Economics of Education Review 30 (2011) 924–937

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Conditional cash penalties in education: Evidence from the Learnfare experiment

Thomas S. Dee [a,b]

[a] University of Virginia, Charlottesville VA 22904, United States
[b] NBER and China Center for Human Capital and Labor Market Research, Central University of Finance and Economics, Beijing, China

ARTICLE INFO

Article history:
Received 11 April 2011
Accepted 23 May 2011

JEL classification:
I2
I3

Keywords:
Education
Welfare
Incentives
Attainment
Experiment

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 3.5 percent (effect size = 0.08) and attendance by 4.5 percent (effect size = 0.10). These results suggest that well-designed financial incentives may be an effective mechanism for improving the school persistence of at-risk students at scale.

"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. A prominent example is the recent, broad interest in using student and family-based financial incentives to improvement academic outcomes. 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 & Lavy, 2008; Angrist, Lang, & Oreopoulos, 2009; Bettinger, 2009; Leuven, Oosterbeek, & van der Klaauw, 2010; 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 & Davis, 2006).

This study presents new evidence on the effectiveness of such initiatives through a re-examination of an incentive program with several unique design features, Wisconsin's seminal Learnfare program. Learnfare, a welfare–waiver reform that sanctioned a family's welfare grant when covered teens (i.e., 13–19 year olds) failed to meet school attendance and completion targets, provides a distinctive and policy-relevant 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

E-mail address: dee@virginia.edu

0272-7757/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2011.05.013

[1] Thirty-eight states now have "Learnfare" policies that link school attendance and welfare receipt (Education Commission of the States, 2007).

program requirements. Because of the evidence that people exhibit an asymmetric aversion to income losses relative to a reference point (Kahneman & 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 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 potentially 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 both within and outside Milwaukee County (Frye & 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 provide a unified framework for assessing both the effects of random-assignment to the Learnfare treatment and potential threats to the internal validity of those inferences. This reanalysis also examines the treatment balance of study attrition, correctly identifies high-school completers (who no longer have attendance data) as non-attriters, and implements several procedures for imputing outcome measures among genuine attriters (e.g., last-observation carry forward, worst-case imputation, and multiple imputation).

The results of this analysis indicate that, in Milwaukee County, the county-based random-assignment procedures did not produce balanced baseline traits. For example, teen mothers in Milwaukee County were significantly less likely to have been assigned 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 in the application of sanctions. For these reasons, this analysis focuses separately 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.[2]

The results based on these counties indicate that random assignment to the Learnfare restrictions did generate statistically significant improvements in both school enrollment (3.5 percent increase, effect size = 0.08) and school attendance (4.5 percent increase, effect size = 0.10). The hypothesis of a common treatment effect sustained throughout the study period cannot be rejected. However, the study data often lack the statistical power necessary to precisely estimate longer-term treatment effects, even in models that utilize alternative methods of imputing missing outcome data for attriters. Furthermore, because the study data are only available for four to six semesters and most study participants entered as 13 year olds, the effects of Learnfare on the probability of completing high-school or an equivalency degree – though positive – cannot be precisely estimated (p-value = 0.147). This study concludes with a discussion of the economic relevance of Learnfare's impact estimates as well as the important policy-design and implementation lessons that Wisconsin's experience with Learnfare has for the ongoing developments in using financial incentives to improve student performance.

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 empirical literature in psychology (Deci, Koestner, & 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.[3] 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 & 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

[2] The ex-ante analytical plan for the Learnfare experiment called for an analysis that similarly separated Milwaukee County from the remaining study counties (Frye et al., 1992). It should be noted that estimates based on this study's preferred panel-data specifications replicate the finding that Learnfare was ineffective in Milwaukee County (Frye & Caspar, 1997). However, the evidence of a randomization failure raises concerns about the internal validity of these impact estimates. Regardless, the comparatively low-fidelity implementation in Milwaukee County still provides a timely cautionary tale with important implications for the design and execution of incentive-based policies, specifically with respect to the role of accurate and timely data systems.

[3] Writing from an economics perspective, Bénabou and Tirole (2003) model this hypothesis in a principal-agent framework 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").

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 & McClafferty, 2001). In a random-assignment study conducted in Kenya, Kremer, Miguel, Thornton, and Ozier (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 et al. (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 et al. (2010) 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.

Three 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. Another recent study by Fryer (2010) presents the results from randomized trials fielded in four cities in which incentives were linked either to outputs (i.e., student achievement) or inputs (e.g., attendance, homework, reading, etc.) did improve student achievement. The results of this large-scale study indicate that only the incentives linked to student inputs were successful in improving student achievement.

In addition to these recent studies, six other random-assignment studies also evaluated programs that, like Learnfare, linked the threat of financial *sanctions* to a specific educational input: 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 (Fein, Long, Behrens, & Lee, 2001; Stoker & Wilson, 1998). 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 (Bos & Fellerath, 1997; Jones, Harris, & Finnegan, 2002; Mauldon, Malvin, Stiles, Nicosia, & Seto, 2000; Maynard, 1993).

Campbell and Wright (2005) argue that the comparative results from these welfare/school-attendance evaluations indicate that financial sanctions are less likely to be effective when used in isolation from related services and case management. However, the Learnfare re-analysis presented here suggests that this interpretation should be reconsidered. Furthermore, the other sanctions-only interventions that appeared to have no school-attendance effects (i.e., Maryland's PPI and Delaware's ABC) were distinctive from Learnfare in ways that may have attenuated their impact. Specifically, both of these programs bundled sanctions for school-attendance violations with sanctions for multiple other behaviors (e.g., preventative health care, child immunizations, job-search activities, and parenting classes). The multi-faceted nature of these financial incentives may have created confusion for participating subjects or the perception that the penalties were not effort-responsive. Also, it should be noted that the sanction amounts in both programs (i.e., \$25 per month in Maryland and \$68 per month in Delaware) were lower than the typical Learnfare sanction.

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 some services designed to assist with school-attendance problems. 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 (e.g. caring for an infant under 45 days old or the lack of reasonably available child care).

The available case records indicate that, 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 during their first four semesters. Over the same window, 9 percent of the teens assigned to Learnfare experienced at least 1 benefit sanction, a rate similar both within and outside Milwaukee County. In the typical semester, the sanction rate among Learnfare teens was less than 5 percent (Frye & Caspar, 1997). The amount of the sanction depended on the contribution of the non-complying teen to the family's AFDC grant. For example, the sanction for a single-parent with two children would be approximately $77 per month, 15 percent of the corresponding basic monthly grant of $517. In contrast, for a teen parent living alone, the sanction would be $194, 44 percent of the basic monthly grant of $440 (Frye & Caspar, 1997).

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 & 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 & 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 vs. 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 & 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 & 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, & 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 & 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 & Caspar, 1997), which did utilize random assignment, indicated that the Learnfare program had at most short-term school-participation effects for certain sub-groups (e.g., 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, & 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, p. 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, 3205 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. Study participants had to meet the basic requirements for the Learnfare program: aged 13–19, 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 13–19 during the previous 12 months were excluded from the study (Frye et al., 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 et al., 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.[4]

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. The unexcused-absence rate provides a potentially useful complement to the key attendance-rate measure. For example, if Learnfare had smaller effects on the attendance rate than on the unexcused-absence rate, it would indicate that the treatment led to parents and guardians *excusing* more absences. As a practical matter, this distinction does not appear empirically relevant as Learnfare had largely symmetrical effects on the two measures.

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 3205 observations for each of the last four study semesters. However, the number of observations with attendance data drops from 2833 in the spring of 1994–2070 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.

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 met Learnfare's requirements by completing high school or a GED equivalency (i.e., roughly 7 percent of the study-by-semester observations outside Milwaukee County). This distinction is not relevant for most study participants because they were only 13 years old when they entered the study and did not have sufficient time for the typical period of high school completion during the study window. 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 = 0.147), effect on high-school/GED completion. This pattern of positive treatment effects implies that the primary evaluation's approach of eliminating high-school/GED completers from the enrollment and attendance analysis biases the estimated treatment effect downward. The non-random attrition of high-school/GED completers from the original analysis may particularly confound identifying the longer-

[4] 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.

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	3205	480	2740	2833	2550	2401	2070

term effects of Learnfare (e.g., four semesters after random assignment).

This study presents regression-based evidence on the determinants of study attrition, focusing particularly on the effects associated with treatment 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 (e.g., multiple imputation, worst-case imputation, and last observation carry forward). However, the preferred results simply rely on a straightforward imputation for the missing enrollment and attendance data of high-school/GED completers. Specifically, in most models, high-school/GED completers are identified as fully enrolled and in attendance rather than as missing the enrollment and attendance results. The study results are not, as is shown, sensitive to this imputation. Furthermore, models that accommodate high-school/GED completers in a truncated-regression framework generate results similar to those reported here.

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 closely replicate to those reported by Frye and Caspar (1997, Table 14). 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.[5] 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 conventional view that Learnfare did not have meaningful effects on its targeted outcomes. However, this interpretation may be inaccurate for a number of reasons. First, an analysis based on the cross-sections in Table 2 fails to exploit the potential precision gains made possible 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 oth-

[5] The fourth-semester enrollment result for Milwaukee County suggests that Learnfare had weakly significant but harmful effects. However, the poor treatment-control balance for the study participants from Milwaukee County suggests that these inferences lack internal validity.

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	1955
Months enrolled	2	−0.0600	0.0410	1859
Months enrolled	3	−0.0556	0.0573	1676
Months enrolled	4	−0.0904*	0.0492	1582
Rate of attendance	1	0.0003	0.0123	1930
Rate of attendance	2	−0.0124	0.0113	1827
Rate of attendance	3	−0.0067	0.0127	1648
Rate of attendance	4	−0.0204	0.0126	1561
Rate of unexcused absences	1	0.0018	0.0112	1930
Rate of unexcused absences	2	0.0109	0.0091	1827
Rate of unexcused absences	3	0.0124	0.0137	1648
Rate of unexcused absences	4	0.0197	0.0121	1561
		Outside Milwaukee County		
Months enrolled	1	0.1072	0.0919	1146
Months enrolled	2	0.1229*	0.0671	1074
Months enrolled	3	0.0504	0.0836	949
Months enrolled	4	0.0037	0.0843	868
Rate of attendance	1	0.0292**	0.0137	1102
Rate of attendance	2	0.0192	0.0134	1024
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	1102
Rate of unexcused absences	2	−0.0118	0.0140	1024
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.
* $p < 0.05$.
** $p < 0.10$.

erwise 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, & 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 are related to the likelihood of being assigned to the Learnfare. Table 3 presents the key results from auxiliary regressions where treatment status is a function of the available baseline traits. Interestingly, the results for Milwaukee County indicate that teens who were parents at baseline were less likely to be subjected to Learnfare's restrictions (p-value = 0.083). Similarly, teens who were overage for their grade (and, therefore, at a high risk of not meeting attendance requirements) were also less likely to have been assigned to Learnfare (p-value = 0.052). The evidence from these weakly significant regression coefficients may be misleading simply because, even when the null hypotheses of no effects on treatment status are all true, we could expect to make some Type I errors. However, a single F-test based on the data from Milwaukee County also indicates that, jointly, the baseline traits have a weakly significant effect on treatment status (p-value = 0.0684).

In contrast, the results in Table 3 also indicate that, outside of the Milwaukee County, none of the baseline traits has a statistically significant relationship with treatment status. Furthermore, the hypothesis that these baseline traits are jointly insignificant determinants of treatment status cannot be rejected as well (p-value = 0.9459). It should be noted that this pattern of treatment balance outside of Milwaukee County generally holds when these auxiliary regressions are estimated separately for each of the nine counties. The modest exceptions involve two of the smallest counties, Douglas (n = 85) and Eau Claire (n = 91) where there were imbalances among low-frequency racial and ethnic sub-groups.[6] However, the study's main results are quite robust to simply excluding observations from these counties. It should also be noted that, even outside Milwaukee County, Table 3 does suggest a strong, though statistically insignificant, partial correlation between being a teen parent at baseline and not being assigned to Learnfare. However, this imprecision appears to be due to the higher multicollinearity between teen-parent status and other baseline traits. For example, simple t-tests (which do not condition on other possibly collinear observables) indi-

[6] In Douglas County, there were only two Hispanic subjects, both of whom were assigned to the control state. In Eau Claire, there were 14 Asian participants, only two of whom were assigned to the treatment condition.

Table 3
Auxiliary regressions, effects of baseline traits on treatment status by counties.

Baseline trait	Milwaukee County	Other counties
Female	0.0009	0.0117
	(0.025)	(0.031)
Black	−0.0451	−0.0694
	(0.031)	(0.047)
Hispanic	0.0397	−0.0403
	(0.040)	(0.059)
Asian	0.0359	−0.0194
	(0.065)	(0.050)
Native American	0.0450	−0.0916
	(0.108)	(0.097)
Baseline age = 14	0.0062	0.0159
	(0.042)	(0.047)
Baseline age = 15	0.0403	0.0122
	(0.042)	(0.052)
Baseline age = 16	0.0494	−0.0073
	(0.047)	(0.057)
Baseline age = 17	−0.0137	0.0162
	(0.053)	(0.061)
Baseline age = 18	0.0903	0.1177
	(0.069)	(0.097)
Baseline age = 19	0.1080	0.1812
	(0.082)	(0.120)
Overage for grade	−0.0599*	−0.0076
	(0.031)	(0.044)
Teen parent	−0.0921*	−0.1261
	(0.053)	(0.086)
Dropout	−0.0336	−0.0273
	(0.039)	(0.050)
Sample size	2022	1183
R^2	0.011	0.014
p-Value	0.0684	0.9459

Notes: The standard errors are reported in parentheses. The "other counties" model also conditions on county fixed effects. The p-value refers to a test of the joint significance of the baseline traits.

* $p<0.1$.

cate that teen parents outside of Milwaukee County were only 2.3 percentage points more likely than other participants to be assigned to the treatment (p-value = 0.5552). In contrast, teen parents within Milwaukee County were 4.9 percentage points more likely to receive the treatment (p-value = 0.0872).

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. For example, one plausible scenario is that, in order to protect young, economically vulnerable welfare recipients who were particularly likely to face Learnfare sanctions (i.e., teen mothers, students who were old for their grades), officials in Milwaukee County disproportionately allocated them to the control group, which was not subject to potential sanctions. However, the true source of this seemingly non-random assignment and the possible role played by discretionary treatment assignment are unclear.[7] The implications of this randomization failure for Learnfare's impact estimates in Milwaukee County are also somewhat unclear. One reasonable conjecture is they will be biased upwards because the teens who were at particular risk for low school attendance and dropping out were more likely to be assigned to the control condition. However, this claim cannot be made with complete certainty because non-assignment to the control condition could have also been based on unobserved traits (e.g., resiliency and likeability) that predict higher educational attainment.

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. Similarly separate analyses of Milwaukee County and the combined group of other study counties were part of the original, ex-ante analytical plan for the Learnfare experiment (Frye et al., 1992). A second, important 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 implications of this choice for external validity and the corresponding policy lessons from Milwaukee County's experience with Learnfare (e.g., the role of timely and accurate data systems in effective implementation) should not be dismissed lightly and are underscored in the concluding discussion of this study.[8]

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 1183 study participants outside of Milwaukee County is 6028. 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, $\boldsymbol{X}_i$. A generalized panel-based specification for these auxiliary regressions takes the following form:

$$A_{icms} = \alpha + \gamma T_i + \beta \boldsymbol{X}_i + \eta_c + \delta_s + \theta_m + \varepsilon_{icms} \quad (1)$$

where η_c, δ_s, and θ_m respectively represent county, semester, and entry-month fixed effects and ε_{icms} represents a mean-zero error term for teen i in county c who entered the study in the month-year combination m and is observed in semester s.[9] A second version of Eq.

[7] Efforts to identify and query officials who might have firsthand knowledge of the random-assignment procedures used in the Milwaukee County office have been unsuccessful.

[8] Furthermore, it should be noted that the panel-data specifications used in this study replicate the finding that Learnfare had small and statistically insignificant effects in Milwaukee County. However, the internal validity of these estimates is uncertain.

[9] 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 × 14 = 126), so the finite-sample bias in

Table 4
Descriptive statistics, Learnfare 9-county panel data.

Variable	Mean	Standard deviation	Sample size
Treatment	0.519	0.500	6028
Female	0.563	0.496	6028
Black	0.161	0.368	6028
Hispanic	0.083	0.275	6028
Asian	0.132	0.338	6028
Native American	0.025	0.156	6028
Baseline age = 13	0.478	0.500	6028
Baseline age = 14	0.123	0.329	6028
Baseline age = 15	0.098	0.298	6028
Baseline age = 16	0.081	0.273	6028
Baseline age = 17	0.076	0.266	6028
Baseline age = 18	0.106	0.308	6028
Baseline age = 19	0.037	0.188	6028
Over age for grade	0.147	0.354	6028
Teen parent at baseline	0.170	0.376	6028
Dropout at baseline	0.150	0.357	6028
Months enrolled	3.591	1.574	4862
Months enrolled, imputation for HS graduates	3.664	1.530	5173
Months enrolled, LOCF imputation	3.518	1.617	5908
Months enrolled, worst-case imputation	3.144	1.909	6028
Rate of attendance	0.749	0.332	4697
Rate of attendance, imputation for HS graduates	0.770	0.325	5030
Rate of attendance, LOCF imputation	0.740	0.347	5826
Rate of attendance, worst-case imputation	0.643	0.412	6028
Rate of unexcused absences	0.187	0.342	4697
High-school/GED completer	0.070	0.256	6028
Attrition Rate	0.221	0.415	6028
Attrition rate|imputation for HS graduates	0.166	0.372	6028
Attrition rate|LOCF imputation	0.042	0.202	6028

(1) conditions on interactions between the county, entry-month, and semester fixed effects. This specification allows for entry-cohort fixed effects specific to each county (i.e., $\eta_c \times \theta_m$), fixed effects specific to a county in a particular semester (i.e., $\eta_c \times \delta_s$), and fixed effects related to the length of time in the study (i.e., $\theta_m \times \delta_s$). Some specifications also condition on a fully general set of interactions among all three fixed effects (i.e., $\eta_c \times \delta_s \times \theta_m$).

The results based on estimates of Eq. (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 and threatens the internal validity of the impact estimates. Table 5 provides direct evidence on this question by reporting the estimated effects of treatment status on different measures of attrition as well as across alternative specifications.

The results in Table 5 consistently indicate that treatment status did not have a statistically significant effect on the probability of attrition. Overall, these results imply that attrition is not likely to threaten the internal validity of the Learnfare's impact estimates. Nonetheless, it is interesting to note that treatment status had a positive (but statistically insignificant) effect when attrition is defined as having no recorded attendance data (i.e., the first row of Table 5). This pattern is partly due to the fact that those assigned to the Learnfare treatment were more likely to complete high or a GED equivalent and, thus, no longer have attendance data. When these high-school/GED completers are not defined as attriters, the attrition rate falls from 22 percent to 17 percent and the estimated effect of treatment status switches to negative (but remains statistically insignificant).

Because those who have completed high school belong in the analytical sample, most of the results presented here rely on a basic and uncontroversial imputation that leverages the fact that the missing outcome data have bounded supports. Specifically, in some models, high-school graduates are identified as fully enrolled and in attendance rather than missing. As noted earlier, an alternative approach based on treating high-school completion in a truncated regression framework leads to results quite similar to those based on this approach.

Attrition from the Learnfare study was still fairly high (i.e., 17 percent) even after setting aside those who actually completed high school. Fortunately, the results in the second row of Table 5 indicate that this attrition is unrelated to treatment status. Nonetheless, as a robustness check, some of the results presented here rely on three alternative imputations for the missing outcome data: "last observation carry forward" (LOCF) imputation, worst-case imputation, and multiple imputation. The LOCF procedure, the most commonly used imputation procedure in medical trials with repeated outcome measures (Wood, White, & Thompson, 2004), simply imputes to missing outcomes the

such cluster adjustments (Angrist & Pischke, 2009) is unlikely to be a concern.

Table 5
Estimated treatment effects on attrition measures.

Attrition measure	(1)	(2)	(3)	Dependent mean
Attrition|recorded data	0.0054 (0.0169)	0.0045 (0.0181)	0.0045 (0.0178)	0.2208
Attrition|HS/GED imputation	−0.0081 (0.0151)	−0.0101 (0.0160)	−0.0101 (0.0158)	0.1656
Attrition|LOCF imputation	0.0044 (0.0107)	0.0054 (0.0114)	0.0054 (0.0112)	0.0425
County FE	Yes	No	No	
Entry-month FE	Yes	No	No	
Semester FE	Yes	No	No	
County/entry-month FE	No	Yes	No	
Semester/entry-month FE	No	Yes	No	
County/semester FE	No	Yes	No	
County/semester/entry-month FE	No	No	Yes	

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

last recorded measure for the given individual.[10] 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., row 3 of Table 5).

The results in Table 5 suggest that attrition is unlikely to confound the impact estimates based on the LOCF imputation. However, the results based on this approach are complemented by two other imputation procedures (i.e., worst-case imputation and multiple imputation) that allow for an analysis based on the *full* set of 6028 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 be particularly misleading with respect to distinguishing short and long-term treatment effects. Multiple imputation (Rubin, 1987) addresses 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 point estimates based on these data sets.

[10] This approach has also been used in the econometric analyses of data from the Project STAR class-size experiment (Dee, 2004; Krueger, 1999). 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.

7. Impact estimates

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

$$Y_{icms} = \alpha + \gamma T_i + \beta \mathbf{X}_i + \eta_c + \delta_s + \theta_m + \varepsilon_{icms} \quad (2)$$

As in the attrition analysis, some results are also based on specifications that introduce unrestrictive interactions between the county, semester, and entry-month fixed effects (i.e., $\eta_c \times \theta_m$, $\eta_c \times \delta_s$ and $\theta_m \times \delta_s$ as well as $\eta_c \times \delta_s \times \theta_m$). These fixed effects provide controls for the unobserved determinants of Y_{icms} that might be unique to entry cohorts from a particular county (i.e., $\eta_c \times \theta_m$), to counties observed at particular points in time (i.e., $\eta_c \times \delta_s$), and to subjects who have participated in the study for a particular amount of time (i.e., $\theta_m \times \delta_s$). The impact estimates based on Eq. (2) should be understood as reduced-form estimates that identify the effects of the original treatment assignment throughout the study period, even after some study participants are no longer actually subject to Learnfare's requirements.[11] Dynamic treatment effects are represented (and their equivalence tested) by introducing interactions between T_i and binary indicators for whether the subject is in their first through sixth semester of study participation.

7.1. Baseline results

Table 6 reports the estimated γ from alternative versions of Eq. (2) applied to each of the three outcome measures. These results consistently indicate that random assignment to the Learnfare program generated statistically significant increases in enrollment and attendance. For example, in the third specification, which allows for

[11] Some study participants became ineligible for Learnfare's restrictions during the study window. For example, Frye and Caspar (1997, p. 16) note that, by their fourth study semester, 45 percent of treatment subjects were not actually Learnfare-eligible. In most cases, this occurred simply because the relevant AFDC case closed.

Table 6
Estimated treatment effects.

Dependent variable	(1)	(2)	(3)
Months enrolled	0.1023**	0.1276**	0.1274**
	(0.0514)	(0.0565)	(0.0588)
Rate of attendance	0.0294**	0.0325**	0.0335**
	(0.0128)	(0.0139)	(0.0144)
Rate of unexcused absences	−0.0265**	−0.0306**	−0.0313**
	(0.0127)	(0.0137)	(0.0143)
County FE	Yes	No	No
Entry-month FE	Yes	No	No
Semester FE	Yes	No	No
County/entry-month FE	No	Yes	No
Semester/entry-month FE	No	Yes	No
County/semester FE	No	Yes	No
County/semester/entry-month FE	No	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. The dependent variables reflect imputations for high-school/GED completers.
** $p<0.05$.

fully unrestrictive interactions between the fixed effects, the implied increase in months enrolled is 0.1274 while the increase in the attendance rate is approximately 0.0335 percentage points. Interestingly, Learnfare's estimated effects on the rate of unexcused absences and the attendance rate are quite symmetrical, suggesting that the treatment-induced growth in attendance came largely from reductions in unexcused absences.

These baseline estimates are based on the preferred outcome measures that include simple imputations for high-school/GED completers (but no other imputations). However, Table 7 reports the estimated effect of Learnfare on the enrollment and attendance measures across models that rely both on the unadjusted outcome measures as well as outcome measures reflecting several alternative imputation procedures. All of these approaches consistently indicate that Learnfare generated broadly similar and statistically significant increases in enrollment and attendance.[12] However, it is interesting to note that ignoring the attrition of study participants who had actually met Learnfare's requirements by completing high school or an equivalency does imply a notable downward bias in the estimated impact of Learnfare on school-attendance rates (i.e., roughly a one-third reduction in the estimated γ).

Table 7
Estimated treatment effects by imputation method.

Imputation method	Months enrolled	Rate of attendance
Recorded data	0.1211**	0.0258**
	(0.0513)	(0.0112)
Imputation for HS/GED completers	0.1274**	0.0335**
	(0.0588)	(0.0144)
LOCF imputation	0.1636***	0.0337**
	(0.0622)	(0.0143)
Worst-case imputation	0.1665**	0.0397**
	(0.0775)	(0.0182)
Multiple imputation	0.1172**	0.0320**
	(0.0519)	(0.0130)

Notes: The standard errors are reported in parentheses and adjusted for heteroscedasticity clustered at the county/entry-month level. All models condition on baseline observables and fixed effects specific to each cell defined by the intersection of county, semester, and entry month.
** $p<0.05$.
*** $p<0.01$.

7.2. Dynamic treatment effects

The seminal analysis of the Learnfare experiment suggested that its effects on enrollment and attendance decayed substantially soon after study entry (e.g., see Frye and Caspar (1997) as well as the replication presented here in Table 2). Tables 8 and 9 present new evidence on this question by evaluating how Learnfare's effects on enrollment and attendance 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. This evidence differs from the prior evaluation by *not* always excluding study participants who met Learnfare's requirements by completing high school. Additionally, this study's panel-data specification provides a setting for testing the hypothesis that Learnfare's treatment effects are the same by length of time in the study.

Tables 8 and 9 present the key results of these panel-data specifications for the enrollment and attendance measures, respectively. The results based on data that excludes high-school/GED completers (i.e., model 1) generally suggest that the treatment-induced increases in enrollment and attendance are largest in the first two semesters of study participation. However, the conventional view that Learnfare had at most short-term effects appears to be overdrawn. By the fourth semester, the Learnfare treatment effects do appear to have fallen somewhat and, in the case of the enrollment measure, to become statistically indistinguishable from zero. However, 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 high-school/GED completers are included in the analysis (i.e., model 2).

More formally, Tables 8 and 9 report, for each outcome measure and imputation method, the *p*-values from *F*-tests of the null hypothesis of a common treatment effect across semesters. The hypothesis that the treatment has the same effect by length of time in the study cannot be rejected in any of these models. Models based on alternative imputations for the remaining outcome data that is missing (i.e., models 3 through 5) generate broadly similar results. Furthermore, this pattern of results is also similar both in truncated-regression models and in specifications that specify level and linear-trend treatment effects in lieu of unrestrictive treatment by time-in-study fixed effects.

[12] Learnfare appears to have been particularly effective among some at-risk subgroups (e.g., dropouts, older teens). However, these smaller-sample, heterogeneous effects are not always estimated with precision.

Table 8
Estimated treatment effects on months enrolled by time in study.

Independent variable	(1)	(2)	(3)	(4)	(5)
Treatment × 1st	0.1239**	0.1141**	0.1151**	0.1403**	0.1128**
semester in study	(0.0596)	(0.0550)	(0.0556)	(0.0585)	(0.0561)
Treatment × 2nd	0.1468*	0.1124	0.1666**	0.1834**	0.1154
semester in study	(0.0743)	(0.0741)	(0.0734)	(0.0911)	(0.0732)
Treatment × 3rd	0.1044	0.1207	0.1411	0.1147	0.1182
semester in study	(0.0962)	(0.0969)	(0.0883)	(0.0973)	(0.0895)
Treatment × 4th	0.0893	0.1372	0.1880**	0.2552**	0.1381
semester in study	(0.0952)	(0.1096)	(0.0925)	(0.1224)	(0.1016)
Treatment × 5th	0.1313	0.1351	0.1679*	0.0976	0.0943
semester in study	(0.0971)	(0.0929)	(0.0921)	(0.1324)	(0.0900)
Treatment × 6th	0.1541	0.2997	0.3886	0.3837	0.2762
semester in study	(0.3789)	(0.3548)	(0.2507)	(0.3403)	(0.2701)
Missing outcome imputation	None	HS/GED	LOCF	Worst-case	Multiple
Sample size	4862	5173	5908	6028	6028
p-Value	0.9961	0.9976	0.8737	0.5048	0.9928

Notes: The standard errors are reported in parentheses and adjusted for heteroscedasticity clustered at the county/entry-month level. All models condition on baseline observables and fixed effects specific to each cell defined by the intersection of county, semester, and entry month. The *p*-values refer to *F*-tests of the null hypothesis of a common treatment effect across the six study semesters.
* $p < 0.1$.
** $p < 0.05$.

7.3. Interpreting effect sizes

Overall, these results indicate that Learnfare generated statistically significant increases in school enrollment and attendance that may have been sustained over the 6-semester study window. However, how can the magnitudes of these treatment effects be understood? One approach is to note that the treatment-induced increase in enrollment (i.e., Table 6, model 3) is equivalent to 3.6 percent of the control-group mean and 0.083 of the control-group standard deviation. Similarly, the increase in the rate of attendance is 4.5 percent of the control group mean (and 0.103 of a standard deviation).

However, one alternative and potentially compelling way to interpret these treatment estimates is to compare them to the policy-relevant achievement gaps evidenced in the data. For example, the estimates from Eq. (2) indicate that being a dropout at baseline implies an enrollment outcome that is 0.99 lower (t-statistic = −6.69) and an attendance rate that is 0.2582 lower (t-statistic = −7.43). The improvements implied by Learnfare's treatment effects are equivalent to 13 percent of these enrollment and attendance gaps. Alternatively, the enrollment measure is 0.1567 higher for females than for males (t-statistic = 2.84). The treatment effect implied by Learnfare equals 81 percent of this gender gap. And those who are "over age" for their grades have an attendance rate that is 0.0748 lower (t-statistic = −3.01). The increase in school attendance implied by Learnfare is equal to 45 percent of this gap. This sort of evidence from benchmarking

Table 9
Estimated treatment effects on attendance rate by time in study.

Independent variable	(1)	(2)	(3)	(4)	(5)
Treatment × 1st	0.0382**	0.0408**	0.0410***	0.0427**	0.0400**
semester in study	(0.0161)	(0.0158)	(0.0152)	(0.0191)	(0.0167)
Treatment × 2nd	0.0230	0.0277*	0.0340**	0.0371*	0.0256
semester in study	(0.0141)	(0.0156)	(0.0161)	(0.0196)	(0.0171)
Treatment × 3rd	0.0114	0.0197	0.0233	0.0287	0.0191
semester in study	(0.0164)	(0.0191)	(0.0176)	(0.0210)	(0.0193)
Treatment × 4th	0.0303*	0.0386	0.0395**	0.0693***	0.0349
semester in study	(0.0176)	(0.0234)	(0.0184)	(0.0255)	(0.0233)
Treatment × 5th	0.0274	0.0353	0.0289	0.0216	0.0267
semester in study	(0.0231)	(0.0242)	(0.0213)	(0.0301)	(0.0231)
Treatment × 6th	0.0045	0.0768	0.0433	0.0268	0.0726
semester in study	(0.0674)	(0.0712)	(0.0517)	(0.0648)	(0.0595)
Missing outcome imputation	None	HS/GED	LOCF	Worst-case	Multiple
Sample size	4697	5030	5826	6028	6028
p-Value	0.7775	0.8009	0.8419	0.1562	0.9180

Notes: The standard errors are reported in parentheses and adjusted for heteroscedasticity clustered at the county/entry-month level. All models condition on baseline observables and fixed effects specific to each cell defined by the intersection of county, semester, and entry month. The *p*-values refer to *F*-tests of the null hypothesis of a common treatment effect across the six study semesters.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

Learnfare's treatment effects against achievement gaps suggests that these impacts are likely to be seen as policy relevant.[13]

Another more speculative way to frame these effect sizes is to consider their possible implications for subsequent labor-market outcomes. An extensive literature provides evidence that exogenously determined increases in secondary-school persistence leads to statistically significant increases in wages that are sustained over the life cycle (e.g., Card, 2001). Under certain assumptions, the estimated wage returns to secondary-school persistence from this literature provide a natural way to monetize the benefits of Learnfare's impact. In particular, the results in Table 6 imply that Learnfare generated approximately 3 additional days of enrollment and attendance per semester.[14] If we maintained the assumption that these gains were sustained over the 6-semester study window, the implication is that Learnfare led to 18 additional school days, an amount equivalent to 10 percent of the 180-day school year in Wisconsin.

How large is the present discounted value of the wage increase implied by having 0.10 additional years of secondary-school persistence? To provide some evidence on this question, I constructed an age-earnings profile using data from 18 to 65 year-old respondents to the March 2007 Current Population Survey (CPS). Under the fairly conservative assumption that a year of additional schooling increases wages by 7 percent, Learnfare implies a wage increase of 0.7 percent (i.e., 0.10×0.07). I assumed that this wage increase began at age 18 and lasted until age 65 and calculated its present discounted value as of age 14. Under the assumption of a 3 percent real discount rate, the benefits of this increased school persistence is nearly $6000. To be clear, this rough estimate may be sensitive to several relevant assumptions. For example, this approach will understate the social gains from school persistence because it ignores how productivity growth will shape earnings trajectories and it ignores the social externalities of increased educational attainment (e.g., productivity spillovers, civic engagement, and crime). In contrast, the assumption that the Learnfare-induced increases in school persistence have the same labor-market consequences as increases in secondary-school attainment may lead to overstated gains. Nonetheless, this evidence suggests that even seemingly modest gains in secondary-school persistence from policies may have meaningful implications for economic outcomes over the life cycle.

8. 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. In particular, 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. Second, 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). 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 & 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 effective in generating policy-relevant increases in both school enrollment and attendance. 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 financially harming economically vulnerable populations.

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 promoting school enrollment and attendance. For example, the "What Works Clearinghouse" maintained by the Institute of Education Sciences has identified other

[13] To be clear, this is not to say that Learnfare would close the enrollment and attendance gaps by these amounts; the Learnfare experiment was not well powered for sub-group analysis. Rather, the point here is that Learnfare's *overall* impact estimates are large relative to the performance gaps that often motivate policy attention.

[14] The assumption of 20 school days in a month implies that 0.1274 additional months is a 2.55-day increase. The assumption of 90 school days in a semester implies that a 0.0335 increase in the attendance rate is 3.02 days.

effective dropout prevention programs (e.g., ALAS, Check and Connect) that rely on intensive case management rather than financial incentives. Small-scale random-assignment evaluations of ALAS and Check and Connect (Larson & Rumberger, 1995; Sinclair, Christenson, Evelo, & Hurley, 1998) found that they generated comparatively large increases in school enrollment (i.e., 15 and 19 percentage-point increases, respectively).

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 exclusively on expensive case management and support services. Furthermore, whether programs like ALAS and Check and Connect can be effective at scale is an open, empirical question (each evaluation had fewer than 50 students in the treatment condition). Similar external-validity concerns are also relevant with regard to the Learnfare results presented here, in no small part because of the apparent randomization failure and low-fidelity implementation in Milwaukee County. Nonetheless, the evidence from the random-assignment evaluation analyzed here provides evidence for the efficacy of Learnfare as a mature policy that had been implemented at a comparatively large scale.

References

Angrist, J. D., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, *1*(1), 136–163.

Angrist, J. D., & Lavy, V. (June 2008). *The effects of high-stakes high school achievement awards: Evidence from a group-randomized trial*. Working paper.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Bénabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies*, *70*, 489–520.

Bettinger, E. P. (March 2009). *Paying to learn: The effect of financial incentives on elementary test scores*. Working paper.

Bos, J. M., & Fellerath, V. (1997). *LEAP: Final report on Ohio's welfare initiative to improve school attendance among teenage parents: Ohio's learning, earning, and parenting program*. New York: MDRC.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*(December (1–3)), 7–42.

Cameron, J., & Pierce, W. D. (2002). *Rewards and intrinsic motivation: Resolving the controversy*. Westport, CT: Bergin and Garvey.

Cameron, J. (2001). Negative effects of reward on intrinsic motivation – A limited phenomenon: Comment on Deci Koestner, and Ryan (2001). *Review of Educational Research*, *71*(Spring (1)), 29–42.

Campbell, D., & Wright, J. (2005). Rethinking welfare school – Attendance policies. *Social Service Review*, (March), 2–28.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, *69*(September (5)), 1127–1160.

Dee, T. S. (2004). Teachers, race and student achievement in a randomized experiment. *The Review of Economics and Statistics*, *86*(February), 195–210.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, *71*(Spring (1)), 1–27.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*, 105–115.

Education Commission of the States (2009). *Student accountability initiatives: Learnfare*. Updated July 30, 2007. http://mb2.ecs.org/reports/Report.aspx?id=1633 Accessed 27.03.2009.

Etheridge, M. E., & Perry, S. L. (1993). A new kind of public policy encounters disappointing results: Implementing Learnfare in Wisconsin. *Public Administration Review*, *53*(July–August (4)), 340–347.

Fein, D. J., Long, D. A., Behrens, J. M., & Lee, W. S. (2001). *The ABC evaluation: Turning the Corner: Delaware's a better chance welfare reform program at four years*. Cambridge, MA: Abt Associates Inc.

Frye, J., & Caspar, E. (1997). *An evaluation of the Learnfare program: Final report*. Madison, WI: State of Wisconsin Legislative Audit Bureau.

Frye, J., Caspar, E., & Merrill, N. (1992). *Research design evaluation of the Learnfare program*. Madison, WI: State of Wisconsin Legislative Audit Bureau.

Fryer, R. (April 2010). Financial incentives and student achievement: Evidence from randomized trials. NBER Working paper No. 15898.

Handa, S., & Davis, B. (2006). The experience of conditional cash transfers in Latin America and the Caribbean. *Development Policy Review*, *24*(September (5)), 513–536.

Hoynes, H. (1997). Work, welfare and family structure: What have we learned? In A. Auerbach (Ed.), *Fiscal policy: Lessons from economic research* (pp. 101–146). Cambridge, Mass: MIT Press.

Jones, L. P., Harris, R., & Finnegan, D. (2002). 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), 222–237.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kremer, M., Miguel, E., Thornton, R., & Ozier, O. (May 2004). *Incentives to Learn*. World Bank Policy Research Working Paper No. 3546.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, *114*(2), 497–532.

Larson, K. A., & Rumberger, R. W. (1995). ALAS: Achievement for Latinos through academic success. In H. Thornton (Ed.), *Staying in school. A technical report of three dropout prevention projects for junior high school students with learning and emotional disabilities*. Minneapolis, MN: University of Minnesota, Institute on Community Integration.

Leuven, E., Oosterbeek, H., & van der Klaauw, B. (2010). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, *8*(6), 1243–1265.

Mauldon, J., Malvin, J., Stiles, J., Nicosia, N., & Seto, E.,Y. (2000). *The impact of California's cal-learn demonstration project, Final Report.* UC Data Archive & Technical Assistance. UC Data Reports: Paper, June 1, 2000.

Maynard, R. (1993). *Building self-sufficiency among welfare-dependent teenage parents: Lessons from the teenage parent demonstration*. Princeton, NJ: Mathematica Policy Research, Inc.

Mead, L. (1986). *Beyond entitlement: The social obligations of citizenship*. New York: Free Press.

Pawasarat, J., Quinn, L., & Stetzer, F. (1992). *Evaluation of the impact of Wisconsin's Learnfare experiment on the school attendance of teenagers receiving aid to families with dependent children*. Employment Training Institute, University of Wisconsin-Milwaukee.

Quinn, L. M., & Magill, R. S. (1994). Politics versus research in social policy. *The Social Service Review*, *68*(December (4)), 503–520.

Richburg-Hayes, L., Brock, T., LeBlanc, A., Paxson, C., Rouse, C. E., & Barrow, L. (2009). *Rewarding persistence effects of a performance-based scholarship program for low-income parents*. New York: Manpower Defense Research Corporation.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: J. Wiley & Sons.

Sinclair, M. F., Christenson, S. L., Evelo, D. L., & Hurley, C. M. (1998). Dropout prevention for youth with disabilities: Efficacy of a sustained school engagement procedure. *Exceptional Children*, *65*(1), 7–21.

Skoufias, E., & McClafferty, B. (July 2001). *Is PROGRESA Working? Summary of the Results of an Evaluation by IFPRI*. International Food Policy Research Institute, FCND Discussion Paper No. 118.

Stoker, R. P., & Wilson, L. A. (1998). Verifying compliance: Social regulation and welfare reform. *Public Administration Review*, *58*(September/October (5)), 395–405.

Thaler, R., & Susstein, C. (2008). *Nudge: Improving decisions about health, wealth and happiness*. New Haven: Yale University Press.

Wiseman, M. (1996). State strategies for welfare reform: The Wisconsin story. *Journal of Policy Analysis and Management*, *15*(Autumn (4)), 515–546.

Wood, A. M., White, I. R., & Thompson, S. G. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials*, *1*, 368–376.