Conditional Cash Penalties in Education:
Evidence from the Learnfare Experiment

Thomas S. Dee
Swarthmore College and NBER
e-mail: dee@swarthmore.edu

June 2009

Abstract - Wisconsin's influential Learnfare initiative is a conditional cash *penalty* program that sanctions a family's welfare grant when covered teens fail to meet school attendance targets. In the presence of reference-dependent preferences, Learnfare provides uniquely powerful financial incentives for student performance. However, a 10-county random-assignment evaluation suggested that Learnfare had no sustained effects on school enrollment and attendance. This study evaluates the data from this randomized field experiment. In Milwaukee County, the Learnfare procedures were poorly implemented and the random-assignment process failed to produce balanced baseline traits. However, in the nine remaining counties, Learnfare increased school enrollment by 3.7 percent (effect size = 0.08) and attendance by 4.5 percent (effect size = 0.10). The hypothesis of a common treatment effect sustained throughout the six-semester study period could not be rejected. These effects were larger among subgroups at risk for dropping out of school (e.g., baseline dropouts, those over age for grade). For example, these heterogeneous treatment effects imply that Learnfare closed the enrollment gap between baseline dropouts and school attendees by 41 percent. These results suggest that well-designed financial incentives can be an effective mechanism for improving the school persistence of at-risk students at scale.

Keywords: Education, welfare, incentives, attainment, experiment
JEL Classifications: I2, I3

"Eighty percent of success is showing up." - Woody Allen

1 - Introduction

The recent growth in economic inequality and the well-established importance of education for economic success have created a focused interest in identifying scalable policies that can promote the human-capital accumulation of at-risk youth. Some of the most fundamental antecedents to cognitive development are "non-cognitive" traits like academic engagement and motivation. However, the deterioration of family environments in recent decades and the relative lack of corresponding school and community supports may disadvantage the neediest children with respect to the development of these instrumentally relevant traits.

Concerns like these have motivated a renewed interest in leveraging student's academic engagement and improving cognitive development through providing performance-based financial incentives for students. In particular, several recent studies have focused on the effects of providing cash incentives linked directly to the test scores and course performance of K-12 and post-secondary students in developed nations (e.g., Angrist and Lavy 2008, Angrist, Lang, and Oreopolous 2009, Bettinger 2009, Leuven, Oosterbeek, and van der Klaauw, forthcoming, and Richburg-Hayes et al. 2009). In developing countries, the proliferation of "conditional cash transfer" (CCT) programs has provided family-based financial incentives for school attendance and the utilization of social services (e.g., Handa and Davis 2006).

However, Wisconsin's seminal Learnfare program - a welfare-waiver reform that sanctioned a family's welfare grant when covered teens failed to meet school attendance targets - provides a distinctive contrast to conventional cash-incentive policies.[1] For example, like CCT programs (e.g., Mexico's PROGRESA), Learnfare linked a family-based grant to meeting attendance targets. But Learnfare could be termed a "conditional cash penalty" (CCP) program in that it *reduces* an extant welfare grant for failure to meet program requirements. Because of the evidence that people exhibit an asymmetric aversion to income losses relative to a reference point (Kahneman and Tversky 1979), this aspect of Learnfare may amplify its behavioral effects.

Notably, Learnfare also differs from other recently studied incentive programs in developed countries by leveraging *family* involvement instead of directly targeting students with

---

[1] Learnfare, which began over two decades ago, has been enormously influential in shaping other state policies. Thirty-eight states currently take advantage of the flexibility created by the 1996 Federal welfare reforms to craft similar policies that link school attendance and welfare receipt (Education Commission of the States, 2007).

cash incentives. At least one other design feature of Learnfare is particularly noteworthy. The psychological literature on the use of extrinsic rewards in education suggests that they can be ineffective or even harmful when students feel they lack the capacity to meet the stated requirements. However, Learnfare targets outcomes that are likely to be viewed as comparatively attainable but still economically and educationally meaningful (i.e., school attendance rather than achievement targets).

Despite these unique and compelling design features, a 10-county random-assignment evaluation of Wisconsin's Learnfare program suggested that it had at best modest and short-term effects on its targeted enrollment and attendance outcomes (Frye and Caspar 1997). In this study, I re-examine the data from that random-assignment study. In particular, I exploit panel-based econometric specifications based on pooling the available enrollment and attendance data from the six-semester study period. These specifications increase the statistical precision of the estimated treatment effects. Furthermore, they provide a unified framework for assessing the impact of study attrition and the quality of the random assignment results. This research design also provides a framework for formal hypothesis tests related to the dynamic treatment effects of Learnfare assignment (i.e., distinguishing short and long-term effects).

The results of this analysis indicate that, in Milwaukee County, the county-based random-assignment procedures did not produce balanced baseline traits. In particular, black teens in Milwaukee County were significantly less likely to be subjected to Learnfare's requirements. Furthermore, legal challenges weakened the Learnfare requirements in this county while logistical challenges related to the accurate tracking of attendance data made the program comparatively slow and capricious. For these reasons, this analysis focuses largely on the nine remaining counties that participated in the study where the program implementation was relatively good and the random-assignment procedures appear to have performed well.

The results based on these counties indicate that random assignment to the Learnfare restrictions generated statistically significant improvements in both school enrollment (3.7 percent increase, effect size = 0.08) and school attendance (4.5 percent increase, effect size = 0.10). Attrition from the study compromises the statistical power of inferences about the longer-term effects of this random assignment. However, the hypothesis that these treatment effects were the same throughout the study period cannot be rejected, even in models that allow for alternative methods of imputing missing outcome data. Furthermore, the estimated treatment

effects of Learnfare were particularly large for at-risk subgroups. For example, Learnfare increased the enrollment of teens identified as baseline dropouts by 25 percent. This study concludes with a discussion of the unique policy design and implementation lessons from Wisconsin's experience with Learnfare.

2 - Financial Incentives for Students

The notion that financial incentives will influence behavior in the expected directions is commonplace in economics. In contrast, an extensive literature in psychology (Deci, Koestner, and Ryan 2001) that began with a classic laboratory experiment by Deci (1971), suggests that extrinsic rewards in education can substantially undermine student performance by decreasing their *intrinsic* interest in the targeted tasks.[2] However, Cameron (2001) argues that this interpretation conflates the heterogeneous effects of extrinsic rewards for individuals with high and low levels of initial intrinsic motivation. When students lack intrinsic motivation, external incentives can improve academic outcomes (Cameron and Pierce 2002). However, for students who already possess intrinsic motivation, there is evidence that external rewards can be harmful. In a review of this literature, Camerer and Hogarth (1999) also underscore the importance of whether the task targeted with financial incentives is "effort responsive." With regard to both of these concerns, Learnfare would appear to be well designed. Because Learnfare applies only to economically disadvantaged families (i.e., those receiving welfare), it may target teens with comparatively low baseline levels of intrinsic motivation. And, because Learnfare is linked to school attendance and not academic performance, most covered teens should feel comparatively capable of avoiding the financial penalties.

A surprisingly large number of recent random-assignment evaluations have examined the effects of extrinsic education-related awards in field settings. Perhaps, the most well-known of these program evaluations involves Mexico's seminal conditional cash transfer (CCT) program, which was originally called PROGRESA. This program, which has been replicated in multiple countries, provided cash payments to parents every two months conditional on children meeting school attendance goals. Evaluations of this program found that it generated significant improvements in school enrollment as well as other outcomes (e.g., Skoufias and McClafferty

---

[2] Writing from an economics perspective, Bénabou and Tirole (2003) explicate the determinants of intrinsic and extrinsic motivation in a principal-agent model where agents infer information about themselves and the task at hand from principal's provision of encouragement and rewards (i.e., the "looking-glass self").

2001). In a random-assignment study conducted in Kenya, Kremer et al. (2004) provided financial awards (i.e., cash grants and school fees) to adolescent girls who met test-score targets. This treatment increased test scores by 0.15 standard deviations and exhibited program externalities in that it also increased the academic performance of boys (who were ineligible) and girls with low baseline scores (who were unlikely to earn rewards).

Several of the studies conducted in developed nations have focused on postsecondary students. For example, Angrist, Lang, and Oreopolous (2009) evaluated the direct and interactive effects of financial rewards linked to GPA performance and academic support services for first-year students at a large Canadian university. The financial rewards, particularly in combination with the offer of support services, improved the performance of female students but not male students. Leuven, Oosterbeek, and van der Klaauw (forthcoming) evaluated the effect of providing cash rewards of different sizes to students at the University of Amsterdam who completed their first-year credit requirements. They found that these rewards improved the performance of students whose measured performance in high school mathematics was high but lowered the performance of students whose prior mathematics achievement was weaker, an effect interpreted as consistent with the degradation of intrinsic motivation. A third random assignment, post-secondary study (Richburg-Hayes et al. 2009) evaluated the effects of providing financial rewards to parents planning to attend or already attending a community college in Louisiana. These financial incentives, which were linked to enrollment and GPA targets, improved the number of credits earned, longer-term college persistence as well as measures of motivation.

Two other recent random-assignment studies in developed countries evaluated the effects of financial incentives at elementary and secondary levels. Angrist and Lavy (2008) examine the effects of a school-level policy providing cash incentives for Israeli students to complete a matriculation certificate required for post-secondary schooling. The results of this cluster-randomized trial indicate that cash incentives increased the performance of girls but had no effects on boys. Bettinger (2009) presents an evaluation of cash incentives linked to performance on standardized tests for elementary-school students in a low-income section of eastern Ohio. These incentives increased scores in mathematics (effect size = 0.15) and did not lower measures of intrinsic motivation but had no detectable effects on reading, social science, and science scores. Similar K-12 studies (Medina 2008, Vargas 2009) are ongoing in several cities where

student-level financial incentives are linked to attendance, behavior, and academic performance (Washington, DC), test scores (New York City), and grades alone (Chicago).

In addition to these recent studies, six other random-assignment studies evaluated programs that, like Learnfare, linked the threat of financial *sanctions* to school attendance. Campbell and Wright (2005) argue that two of these programs (Maryland's Primary Prevention Initiative and Delaware's A Better Chance program) particularly resembled Wisconsin's seminal Learnfare program in that they targeted teen welfare recipients and relied primarily on the threat of sanctions rather than an expansion of case-management or support services. These two programs appeared to have negligible effects on school enrollment and attendance (Stoker and Wilson 1998, Fein et al. 2001). The four other programs (i.e., the Teenage Parent Demonstration Program, Ohio's Learning, Earning, and Parenting Program, California's Cal-Learn Demonstration Project, and San Diego County's School Attendance Demonstration Project) largely targeted teen parents on welfare and blended the threat of sanctions with program features such as intensive case management, support services and financial bonuses for performance. Evaluations of these initiatives suggest that they did increase school enrollment and, to a lesser extent, attendance (Maynard 1993, Bos and Fellerath 1997, Mauldon et al. 2000, and Jones et al. 2002). However, Campbell and Wright (2005) suggest that these comparative results imply that financial sanctions are less likely to be effective when used in isolation from related services and case management.

Taken as a whole, the field-experimental literature on extrinsic rewards in education provides virtually no evidence that such policies have unintended negative consequences, contradicting the concerns that have dominated the lab-experimental literature from psychology. However, the evidence that extrinsic rewards and penalties are consistently effective in promoting targeted outcomes is decidedly mixed. This pattern of robust treatment effects and null findings suggests that program-design details, implementation quality and participant targeting are important policy parameters. In the next section, I describe Wisconsin's seminal Learnfare program in more detail.

3 - Wisconsin's Learnfare Program

In mid 1980s, the state of Wisconsin was in the vanguard of states that utilized increased Federal flexibility (i.e. waivers) to experiment with the design and implementation of its welfare

programs. Wisconsin's "first wave" of waiver demonstrations both reduced the work disincentives for welfare recipients and expanded existing job-search and training requirements to the mothers of pre-school children. However, the "centerpiece of the first round of Wisconsin initiatives" (Wiseman 1996) was the new Learnfare policy that linked welfare receipt to the school attendance of covered teens. The philosophical motivation for these changes was rooted in an interpretation of social-contract theory (e.g., Mead 1986) which argues that the receipt of welfare creates an implicit obligation for the recipient to undertake activities (e.g., employment, job training, and school attendance) that can break cycles of economic dependency. Learnfare required that teens in families receiving welfare, including teen parents, attend school regularly if they had not graduated from high school or completed an equivalency degree. Specifically, school attendance records were reviewed upon initial application for welfare and twice a year thereafter. Teens who were not enrolled in school (and who had not graduated from high school, completed an equivalency degree or shown good cause) were removed from their family's welfare grant until school enrollment was established.

If a review indicated that an enrolled teen had 10 or more unexcused full-day absences in a semester, they were designated as having poor attendance and were subjected to monthly monitoring. Families on monthly monitoring received monthly notices that reminded them of Learnfare's attendance requirement and offered services designed to assist with school-attendance problems.[3] However, when monthly monitoring indicated that a student had more than 2 unexcused, full-day absences in a month, the family was informed that it would face a 1-month benefit sanction unless it could show good cause for the absences. The amount of the sanction depended on the family's status. For example, the sanction for a single-parent with two children would be approximately $80 per month while, for a teen parent living alone, the sanction would be $190 (Quinn and Magill 1994). According to Frye and Caspar (1997), these sanction amounts 2 generally ranged from $60 to $190.

The actual application of sanctions appears to have been relatively infrequent. For example, in the 10-county random assignment evaluation that is the focus of this study, 26 percent of the teens assigned to Learnfare were subjected to monthly monitoring at least once

---

[3] However, Wisconsin secured a waiver from Federal requirements for assessment and identification of supportive services prior to sanctioning. Wisconsin was also exempted from Federal requirements for a "conciliatory procedure" to resolve disputes prior to sanctioning, though a 1990 court decision restored some "due process" requirements (Quinn and Magill 1994).

during their first four semesters and only 9 percent were ever sanctioned. In the typical semester, the sanction rate among Learnfare teens was less than 5 percent (Frye and Caspar 1997, page 18). Given the relatively modest sanction rate, it is not surprising that assignment to the Learnfare treatment did not have a statistically significant effect on the likelihood or magnitude of AFDC receipt.

Learnfare was implemented for teen parents and 13-14 year olds in March of 1988 and extended to all covered teens by September 1988 (Etheridge and Perry 1993). Governor Tommy Thompson advocated the early implementation of Learnfare. Wisconsin's early experience with Learnfare was characterized as an "administrative disaster" (Wiseman 1996) because of the difficulties of establishing new, reliable and accurate links between schools and welfare offices for attendance monitoring. While the quality of Learnfare monitoring had largely improved throughout the state by the time of the random-assignment evaluation, Milwaukee County is a notable exception. This county contains both the largest school district in the state (Milwaukee Public Schools) and roughly 50 percent of the state's Learnfare-eligible population (Frye and Caspar 1997).

Milwaukee County effectively had a separate set of Learnfare procedures that included an additional attendance verification check that delayed the time that lapsed between attendance violations and benefit sanctions. This procedure was adopted in 1992 as a part of a settlement to a lawsuit (Kronquist v. Whitburn), which alleged that Learnfare procedures violated due process because of the exceptionally poor quality of the attendance data in Milwaukee County schools. These procedures created an "appreciably longer" time between poor attendance and a sanction (Frye and Caspar 1997). Outside of Milwaukee County, poor attendance could trigger a processed sanction in as little as 2 months. In Milwaukee County, the lapsed time to a sanction would be at least twice as long.

Furthermore, as a practical matter, a 1995 review found that the average time between poor attendance and the resulting sanction was actually 6.6 months in Milwaukee Public Schools (Frye and Caspar 1997). This review also found that poor data quality and processing errors in Milwaukee Public Schools led to false negatives: the absence of sanctions in situations when the school attendance of covered teens failed to meet Learnfare standards. Because of these concerns, both the primary analysis of Learnfare's experimental evaluation and this re-analysis treat Milwaukee County separately from the other participating counties.

4 - A Random-Assignment Learnfare Evaluation

The Federal waivers that allowed Wisconsin to introduce a policy like Learnfare also required that comprehensive evaluations were conducted. An early non-experimental evaluation based on administrative data from six school districts prior to and after the introduction of Learnfare (Pawasarat, Quinn, and Stetzer 1992) found no evidence that Learnfare improved school attendance. The quality of these inferences was hotly debated by state officials and the evaluation team (Quinn and Magill 1994) Nonetheless, the report in question acknowledged itself that "Given the limitations of the control group populations and problems of identifying AFDC and non-AFDC teen parents, the Learnfare hypothesis testing lacks the strength of an experimental design using random assignment." However, a subsequent evaluation (Frye and Caspar 1997), which did utilize random assignment, indicated that the Learnfare program had at most short-term school-participation effects for certain sub-groups (Education Week, 1997). That random-assignment evaluation is the focus of the re-analysis presented here.

4.1 Study Design

The random-assignment evaluation of Learnfare was based on data from 10 counties. These 10 counties were chosen from Wisconsin's 72 counties by a procedure that sought both representativeness of the statewide Learnfare population and a balance of other programmatic concerns. Specifically, counties with fewer than 125 Learnfare teenagers were excluded from consideration because of the impracticality of monitoring attendance for small numbers of welfare recipients (Frye, Caspar, and Merrill 1992). Other counties (with the exception of Milwaukee County) were excluded because they were participating in a contemporaneous evaluation of the Parental and Family Responsibility program, which influenced the incentives of teen mothers receiving welfare to marry and abstain from having further children (Hoynes 1997, page 133). These exclusions left 29 counties as potential participants in the Learnfare evaluation.

Ten counties were randomly selected from this pool with probabilities proportional to their share of the statewide Learnfare population (Milwaukee, Brown, Douglas, Eau Claire, Kenosha, La Crosse, Marathon, Marinette, Portage, and Racine). However, stratification insured the participation of 3 rural counties (i.e., Marathon, Portage, and Marinette). Between March of 1993 and April of 1994, 3,205 teenagers from these 10 counties were selected for the study. Selection into the study occurred at the time when a teenager was scheduled to be introduced to Learnfare. This usually occurred when a member of an ongoing AFDC case turned 13 or when a

new AFDC case opened.[4] Study participants had to meet the basic requirements for the Learnfare program: aged 13 to 19, either a parent or living with natural or adoptive parents, and having neither graduated from high school nor completed an equivalency degree. Teens with a sibling who had been on the AFDC case and aged 13 to 19 during the previous 12 months were excluded from the study (Frye, Caspar, and Merrill 1992).

Once baseline data had been collected and a teen had been determined as eligible for the study, they were randomly assigned a treatment status. A statewide specialist was available to review the eligibility determination and to conduct the random assignment. However, another option was for county staff to make these designations (Frye, Caspar, and Merrill 1992). Teens assigned to the treatment received the usual introduction to Learnfare and were subject to its sanctions. Those assigned to the control group were not introduced to Learnfare and were exempted from its restrictions for the duration of the study.[5]

4.2 Outcome Measures

For each study participant, school enrollment and attendance data were collected over a six-semester study period (i.e., spring 1993 through fall 1995). Both the original analysis and this study's re-analysis focus on 3 distinct school enrollment and attendance measures. First, school enrollment is measured by the number of months in the semester for which a student's enrollment was verified. This measure varies from 0 to 4.5 in increments of 0.5. Second, the attendance rate identifies the fraction of school days in the teen's school district for which the student was in attendance. A third measure identifies the fraction of school days for which the student had an unexcused full-day absence. These last two measures are not fully symmetrical because of *excused* student absences. Identifying the comparative effects of Learnfare on the attendance rate and the rate of unexcused absences provides a direct way to assess whether Learnfare generated genuine increases in attendance or merely increased the use of excused absences. For the full student-by-semester sample, the mean value of the months-enrolled

---

[4] A teenager who had not previously been participating in Learnfare could also enter the study upon moving to the home of a parent receiving welfare support.

[5] One potential issue with welfare demonstrations of this sort is that their limited duration may bias the inferences towards finding no effect by weakening the treatment contrast (e.g., Hoynes 1997). However, in this instance, the study window of four to six semesters covers a substantial portion of the period during which Learnfare would be binding for an AFDC recipient.

measure is 3.4 (SD = 1.68). The mean attendance rate is 0.687 (SD = 0.352) while the mean rate of unexcused absences is 0.257 (SD = 0.366).[6]

Table 1 illustrates the basic panel structure of the available data by showing the number of study participants by month of entry and the number of subjects with valid attendance data by each of the six available semesters. This table also suggests the extent of attrition from the sample used in the original analysis (i.e., observations of attendance data). In the absence of attrition, we would expect to see 3,205 observations for each of the last four study semesters. However, the number of observations with attendance data drops from 2,833 in the spring of 1994 to 2,070 in the fall of 1995. That is, by the last semester of the study, attendance data were not available for over a third of the study participants.

This attrition is due in large part to the difficulty of tracking study participants who moved. The absence of outcome data for some study participants could compromise both the internal and the external validity of the impact analysis. For example, the estimated effect of Learnfare on the enrollment and attendance measures would be biased upwards if study participants who were assigned to the treatment but unlikely to meet Learnfare's restrictions were more likely to move away.[7]

However, there was also an unconventional dimension to the missingness of some outcome data in the original Learnfare analysis. The enrollment and attendance data are not defined for study participants who completed high school or a GED equivalency. Most of the study participants (i.e., slightly more than half) were only 13 years old when they entered the study so they did not have sufficient time for the typical period of high school completion during the study window.[8] Therefore, this study cannot provide a strong test of whether the Learnfare restrictions improved the probability of completing high school.

Nonetheless, inferences based on the preferred specifications applied to the data outside Milwaukee County suggest that random assignment to the Learnfare restrictions had a positive, though not quite statistically significant (p-value = 0.122), effect on high school completion. This pattern of positive treatment effects implies that the primary evaluation's approach of

---

[6] While most of the analysis presented here utilizes these measures, the results based on several alternative outcome measures (e.g., binary indicators for no enrollment, full-time enrollment, no unexcused absences, etc.) are also presented.

[7] Because Learnfare was in place statewide during the evaluation, only an out-of-state relocation could circumvent its restrictions.

[8] Only 5.1 percent of the student-by-semester observations were identified as high school completers.

eliminating high school completers from the enrollment and attendance analysis biases the estimated treatment effect downward. The attrition of high-school completers from the original analysis may particularly complete identifying the longer-term effects of Learnfare (e.g., four semesters after random assignment).

This study presents new evidence on the determinants of attrition from the Learnfare evaluation and, in particular, on the effects of random-assignment status. The empirical relevance of study attrition is also examined by presenting impact estimates based on several alternative procedures for imputing the missing outcome data.

4.3 Replicating Frye and Caspar (1997)

Before moving to an independent analysis of the Learnfare data, this section establishes an important baseline by describing and replicating the key evaluation results reported by Frye and Caspar (1997). This primary evaluation estimated the effects of random assignment to Learnfare on the 3 enrollment and attendance measures (i.e., months enrolled, rate of attendance, rate of unexcused absences) using separate cross-sections of study participants defined by whether they were in their first, second, third, or fourth study semester. So, for example, the "first-semester" results are based on pooling outcome data from the spring 1993, fall 1993 and spring 1994 semesters.

I report regression results based on the same sample selection and a similar regression specification in Table 2. These results are similar to those reported by Frye and Caspar (1997, Table 14).[9]. For the study participants from Milwaukee County, random assignment to Learnfare appears to have had small and statistically insignificant effects on enrollment and attendance across all 3 outcome measures and regardless of the length of time in the study.[10] Outside of Milwaukee County, where the randomization procedures appear to have performed well, Learnfare appears to have generated significant increases in enrollment and attendance (e.g., a 3 percentage-point increase in attendance) but only in either the first or second semester.

This apparent lack of persistent treatment effects is the basis for the widespread view that Learnfare did not have meaningful effects on its targeted outcomes. However, this interpretation

---

[9] The sample sizes match exactly for all 24 subgroups. However, the estimated treatment effects reported here differ slightly because of modest differences in the regression controls. For example, the results in Table 2 condition on unrestrictive county and semester fixed effects.

[10] The fourth-semester enrollment result for Milwaukee County suggests that Learnfare reduced enrollment. This weakly significant effect suggests the harmful effects of cash incentives on intrinsic motivation. However, the poor treatment-control balance for the study participants from Milwaukee County suggests that these inferences lack internal validity.

may be inaccurate for a number of reasons. First, an analysis based on the cross-sections in Table 2 fails to exploit the statistical precision made available by the panel structure of the available study data. Second, a panel-data approach to this analysis would also provide a framework for explicit tests of whether the treatment effects have statistically significant differences across semesters.

Third, while it is true that the estimated treatment effects appear to decline with the length of time in the study, these longer-term effects are also estimated with comparatively less precision because study attrition from the study substantially reduces the number of observations observed for multiple semesters. And the lack of precision associated with longer-term effects may be meaningful. For example, the 95-percent confidence intervals for the fourth-semester treatment effects for each of the 3 outcome variables include the corresponding first-semester point estimate. Statistical tests based on the pooled data can indicate more formally whether the data reject the hypothesis of a common treatment effect across the length of time in the study.

5 – Treatment-Control Balance

The fundamental rationale for using random assignment to choose the Learnfare status of these study participants was to break the correlation that might otherwise exist between the determinants of the outcomes under study and assignment to Learnfare. However, it is possible (though unlikely) that, merely by chance, random assignment failed to balance the observed and unobserved traits of study participants across the treatment and control conditions. Furthermore, in the Learnfare evaluation, county officials (as opposed to a trained state officer) had the autonomy to conduct the random assignment by themselves (Caspar, Frye, and Merrill 1992). This potential decentralization of the random assignment process suggests the possibility that the fidelity of the procedures could have been inconsistent or even subject to some discretion.

A straightforward way to assess the quality of the random-assignment results is to examine whether the observed baseline traits appear to differ across those assigned to the treatment and control groups. Table 2 presents descriptive statistics on nine baseline traits of the 3,205 study participants, separately for Milwaukee County and the other nine counties and by treatment status. These measures include binary indicators for sex, race, and ethnicity. They also include age measured in years and binary indicators for being "over age" for their grade (e.g., 15+ years old while in grade 8, 16+ years old while in grade 9, etc.), a teen parent, and a school

dropout. Nearly 80 percent of the participating teens from Milwaukee County were Black or Hispanic while under 4 percent were Asian. In the other nine counties, over 13 percent of the participants were Asian and just under 25 percent were Black or Hispanic. However, the remaining baseline traits were relatively similar across Milwaukee County and the remaining counties. For example, 15 to 17 percent of participants were defined as school dropouts when they entered the study. And 17 to 19 percent of participants were identified as teen parents at baseline.

Table 2 also presents the probability values from t tests of treatment-control comparisons for each baseline trait. The results for Milwaukee County indicate that black participants were significantly less likely to be subjected to Learnfare's restrictions (p-value = 0.0021) while Hispanics were significantly more likely (p-value = 0.0103). Furthermore, within Milwaukee County, there were weakly significant differences in the likelihood of being "over age" and a teen parent across the treatment and control conditions. Specifically, both those who were over age and those who were teen parents were more likely to be exempted from Learnfare's restrictions.

The evidence from these "multiple comparisons" may be misleading simply because, even when the null hypotheses of no treatment-control differences are all true, we could expect to make some Type I errors.[11] The procedure developed by Benjamini and Hochberg (1995) provides a powerful way to adjust for the false discovery rate (FDR) associated with such multiple comparisons. Table 2 reports the p-values based on this correction.[12] These results indicate that the imbalance of Black and Hispanic study participants across the treatment and control conditions is still statistically significant at the 5 percent level. However, the imbalances associated with study participants who were over age or teen parents at baseline are no longer statistically significant.

In contrast to the results for Milwaukee County, the baseline traits of the study participants in the nine other counties appear to be consistently well balanced across the

---

[11] See Schochet (2008) for a discussion of the multiple-comparisons problem in the context of educational interventions.

[12] The p-values for the nine hypothesis tests are ordered from smallest to largest (i.e., $p_1 < p_2 < \ldots < p_9$) and each conventional p-value is inflated by a factor equal to the rank of the original p-value divided by the number of comparisons conducted. Interestingly, the adjusted p-value for the treatment-control comparison of Blacks is equivalent under both the Benjamini-Hochberg correction and the less powerful Bonferroni correction (i.e., 0.0021 x 9 = 0.0189). However, the imbalance of Hispanic study participants would only be weakly significant under a Bonferroni correction (i.e., 0.0103 x 9 = 0.0927).

treatment and control conditions. Auxiliary regressions that model treatment status as a function of all of these baseline traits imply similar results. Within Milwaukee County, such regressions suggest that teen-parent status has a particularly robust negative effect on being assigned to Learnfare. However, outside of Milwaukee County, these baseline traits are neither individual nor jointly significant determinants of treatment status.

One candidate explanation for the treatment-control imbalance observed in Milwaukee County is that it simply occurred by chance (i.e., an unintended randomization "failure"). Another possibility is that this pattern reflects discretion on the part of the state or county officers who identified each participant's treatment assignment. More specifically, in order to protect study participants who were thought to be particularly likely to face Learnfare sanctions, officials in Milwaukee County may have been more likely to designate them as being in the control group which was not subject to potential sanctions.

However, both the source of this non-random assignment and the direction of the implied bias in the estimated treatment effects for participants from Milwaukee County are unknown. To examine the effects of the Learnfare restrictions in an unbiased manner, the remaining analysis will focus on the nine other counties where the treatment-control balance suggests that the random assignment procedures worked well. An additional rationale for this focus is the evidence that the Learnfare sanctions were implemented with substantially higher fidelity (i.e., more quickly and accurately) outside of Milwaukee County. Nonetheless, the potential policy lessons from Milwaukee County's experience with Learnfare (e.g., the role of data systems in effective implementation) should not be dismissed and are underscored in the concluding discussion of this study.

6 - Study Attrition

Table 4 presents descriptive statistics for the nine-county, student-by-semester panel data. The number of potential panel observations from the 1,183 study participants outside of Milwaukee County is 6,028. However, study attrition implies that attendance data are missing for over 22 percent of these observations. This attrition, which was not comprehensively addressed in the original Learnfare analysis, constitutes a potential threat to both internal and external validity. A straightforward way to examine the study attrition is to model an attrition indicator,

A_{icms}, as a function of treatment assignment, T_i, and other baseline observables, X_i. A generalized panel-based specification for these auxiliary regressions takes the following form:

$$A_{icms} = \alpha + \gamma T_i + \beta X_i + \eta_c + \theta_m + \delta_s + \varepsilon_{icms} \tag{1}$$

where η_c, θ_m, and δ_s respectively represent county, entry month and semester fixed effects and ε represents a mean-zero error term for teen i in county c who entered the study in the month-year combination a and is observed in semester s.[13] A second version of equation (1) conditions on fully general interactions between the county, entry-month, and semester fixed effects. This specification allows for entry-cohort fixed effects specific to each county (i.e., η_s x θ_m), fixed effects specific to a county in a particular semester (i.e., η_c x δ_s) and fixed effects related to the length of time in the study (i.e., θ_m x δ_s).

The results based on estimates of equation (1) indicate that attrition is significantly more likely among Hispanics, older teens, and teen parents (and less likely among Asians). The attrition of these subgroups compromises the generalizability of the Learnfare evaluation. However, a more central concern is whether random assignment to Learnfare increased the likelihood of attrition. The first two columns of Table 5 report the estimated effects of the treatment assignment on the probability of attrition, both for the full sample and for models based on several subgroups. The results indicate that assignment to Learnfare had a positive but small and statistically insignificant effect on attrition in the full sample.

Furthermore, random assignment to Learnfare's restrictions did not have a statistically significant effect on study attrition among most subgroups. However, one notable exception involves those who were teen parents at baseline. For this subgroup, assignment to Learnfare increased the probability of study attrition by nearly 10 percentage points.

Because of both the large amount of study attrition and the limited evidence that attrition was influenced by treatment status, some of the results presented in this re-analysis rely on imputations for missing outcome data. One basic and uncontroversial imputation is to define enrollment and attendance outcomes for those who have met Learnfare's requirements by completing high school or a GED equivalency. Specifically, in some models, high-school

---

[13] The standard errors in this specification are adjusted for heteroscedasticity clustered at the county/entry-month level. This approach appears to generate the most conservatively large measures of precision relative to several sensible alternatives (e.g., classical and robust standard errors as well as standard errors clustered at either the individual, county, entry month, semester, semester/entry-month, or county/semester levels). Clustering based on county/entry-month cells also implies a fairly large number of clusters (i.e., $9 \times 14 = 126$), so the finite-sample bias in such cluster adjustments (Angrist and Pischke 2009) is unlikely to be a concern.

graduates are identified as fully enrolled and in attendance rather than missing. This simple imputation reduces the attrition rate from 22.1 percent to 16.6 percent (Table 4). The estimated effects of treatment status on this alternative attrition measure (i.e., columns (3) and (4) of Table 5) are similar. In particular, teen parents assigned to Learnfare's restrictions were significantly more likely to leave the study.

This study also utilizes three alternative imputation procedures for the remaining outcome measures that are missing: "last observation carry forward" (LOCF) imputation, worst-case imputation, and multiple imputation. The LOCF procedure, which is the most commonly used imputation procedure in medical trials with repeated outcome measures (Wood, White, and Thompson 2004), simply imputes to missing outcomes the last recorded measure for the given individual.[14] Applying a LOCF imputation to the Learnfare data reduces the attrition rate to 4.2 percent (Table 4). The attrition that remains following the LOCF imputation reflects study participants for whom outcome data were never observed. Auxiliary regressions indicate that treatment status does not have a statistically significant effect on this post-LOCF attrition measure (i.e., columns (5) and (6) in Table 5). In particular, in models that allow for interactions between the county, semester and entry-month fixed effects, treatment status has no statistically significant effect on attrition either for the full sample or for any of the subgroups.

The results from column (6) in Table 5 suggest that attrition is unlikely to confound the impact analyses based on the LOCF imputation. However, the robustness of the results based on this approach is examined by utilizing two other imputation procedures (i.e., worst-case imputation and multiple imputation) that allow for an analysis based on the full set of 6,028 potential panel observations.

Under worst case imputation *all* missing outcome data are assumed to reflect school dropouts (i.e., no enrollment or attendance). One of the drawbacks of both the LOCF and worst-case imputations is that the resulting standard errors may be misleading because the imputed outcome measures, which are constant, understate the true variation in the dependent variables. The time-invariant nature of these imputations may also be misleading with respect to distinguishing short and long-term treatment effects.

---

[14] This approach has also been used in the econometric analyses of data from the Project STAR class-size experiment (Krueger 1999, Dee 2004). For ease of interpretation, the LOCF imputation used here is based on the cardinal value of the enrollment and attendance measures. However, LOCF imputations based on the *percentile rank* of these measures (i.e., preserving the rank position of attriters in each outcome distribution) return similar results.

Multiple imputation (Rubin 1987) addresses both of these concerns. The multiple imputation (MI) technique is a Monte Carlo procedure in which all missing values of the outcome measures are imputed by the predicted values from regressions fitted to the observed data and combined with a randomly generated error term. Multiple versions of complete data sets are generated in this fashion and the estimated coefficients are the means of the estimates based on these data sets.[15] While the impact of study attrition cannot be definitively addressed, the comparative results from the LOCF, worst-case and multiple-imputation procedures should suggest the extent to which study attrition is a confounding source of either bias or imprecision.

6 - Impact Estimates

The basic econometric specification applied to the pooled nine-county data from the Learnfare evaluation takes the following form:

$$Y_{icms} = \alpha + \gamma T_i + \beta X_i + \eta_c + \theta_m + \delta_s + \varepsilon_{icms} \qquad (2)$$

As in the attrition analysis, some results are based on specifications that introduce unrestrictive interactions between the county, entry month and semester fixed effects (i.e., η_s x θ_m, η_c x δ_s and θ_m x δ_s).

6.1 Baseline Results

Table 6 reports the estimated γ from versions of equation (2) applied to each of the three outcome measures and using both the observed data and data based on different imputation procedures. These results consistently indicate that random assignment to the Learnfare program generated statistically significant increases in enrollment and attendance. In the preferred specifications, which condition on interacts between the fixed effects and impute data for those who have completed high school, the implied increase in months enrolled is 0.1325 while the increase in the attendance rate is approximately 0.0339 percentage points.

The treatment-induced increase in enrollment is equivalent to 3.7 percent of the control-group mean and 0.083 of the control-group standard deviation. The increase in the rate of attendance is 4.5 percent of the control group mean (and 0.101 of a standard deviation).

---

[15] Rubin (1987) shows that, for the amount of data missing in this context, there is little efficiency gain to conducting more than 5 to 10 imputations. The results reported here are based on 10 imputations. The standard errors based on this procedure adjust for the within-imputation variance, the between-imputation variance and the number of imputations.

Alternatively, these full-sample treatment estimates imply approximately 3 additional days of enrollment and attendance per semester.[16]

Another compelling way to interpret these treatment estimates, which circumvents the methodological issues surrounding effect-size calculations, is to compare them to policy-relevant achievement gaps. For example, the estimates from equation (2) indicate that being a dropout at baseline implies an enrollment outcome that is 1.02 lower (t-statistic = -6.95) and an attendance rate that is 0.2610 lower (t-statistic = -7.69). The improvements implied by Learnfare's full-sample treatment effects are equivalent to 13 percent of these enrollment and attendance gaps. Alternatively, the enrollment measure is 0.1632 higher for females than for males (t-statistic = 2.85). The treatment effect implied by Learnfare equals 81 percent of this gender gap. And those who are "over age" for their grade have an attendance rate that is 0.0722 lower (t-statistic = -2.81). The increase in school attendance implied by Learnfare is equal to 47 percent of this gap.

Interestingly, the Learnfare effects on the rate of unexcused absences and the attendance rate are quite symmetrical, which suggests that Learnfare did not merely increase the number of absences that were excused. Furthermore, the impact estimates based on alternative imputation procedures are quite similar. However, ignoring the attrition of study participants who had actually met Learnfare's requirements by completing high school does imply a notable downward bias in the estimated impact of Learnfare on school-attendance rates (i.e., a one-third reduction in the estimated γ).

Table 7 identifies, for each of the three outcome measures, how the estimated effects of Learnfare evolved by participants' length of time in the study. More specifically, the indicator for random assignment to the Learnfare treatment is interacted with binary indicators for whether the participant is in their first through sixth semester of study participation. All of these specifications condition on interactions between county, entry-month and semester fixed effects. The results based on the observed data as well as on data sets that include imputations for high school graduates and the LOCF imputation are also reported.

These results based on the observed data generally suggest that the treatment-induced increases in enrollment and attendance are largest in the first two semesters of study participation. By the fourth semester, the Learnfare treatment effects appear to have fallen

---

[16] The assumption of 20 school days in a month implies that 0.1325 additional months is a 2.7 day increase. The assumption of 90 school days in a semester implies that a 0.039 increase in the attendance rate is 3.1 days.

somewhat and to have become statistically indistinguishable from zero. However, the conventional view that Learnfare had at most short-term effects appears to be overdrawn. The fourth-semester effects are generally within a fraction of the standard errors associated with the larger first and second-semester effects. Furthermore, even the casual appearance of decaying treatment effects is substantially diminished after imputing for the absence of high-school completers in a naïve analysis of the observed attendance and enrollment data. More directly, for each outcome measure and imputation method, the hypothesis that the treatment has the same effect by length of time in the study cannot be rejected.

6.2 Alternative Outcome Measures

The results in Tables 6 and 7 indicate that Learnfare generated meaningful and sustained increases in school enrollment and attendance. Figures 1, 2, and 3 provide visual, non-parametric evidence of these treatment effects by showing the kernel density estimates for each outcome measure by treatment status. Figures 1 and 2 indicate that, for those assigned to the treatment, the probability mass for these enrollment and attendance measures is concentrated in higher values. Similarly, Figure 3 indicates that, for those assigned to Learnfare, the rate of unexcused absences tend to be concentrated in the lower values.

However, these kernel densities also illustrate that the three continuous outcome measures used in the original evaluation have skewed and bimodal distributions. These figures suggest that a more natural way to interpret the effects of Learnfare would be to identify how it influences the probabilities that the enrollment and attendance measures exceed particular values. Table 8 reports the key results of such an exercise using a preferred specification and 20 different binary outcome measures defined for each enrollment and attendance measure and multiple cut points.[17] The results indicate that Learnfare increased the probability of full-time enrollment (i.e., months enrolled equal to 4.5) by 4.24 percentage points (i.e., 6.6 percent of the control-group mean). Similarly, Learnfare increased the probability of having any enrollment for the entire semester (i.e., months enrolled > 0) by 3.4 percentage points (i.e., 3.9 percent of the control-group mean).

Though Learnfare generated consistent increases throughout the distribution of the enrollment variable (i.e., the extensive margin), the treatment effects with respect to the

---

[17] Because study induction began in March of 1993, the months-enrolled measure for the spring 1993 semester takes on values of 0, 1, 2, and 3. For purposes of defining these binary outcome measures, these values are redefined as 0, 1.5, 3.0 and 4.5, respectively.

attendance measures (i.e., the intensive margin) were somewhat more heterogeneous. For example, the attendance results indicate that Learnfare did not generate statistically significant increases in the probability of perfect attendance or in the probability of an attendance rate $\geq$ 0.10. However, Learnfare did generate statistically significant increases in the probability of near-perfect attendance (i.e., attendance $\geq$ 0.90) as well as increases in attendance on more modest margins (e.g., attendance $\geq$ 0.50). For example, Learnfare increased the probability of school attendance $\geq$ 0.90 by 5.41 percentage points (or 11.3 percent relative to the control-group mean). Overall, these full-sample results indicate that Learnfare was consistently successful in promoting both school enrollment and high-to-moderate levels of school attendance.

6.3 Subgroup Results

Table 9 presents the estimated effects of Learnfare for each of the three outcome measures and for sub-groups of study participants defined by policy-relevant baseline traits. The estimated treatment effects are roughly similar for males, females, minorities (i.e., black or Hispanic teens), and non-minorities. However, these results also suggest that Learnfare had substantially larger effects for subgroups that are at particular risk of academic failure (e.g., those who were over age for their grade, teen parents, or school dropouts at baseline).

For example, Learnfare increased the months-enrolled measure for baseline dropouts by an amount (i.e., 0.5370). The control-group mean of the enrollment measure among baseline dropouts was 2.13 so this treatment effect constitutes a 25 percent increase. This estimated treatment effect is also nearly five times as large as the treatment effect for those enrolled at baseline (i.e., 0.1164). Similarly, the increased enrollment among those who were over age for their baseline grade (i.e., 0.4829) is nearly four times as large as the effect for those who were not over age (i.e., 0.1430). The treatment effects on attendance were also larger for these subgroups. However, because of the comparatively small size of these subgroups, these estimates generally had less precision. These heterogeneous treatment effects imply that Learnfare policies policy-relevant gaps in school persistence. For example, because Learnfare increased the enrollment of dropouts by 0.5370 and that of non-dropouts by 0.1164, it closed the enrollment gap between these two groups by 41 percent (i.e., (0.5370-0.1164)/1.02).

7 - Conclusions

Wisconsin's influential Learnfare program sanctioned the welfare benefits of families where covered teens did not meet school attendance requirements. The design features of Learnfare are distinct from other recent and ongoing initiatives to provide students with financial incentives for academic performance in several ways. For example, unlike the recent student-incentive programs in developed countries, Learnfare leveraged family-based financial incentives to improve student outcomes (as in the conditional cash transfer programs that have proliferated in developing countries). Second, Learnfare provided sanctions against an existing transfer rather than rewards. In the presence of reference-dependent preferences (e.g., loss aversion), this aspect of Learnfare should amplify its behavioral impact. Third, the extant psychological literature suggests that, to avoid harming intrinsic motivation, financial incentives should be based on requirements that participants feel they have the capacity to meet (i.e., tasks which are "effort responsive"). Learnfare may have been particularly likely to satisfy this condition because it targeted attendance rather than grades or test performance. These psychologically informed design features suggest that Learnfare is a novel example of using "choice architecture" to increase the desired impact of a policy (Thaler and Susstein 2008).

The conventional understanding of Learnfare has been that it was unsuccessful in influencing its targeted outcomes. However, the results presented here indicate that Learnfare was highly effective in improving both school enrollment and attendance. In fact, the benefits of Learnfare in promoting school attendance were concentrated among some of the most at-risk students (i.e., those who were school dropouts at baseline). The effectiveness of Learnfare suggests that its unique design parameters merit further scrutiny and consideration. It should be noted that these design features can be utilized in ways that attenuate the pejorative, normative consequences of sanctioning the welfare grants of economically disadvantaged youths. For example, the creation of a new grant or scholarship that could be subjected to performance-related sanctions could leverage reference-dependent preferences to improve student outcomes without lowering overall income.

However, another notable and important lesson from Wisconsin's Learnfare experience involves the serious implementation challenges that occurred within Milwaukee County. The failure of the random assignment procedures within Milwaukee County to balance the baseline traits of study participants across the treatment and control states strongly qualifies any

conclusions based on the experimental evaluation that occurred there. Nonetheless, the comparative difficulty of producing timely and accurate attendance data within Milwaukee County serve as a compelling reminder that any policy linking financial incentives tied to school attendance is likely to require high-performance data systems that can provide quick and accurate feedback to students and their families. The growing sophistication of data systems in public schools may, therefore, provide an important complement to future policies like Learnfare.

Any future consideration of Learnfare-like policies should also consider how a program of extrinsic rewards compares to other rigorously evaluated policy alternatives. For example, the "What Works Clearinghouse" maintained by the Institute of Education Sciences has identified other effective dropout prevention programs (e.g., ALAS, Check and Connect) that rely on intensive case management rather than financial incentives. The comparative desirability of such programs is an open question whose answer is likely to depend in part on the amount of intrinsic motivation that exists in the targeted population.

However, two other highly policy-relevant criteria for comparing dropout prevention strategies are cost-effectiveness and scalability. With respect to both of these desiderata, Learnfare-like policies may provide an attractive contrast to initiatives that focus on case management and support services. For example, the development of a Learnfare-like policy implies new fixed and operating expenditures. And the Learnfare experience suggests that there is relatively little revenue gain from imposing sanctions, which occurred at a fairly low rate (i.e., typically less than 5 percent). Nonetheless, Learnfare-like initiatives are likely to be to be substantially more cost-effective than comparatively labor-intensive case-management programs. Furthermore, the evidence from the random-assignment evaluation analyzed here provides strong evidence for the efficacy of Learnfare as a mature policy that had been implemented at scale statewide.

References

Angrist, Joshua D., Daniel Lang and Philip Oreopoulos. "Incentives and Services for College Achievement: Evidence from a Randomized Trial," American Economic Journal: Applied Economics, 2009.

Angrist, Joshua D. and Victor Lavy. "The Effects of High-Stakes High School Achievement Awards: Evidence from a Group-Randomized Trial," working paper, June 2008.

Angrist, Joshua D. and Jörn-Steffen Pischke. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, 2009.

Bénabou, Roland and Jean Tirole. "Intrinsic and Extrinsic Motivation" Review of Economic Studies 70, 2003, pages 489-520.

Benjamini, Yoav and Yosef Hochberg. "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing," Journal of the Royal Statistical Society B 57(1), , 1995, pages 289-300.

Bettinger, Eric P. "Paying to Learn: The Effect of Financial Incentives on Elementary Test Scores," working paper, March 12, 2009.

Bos, Johannes M. and Veronica Fellerath. "LEAP: Final Report on Ohio's Welfare Initiative to Improve School Attendance among Teenage Parents:

Ohio's Learning, Earning, and Parenting Program." Manpower Demonstration Research Corporation, New York: January 1997.

Camerer, Colin F. and Robin M. Hogarth. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework," Journal of Risk and Uncertainty 19(1-3), December 1999, 7-42.

Cameron, Judy, and W. David Pierce. Rewards and Intrinsic Motivation: Resolving the Controversy. Westport, CT: Bergin and Garvey, 2002.

Cameron, Judy. "Negative Effects of Reward on Intrinsic Motivation – A Limited Phenomenon: Comment on Deci, Koestner, and Ryan (2001)" Review of Educational Research 71(1), Spring 2001, pages 29-42.

Campbell, David, and Joan Wright. "Rethinking Welfare School- Attendance Policies." Social Service Review, March 2005, pages 2-28.

Dee, Thomas S. "Teachers, Race and Student Achievement in a Randomized Experiment," The Review of Economics and Statistics 86(1), February 2004, pages 195-210.

Deci, Edward L., Richard Koestner, and Richard M. Ryan. "Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again," Review of Educational Research 71(1), Spring 2001, pages 1-27.

Deci, Edward. L. "Effects of externally mediated rewards on intrinsic motivation". Journal of Personality and Social Psychology 18, 1971, pages 105-115.

Education Commission of the States. Student Accountability Initiatives: Learnfare. Updated July 30, 2007, Accessed March 27, 2009, http://mb2.ecs.org/reports/Report.aspx?id=1633.

Etheridge, Marcus E. and Stephen L. Perry. "A New Kind of Public Policy Encounters Disappointing Results: Implementing Learnfare in Wisconsin," Public Administration Review 53(4), July-August 1993, pages 340-347.

Fein, David J., David A. Long, Joy M. Behrens, and Wang S. Lee. The ABC Evaluation: Turning the Corner: Delaware's A Better Chance Welfare Reform Program at Four Years. Abt Associates Inc., Cambridge, MA: January 2001.

Frye, Judith, and Emma Caspar. "An Evaluation of the Learnfare Program: Final Report," State of Wisconsin Legislative Audit Bureau, Madison, WI: 1997.

Frye, Judith, Emma Caspar and Nancy Merrill. "Research Design Evaluation of the Learnfare Program," State of Wisconsin Legislative Audit Bureau, Madison, WI: December 1992.

Handa, Sudhanshu and Benjamin Davis. "The Experience of Conditional Cash Transfers in Latin America and the Caribbean.," Development Policy Review 24(5), September 2006, pages 513-36

Hoynes, Hilary. "Work, Welfare, and Family Structure: What Have We Learned?" in Fiscal Policy: Lessons From Economic Research, edited by Alan Auerbach. MIT Press: Cambridge, Mass, 1997, 101-146.

Jones, Loring P. Ron Harris, and Daniel Finnegan. "School Attendance Demonstration Project: An Evaluation of a Program to Motivate Public Assistance Teens to Attend and Complete School in an Urban School District," Research on Social Work Practice 12(2), 2002, pages 222-37.

Kahneman, Daniel & Amos Tversky. "Prospect Theory: An Analysis of Decision under Risk," Econometrica 47, 1979, 263-291.

Kremer, Michael, Miguel, Edward, Thornton, Rebecca and Ozier, Owen. "Incentives to Learn" World Bank Policy Research Working Paper No. 3546., May 2004.

Krueger, Alan B., "Experimental Estimates of Education Production Functions,' Quarterly Journal of Economics 114(2), 1999, pages 497-532.

Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw. "The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment," Journal of the European Economic Association, forthcoming.

Mauldon, Jane, Jan Malvin, Jon Stiles, Nancy Nicosia, and Eva Y. Seto. "The Impact of California's Cal-Learn Demonstration Project, Final Report." UC Data Archive & Technical Assistance. UC Data Reports: Paper, June 1, 2000.

Maynard, Rebecca. Building Self-Sufficiency Among Welfare-Dependent Teenage Parents: Lessons from the Teenage Parent Demonstration. Princeton, NJ: Mathematica Policy Research, Inc., Princeton, NJ: June 1993.

Mead, Lawrence. Beyond Entitlement: The Social Obligations of Citizenship. New York: Free Press, 1986.

Medina, Jennifer. "Next Question: Can Students Be Paid to Excel?" The New York Times, March 5, 2008.

Pawasarat, John, Lois Quinn, and Frank Stetzer. "Evaluation of the Impact of Wisconsin's Learnfare Experiment on the School Attendance of Teenagers Receiving Aid to Families with Dependent Children," Submitted to the Wisconsin Department of Health and Social Services and the U.S. Department of Health and Human Services, Milwaukee, WI: Employment Training Institute, University of Wisconsin-Milwaukee, February 5, 1992.

Quinn, Lois M. and Robert S. Magill. "Politics versus Research in Social Policy," The Social Service Review 68(4), December 1994, 503-520.

Richburg-Hayes, Lashawn, Thomas Brock, Allen LeBlanc, Christina Paxson, Cecilia Elena Rouse, and Lisa Barrow. "Rewarding Persistence Effects of a Performance-Based Scholarship Program for Low-Income Parents," Manpower Defense Research Corporation, New York: January 2009.

Rubin, Donald B. (1987) *Multiple Imputation for Nonresponse in Surveys*. J. Wiley & Sons, New York

Schochet, Peter V. "Guidelines for Multiple Testing in Impact Evaluations of Educational Interventions," Submitted to the Institute of Education Sciences by Mathematica Policy Research, Inc., Contract No. ED-04-CO-0112/0006, May 2008.

Skoufias, Emmanuel and Bonnie McClafferty. "Is PROGRESA Working? Summary of the Results of an Evaluation by IFPRI" International Food Policy Research Institute, FCND Discussion Paper No. 118, July 2001.

Stoker, Robert P. and Laura A.Wilson. "Verifying Compliance: Social Regulation andWelfare Reform," Public Administration Review 58(5), September/October 1998, pages 395-405.

Thaler, Richard and Cass Susstein. Nudge: Improving Decisions about Health, Wealth, and Happiness, New Haven, Yale University Press, 2008.

Vargas, Theresa. "Cash Incentives Create Competition," The Washington Post, March 22, 2009, C01.

Wiseman, Michael. "State Strategies for Welfare Reform: The Wisconsin Story," Journal of Policy Analysis and Management 15(4), Autumn 1996, pages 515-546.

Wood, Angela M., Ian R. White, and Simon G. Thompson. "Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals," Clinical Trials 1, 2004, pages 368-376.

Table 1 - Study Participants by Entry Month and Semester with Attendance Data

Entry Month	Study Participants	Participants with attendance data					
		Spring 1993	Fall 1993	Spring 1994	Fall 1994	Spring 1995	Fall 1995
March 1993	103	96	84	80	70	61	47
April 1993	203	187	173	158	143	136	117
May 1993	209	197	189	174	164	153	127
June 1993	294	-	269	248	222	208	184
July 1993	297	-	273	256	235	230	200
August 1993	306	-	283	264	236	222	202
September 1993	362	-	350	319	276	260	223
October 1993	341	-	330	312	280	258	232
November 1993	282	-	272	263	245	231	184
December 1993	296	-	288	274	243	230	196
January 1994	235	-	229	227	194	182	149
February 1994	206	-	-	189	176	166	151
March 1994	60	-	-	59	57	55	51
April 1994	11	-	-	10	9	9	7
Total in Study	3,205	480	2,740	2,833	2,550	2,401	2,070

Table 2 - Estimated Treatment Effects by County and Time in Study

Dependent variable	Semesters in Study	Estimated Effect	Standard Error	Sample Size
		Milwaukee County		
Months Enrolled	1	-0.0009	0.0738	1,955
Months Enrolled	2	-0.0600	0.0410	1,859
Months Enrolled	3	-0.0556	0.0573	1,676
Months Enrolled	4	-0.0904*	0.0492	1,582
Rate of Attendance	1	0.0003	0.0123	1,930
Rate of Attendance	2	-0.0124	0.0113	1,827
Rate of Attendance	3	-0.0067	0.0127	1,648
Rate of Attendance	4	-0.0204	0.0126	1,561
Rate of Unexcused Absences	1	0.0018	0.0112	1,930
Rate of Unexcused Absences	2	0.0109	0.0091	1,827
Rate of Unexcused Absences	3	0.0124	0.0137	1,648
Rate of Unexcused Absences	4	0.0197	0.0121	1,561
		Outside Milwaukee County		
Months Enrolled	1	0.1072	0.0919	1,146
Months Enrolled	2	0.1229*	0.0671	1,074
Months Enrolled	3	0.0504	0.0836	949
Months Enrolled	4	0.0037	0.0843	868
Rate of Attendance	1	0.0292**	0.0137	1,102
Rate of Attendance	2	0.0192	0.0134	1,024
Rate of Attendance	3	0.0026	0.0158	925
Rate of Attendance	4	0.0133	0.0165	846
Rate of Unexcused Absences	1	-0.0257*	0.0133	1,102
Rate of Unexcused Absences	2	-0.0118	0.0140	1,024
Rate of Unexcused Absences	3	-0.0028	0.0164	925
Rate of Unexcused Absences	4	-0.0110	0.0162	846

Notes: These models condition on the eight baseline observables and semester FE. The standard errors are adjusted for heteroscedasticity clustered at the county/entry-month level.
***$p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3 - Baseline Traits by Treatment Status and County, Learnfare Evaluation

	Milwaukee County				Other Counties			
Baseline trait	Treatment	Control	p-value	B-H adjusted p-value	Treatment	Control	p-value	B-H adjusted p-value
Female	0.599	0.609	0.6509	0.6509	0.570	0.559	0.6993	0.8991
Black	0.603	0.669	0.0021	0.0189	0.147	0.176	0.1725	0.9999
Hispanic	0.174	0.133	0.0103	0.0464	0.080	0.086	0.6943	0.9999
Asian	0.040	0.033	0.3804	0.6847	0.122	0.141	0.3479	0.9999
Native American	0.013	0.010	0.5140	0.6609	0.023	0.026	0.6926	0.9999
Age	14.344	14.383	0.6491	0.7302	14.591	14.612	0.8580	0.9653
Over age for grade	0.147	0.178	0.0588	0.1764	0.142	0.144	0.9055	0.9055
Parent	0.174	0.204	0.0872	0.1962	0.163	0.176	0.5552	0.9999
Dropout	0.157	0.170	0.4221	0.6332	0.140	0.156	0.4291	0.9999
Sample Size	1,006	1,016			614	569		

Notes: The treatment and control columns identify the mean value of the baseline trait by treatment status and county. The p-value refers to t-test of the hypothesis that the mean value of the baseline trait is the same across treatment and control states. The Benjamini-Hochberg (B-H) adjusted p-values reflect an inflation factor that adjusts for false discoveries in multiple comparisons.

Table 4 - Descriptive Statistics, Learnfare 9-County Panel Data

Variable	Mean	Standard Deviation	Sample Size
Treatment	0.519	0.500	6,028
Female	0.563	0.496	6,028
Black	0.161	0.368	6,028
Hispanic	0.083	0.275	6,028
Asian	0.132	0.338	6,028
Native American	0.025	0.156	6,028
Age	14.62	1.957	6,028
Over age for grade	0.147	0.354	6,028
Teen parent at baseline	0.170	0.376	6,028
Dropout at baseline	0.150	0.357	6,028
Months Enrolled	3.591	1.574	4,862
Months Enrolled, HS-graduate Imputation	3.664	1.530	5,173
Months Enrolled, LOCF Imputation	3.518	1.617	5,908
Months Enrolled, Worst-Case Imputation	3.144	1.909	6,028
Rate of Attendance	0.749	0.332	4,697
Rate of Attendance, HS-Graduates Imputation	0.770	0.325	5,030
Rate of Attendance, LOCF Imputation	0.740	0.347	5,826
Rate of Attendance, Worst-Case Imputation	0.643	0.412	6,028
Rate of Unexcused Absences	0.187	0.342	4,697
Rate of Unexcused Absences, HS-Graduate Imputation	0.171	0.331	5,030
Rate of Unexcused Absences, LOCF Imputation	0.202	0.357	5,826
Rate of Unexcused Absences, Worst-Case Imputation	0.309	0.432	6,028
HS Graduate	0.070	0.256	6,028
Attrition Rate	0.221	0.415	6,028
Attrition Rate	HS-Graduate Imputation	0.166	0.372	6,028
Attrition Rate	LOCF Imputation	0.042	0.202	6,028

Table 5 - Estimated Treatment Effects on Attrition Measures, Full Sample & Subgroups

	Dependent Variable						
	Attrition		Attrition | Imputation for HS Graduates		Attrition | LOCF Imputation		Sample
Sample trait	(1)	(2)	(3)	(4)	(5)	(6)	Size
Full Sample	0.0083	0.0082	-0.0048	-0.0083	0.0056	0.0064	6,028
	(0.0170)	(0.0182)	(0.0160)	(0.0171)	(0.0108)	(0.0115)	
Female	0.0261	0.0241	-0.0011	-0.0092	-0.0067	-0.0048	3,395
	(0.0217)	(0.0243)	(0.0199)	(0.0218)	(0.0158)	(0.0169)	
Male	-0.0045	-0.0091	-0.0044	-0.0071	0.0225*	0.0190	2,633
	(0.0222)	(0.0264)	(0.0220)	(0.0260)	(0.0122)	(0.0148)	
Minority	0.0006	0.0021	-0.0019	-0.0062	-0.0001	-0.0036	1,471
	(0.0273)	(0.0298)	(0.0301)	(0.0328)	(0.0267)	(0.0257)	
Not a Minority	0.0095	0.0109	-0.0047	-0.0085	0.0047	0.0095	4,557
	(0.0191)	(0.0210)	(0.0175)	(0.0190)	(0.0111)	(0.0125)	
Teen Parent	0.0960**	0.0966*	0.0771**	0.0729*	0.0673	0.0433	1,024
	(0.0365)	(0.0484)	(0.0346)	(0.0375)	(0.0455)	(0.0522)	
Not a Teen Parent	-0.0119	-0.0173	-0.0158	-0.0207	-0.0062	-0.0046	5,004
	(0.0176)	(0.0187)	(0.0173)	(0.0184)	(0.0089)	(0.0093)	
Dropout	0.1003**	0.0215	0.1016**	0.0260	0.0508	0.0184	904
	(0.0432)	(0.0570)	(0.0436)	(0.0557)	(0.0315)	(0.0468)	
Not a Dropout	-0.0056	-0.0049	-0.0207	-0.0246	0.0001	0.0033	5,124
	(0.0170)	(0.0184)	(0.0149)	(0.0159)	(0.0104)	(0.0118)	
County FE	yes	no	yes	no	yes	no	
Entry-Month FE	yes	no	yes	no	yes	no	
Semester FE	yes	no	yes	no	yes	no	
County/Entry-Month FE	no	yes	no	yes	no	yes	
Semester/Entry-Month FE	no	yes	no	yes	no	yes	
County/Semester FE	no	yes	no	yes	no	yes	

The standard errors are reported in parentheses and adjusted for heteroscedasticity clustered at the county/entry-month level. All models condition on the nine baseline observables.
***p<0.01, ** p<0.05, * p<0.1.

Table 6 - Estimated Treatment Effects by Imputation Method

Imputation Method	Dependent variable					
	Months Enrolled		Rate of Attendance		Rate of Unexcused Absences	
Observed data	0.0875*	0.1165**	0.0201*	0.0230**	-0.0180*	-0.0219*
	(0.0482)	(0.0530)	(0.0102)	(0.0110)	(0.0106)	(0.0114)
HS-Graduate Imputation	0.1061**	0.1325**	0.0302**	0.0339**	-0.0271**	-0.0315**
	(0.0531)	(0.0589)	(0.0128)	(0.0139)	(0.0128)	(0.0139)
LOCF Imputation	0.1294**	0.1667***	0.0300**	0.0339**	-0.0263**	-0.0315**
	(0.0586)	(0.0626)	(0.0129)	(0.0141)	(0.0129)	(0.0140)
Worst-Case Imputation	0.1386*	0.1674**	0.0364**	0.0405**	-0.0344**	-0.0396**
	(0.0741)	(0.0777)	(0.0169)	(0.0180)	(0.0171)	(0.0181)
Multiple Imputation	0.0951*	0.1191**	0.0302**	0.0336**	-0.0280**	-0.0319**
	(0.0510)	(0.0542)	(0.0123)	(0.0129)	(0.0124)	(0.0131)
County FE	yes	no	yes	no	yes	no
Entry-Month FE	yes	no	yes	no	yes	no
Semester FE	yes	no	yes	no	yes	no
County/Entry-Month FE	no	yes	no	yes	no	yes
Semester/Entry-Month FE	no	yes	no	yes	no	yes
County/Semester FE	no	yes	no	yes	no	yes

The standard errors are reported in parentheses and adjusted for heteroscedasticity clustered at the county/entry-month level. All models condition on the nine baseline observables.
***p<0.01, ** p<0.05, * p<0.1.

Table 7 - Estimated Treatment Effects by Time in Study and Imputation Method

| | Dependent Variable | | | | | | | | |
| | Months Enrolled | | | Rate of Attendance | | | Rate of Unexcused Absences | | |
Independent Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
							-	-	-
Treatment x 1st Semester in Study	0.1067*	0.1096**	0.1157**	0.0366**	0.0440***	0.0436***	0.0348**	0.0410***	0.0407***
	(0.0540)	(0.0511)	(0.0525)	(0.0148)	(0.0149)	(0.0147)	(0.0142)	(0.0141)	(0.0140)
Treatment x 2nd Semester in Study	0.1661**	0.1468**	0.1886**	0.0253*	0.0333**	0.0368**	-0.0193	-0.0246*	-0.0296*
	(0.0699)	(0.0709)	(0.0720)	(0.0131)	(0.0146)	(0.0154)	(0.0139)	(0.0146)	(0.0158)
Treatment x 3rd Semester in Study	0.1135	0.1353	0.1618*	0.0141	0.0248	0.0273	-0.0144	-0.0237	-0.0249
	(0.0886)	(0.0890)	(0.0828)	(0.0156)	(0.0178)	(0.0170)	(0.0164)	(0.0181)	(0.0174)
Treatment x 4th Semester in Study	0.0885	0.1380	0.1840**	0.0227	0.0362*	0.0369**	-0.0217	-0.0347	-0.0347**
	(0.0905)	(0.1015)	(0.0899)	(0.0166)	(0.0218)	(0.0178)	(0.0163)	(0.0212)	(0.0172)
Treatment x 5th Semester in Study	0.0762	0.1049	0.1522*	0.0119	0.0245	0.0246	-0.0137	-0.0262	-0.0246
	(0.0899)	(0.0898)	(0.0895)	(0.0204)	(0.0227)	(0.0203)	(0.0203)	(0.0223)	(0.0200)
Treatment x 6th Semester in Study	0.2295	0.3162	0.3392	0.0121	0.0640	0.0362	-0.0314	-0.0777	-0.0474
	(0.3357)	(0.3104)	(0.2171)	(0.0603)	(0.0635)	(0.0449)	(0.0674)	(0.0677)	(0.0434)
Missing outcome imputation	None	HS-Grads	LOCF	None	HS-Grads	LOCF	None	HS-Grads	LOCF
Sample size	4,862	5,173	5,908	4,697	5,030	5,826	4,697	5,030	5,826
R-squared	0.518	0.445	0.460	0.469	0.308	0.385	0.491	0.345	0.423
p-value	0.8747	0.9581	0.7327	0.8636	0.8354	0.8503	0.8526	0.7463	0.8426

The standard errors are reported in parentheses and adjusted for heteroscedasticity clustered at the county/entry-month level. All models condition on baseline observables, county/entry-month FE, county/semester FE, and semester/entry-month FE.

***$p<0.01$, ** $p<0.05$, * $p<0.1$.

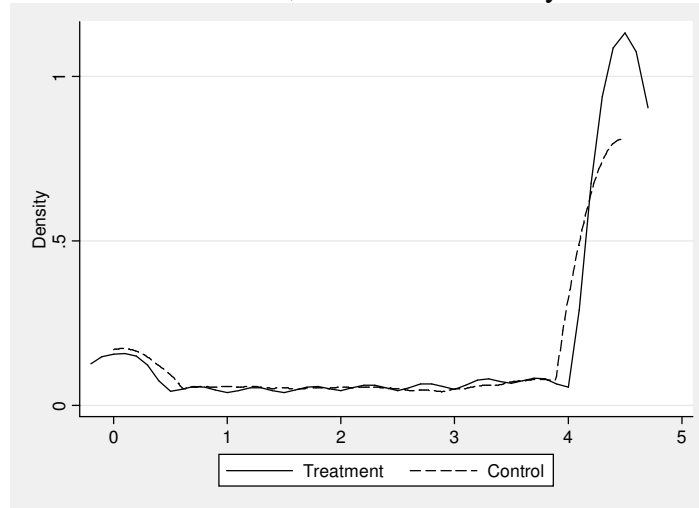Figure 1 – Kernel Densities, Months Enrolled by Treatment Status

Figure 2 – Kernel Densities, Rate of Attendance by Treatment Status
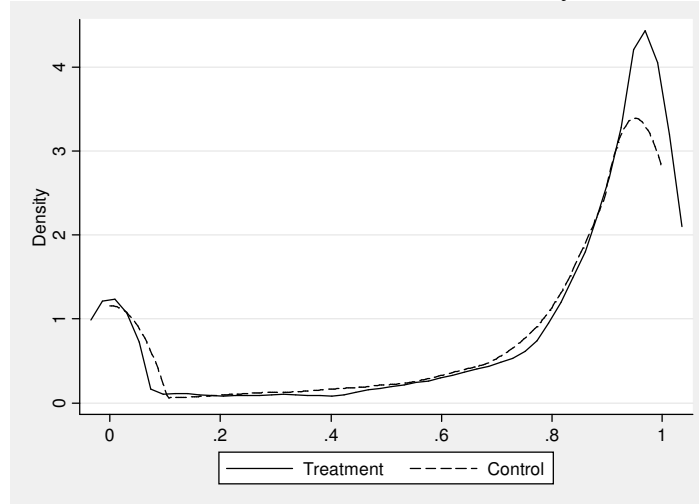

Figure 3 – Kernel Densities, Rate of Unexcused Absences by Treatment Status
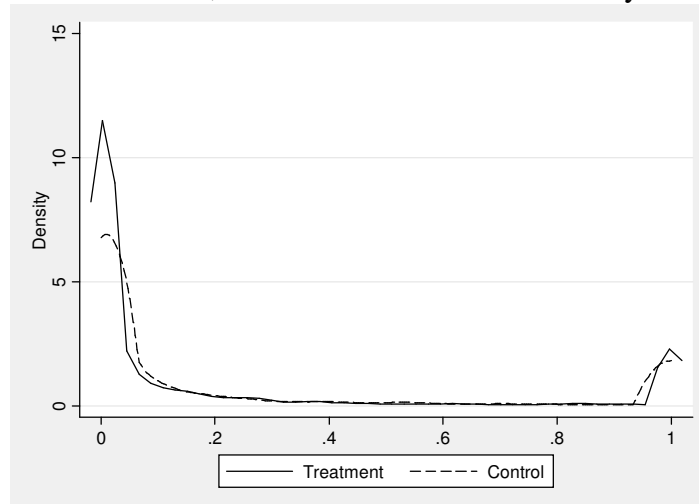
Table 8 - Estimated Treatment Effects, Alternative Outcome Measures

Binary Outcome Variable	Treatment Estimate	Standard Error	Control-Group Mean
Months enrolled = 4.5	0.0424**	0.0185	0.647
Months enrolled ≥ 4.0	0.0373**	0.0164	0.684
Months enrolled ≥ 3.0	0.0379**	0.0163	0.746
Months enrolled ≥ 2.0	0.0396***	0.0130	0.791
Months enrolled ≥ 1.0	0.0326**	0.0140	0.847
Months enrolled > 0	0.0340**	0.0139	0.870
Attendance rate = 1.0	0.0267	0.0177	0.160
Attendance rate ≥ 0.9	0.0541**	0.0236	0.477
Attendance rate ≥ 0.75	0.0400**	0.0173	0.693
Attendance rate ≥ 0.50	0.0363**	0.0157	0.784
Attendance rate ≥ 0.25	0.0236*	0.0141	0.823
Attendance rate ≥ 0.10	0.0216	0.0147	0.836
Attendance rate = 0	-0.0246*	0.0148	0.160
Rate of Unexcused Absences = 0	0.0428*	0.0222	0.455
Rate of Unexcused Absences ≤ 0.10	0.0363*	0.0186	0.685
Rate of Unexcused Absences ≤ 0.25	0.0363**	0.0154	0.761
Rate of Unexcused Absences ≤ 0.50	0.0331**	0.0149	0.803
Rate of Unexcused Absences ≤ 0.75	0.0212	0.0141	0.829
Rate of Unexcused Absences ≤ 0.90	0.0244	0.0148	0.837
Rate of Unexcused Absences = 1	-0.0259*	0.0147	0.159

The standard errors are reported in parentheses and adjusted for heteroscedasticity clustered at the county/entry-month level. All models condition on baseline observables, county/entry-month FE, county/semester FE, and semester/entry-month FE. The outcome measures reflect LOCF imputations for missing values. ***$p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 9 - Estimated Treatment Effects by Subgroup

Subgroup	Months Enrolled	Rate of Attendance	Rate of Unexcused Absences
	Dependent variable		
Female	0.1408	0.0364	-0.0319
	(0.1032)	(0.0258)	(0.0248)
Male	0.1558**	0.0267	-0.0266
	(0.0772)	(0.0196)	(0.0184)
Minority	0.2089	0.0300	-0.0297
	(0.1622)	(0.0269)	(0.0276)
Non-minority	0.1582**	0.0356**	-0.0311*
	(0.0628)	(0.0166)	(0.0158)
Over age for grade	0.4829**	0.0593	-0.0736
	(0.2174)	(0.0490)	(0.0505)
Not over age for grade	0.1430**	0.0324**	-0.0297**
	(0.0596)	(0.0149)	(0.0143)
Teen parent	0.5045	0.0668	-0.0772
	(0.3553)	(0.0864)	(0.0887)
Not a teen parent	0.0835*	0.0203	-0.0173
	(0.0503)	(0.0130)	(0.0123)
Dropout	0.5370*	0.0626	-0.0815
	(0.2725)	(0.0609)	(0.0627)
Not a dropout	0.1164**	0.0265**	-0.0209*
	(0.0558)	(0.0119)	(0.0114)

The standard errors are reported in parentheses and adjusted for heteroscedasticity clustered at the county/entry-month level. All models condition on baseline observables, county/entry-month FE, county/semester FE, and semester/entry-month FE. The outcome measures reflect LOCF imputations for missing values.
***p<0.01, ** p<0.05, * p<0.1.