades and achievement test performance, most incentive systems in this study also cover basic learning behaviors such as sitting quietly, tracking the teacher, completing class work and general decorum, which have been shown to positively correlate with academic outcomes.[6] The lag between desired result and payoff is shorter, creating strong reinforcement of targeted behavior. The outlines of operant conditioning in token economies are apparent. One note of particular interest from prior work is the tendency for a significant share of individuals not to respond to the incentive program at all. In addition, closer proximity between the student and those who evaluate behavior and dole out rewards provides a secondary source of reinforcement between students and educator role models. [7]

Incentive systems differ from each other in other ways. One way is in the interpersonal proximity of subject and the individual administering the reward. The same reward can have stronger or weaker effect depending on the degree of closeness in the interpersonal dyad. Clearly, family members would be the strongest actors in shaping behavior, but teachers and other adults that a child regularly encounters also can have a strong influence on the sustainability of reinforcement.[8] As one moves further away from the center of an individual's world, their implicit endorsement of the reward diminishes in magnitude. Reward systems can materially affect their results if the are delivered by individuals that are closely related to the subjects they aim to influence.

Little empirical attention has been paid to the impact of the more recent conceptions of incentives and rewards. In two studies with hyperactive students, token economies influenced behavior but did little to improve academic performance. Eric Bettinger has analyzed a randomized experiment in Coshocton Ohio in which students were given cash rewards if they performed well on state achievement tests; he found a small significant increase in Math performance but not in reading.[9] The evidence is modestly positive and contradicts the opposing view that rewards impair the intrinsic motivation of students.[10]

Figure 1 presents a spatial taxonomy that distinguishes the types of programs described above along two dimensions. The first is the time lag between the behavior that is identified as worthy of reward and the receipt of the reward itself. The longest lags occur with college scholarships that have an uncertain payoff many years in the future. The second dimension is the degree of social distance between student and the individual(s) empowered to assess the student behavior and award points, credits or direct rewards. In the case of college scholarships, the individuals deciding the worthiness of a student are typically anonymous. The programs included in this study are located closer to the origin than the traditional incentive programs models. They have the shortest time between behavior and rewards, ranging from a few days to several months, and they are directly familiar with the individuals making the decisions about their behavior. This taxonomy provided a useful way to translate survey responses into an intensity measure for the modeling described below.

Study Approach

Data. In October 2007, an electronic survey was sent to 250 charter schools in seventeen states that had previously agreed to participate in a longitudinal study of overall charter school effectiveness. The survey was sent to school principals asking for information about the use of student incentives or rewards systems in their schools. If a principal reported that the school did not use such a system, the survey ended and the remaining questions were unanswered. If the response was affirmative, additional information was gathered about the system, including the frequency of assessment, the frequency of resetting the system, the areas of student life that were covered by the system including individual behavior or academic performance, citizenship or school-wide behavior or academic performance. Principals were asked to identify who in the school was empowered to assess student behavior and in what areas. The nature of rewards was also explored. Finally the principal was asked to rate separately the degree to which they, the

teachers and other staff viewed the reward system as fair and effective; these data were used to create measures of cohesion in perceived effectiveness and fairness in each school.

Completed surveys were received from 186 schools, a 74 percent response rate. The full sample is used in the following section to describe the prevalence of student reward systems and outline the dimensions of those that sample schools currently operate. The unit of analysis in the descriptions is the school.

The survey information was used to create three different variables to represent the incentive/reward systems. First, a simple binary variable represents the presence or absence of such a system in each school. Second, separate binaries were constructed to reflect if the system focused on academics (such as grades of state test performance) or behavior (classroom behavior, citizenship or social interactions). Finally, a continuous variable was developed to reflect the intensity of attention on the student in each program.[1]

To assess the effect of these programs, a subset of the responding schools was studied. To be included in the subset, schools had to have been open for two or more years and have students enrolled in 4th grade or higher in 2007. Demographic data and state achievement test scale score data for the academic years 2004/05, 2005/06 and 2006/07 were compiled from the Core of Common Data and state education departments. Within each school, achievement test scale scores for each grade and subject were standardized using grade-level means and standard deviations for each test year that were obtained from the state education agencies. Additional information such as the year the school was opened, the specific location of the school, its membership in a larger network of schools and the grades they served in each year of study were obtained from the schools.

A panel dataset was created whose observations consist of a cohort in a school for two of the three academic years, i.e. 2004/05 to 2005/06 and 2005/06 to 2006/07. Standardized cohort gain scores were created for each grade and school and used as the dependent variable in the impact estimation.

Analysis. The analysis consists of two parts. In the first, simple statistics are used to describe the prevalence of rewards systems and the attributes of the ones that were reported.

The impact analysis in the second part is conducted separately for reading and mathematics performance. An initial set of basic models test the premise that use of a reward system makes a significant contribution to student learning. Initial models use

[1] The reported frequency of assessment for the programs was transformed into a proportion ranging from .01 (assessment occurs once each semester) to 1.0 (assessment occurs in each period of the school day or continuously), with intervening values of 0.2 (once a week) 0.1 (once a fortnight).0 .05 (once a month) and so on. This factor was then multiplied by the average rating of the system's effectiveness for the school. This manipulation resulted in a variable that ranges from 0 (no system) to 1.0 (a system with continuous assessment and total belief in its effectiveness) with 46 intervening values.

OLS regression to estimate the impact of incentive/reward systems on student achievement gains, while controlling for differences in student population, level of poverty as reflected in Free and Reduced Price Lunch program eligibility, and special populations such as English Language Learners or students identified as needing special education. An additional independent variable is the baseline year performance; it is included to reflect the diminishing effects of prior learning over time. Variations of the OLS models included interactions of reward systems with grade span (elementary, middle school or high school) and calendar year.

A second set of models test variations on the use of an incentive/reward system variable. Models are developed to test the importance of strong belief in the effectiveness of the system, as captured in the average rating of effectiveness across principals, teachers and staff; values ranged from 0 (no confidence in system effectiveness) to 1.0 (complete confidence). One model tests the combined effects of having a system and the average ratings of effectiveness while a related model uses only the average ratings of effectiveness. In the same vein, a dummy variable is added to the basic model to reflect schools in which the adults have high cohesion in their assessments of the benefits of a reward system, as measured by an average effectiveness rating above .8. The cohesion variable is also tested as a replacement for the broader dummy of simple presence of a reward system. Finally, the composite variable developed to capture the intensity of the system is added to the basic model and subsequently tested as a substitute for the dummy variable for presence of a reward system.

The third set of estimates focus on the role of being part of a charter network. Specifically, some charter schools are stand alone schools, while others are parts of larger organization such as KIPP or Aspire or Uncommon Schools. These networks offer support and structure to their schools and this set of models investigates whether the incentive systems are proxies for larger network effects. The estimation builds on previous estimates by adding network membership. Thus, the effects of more refined characterizations of the reward systems were tested in models that also included a factor for charter network membership.

The final phase of the analysis extended the model beyond simple ordinary least squares regression. The concern is omitted variables bias in OLS, arising if there is an unobserved factor that is strongly associated with adoption and with performance. The correction for this effect -- two-stage least squares using an instrumental variable – requires an exogenous variable that can provide predictive power to the independent variable of concern but is not correlated with the outcome. We need exogenous variation in the use of incentive schemes. Selection of suitable instruments is always a challenge, as it was here. Here, we consider variation in the educational attainment of the communities surrounding the charter school. Specifically, if school leaders use neighborhood characteristics such as low education levels to decide on whether to introduce incentive programs – and these do not separately affect achievement – we can instrument incentive systems with this measure. Using the location of the charter school, the education profile for adults aged 18 – 64 residing in the zip code surrounding the school was obtained from the US Census Bureau website. From these, the proportion of

the adult population with a high-school education or less was calculated. This approach to the estimation was explored in the fourth set of models.

Incentive/Reward Systems

Since information on the design and function of incentive/reward systems is scarce, this section presents a brief overview of the plans. Fifty-seven percent of the charter school respondents (106 of 186) report using an incentive/reward system with their students. Where a system is used, it applies to all the grades in a school. About 40 percent use a scheme of points that accumulate over time that are then redeemed for rewards, while about a third use negative incentives such as subtracting points from a weekly starting total. The remaining schools use a combination or devise occasional short-term efforts such as before formative assessments.

Among the school that use an incentive/reward system, wide variation emerges in which school personnel participates in reviewing students, how often the students are assessed and how frequently the tally for each student resets.

Of the 106 schools that use an incentives/reward system, 99 (or 93 percent) cover both academic effort and behavior in their rubrics. Three schools focus only on behavior and four schools only on academics. A more detailed breakdown of what areas of student life these systems cover appears in the table below.

Academic Concerns (103 schools)	**Percent of All Schools Using Reward Systems That Cover this Area**
Completion of school work	89
Attendance	70
Grades	42
School- or grade-wide academic performance	33
Behavior Concerns (102 schools)	
Classroom behavior	93
Student-adult conduct	86
Citizenship	80
Peer-to-peer conduct	77
Extra effort	58
School- or grade-wide behavior	36

Most often, schools that use a system focus on classroom behavior, followed by completion of assignments (89 percent) and student-adult interactions, such as speaking respectfully or accepting direction (86 percent).

Teachers are relied on most heavily to implement the reward systems in the schools that use them as displayed in the table below. This makes sense considering the closer proximity of the teacher to students during the day. But the proportions also illuminate

that the role of principals in these schemes skews more to the disciplinarian rather than the academic leader: Larger shares of schools report the principal's involvement in behavioral areas than in academic ones.

		Who Assesses Students?*		
Academic Concerns	**Number of Schools**	**Principals**	**Teachers**	**Staff**
Completion of school work	94	17.0%	97.9%	16.0%
Attendance	74	44.6%	79.7%	58.1%
Grades	45	55.6%	91.1%	24.4%
School- or grade-wide academic performance	33	97.0%	93.9%	42.4%
Behavior Concerns				
Classroom behavior	99	37.4%	98.0%	26.3%
Peer-to-peer conduct	82	68.3%	100.0%	59.8%
Student-adult conduct	91	71.4%	97.8%	65.9%
Citizenship	85	65.9%	96.5%	63.5%
Extra effort	62	54.8%	96.8%	48.4%
School- or grade-wide behavior	38	89.5%	97.4%	47.4%

* Values are the percentages of schools using Principals, Teachers or Staff to rate each area.

Even if systems focus on the same things, they can differ in their periodicity. System-using schools reported how frequently the systems are designed to assess student behavior. For schools that use different periods for different elements of their plan, the most frequent interval was recorded[2]. The table below illustrates that slightly more than half the schools are structured to monitor students throughout the school day, either for the time a student is in classes or continuously from beginning to end of a school day. The remaining schools use a less frequent approach, including some that only conduct their reviews at the end of grading periods (every 5 weeks or quarterly), semesters or academic years.

Frequency of Review	**Number of Schools**	**Percentage**
Once/Semester	2	0.67%
Once/Quarter	2	0.67%
Once/5 weeks	2	0.67%
Monthly	61	20.40%

[2] For example, a school may review grades at the end of a grading period for achievement but also monitor students weekly for completion of assignments; in such cases, the frequency of review was coded as weekly.

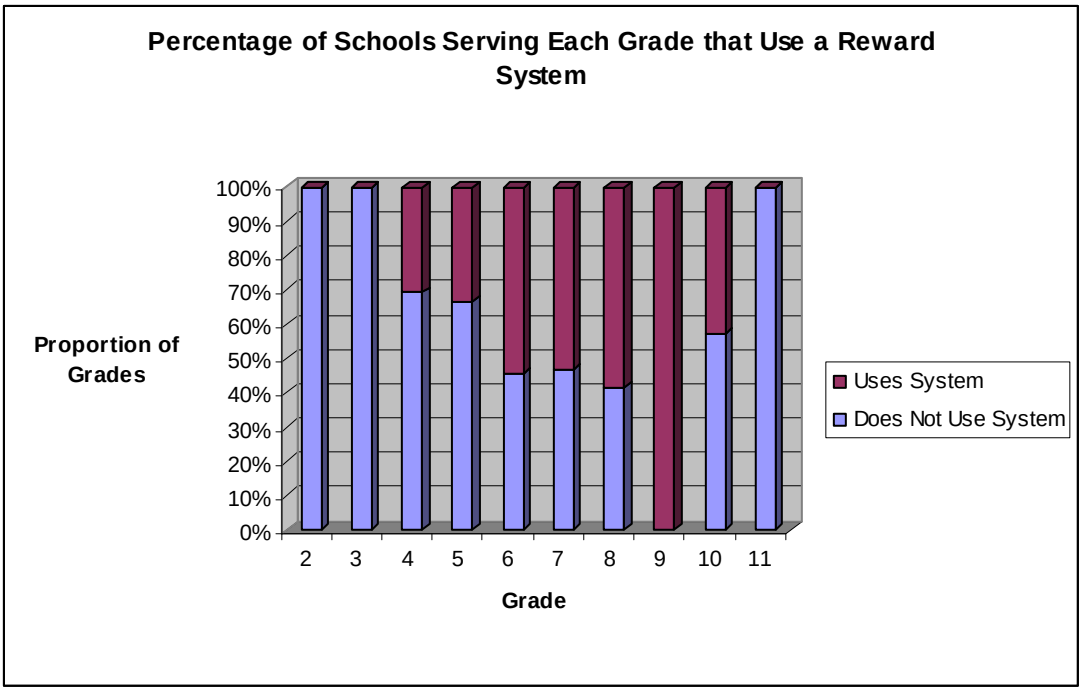
Biweekly	2	0.67%
Weekly	62	20.74%
During Classes	36	12.04%
Constantly	132	44.15%

The range of available rewards is slanted to those with an immediate benefit or perceived value to students. Brightly colored pens or notebooks at the school store have both utility and notability. Earning the right to wear the specially designated colored polo shirt is less prevalent only because many schools are sensitive to family budget concerns or do not have the resources to provide the shirts themselves. Resource constraints also limit the use of cash or college fund investments as rewards in most of the schools.

Reward	Number of Schools	Percentage
Access to select activities	87	82.1%
Certificate of merit	67	63.2%
Purchase items at school store	57	53.8%
Special uniform	45	42.5%
Cash	9	8.5%
College fund contribution	4	3.8%
Other rewards	14	13.2%

Charter schools have all been created relatively recently and are typically dynamic in their early years in size and structure. Since charter schools use different approaches to build their enrollment, they are difficult to classify using typical grade spans of elementary, middle and high school ranges. For example, a school may ultimately aim to serve grades K – 8 but start by enrolling grades 1 and 6 and fill in grades year by year. Accordingly, the descriptive data that follows applies to grades, not schools.

The chart below shows the proportion of schools that serving each grade that make use of an incentive reward system. The prevalence of reward systems rises in upper elementary to include all the schools with 9th grade before tapering off in the senior high grades.



Despite higher prevalence of reward systems in higher grades, ratings by principals of the effectiveness of the system decline as the grade level increases. Principals were asked to rate their sense of how effective the system is on a 1 (Low) to 10 (High) scale. When the ratings are examined by grade show as in the table below, the systems are considered more effective in early grades, and the rating declines over successive grades. The increased rating in 10th grade is partially explained by the fact that one school serves high school dropouts and features the reward system as a central element of their model; were that school excluded, the mean would be 5.33.

Grade	Average Principal Rating of System Effectiveness	Standard Deviation
2	8.55	1.34
3	8.72	1.27
4	8.71	1.49
5	8.58	1.41
6	8.52	1.4
7	8.17	1.6
8	7.98	1.61
9	6.5	2.12
10	6.55	1.99

Analysis of Impact

The basic models for both reading and mathematics regress one-year gains in achievement on baseline performance, demographics, poverty status, language proficiency, special education status and a binary variable to indicate whether a school employs a reward/incentive system. A description of the variables used in the analysis is presented in Appendix A. The basic model is then expanded with additional refinements to measurement of reward system.

Reward System Impact on Reading Performance.

Basic Models The results for the basic Reading models are found in Table 1. The basic model reflects the expected relationships of baseline scores, ethnicity and special education status. In the basic model, the simple presence of an incentive system is positive and significant (Model 1). The effect, when represented by a binary variable, is sensitive to model specification, though the coefficient remains positive through all the model variations. Its significance does not change with the addition of calendar year (Model 2), which was added to consider the increased urgency of performance under their charter and the No Child Left Behind accountability requirements.[3] As seen in Model 2, the time variable is insignificant. The addition of interaction variables to test whether there was a difference in effect by grade span (Model 3) revealed an unexpected negative and significant coefficient for the elementary grades interaction. Taken on its face, the coefficient would suggest that all else being equal, elementary schools that use reward systems produce lower learning gains compared to those that do not use them.

Variations to Basic Model As discussed earlier, we wanted to explore the impact of reward systems more deeply to see if they had common attributes that systematically associated with greater learning results. Three variables were developed: a measure of the perceived effectiveness of the system by the adults using it (Average Ratings of Effectiveness), a binary that subsets those schools that use a system that also have strong alignment among adults about the usefulness of a reward system (Cohesion), and the intensity of intervention that the reward system creates through its structure and function (Intensity). These system qualities were studied to see if more nuanced relationships might better explain the effect of these reward systems on student academic gains. The results of these estimations are presented in Table 2.

[3] For all public schools, The No Child Left Behind Act imposes performance expectations by requiring each school to demonstrate that all students attain Adequate Yearly Progress on state standardized achievement tests or face increasingly onerous consequences. Under the law, after six years of failing to meet the targeted performance levels, the legislation called for a school to be significantly reconstituted, or taken over by the state. Charter schools face an additional layer of accountability with the agency that grants their charter, and in most cases a shorter time frame to prove in. So time would be expected to have more impact on charter schools than on traditional public schools.

One possibility is that the perceptions among those who administer the system might make a difference in how well the systems operate. Lagging confidence in the approach, for example, might lead to lesser impacts. As shown in Model 4, adding the variable 'Average Rating of Effectiveness' whose values range from 0 (has no system) to 1.0 (all adults give the higest possible rating) does not significantly contribute to student gains when added to the Basic Model. The effectiveness variable is used as a substitute for the binary factor in the model in Model 5, and the coefficient of 0.15 is significant to roughly the same degree as the binary.

A related refinement examined the impact in schools where the average rating of effectiveness was an 0.8 or higher, indicating that there was strong alignment and support for the system among principal, teachers and staff. Using this specification limits consideration to those schools that create a unified front to students and hold the expectation that the systems work. Added to the basic model, as appears in Model 6, the binary for strong cohesion is not significant when added to the basic model, but when substituted for the more widespread system use variable in Model 7, produces one of the strongest effects, a coefficient of 0.156, an even greater magnitude than the coefficient on the simple measure.

To cover the possibility that the success of these systems is a function of the "contact" with students, a variable was created that combines two separate dimensions: how frequently the system design rates student conduct and the average ratings of effectiveness among school personnel– the intensity variable. The Intensity variable with continuous values from 0 to 1.0 might more accurately portray the differences across schools in cohesion or granularity of review of student behavior. The initial estimation of this relationship included Intensity as an additional exogenous factor to explain student academic gains; the results shown in Model 8 indicate that use of an incentive/reward system loses its significance when an additional factor for intensity is incorporated into the model.

However, the fact that the intensity variable is continuous serves two functions. First, it provides a way of distinguishing among the schools using a reward system, which is expected to translate to more precise estimations in the models than a simple binary. Second, it consolidates two strongly correlated factors into a single metric in a way that preserves the explanatory power of each. If the intensity variable has a value greater than zero, it means that there is a system operating. If the value is high, it means that there is both a shared view that the system is effective and that it is comprehensively applied. As one respondent noted, "We're all over the students all the time." Middling values arise if the school personnel are less convinced of the value of the approach or less unified in their viewpoints which could led to inconsistency or when the review of student conduct is sporadic, which could lead to inconstancy. So instead of including both frequency and cohesion into the model with their correlation of r = .67, the consolidated variable offers an alternative to the binary factor of "has/does not have a reward system". When the intensity variable substitutes for the binary use factor, as seen in Model 9, it produces the largest coefficient of all the models, 0.18, which is statistically significant. Thus, a

change from no program one that is both continuous and strongly supported by school staff produces a .18 standard deviation increase in reading gains.

Best Models The incremental modifications in the preceding phase of analysis helped form an understanding of the attributes of reward systems that materially affect their utility as a driver of student learning gains. The analysis then turned to the task of using the available data to identify the conditions of maximum impact for reward systems. A large number of models were explored in the process; the best resulting models appear in Table 3.

Many of the schools that participated in this study are members of a charter school network or Charter Management Organization (CMO). Unlike single charter schools, network charter schools are thought to derive benefits from the larger organizational structure, ranging from shared services such as human resources or information technologies. Another possible advantage stems from the notion that networks typically grow from successful original schools and therefore transfer a proven education model to the new environment, lessening the chance that the school is unsuccessful. Some charter networks include reward systems in their education model and some don't, so it is important to consider the effects of network participation separately from the effects of reward systems.

When the school's membership in a charter network is considered (Model 10), the reward system variable loses its significance by a small margin but is similar in magnitude to the simple cases. Importantly, the fact that a school is part of a network produces a shift of 0.13 standard deviations in student achievement gains.

The greatest impact for reward systems was found to occur in conjunction with school participation in charter networks. Models 11 and 12 show that similar results are obtained when reward systems are strongly supported by the adults who administer them or when the system has a high degree of intensity. Since these models consider the effects of network participation and reward system operations separately, they provide the largest degree of insight into the value of reward systems for creating student gains in reading. They show that the effects of charter network membership and use of a reward system are cumulative.

Extensions to the Model The models presented above assume a straightforward relationship between schools' use of a reward system and the gains that result in the students in those schools. However, there remains the concern that schools that adopt incentive and reward systems do so in systematic ways that are not reflected in the model. If schools that adopt a reward system have other characteristics not captured in the included exogenous variables, then importance of reward systems may be biased. Indeed, the surprising finding of a negative coefficient on the elementary grade – reward system interaction suggests the possibility of omitted variables. To address this question, a two-staged least squares model was developed. The results appear in Table 4.

Preliminary investigation of the instrumental variable shows that it is highly correlated with the use of reward systems and insignificant in creating achievement gains. The first column of Table 4 shows the simple regression of low education attainment on use of reward systems.

In the IV model, the coefficient on the instrumental variable construction of the reward system factor was not significant. The loss of significance suggests that the choice to implement a reward system is influenced by the lack of strong support for educational achievement in the communities where the schools locate, and that after controlling for that effect, the systems do not add much to student learning. There are, however, reasons to be concerned about these estimates. The samples are limited and the choice of instrument raises questions about persistent omitted relationships between the use of reward systems and determinants of achievement gains.

Reward System Impact on Mathematics Performance. The same array of models developed for the assessing the impact of incentive/reward systems on reading was tested on mathematics gains. The results appear in Appendix B. Unlike the models for Reading which showed a consistent positive effect from the adoption of a reward system, the performance of these systems on mathematics learning gains was found to be uniformly poor. Regardless of how the presence of a reward system is represented, the effect was insignificant. Moreover the rest of the variables in the models did little to explain the observed results in Math learning either. Contrary to the results obtained by Bettinger, this analysis does not support use of an incentives or reward system to foster gains in student learning in Mathematics.

Discussion

This exploratory research examined the role that incentive/reward programs in K-12 charter schools have on the academic learning of students. It is novel in a number of ways. It is one of the first times that charter schools as a group provide insight on the efficacy of one of their innovations; the maturity of the sector has reached a point where their potential as Research and Development organizations for public education can be realized. Further, since similar programs are growing in both number and complexity around the country, the lessons gleaned from this analysis provide immediately useful input for decisions regarding their use. As well, the results offer current users of reward systems insight into how to improve the impact of their designs.

The role for reward systems in accelerating mathematics learning is not informed by this analysis. Contrary to other work that found positive effects of a reward system on mathematics learning (but not reading), this research was unable to identify similar effects. Differences in study design may account for the differences in results; Bettinger used a random assignment design in one community, whereas this work includes a broad cross-section of schools and is not randomized. Unfortunately, these results are consistent with recent national findings that the parameters of sound instruction in mathematics remain obscure.[11]

Across a variety of model specifications, reward systems are found to have stable and consistent positive impacts for student learning in reading. The effect holds across grades and across network and non-network charter schools. Moreover, the success of incentive/reward systems appears to hinge on a number of factors that are under the control of the school.

Schools whose personnel are strongly aligned in the view that reward systems are effective fare better than schools where the support is weak or where the adults are less aligned. Where the adults align, several important conditions arise. The rules of the reward system are more consistently applied, which in turn leads to a more uniform signal of expectations for students both for behavior and for learning, which are well documented antecedents to achievement.[12] At the same time, having strong confidence in the effectiveness of a reward system is likely also to improve its impacts by reinforcing the expectations of the adults who employ it with students.

Additionally, schools in which there is continuous or near-continuous assessment of student conduct produce larger gains in reading than schools that have reward systems but tally up less frequently. This finding is consistent with established wisdom about continuity in reinforcement and reward: the more constantly behavior is reinforced, the faster and more enduring the learning. So schools might benefit from a redesign of their systems to provide their students more consistent and continuous feedback.

It is difficult to interpret the meaning of the instrumental variable model. One possibility is that the concern for endogeneity is unfounded and that the simple regression models are sufficient to explain the relationship between reward systems and student learning.

Another is that there *is* endogeneity but the limitations of either data or the choice of instrument fail to reveal the underlying relationship. The strong correlation between the instrumented reward system variable and student poverty hints at the complexity of the underlying dynamics. Other approaches to the instrument may do a better job of teasing out these effects. Adding another year of data and expanding the number of schools in the sample will also provide a more solid empirical base on which to test these relationships.

Recognizing that this analysis is preliminary, the findings are still promising. The use of a reward system was examined in a variety of ways and was found to be a significant predictor of student learning gains in reading. Further, the best model results reveal an important insight: participation in a network produces an independent effect on student learning that can be enhanced if the network also has a system that all the adults can strongly get behind. Likewise, the bump in performance that comes from being part of a network can be enhanced by increasing the intensity of the system that is in place.

Viewed in the larger context of school performance and improvement, this analysis illuminates the value of incentive/reward systems when structured and used appropriately. The average effect size on student learning gains associated with the adoption of a reward system across all the systems studied was 0.11 standard deviations. When systems operate under conditions of cohesive and strong support among school personnel and have continuous feedback built into their design, the effect size increases to 0.16. These compare favorably to other school attributes that are considered indicators of better quality; these models revealed a consistent effect size for a charter school being part of a larger network of .16 standard deviations. This comparison suggests that when done right, reward systems can produce a jump in learning gains for significantly less resources.

Areas of Further Work

As this is a first look at the structure and impact of incentive/reward systems, much work remains to be done. Data refinements to examine more schools and to examine the effectiveness of systems using student-level data are two obvious directions for future research. As well, additional development of the instrumental variable models may shed additional light on the conditions for which system adoption holds the highest promise. One option is to consider whether market strength of charter schools in a community affects the possibility that schools choose these programs as a competitive differentiator. And finally, inclusion of new data on the longer term impacts of such systems on students would further an understanding of the durability of the benefits these systems are found to deliver.

Figure 1
Taxonomy of Academic Incentive Programs

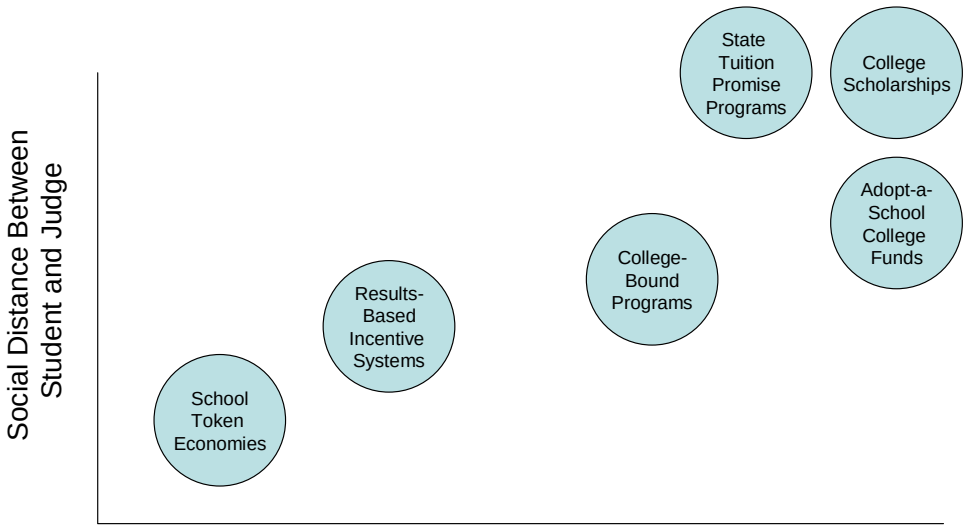


Table 1
Incentive System Impact Analysis- Reading

Baseline Models

Dependent Variable: Standardized Reading Gains	1 Reading with Baseline	2 Model with Time	3 Model with Grade Span
Base Year Performance	-0.325 (0.055)**	-0.323 (0.056)**	-0.326 (0.052)**
Uses Reward System (Binary)	0.116 (0.054)*	0.11 (0.053)*	0.176 (0134)
Percent Free/Reduced Lunch	0.07 (0.138)	0.048 (0.138)	0.011 (0.131)
Percent English Learners	-0.516 (0.137)**	-0.496 (0.134)**	-0.410 (0.132)**
Percent Special Education	-1.335 (0.374)**	-1.34 (0.380)**	-1.297 (0.378)**
Percent Black	-0.285 (0.143)*	-0.258 (0.144)	-0.23 (0.135)
Percent Hispanic	-0.055 (0.131)	-0.042 (0.130)	-0.066 (0.129)
Year		-0.096 -0.057	
Elementary School Grade* Uses System			-0.378 (0.169)*
Middle School Grade * Uses System			-0.024 (0.130)
Constant	0.244 (0.052)**	192.935 (114.257)	0.253 (0.051)**
Observations	119	119	119
R-squared	0.32	0.35	0.37

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 2
Incentive System Impact Analysis- Reading

Model Variations

Dependent Variable: Standardized Reading Gains	4 Effectiveness/Use Model	5 Effectiveness Only Model	6 Cohesion/Use Model	7 Cohesion Only Model	8 Intensity/Use Model	9 Intensity Only Model
Base Year Performance	-0.340	-0.336	-0.354	-0.356	-0.329	-0.327
	(0.053)**	(0.055)**	(0.055)**	(0.056)**	(0.054)**	(0.054)**
Uses Reward System (Binary)	-0.059		0.027		0.050	
	(0.231)		(0.069)		(0.085)	
Average Effectiveness Rating	0.210	0.150				
	(0.027)	(0.006)*				
Cohesion			0.137	0.156		
			(0.070)	(0.054)**		
Intensity					0.130	0.180
					(0.011)	(0.007)*
Percent Free/Reduced Lunch	0.092	0.082	0.079	0.092	0.040	0.048
	(0.136)	(0.138)	(0.136)	(0.134)	(0.134)	(0.132)
Percent English Learners	-0.536	-0.528	-0.558	-0.570	-0.440	-0.418
	(0.136)**	(0.136)**	(0.129)**	(0.129)**	(0.141)**	(0.137)**
Percent Special Education	-1.439	-1.412	-1.526	-1.531	-1.372	-1.350
	(0.374)**	(0.376)**	(0.370)**	(0.372)**	(0.389)**	(0.396)**

Percent Black	-0.310	-0.301	-0.297	-0.303	-0.247	-0.241
	(0.142)*	(0.143)*	(0.139)*	(0.140)*	(0.139)	(0.138)
Percent Hispanic	-0.065	-0.062	-0.040	-0.037	-0.057	-0.056
	(0.130)	(0.131)	(0.132)	(0.131)	(0.131)	(0.131)
Constant	0.251	0.248	0.253	0.255	0.239	0.238
	(0.053)**	(0.053)**	(0.052)**	(0.053)**	(0.053)**	(0.053)**
Observations	119.000	119.000	119.000	119.000	119.000	119.000
R-squared	0.330	0.330	0.350	0.350	0.330	0.330

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 3
Incentive System Impact Analysis - Reading

Best Model Solutions

Dependent Variable: Standardized Reading Gains	10 Network Model	11 Best Model 1	12 Best Model 2
Base Year Performance	-0.342 (0.055)**	-0.383 (0.054)**	-0.348 (0.054)**
Uses Reward System (Binary)	0.095 (0.053)		
Member of charter network	0.133 (0.067)*	0.171 (0.068)*	0.147 (0.069)*
Average Effectiveness Rating			
Cohesion		0.163 (0.052)**	
Intensity			0.16 (0.007)*
Percent Free/Reduced Lunch	-0.064 (0.152)	-0.106 (0.142)	-0.111 (0.146)
Percent English Learners	-0.527 (0.135)**	-0.574 (0.123)**	-0.432 (0.134)**
Percent Special Education	-1.263 (0.358)**	-1.497 (0.352)**	-1.292 (0.368)**
Percent Black	-0.306 (0.145)*	-0.318 (0.140)*	-0.261 (0.140)
Percent Hispanic	-0.061 (0.130)	-0.047 (0.127)	-0.064 (0.130)
Constant	0.263 (0.052)**	0.278 (0.053)**	0.258 (0.052)**
Observations	119	119	119
R-squared	0.35	0.39	0.36

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 4
Incentive System Impact Analysis - Reading

Instrumental Variable Model

	Logistic Prediction	IV Model
Low Education	1.047 (.008)**	
Base Year Performance		-0.319 (0.055)**
Linear Predictor – Uses Reward System		0.069 (0.173)
Percent Free/Reduced Lunch		0.11 (0.197)
Percent English Learners		-0.534 (0.177)**
Percent Special Education		-1.261 (0.486)*
Percent Black		-0.303 (0.141)*
Percent Hispanic		-0.051 (0.137)
Constant	(1.892) (5.71)**	0.247 (0.062)**
Observations	577	119
R-squared		0.32

Absolute value of z statistics in parentheses
* significant at 5%; ** significant at 1%

APPENDIX A

Description of Variables

Variable	Description	Min Value	Max Value
Reading Gain	Change in standardized scale score on state achievement tests $Y_2 - Y_1$	-.564	1.097
Base Year Score	Standardized score on state achievement test in Y_1	-1.191	.980
Use Reward System	Binary reflecting adoption of an incentive/reward system in school	0	1.0
Linear Predictor	Logistical prediction of likelihood of having reward system in school (based on Low Education IV)	0	1.0
Low Education	Percent of adults 18 – 64 in school's zip code with HS education or less	0	100
Year	Calendar year in Y_2	2004	2007
Network	Binary reflecting school's membership in charter school network or Charter Management Organization	0	1.0
Elementary Grade	Binary representing cohort enrolled in elementary school grade in Y2	0	1.0
Middle School Grade	Binary representing cohort enrolled in middle school grade in Y2	1	1.0
Average Rating of System Effectiveness	Average of ratings of Principal, teachers and staff perceptions of how effective the incentive/reward system is in their school	0	1.0
Cohesion	Binary reflecting schools with a reward system and an average rating of effectiveness >.8	0	1.0
Intensity	Index variable equal to the product of the average rating of system effectiveness and the a measure of how frequently the system design calls for review of students (once a semester to continuously)	0	1.0

References

[1] Ehrenberg, Ronald G., Liang Zhang and Jared M. Levin. "Crafting a Class: The Trade-off Between Merit Scholarships and Enrolling Lower-Income Students.", NBER Working Paper 11437, June 2005.

[2] Dynarski, Susan. "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion." *American Economic Review* 93.1 (March 2003): 279-288.

"From 6 New Benefactors, 425 College Dreams", New York Times, June 21, 1986.

[4] Frederick, Shane, George Lowenstein and Ted O'Donoghue. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature,* Vol. XL,(June, 2002), pp. 351-401. Bishop, John. Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn -- Externalities, Information Problems and Peer Pressure in Eric A. Hanushek and Finis Welch (eds.) Handbook of the Economics of Education, Volume 2, 2006, Elsevier B.V. pp. 909 - 944.

[5] Agresta, Kate, Jeff Buenrostro, Jenna Lawrence, Abigail Rubin. "ON TRACK: Factors Influencing Student Retention in the College Track Program", unpublished, January 2007.

Pryor, Donald and Rebecca Sumner. "The Hillside Work-Scholarship Connection: Charting a Course for the Future", Rochester, NY: The Center for Governmental Research, 2004.

[6] Harris Cooper. Homework. White Plains, NY: Longman Publishers. 1989.

[7] Power, Thomas J, James L. Karaustis, and Dina F. Habboushe. Homework Success for Children with ADHD: A Family-School Intervention Program. New York: The Guilford Press, 2001.

Reitman, David, et al. "The Influence of a Token Economy and Methylphenidate on Attentive and Disruptive Behavior During Sports With ADHD-Diagnosed Children." *Behavior Modification* 25.2 (2001): 305-323.

Students in Cash-Incentives Study Score Higher in Math, Education Week, February 27, 2008

[7] Angrist, J., and V. Lavy, (2004) "The Effect of High Stakes High School Achievement Awards: Evidence from a School-Centered Randomized Trial", IZA Discussion Papers Number 1146.

Jackson, C. Kirabo "A Little Now for a Lot Later: A Look at a Texas Advanced Placement Incentive Program", Dec 7, 2007

Kremer, Michael , Edward Miguel, Rebecca Thornton. "Incentives to Learn", NBER Working Paper No. 10971, December 2004.

[8] Lubbers, Miranda J. Group composition and network structure in school classes: a multilevel application of the p^* model. *Social Networks, Volume 25, Issue 4, October 2003, Pages 309-332.*

9 Bettinger, Eric. Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores. Paper Presented at CESifo Conference "Economic Incentives: Do They Work in Education? Insights and Findings from Behavioral Research," Munich, 16-17 May 2008.

[10] Corpus, J., Lepper, M., & Iyengar, S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences
and academic correlates. Journal of Educational Psychology, 97(2), pp. 184-195.

Guthrie, J. T., Wigfield, A. L., Humenick, N. M., Perencevich, KC., Taboada, A., & Barbosa, P. (2006). Influences of stimulating
tasks on reading motivation and comprehension. *The Journal of Educational Research, 99*(4), 232-245.

Jalenko, Mary Renck. Beyond Benchmarks and Scores: Reasserting the Role of Motivation and Interest in Children's Academic Achievement. *Childhood Education* International Issue, 2007. pp. 395-407.

[11] National Mathematics Advisory Panel. *Foundations for Success: The Final Report of the National Mathematics Advisory Panel,* U.S. Department of Education: Washington, DC, 2008. page 35.

[12] Bishop, John. Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn -- Externalities, Information Problems and Peer Pressure in Eric A. Hanushek and Finis Welch (eds.) Handbook of the Economics of Education, Volume 2, 2006, Elsevier B.V. pp. 912.