capsee
CENTER FOR ANALYSIS OF POSTSECONDARY EDUCATION AND EMPLOYMENT

Performance Standards in Need-Based Student Aid

A CAPSEE Working Paper

Judith Scott-Clayton
Community College Research Center
Teachers College, Columbia University

Lauren Schudde
The University of Texas at Austin

December 2016

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C110011 to Teachers College, Columbia University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. We are grateful to staff and administrators at the community college system that facilitated data access and interpretation. For helpful feedback, we are grateful to colleagues at the Community College Research Center, attendees of the NBER education meetings, and seminar participants at the Federal Reserve Board, the University of Virginia, and the University of Connecticut. Rina S. E. Park provided excellent research assistance.

For information about CAPSEE, visit capseecenter.org

Abstract

College attendance is a risky investment. But students may not recognize when they are at risk for failure, and financial aid introduces the possibility for moral hazard. Academic performance standards can serve three roles in this context: signaling expectations for success, providing incentives for increased student effort, and limiting financial losses. Such standards have existed in federal need-based aid programs for nearly 40 years in the form of Satisfactory Academic Progress (SAP) requirements, yet have received virtually no academic attention. In this paper, we sketch a simple model to illustrate not only student responses to standards but also the tradeoffs faced by a social planner weighing whether to set performance standards in the context of need-based aid. We then use regression discontinuity and difference-in-difference designs to examine the consequences of SAP failure. In line with theoretical predictions, we find heterogeneous effects in the short term, with negative impacts on persistence but positive effects on grades for students who remain enrolled. After three years, the negative effects appear to dominate. Effects on credits attempted are 2–3 times as large as effects on credits earned, suggesting that standards increase the efficiency of aid expenditures. But it also appears to exacerbate inequality in higher education by pushing out low-performing low-income students faster than their equally low-performing, but higher-income peers.

Table of Contents

1. Introduction

College is a risky investment. Not only do prospective students face uncertainty about their likely income conditional on graduation (Wiswall & Zafar, 2015), but they also face significant uncertainty regarding how long they will persist and whether or not they will actually graduate. Uncertainty about completion exists because college is an experience good; prospective students may not discover their own tastes and abilities for college-level work until they try it (Altonji, 1993; Manski, 1989). College dropout is in part a manifestation of this uncertainty; scholars have long noted that some level of dropout is to be expected even in a well-functioning postsecondary system, as students learn more about their preferences and abilities (Fischer, 1987; Manski, 1988, 1989; Manski & Wise, 1983).

Even after enrolling and beginning this learning process, however, students may make suboptimal decisions about dropout for a number of reasons. Large and growing gaps in educational attainment by family income, which remain even after controlling for prior measures of ability, are consistent with credit constraints leading some students to end their schooling too soon (Bailey & Dynarski, 2011; Belley & Lochner, 2007). Moreover, students, like other people, may overly discount future payoffs when costly actions are required in the present (Lavecchia, Liu, & Oreopoulos, 2014). Finally, persistence and completion may generate positive social externalities in addition to the private benefits valued by students. Any of these factors could lead to suboptimal college enrollment and persistence, and form the justification for substantial public subsidies in higher education.

Far less attention has been devoted to the question of whether some low-performing students might persist *longer* than they should. Yet the ubiquitous nature of academic performance standards at postsecondary institutions—which typically require students to maintain a minimum grade point average (GPA) or risk being placed on probation and eventually dismissed—suggests a widespread belief that they are needed. Students may not be well-informed regarding institutional expectations for graduation and how their performance compares to it (Scott-Clayton, 2013). Alternatively, they may be informed procrastinators, always planning to increase effort next semester rather than in the present. In addition, evidence suggests that students are slow to update beliefs about completion probability even after a period of poor performance, because they underestimate the role of persistent rather than transitory factors (Stinebrickner & Stinebrickner, 2012). Finally, the same public subsidies intended to promote enrollment and completion also introduce the possibility of moral hazard. Some students might persist in the “college experiment” beyond the socially optimal point.

Academic performance standards can serve three roles in this context: sending a clear signal about institutional expectations for success, providing incentives for increased student effort early in college, and limiting financial losses (potentially for overoptimistic students themselves, as well as for taxpayers). Academic performance standards of one form or another apply to all students, regardless of financial aid status. But the stakes are arguably highest in the context of financial aid policy, where performance standards are commonplace even in purely need-based programs.

In this paper, we focus on the consequences of Satisfactory Academic Progress (SAP) requirements as established by the federal student aid system ("Title IV" aid). While eligibility for Pell Grants and student loans is initially based purely on financial need, recipients must meet SAP standards in order to continue receiving aid (state and institutional need-based programs often follow the federal rules as well). Institutions have flexibility regarding how they define and enforce SAP, but they commonly require students to maintain a cumulative GPA of 2.0 or higher and to complete at least two-thirds of the course credits that they attempt. Meeting SAP is a non-trivial hurdle for many students: in earlier work, we find that 25–40 percent of first-year Pell recipients at public institutions have performance low enough to place them at risk of losing financial aid, representing hundreds of thousands to over a million college entrants each year (Schudde & Scott-Clayton, 2016).

Though minimum performance standards have existed in the need-based federal student aid programs for nearly 40 years—and have become increasingly strict in recent years—we have found very little academic research specifically relating to their consequences, either theoretical or empirical. Bénabou and Tirole (2000) provide a model of how students react to performance standards in general, but do not consider interactions with financial aid policy. While there is a substantial literature on the impacts of performance-based scholarships, such scholarships typically focus on GPA thresholds well above 2.0 (Carruthers & Ozek, 2014; Cornwell, Lee, & Mustard, 2005; Cornwell, Mustard, & Sridhar, 2006; Dynarski, 2008; Scott-Clayton, 2011) and examine the marginal effect of receiving extra aid rather than the effect of losing foundational need-based assistance (Angrist, Lang, & Oreopoulos, 2009; Patel & Valenzuela, 2013; Barrow, Richburg-Hayes, Rouse, & Brock, 2014; Barrow & Rouse, 2013; Richburg-Hayes et al., 2009). Nonetheless, this literature generally finds that students are responsive to performance incentives. A series of experiments with performance-based scholarships (provided as a supplement to Pell Grants) conducted by MDRC are particularly relevant given their low-income target population and 2.0 GPA threshold, which corresponds to the SAP threshold. In these studies, students assigned to the treatment group increased the time they spent on academic activities, earned more credits, and were more likely to persist (Patel & Valenzuela, 2013; Barrow et al., 2014; Barrow & Rouse, 2013).

The most closely related empirical work is a study by Lindo, Sanders, and Oreopoulos (2010) of the causal effects of academic probation for students (regardless of financial aid status) at one large baccalaureate institution in Canada. Using regression discontinuity, the authors compare persistence, grades, and graduation rates for students just above and below the first-year GPA threshold for placement onto academic probation. The authors find that being placed on probation induces some students to drop out but increases GPA for those that return, with a net negative impact on graduation for students near the cutoff. A recent paper by Casey, Cline, Ost, and Qureshi (2015) replicates these findings using data from a U.S. public four-year institution but finds that some of the increase in GPA is due to strategic course selection. Finally, our own recent work documenting the prevalence of SAP failure (Schudde & Scott-Clayton, 2016) includes an analysis of the effects of standards for Pell recipients in a different state than examined here, with our preferred strategy suggesting that the increases in dropout may be larger for Pell recipients than non-recipients. Taken together, these prior studies provide support for the Bénabou and Tirole model of student behavior, yet provide little guidance for evaluating standards in the context of financial aid policy.

Our paper makes three contributions. First, we sketch out a simple model that not only illustrates students' likely responses to performance standards but also highlights the tradeoffs faced by a social planner weighing whether to set performance standards in the context of need-based aid. Our framework draws upon elements of Manski's (1988, 1989) "schooling as experimentation" model as well as Bénabou and Tirole's (2000, 2002) model of student behavior under performance standards. The model has three periods: an evaluation period, a warning period, and an enforcement period. Second, we use this framework to guide an empirical examination of the consequences of minimum performance standards for a high-risk population: community college entrants. Utilizing administrative records on aid recipients at more than 20 community colleges in one state, we apply a regression discontinuity (RD) design similar to that used in prior work. However, our preferred estimates use a difference-in-difference (DID) design which uses unaided students as a control group to net out any effects of academic probation in general (that is, any effects of being below the performance threshold that affect all students, not just those receiving financial aid) and also allows us to estimate effects for students further away from the performance threshold. While the RD results are causally cleaner, the DID estimates are of greater policy relevance in terms of determining whether SAP requirements are effective overall. Finally, we examine a broader range of outcomes—including measures of student labor supply and the estimated value of foregone financial aid—that provide a fuller picture of the costs and benefits of the policy.

In line with theoretical predictions and consistent with Lindo, Sanders, and Oreopoulos (2010), we find a clear pattern of heterogeneous effects of SAP failure during the warning period (when students should realize they are at risk but have not yet faced consequences), with students just below the GPA threshold being less likely to return in the second year, but with positive effects on grades for students who do return. Also consistent with our model, we find that discouragement effects are larger for students further below the threshold while encouragement effects appear larger for those nearest the threshold.

Despite these heterogeneous effects in the short term, negative effects appear to dominate by the end of our three-year follow-up window, after the point at which consequences of continued SAP failure would have been enforced. Both the RD and DID specifications indicate significant reductions in the likelihood that students remain enrolled by that point, with no improvement in cumulative GPA. Interestingly, we find significant negative effects on credits attempted, but much smaller effects on credits earned, suggesting that the marginal credit no longer attempted had a low probability of being completed. Effects on degree completion are difficult to interpret because of potential floor effects, but the most consistent degree result is a small negative effect on certificate completion. We find no effect on student earnings in most specifications, though coefficients are generally negative.

Evaluating SAP policy as a whole requires weighing the value of human capital foregone against the cost of continuing to subsidize enrollment. While a complete assessment of net benefits is beyond the scope of this paper, if we assume that credits attempted but not completed have no value, then SAP does appear to improve the efficiency of aid dollars. The reduction of three credits attempted translates into a reduction of only one credit completed on average, yet saves at least $444–$539 in financial aid expenditures per student.[1] Still, the small negative effects on certification completion may be cause for concern, since such credentials have been found to increase individuals' post-enrollment earnings (Belfield, Liu, & Trimble, 2014; Jacobson & Mokher, 2009; Jepsen, Troske, & Coomes, 2014; Xu & Trimble, 2014). Moreover, policymakers may be concerned about the equity implications, as our results clearly indicate that aid recipients with low GPAs leave college more quickly than their similar but financially unassisted peers.

The remainder of the paper proceeds as follows: Section 2 provides additional background on SAP policy. Section 3 introduces the theoretical framework and key predictions. Section 4 describes our data, section 5 describes our empirical strategy, and section 6 presents our main results. Section 7 concludes with a discussion of policy implications and unanswered questions.

2. Policy Background

SAP regulations have been a part of federal student aid since 1976 when an amendment to the Higher Education Act of 1965 stipulated that students must demonstrate "satisfactory progress" toward a degree in order to continue receiving aid (Bennett & Grothe, 1982). The regulations give institutions flexibility regarding how they define SAP, though in practice it appears typical for institutions to require a cumulative grade point average (GPA) of 2.0 or higher and completion of at least two thirds of the course credits that students attempt (Schudde & Scott-Clayton, 2016). Our analysis will focus on the GPA criterion, as it is highly correlated with credit completion percentage in our sample ($\rho = 0.79$), but much more continuously distributed.

[1] For comparison, note that a credit is 1/24th of a full-time school year, and in-state tuition in this system was less than $115/credit over this time period. This is a conservative estimate that assumes students who continue to enroll continue to receive aid (when in fact some students who remain enrolled may have lost their aid eligibility). We do not have measures of actual aid received beyond the first year. Instead, we use actual first-year aid receipt to estimate that students in our sample would qualify for approximately $140–170 in aid per credit attempted in subsequent years.

SAP policy applies to federal Pell Grant recipients, student loan borrowers, and work-study participants; and state and institutional need-based aid programs often piggyback their minimum performance rules on the federal standards.[2] While eligibility for federal aid is initially based purely on financial need, recipients must meet satisfactory academic progress (SAP) requirements in order to remain eligible beyond the first year. The federal Pell Grant is the single largest source of need-based financial aid in the country and the dominant form of aid received at most community colleges, both in terms of frequency and magnitude of awards (Baum & Payea, 2013). Thus, while our empirical analysis groups together students receiving any type of aid,[3] for interpretation purposes we recognize that most financial aid recipients in our sample are Pell recipients. For example, among the aid recipients in our community college sample, 70 percent received Pell, compared to 44 percent receiving state grants and 20 percent receiving loans. Average amounts among all aid recipients were $2,385 in Pell compared to just $407 in state grants and $352 in student loans.

Until recently, institutions have had a great deal of flexibility in terms of how frequently they evaluate SAP and thus how quickly consequences are enforced. Prior to 2011, institutions were only bound to enforce SAP at the end of the second year; at that point, students whose GPA fell below the threshold for graduation could no longer receive aid.[4] Many institutions nonetheless opted to evaluate SAP more frequently, often by the end of the first year, giving students another full year under either "warning" or "probation" status to try to meet the standard. A "warning" generally means a student is notified of their precarious status but no formal action is taken; "probation" means that the student failed to meet SAP requirements during the warning period and filed an appeal explaining their poor performance in order to continue receiving aid—though these terms were not always consistently applied.

SAP standards layer on top of "academic good standing" policies that apply to all students regardless of aid status. Though community colleges are "open access" institutions, meaning that any individual with a high school diploma or its equivalent may enroll initially, they still have minimum academic standards that students must meet in order to remain enrolled and to earn a degree. In the state community college system (SCCS) that we examine, a 2.0 cumulative GPA, or a C average, is required to earn a credential (this appears to be a fairly typical graduation standard at public institutions nationally). Students who fall below a 2.0 in any semester are placed on academic warning status and a notification will appear on their

[2] In the state we examine, need-based state aid is the second largest source of student grants, and aid administrators confirmed that the state programs follow the same rules used for federal SAP.
[3] A prior version of this paper focused exclusively on Pell recipients and generated a substantively identical pattern of results; however, moving other aid recipients from the control group to the treated group helps improve power and is most consistent with how the policy is implemented in practice.
[4] Even defining the "end of the second year" is not particularly straightforward in the context of community colleges, where many students attend part-time. We found many policies were defined in terms of credits attempted, where 48+ would correspond to the end of the second year.

transcript.[5] However, besides warning students that they are not on track to graduate, academic warning in this system has little durable consequence except for students on financial aid (who, prior to 2011, would lose aid after the end of two years). More immediate consequences are reserved for students who fall below a 1.5 GPA: these students may be immediately placed on probation and required to file an appeal in order to continue enrolling or receiving aid. Still, the system's guidelines emphasize that even if an academic warning does not itself lead to dismissal, students will not be able to earn a degree with a less than 2.0 GPA.[6]

Our review of college catalogs describing SAP and academic good standing policies suggests performance standards may be less than perfectly transparent to students; indeed, we found them challenging to decipher ourselves. In the years pertaining to our sample, policies appear to vary somewhat across colleges, and the thresholds and timelines for SAP evaluation do not always seem to correspond to the thresholds and timelines for broader institutional academic standards. We will return to this complexity in our discussion section. Nonetheless, for the period and sample under consideration, all students with below a 2.0 received at least some notification that they were at academic risk by the end of their first year; but students were unlikely to face binding consequences (such as financial aid loss or dismissal) before the end of their second year unless they fell below a 1.5 GPA.

3. Theoretical Framework

Basic Model of Performance Standards

Bénabou and Tirole (2000) present a simple principal–agent model in which agents choose between shirking, a low-effort/low-benefit task, and a high-effort/high-benefit task. Lindo, Sanders, and Oreopoulos (2010) use an even simpler version of this model in their analysis of academic probation, focusing on the agent's decision. Because of our interest in optimal policy, we utilize Bénabou and Tirole's original model and examine its predictions after incorporating financial aid.

In the original model, individuals choose one of these: shirking, which yields no costs or benefits; a low-effort/low-benefit task (Task 1) with a private benefit of V_1 and a private effort cost of c_1; or a high-effort/high-benefit task (Task 2) with private benefit of V_2 and a private effort cost of c_2. The principal bears no costs but receives a benefit of W_1 from Task 1 and a benefit of W_2 for Task 2. In other words, the ranking of costs and benefits is:

[5] We find some conflicting information between policy manuals produced at the system level, which suggest students below 2.0 are placed on academic warning, and catalogs at the college level, at least one of which describes students as being in "good standing" as long as they maintain a 1.5 GPA (though they will not be able to graduate).
[6] Information on institutional academic good standing and SAP policies are taken from course catalogs for years prior to 2011.

$$0 < V_1 < V_2, \quad 0 < W_1 < W_2, \quad and \quad 0 < c_1 < c_2 \tag{1}$$

Ability is conceptualized as an exogenously determined probability of success at either task, θ, which for now we assume the agent knows but the principal does not. To ensure that the problem does not degenerate and that at least some individuals choose each option, the following assumption is made (intuitively, this assumes that marginal cost of Task 2 relative to Task 1 is less than the marginal benefit, but the ratio of marginal costs to marginal benefits is higher for Task 2 than for Task 1):

$$\frac{c_1}{V_1} < \frac{c_2 - c_1}{V_2 - V_1} < 1 \tag{2}$$

The agent chooses the course of action that maximizes her individual outcome:

$$\max\{0, \theta V_1 - c_1, \theta V_2 - c_2\} \tag{3}$$

So the individual will choose:

$$to\ shirk:\ if\ 0 \leq \theta < \frac{c_1}{V_1} \equiv \theta_1$$

$$Task\ 1:\ if\ \theta_1 \equiv \frac{c_1}{V_1} \leq \theta < \frac{c_2 - c_1}{V_2 - V_1} \equiv \theta_2 \tag{4}$$

$$Task\ 2:\ if\ \frac{c_2 - c_1}{V_2 - V_1} \equiv \theta_2 \leq \theta$$

If the principal removes the low-effort/low-benefit option, individuals who would otherwise have chosen that option now are forced to choose between shirking or increasing their effort. Now, individuals will shirk only if:

$$\theta < \frac{c_2}{V_2} \equiv \theta^* \tag{5}$$

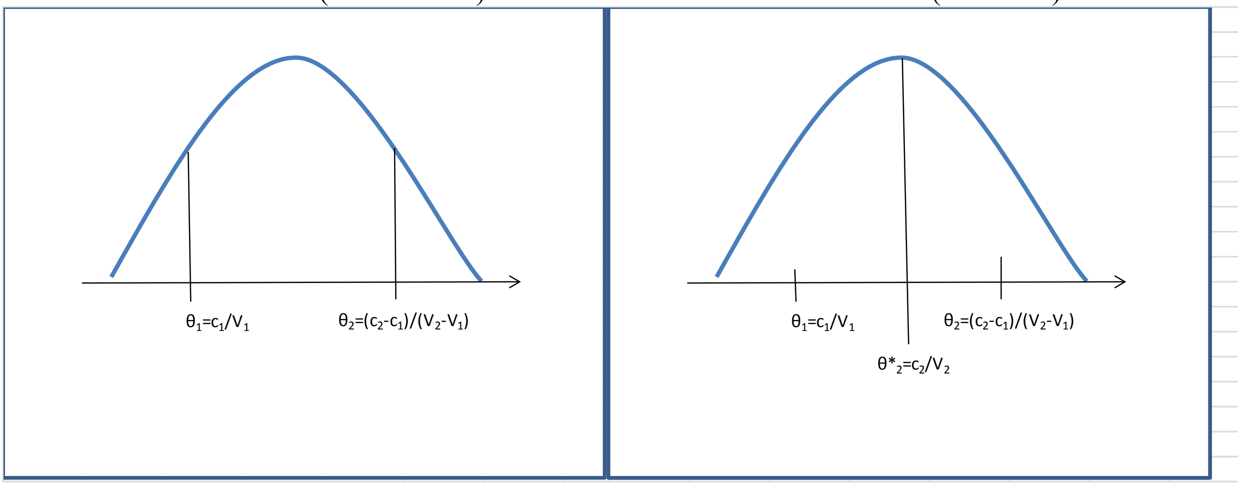
The key insight of this model, which Lindo et al. (2010) emphasize, is the heterogeneous impact that results when performance standards are applied. Higher ability individuals are motivated to work harder, while lower ability individuals are discouraged and drop out. See Figure 1 for a graphical illustration.

For our analysis, we are also interested in the principal's perspective. Imposing a standard is only worthwhile for the principal if the increase in value coming from those induced to work harder exceeds the loss of value attributable to those induced to shirk. This depends not only on the parameters discussed above but also on the distribution of ability *f(θ)* in the population (i.e., how many individuals are in the affected ranges):

$$S(\theta_1, \theta_2) = \left(\int_{\theta^*}^{\theta_2} \theta f(\theta) d\theta\right)(W_2 - W_1) - \left(\int_{\theta_1}^{\theta^*} \theta f(\theta) d\theta\right)(W_1) > 0 \tag{6}$$

Figure 1: Cutoff Values for Choosing Shirking, Task 1, or Task 2 in the Distribution of Ability

Panel A. Task 1 Permitted (no standards) Panel 2. Task 1 Forbidden (standards)



Introducing Financial Aid Without Standards

The Pell Grant and other scholarship programs provide another means by which a social planner could encourage greater investment in education (either because of perceived externalities or because individuals systematically underestimate the true private benefit). If the social planner provides an upfront scholarship, *P*, based on enrollment but not outcomes (i.e., available to those who choose either Task 1 or Task 2 in the model), it is straightforward to show that this will result in a new $\theta_1^P < \theta_1$, but $\theta_2^P = \theta_2$ (see Figure 2). In other words, the scholarship induces more individuals into the low-effort/low-benefit Task 1, but no more individuals into Task 2. Manski (1988) also highlights a similar conclusion in his analysis of the effects of upfront, non-contingent enrollment subsidies: these "can induce students to change their enrollment decisions but cannot induce changes in completion decisions" (p. 13). As such, enrollment subsidies cannot guarantee the socially optimal outcome, but they may still be desirable (relative to no aid with no standards) if:

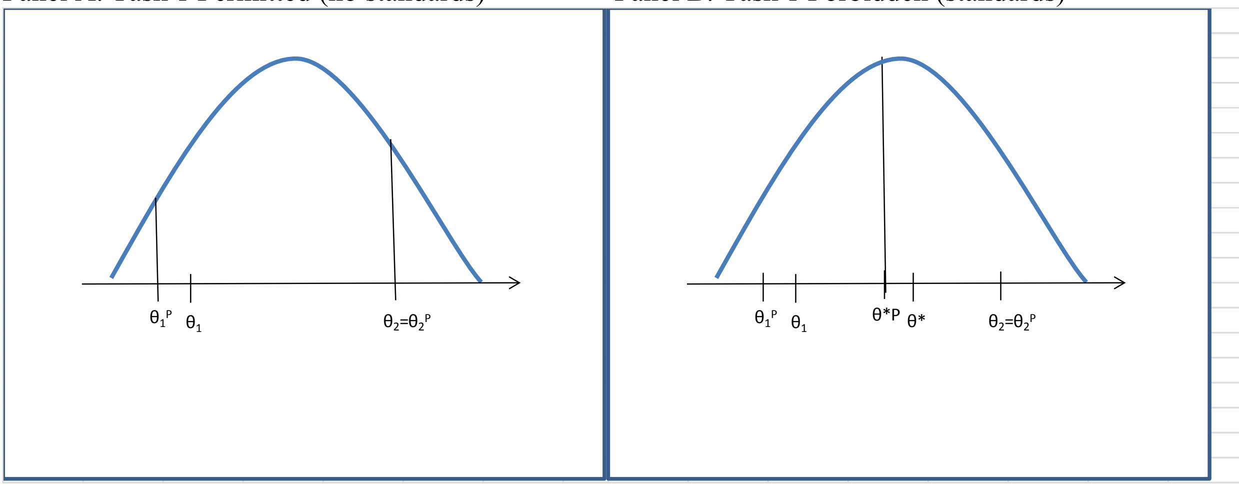
$$\left(\int_{\theta_1^P}^{\theta_1} \theta f(\theta)d\theta\right)(W_1) - \left(\int_{\theta_1^P}^{\overline{\theta}} f(\theta)d\theta\right)(P) > 0 \quad (7)$$

In essence, the new benefits attributable from those induced to enroll with low effort must more than cover the costs of providing the scholarship to all those who enroll, including potentially many individuals whose enrollment and effort are unaffected by the scholarship.[7]

[7] If we consider Task 1 to be enrolling but dropping out and Task 2 to be persisting to completion, it is worth noting that a grant program could be worthwhile even if it increases dropout rates. Manski emphasizes that "dropout statistics per se carry no normative message…among [some] students, society prefers a *higher* dropout rate than that generated privately" (1989, p. 310, italics in original).

Figure 2: Cutoff Values With Financial Aid *P*

Panel A. Task 1 Permitted (no standards) | Panel B. Task 1 Forbidden (standards)



Adding Performance Standards in the Context of Financial Aid

If the social planner can forbid low-effort enrollment in the context of financial aid, the threshold value for choosing the high-effort option declines to:

$$\frac{c_2 - P}{V_2} \equiv \theta^{*P} < \theta^* \tag{8}$$

Relative to providing a given P *without* performance standards, this policy is worthwhile if:

$$\left(\int_{\theta^{*P}}^{\theta_2} \theta f(\theta) d\theta\right)(W_2 - W_1) - \left(\int_{\theta_1^P}^{\theta^{*P}} (\theta W_1 - P) f(\theta) d\theta\right) > 0 \tag{9}$$

See Figure 2 for a graphical illustration. If W_1, the social value of the forbidden low-effort option, is lower than the value of the scholarship P, then aid-with-standards is unambiguously better than aid-without-standards. Assuming that there is some level of low effort that generates social value less than P, then performance standards are always desirable; the only question is where to set the dividing line between the acceptable W_2 and the unacceptable W_1. Intuitively, the optimal dividing line will depend upon the relative benefits of high-effort versus low-effort enrollment, the relative benefits of low-effort enrollment versus no enrollment, the shape of the ability distribution, and the magnitude of the scholarship.[8]

[8] Note that it is not obvious that larger scholarships should necessarily warrant higher standards, because they enter into equation (9) not only directly but also indirectly: both the positively and negatively affected ranges shift lower in the ability distribution.

Timing and Other Unresolved Issues

One ambiguity in the Bénabou & Tirole (2000) model is the timing of assessment and enforcement, and why we would ever observe students, even in the absence of aid, enrolling only to perform the "prohibited" low-effort task. Manski's (1989) model of education as experimentation is useful here: students themselves may not know their ability until they enroll. We can thus think of the model being spread over three periods, with the first period being an evaluation period in which students learn about their ability, the second period being a warning period (in which individuals know their ability but still receive unconditional aid), and the third period being an enforcement period (in which aid recipients who do not improve above standard lose their aid). These periods are reasonably well-defined for aid recipients in our sample (as the first, second, and third year of follow-up). Note, however, that unaided students in our sample do not really face a defined enforcement period: unlike the probation policies examined by Lindo et al. (2010) and Casey et al. (2015), unaided students can continue taking classes under a warning status indefinitely, as long as their GPA does not fall below 1.5 (even though they will be unable to graduate with a GPA below 2.0).

The model sketched above highlights how the interests of policymakers and students may sometimes diverge, and how performance standards may be socially desirable even while they necessarily reduce utility for at least some students. However, it is possible to imagine scenarios under which students themselves benefit from the enforcement of standards. For example, if students are slow to update beliefs about their own ability (as found by Stinebrickner & Stinebrickner 2012), they may actually benefit from leaving school earlier: they may reallocate their time from unproductive studies to more productive work in the labor market. This suggests examining changes in students' labor supply during the warning and enforcement period.

Implications

In the context of SAP policy, we can think of earning less than a 2.0 GPA in college as the low-effort option. Our empirical analysis will not correspond directly to the model above, since individuals have a continuum of effort levels to choose from, and we do not have direct measures of ability or of the benefits associated with different effort levels. Nonetheless, the model suggests the following implications:

1. Being placed on warning status will induce some individuals to drop out of school, while those who return will increase effort (as shown in Lindo et al. 2010).
2. The discouragement (dropout) effects will be concentrated among those lower in the ability distribution, while the encouragement (improved GPA) effects should be concentrated among those near the threshold.

3. All else equal, encouragement effects of receiving a warning will be bigger in the presence of financial aid (evident by comparing affected regions of the distribution in Figure 2).
4. All else equal, whether discouragement effects are bigger or smaller for the aid-eligible population during the warning period is ambiguous.[9]
5. In the enforcement period, effects of failing the standard for aided students should be unambiguously more negative than for unaided students, as aided students experience the loss of aid.

In addition to these hypotheses, the model also provides some insight regarding the key outcomes to consider from either a student or a social planner perspective. While persistence rates and GPA may be useful for testing the behavioral implications of the model, they are not of direct use in terms of evaluating the net benefits of the policy. For that purpose, we will consider summary measures of human capital accumulated, including total credits completed, cumulative GPA at the end of the follow-up period, and degree/transfer outcomes. We can then weigh the value of any impacts on human capital attained against the impacts on estimated scholarship outlays as well as on students' foregone earnings.

4. Data

We utilize de-identified state administrative data on first-time students who entered one of more than 20 community colleges in a single eastern state between 2004 and 2010, and follow all students' outcomes for three years after initial entry (unless otherwise noted). Only fall entrants are included in the data; however, the community college system classifies a small number of students who begin coursework in the summer as fall entrants. We focus on students who enrolled full-time in their first semester to ensure enough courses are attempted to compute a reliable first-year GPA (an additional justification is that performance standards may not be implemented until students have attempted at least 12 credits). The data include limited demographic information, detailed transcripts, placement test scores, information on financial aid received in the first year, and credentials earned. The state system also links these institutional data to two additional databases: the National Student Clearinghouse (NSC), which captures enrollment and credentials at institutions outside the state community college system, and individual Unemployment Insurance (UI) records, which indicate quarterly employment and earnings. We use the UI data to construct measures of student labor supply during the second and third years post-entry, to capture potential earnings foregone by students who remain enrolled.

[9] Aid induces some low-ability students to enroll who are likely to drop out once standards are introduced; however, some moderate-ability students who would have enrolled even without aid are induced into the high-effort task when standards are introduced.

One thing the data do not include is any explicit measure of academic warning or probation. Instead, we infer who is subject to these labels based upon students' cumulative grade point averages (GPAs). It is possible that not all students below the 2.0 GPA threshold were placed on a warning status, and that some students above the threshold were.[10] To the extent this occurs, it will tend to bias our estimated effects towards zero. Similarly, we are not able to explicitly identify students who lost Pell eligibility as a result of SAP failure; while this would be interesting, it is not necessary for our analysis, which includes the initial *threat* of losing Pell (based on first-year GPA) as a key aspect of treatment rather than only considering students who experience the actual loss of aid.

Table 1 describes our full sample and provides mean outcome levels for all full-time entrants to this state system as well as for reference groups relevant to our RD strategy (aid recipients and non-recipients +/- 0.25 around the cutoff) and DID strategy (recipients and non-recipients with a GPA between 1.0 and 2.0).

5. Empirical Strategy

Because all students in our sample face performance standards by the end of their first year, and about 54 percent of our sample receives financial aid in their first year, we can compare the effects of introducing performance standards for aided and unaided students. The data suggest two possible approaches: regression discontinuity (RD) analysis for students just above and below the GPA threshold to remain in good standing, and/or a difference-in-difference (DID) analysis comparing students above and below this threshold for aid recipients versus non-recipients.

Regression Discontinuity

The only assumption required is that the underlying relationship between first-year GPA and the outcome of interest (in the absence of the performance standard) is continuous through the threshold. Following Imbens and Lemieux (2008), we use a local linear specification with the bandwidth restricted to a narrow range around the 2.0 threshold. We focus on a bandwidth of 0.5, guided both by graphical plots as well as by the concern that other school policies may come into play below 1.5 or above 2.5. Still, we test for sensitivity to bandwidth selection by using the preferred bandwidth (0.5), half this bandwidth (0.25), and twice this bandwidth (1.0). For the 1.0 bandwidth, we use a more flexible local quadratic specification (though it makes little difference if we stick with a local linear model).

[10] As noted above, our analysis focuses on the GPA criterion, as it is highly correlated with credit completion percentage in our sample ($\rho = 0.79$) but much more continuously distributed. Thus, some students above 2.0 may still have received a warning. Students below the threshold should have received a warning, though there is always the possibility of some noise in the GPA calculation.

Table 1: Descriptive Statistics, State CC Sample (2004–2010 First-Time Full-Time Entrants)

	Full Sample		Near Threshold		Below Threshold	
Variable	Aided	No Aid	Aided	No Aid	Aided	No Aid
Background variables						
Age (Years)	21.0	19.9	19.9	19.1	20.1	19.0
Female (%)	59%	46%	56%	42%	55%	40%
White (%)	60%	73%	62%	74%	55%	72%
Black (%)	29%	11%	29%	11%	35%	12%
Hispanic (%)	5%	7%	6%	7%	5%	8%
Avg. Total Aid Yr 1 ($)	$4,083	$0	$4,158	$0	$4,111	$0
Pell Recipient (%)	70%	0%	71%	0%	73%	0%
Avg. Pell Amt. ($)	$2,385	$0	$2,491	$0	$2,530	$0
Cum. GPA<2.0, Yr 1	29%	32%	45%	45%	100%	100%
Comp. < 67% creds, Yr 1	37%	36%	41%	35%	72%	67%
Failed SAP, Yr 1	42%	42%	61%	59%	100%	100%
Ever failed/wth, Yr 1	69%	68%	91%	90%	98%	98%
Credits attempted, Yr 1	27.9	26.5	29.4	28.1	26.8	26.1
Credits earned, Yr 1	20.6	19.7	20.8	20.5	14.9	15.3
Took placement test	76%	71%	76%	73%	79%	75%
Needs remed (predicted)	*74%*	*64%*	*74%*	*63%*	*78%*	*66%*
Ever dual-enrolled	21%	14%	27%	18%	21%	15%
Intent: Occ. Associate's	34%	33%	34%	32%	35%	33%
Intent: Occ. Cert.	16%	10%	14%	8%	14%	9%
Intent: Liberal arts AA/AS	50%	57%	52%	59%	51%	58%
Outcome variables						
Enrolled, Fall Year 2	65%	68%	68%	76%	53%	66%
Term GPA, Fall Year 2	2.25	2.25	1.95	1.99	1.60	1.63
Credits Attempted, Fall Y2	9.3	9.6	9.0	10.4	6.3	8.3
Credits Earned, Fall Y2	7.1	7.3	5.9	7.0	3.6	4.8
School-Year Earnings, Y2	$2,043	$2,048	$2,080	$1,972	$2,013	$2,025
Any Earnings, Y2	41%	40%	44%	41%	42%	42%
Enrolled, Fall Year 3	36%	39%	41%	49%	30%	42%
Term GPA, Fall Year 3	2.28	2.30	2.03	2.05	1.65	1.73
Total Credits Attempted, Y3	8.23	8.67	8.91	10.95	6.25	8.94
Total Credits Earned, Y3	6.37	6.65	6.40	7.77	4.24	6.05
School-Year Earnings, Y3	$2,557	$2,541	$2,548	$2,543	$2,448	$2,588
Any Earnings, Y3	41%	39%	43%	42%	42%	42%
Total Credits Attempted, Y2-Y3	23.9	24.9	24.2	28.6	16.8	22.6
Total Credits Earned, Y2-Y3	18.4	19.1	16.5	19.7	10.4	14.2
Cumulative GPA, end of Y3	2.32	2.31	2.02	2.04	1.62	1.66

Table 1: (cont.) Descriptive Statistics, State CC Sample (2004–2010 First-Time Full-Time Entrants)

	Full Sample		Near Threshold		Below Threshold	
Variable	Aided	No Aid	Aided	No Aid	Aided	No Aid
Earned Certificate, by Y3	10%	7%	6%	5%	2%	2%
Earned AA/AS, by Y3	17%	17%	7%	9%	2%	4%
Transferred to 4Yr, by Y3	22%	27%	14%	19%	10%	12%
Still enrolled, Spring Y3	31%	32%	35%	42%	25%	35%
Earnings, Y2-Y3	$7,111	$7,161	$7,183	$6,997	$6,874	$7,178
Any Earnings, Y2-Y3	57%	53%	59%	55%	58%	57%
Sample size	60,482	52,141	5,111	4,699	8,716	8,008

Note. Source is author's calculations using restricted SCCS administrative data, 2004–2010 first time fall entrants who initially enrolled full-time. "Needs remediation" is predicted based on student scoring below typical remedial cutoff scores in any of three possible tests. These may not correspond to actual cutoffs in use for a given school/cohort. Percentages computed only for those with at least one test score.

The basic model, which we run on the sample restricted to aid recipients within the given bandwidth, takes the form:

$$Y_i = B_0 + \beta_1 (Below_i) + \beta_2 (GPADistance_i * Below_i) + \beta_3 (GPADistance_i * Above_i)$$
$$+ CollegeFE + CohortFE + \beta_n X_i + \varepsilon_I \quad (10)$$

where Y_i represents the outcome for student *i*, and β_1 is the estimate of the effect of falling below the SAP cutoff on the outcome. *CollegeFE* is a vector of institutional fixed effects (entered as a set of dummy variables indicating the institution initially attended, with one institution excluded), important because the financial aid officers responsible for enforcing performance standards are nested within institutions. *CohortFE* is a vector of cohort fixed effects, a necessary inclusion because of potential changes in the student population over time. X_i represents a vector of individual-level covariates including race, gender, age at initial enrollment, whether students were exempt from placement testing in reading and math (indicators of prior achievement), placement test scores for those who were not exempt, whether the student was predicted to be assigned to remedial coursework, whether the student had previously enrolled as a high school student, and dummies for the student's degree intent at entry (occupational associate's degree, occupational certificate, or academic associate's degree).[11] We also include additional controls to capture elements of the student's first year experience, including total credits attempted in the

[11] Note we do not include controls for family income or students' dependency status because these measures are missing for the 45 percent of students who did not file a FAFSA, including 75 percent of Pell non-recipients.

first year, whether the student worked for pay, and how much the student earned during the school year.[12] We test the sensitivity of our results to models with and without covariates.

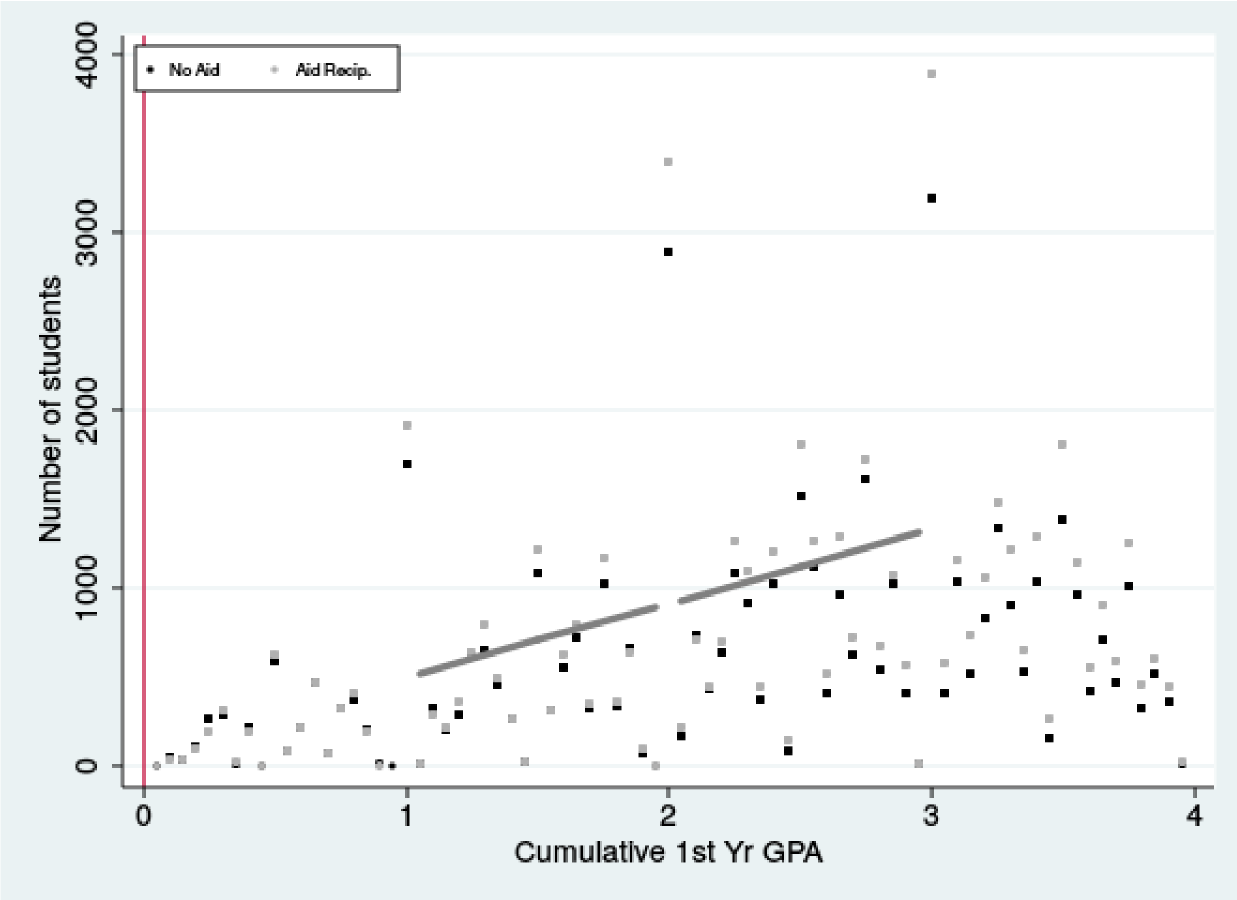
Grade Heaping

Using GPAs as the running variable in an RD introduces an additional challenge. Given the nature of the state's grading system—in which only whole letter grades are awarded—we find "heaping" in whole number GPAs across the distribution, including at our cutoff value of 2.0. Moreover, the fewer credits a student has attempted, all else equal, the more likely they are to have a whole-number GPA. This problem, however, is surmountable.[13] Excluding students with precisely a 2.0, a McCrary test (2008) indicates the distribution of cumulative GPAs is continuous around the threshold. This suggests that the observed heaping is due to grading policy and not to students precisely manipulating whether they fall above or below the threshold. Thus, following the recommendations of Barreca, Lindo, and Waddell (2011; see also Barreca, Guldi, Lindo, & Waddell, 2011), we rely on "donut-RD" estimates, dropping observations with precisely a 2.0 GPA. Figure 3 shows the distribution of first-year GPAs for aid recipients and non-recipients before removing whole number GPAs.

[12] These additional first-year controls make virtually no difference to the point estimates. Results available on request.

[13] Lindo, Sanders, and Oreopoulos (2010), Schudde and Scott-Clayton (2016), and Casey et al. (2015) all find similar patterns in their samples and implement similar donut-RD designs.

Figure 3: Distribution of GPA Before Eliminating Heaping at GPA = 2.0



Note: Fitted lines are shown for aid recipients above and below the cutoff after disregarding GPA = 2.0.

RD-Difference-in-Difference

The RD estimates capture the total effect of performance standards for aid recipients, including general standards at the institution as well as the effects of SAP policy specifically. To directly test whether the estimated effects of performance standards are larger for aid recipients, we run the following pooled regression:

$$Y_i = B_0 + \beta_1 (Below_i * Aid_i) + \beta_2 (Below_i) + \beta_3 (Aid_i) + \beta_4(GPADistance_i * Below_i * Aid_i)$$
$$+ \beta_5 (GPADistance_i * Above_i * Aid_i) + \beta_6(GPADistance_i * Below_i * NoAid_i) + \beta_7$$
$$(GPADistance_i * Above_i * NoAid_i) + CollegeFE + CohortFE + \beta_n X_i + \varepsilon_I \qquad (11)$$

The β_1 in this regression provides an estimate of the difference in the two RD estimates; we will refer to this as the RD-DID estimate. A drawback of this approach is that it is somewhat weakly powered.

Difference-in-Difference

The conceptual model suggests that encouragement effects should be strongest for those nearer to this threshold, while discouragement effects should be larger for individuals further below the threshold. Unfortunately, the RD is ill-equipped to test this important implication, because the RD estimates effects only for those right at the cutoff.[14] The difference-in-difference allows us to examine the effect of performance standards for Pell recipients, over and above the effects for non-recipients, for a wider range of students affected by the policy. It also provides much greater power to detect effects than either the RD or the RD-DID. The cost of obtaining this broader estimate is that we must make stronger assumptions about the relationship between first-year GPA and subsequent outcomes, namely by assuming that the relationship between *GPADistance* and potential outcomes is the same for aided and non-aided recipients. We still allow for a very flexible relationship between GPA and potential outcomes by replacing the *Below* and *GPADistance* interactions from equation (11) with a set of fixed effects for GPA bins (with width of 0.05):

$$Y_i = B_0 + \beta_1 (Below_i * Aid_i) + \beta_2 (GPABin05_i) + \beta_3 (Aid_i) + CollegeFE + CohortFE + \beta_n X_i + \varepsilon_I \qquad (12)$$

For the difference-in-difference, we continue to limit the bandwidth above the cutoff to +0.5 GPA points; however, we vary the bandwidth below the cutoff from -0.15 to -1.0. This allows us to test our hypothesis that discouragement effects will be bigger in the DID relative to the RD-DID as we include students further below the threshold, while encouragement effects may be smaller.

6. Main Results

Graphical Analysis and Covariate Checks

Figures 4–7 show average outcomes by first-year GPA in bins of 0.05 and with the size of the circles reflecting numbers of observations. In several of these graphs, the outcomes of aid recipients versus non-recipients appears to be more similar above the cutoff than below the cutoff, although it is difficult to discern sharp discontinuities visually. For example, reenrollment rates and credits attempted in Year 2 (Figure 4) and Year 3 (Figure 5) separately, as well as

[14] Lindo et al. (2010) examine subgroups by HSGPA to get at this question, but are necessarily limited to the variation in HSGPA that exists for students around the specified college GPA performance threshold. Given the generally strong correlation between high school and college GPAs, such an analysis may be particularly susceptible to issues of measurement error.

measured cumulatively after 3 years (Figure 6), appear to fall more for aided students below the cutoff than for unaided students. In the bottom right panel of Figure 6 (showing whether students were still enrolled at the end of the follow-up), a clear negative discontinuity is evident for aided students but not unaided students.

Figure 4. Fall Year 2 Outcomes

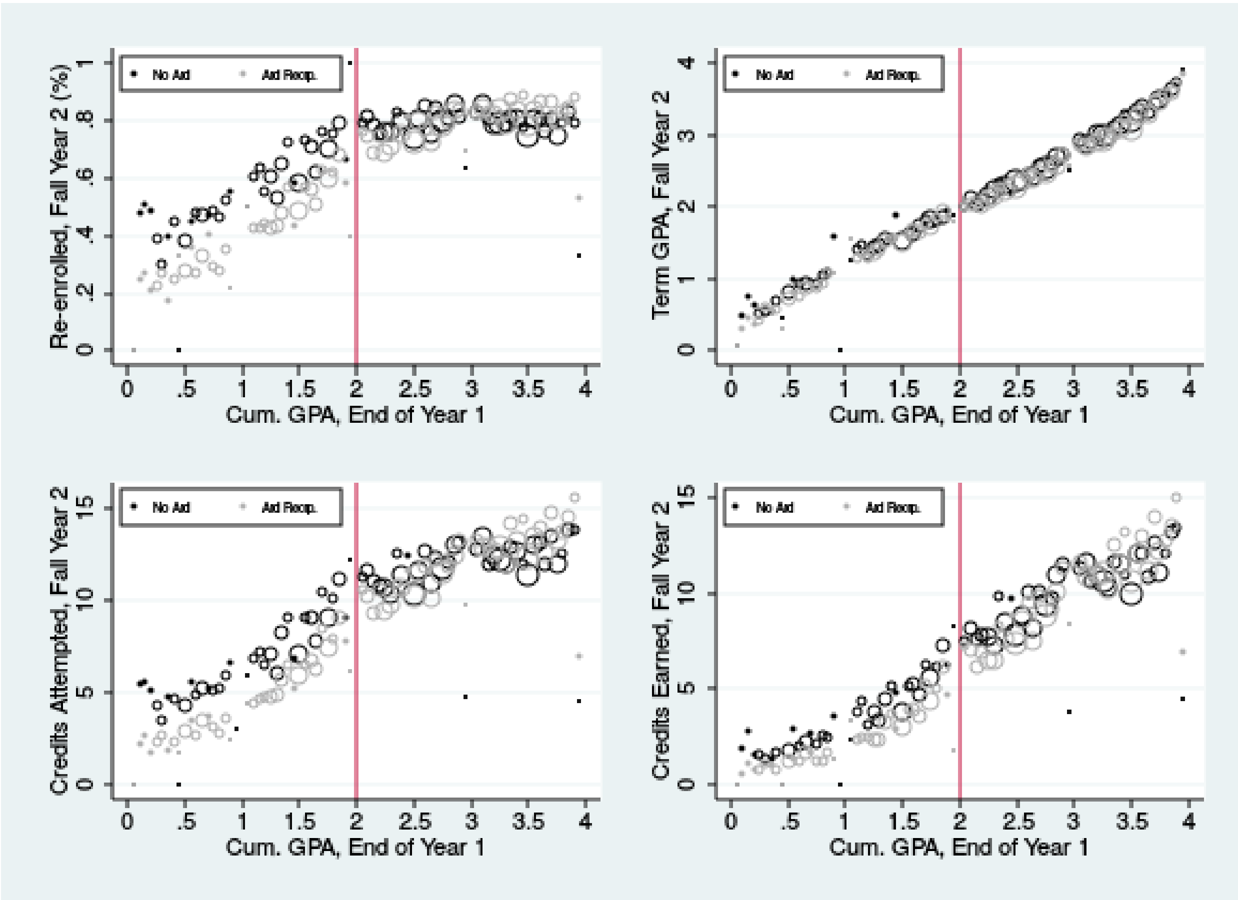


Figure 5. Year 3 Outcomes

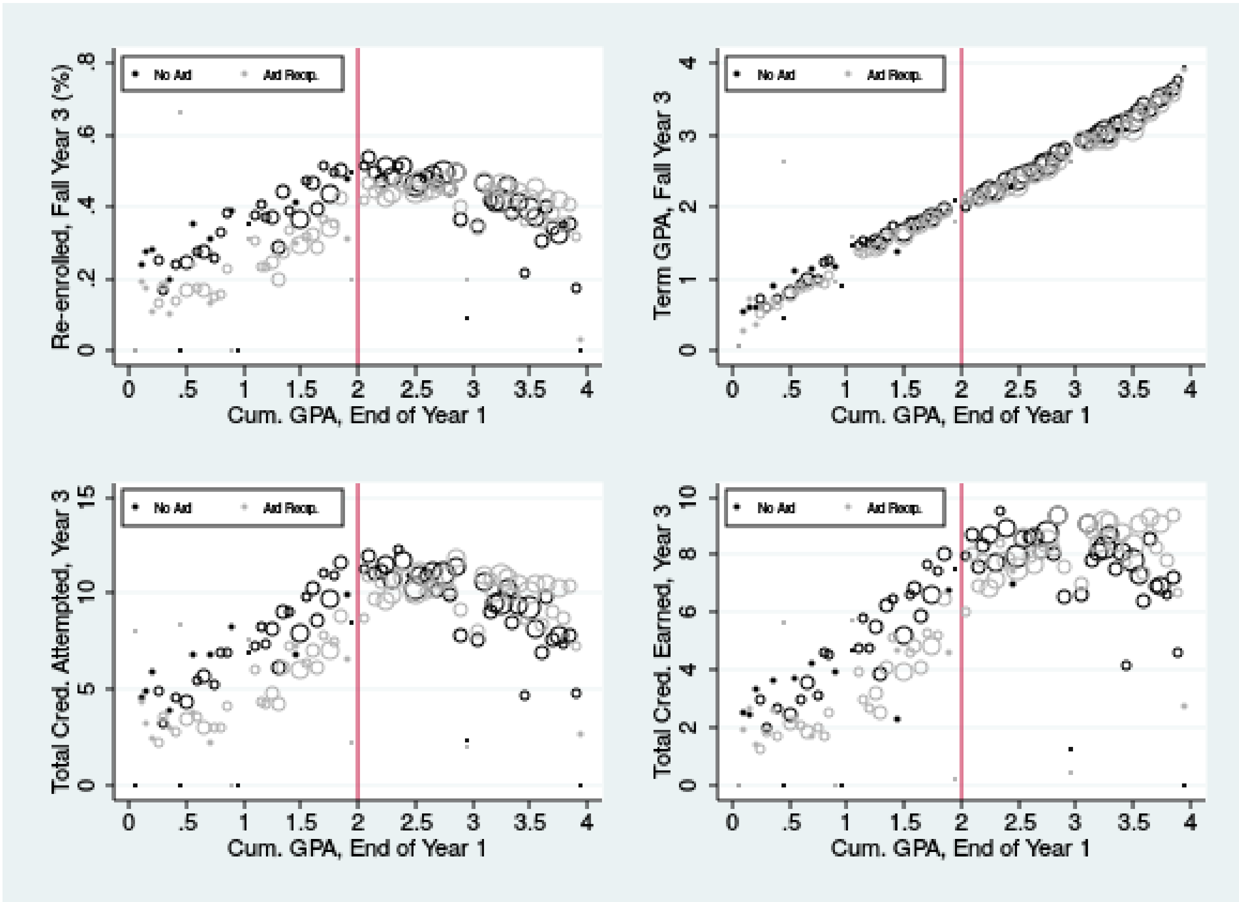
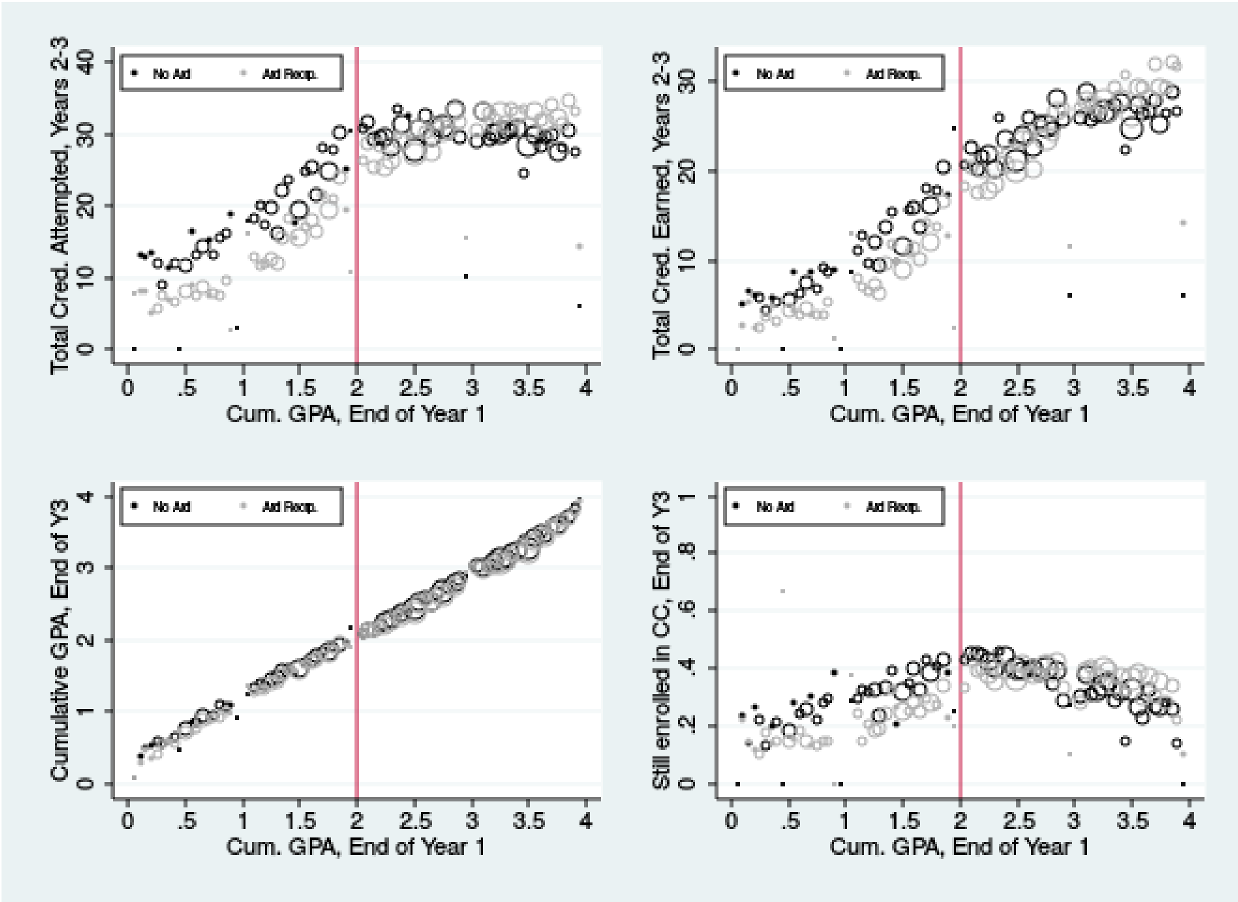


Figure 6. Cumulative Outcomes After 3 Years



In Table 2, we test for significant differences in pretreatment observable covariates under each of our main estimation strategies. The three columns of this table present "impact" estimates obtained by running equations (10), (11), and (12) above with the relevant background characteristic as the dependent variable. While we find no differences for most covariates, we find small positive difference in placement test scores for treated students in each specification. However, only about two-thirds of individuals have these placement test scores and the differences are small in magnitude (0.1–0.2 of a standard deviation). Gender, race, and ever dual-enrolled also emerge as significant in at least one specification. Importantly, despite these small differences in covariates, we will show that it makes virtually no difference to our point estimates whether or not they are included as controls.

Table 2: Covariate Balance Checks, Key Specifications

	RD (+/- 0.5)		RD-DID (+/- 0.5)		DID (+/- 0.5)	
Outcome	Coef. (S.E)		Coef. (S.E)		Coef. (S.E)	
Age	0.15 (0.20)		0.24 (0.24)		0.02 (0.09)	
Female	0.06 (0.02)	**	0.02 (0.03)		0.01 (0.01)	
White	0.02 (0.02)		0.04 (0.03)		-0.01 (0.01)	
Black	0.00 (0.02)		-0.01 (0.03)		0.02 (0.01)	**
Hispanic	-0.02 (0.01)	*	-0.01 (0.02)		0.00 (0.01)	
Missing race	0.00 (0.01)		-0.01 (0.01)		0.00 (0.00)	
Took reading test	0.01 (0.02)		0.00 (0.03)		-0.01 (0.01)	
Took writing test	0.01 (0.02)		0.00 (0.03)		0.00 (0.01)	
Took math test	0.01 (0.02)		0.00 (0.03)		0.00 (0.01)	
Reading score	*1.40 (0.69)*	**	*1.14 (0.96)*		*0.88 (0.37)*	**
Writing score	*3.95 (1.41)*	***	*2.71 (1.93)*		*0.30 (0.74)*	
Math score	*2.11 (0.98)*	**	*3.42 (1.53)*	**	*1.13 (0.57)*	**
Predicted to need remediation	-0.05 (0.02)	**	-0.04 (0.03)		-0.01 (0.01)	
Ever dual enrolled	-0.02 (0.02)		-0.04 (0.03)		-0.02 (0.01)	**
Intent: Occ AA/AS	0.01 (0.02)		0.01 (0.03)		0.02 (0.01)	
Intent: Occ certif.	-0.01 (0.02)		0.00 (0.02)		-0.01 (0.01)	
Credits attempted, Yr 1	-0.06 (0.43)		0.07 (0.59)		-0.11 (0.20)	
Any earnings, Yr 1	-0.01 (0.02)		-0.01 (0.03)		0.00 (0.01)	
Earnings, Yr 1	-$112 (116)		-$18 (167)		$38 (63)	
Sample size	13,506		25,557		25,557	

Note. Source is author's calculations using restricted SCCS administrative data, 2004–2010 first time fall entrants who initially enrolled full-time. Test score rows are italicized because they are calculated only for those students who have test scores available (approximately 75 percent took at least one test).

*** $p < .01$. ** $p < .05$. * $p < .1$

RD and RD-DID Results

Table 3 provides the results from the RD and RD-DID specifications for several aspects of student behavior that our model suggests should be affected, measured separately during the fall of Year 2 (the first warning period) and Year 3 (the enforcement period when individuals could be actually prohibited from reenrolling or receiving aid) as well as cumulatively. The first column of the table shows our preferred RD specification, while the subsequent rows show alternative specifications. The final column shows the RD-DID results.

We first examine reenrollment rates, term GPAs, and credits in the fall of Year 2. Note that for dropouts, term GPAs are imputed to the last known cumulative GPA (this ensures that impacts only come from those who reenroll, without introducing attrition bias). For aid recipients near the cutoff, failing SAP appears to have little effect on reenrollment decisions, or on continuous measures of credits attempted and completed (coefficients are consistently negative but generally very small in magnitude). However, we do find a significant increase of 0.07 GPA points ($p = .06$) in term GPA in the fall of the second year. Because this increase can only come from the 68 percent of students who reenroll, this implies a 0.10 GPA improvement among students who reenroll. This pattern of findings is quite stable across the RD specifications, though more pronounced in the narrow-bandwidth estimation. Notably, none of the Year 2 estimates are statistically significant in the RD-DID, although the signs are generally consistent.

However, by the fall of Year 3, any short-term GPA effects have washed out and more negative effects begin to appear. Most notably, we find a significant reduction of about 1.5 credits attempted (about a 15 percent reduction) and about 0.8 credits completed (a 12 percent reduction). This is consistent with some students facing actual loss of financial aid during this year. The negative effects grow over this year to a large, highly significant 8 percentage point reduction in the likelihood of still being enrolled in the spring of Year 3, consistent across all specifications. Cumulatively, aid recipients just below the GPA cutoff in Year 1 attempt 2.2 fewer credit and complete 1.4 fewer credits over three years than their counterparts just above the cutoff. We also find reductions of 2–3 percentage points in both certificate and associate's degree completion, though statistical significance varies across specification.

Finally, we examine school year earnings (measured in calendar Q4 and Q1 corresponding to the relevant academic year, with the cumulative measure also including Q2 and Q3 between the second and third academic year) as a measure of the possible opportunity cost of enrollment (i.e., do students who drop out go back to working instead?). Effects here are positive in the preferred RD specification but not statistically significant in any specification, and even the sign varies across specification. The standard errors are quite large (often equivalent to about 10 percent of mean earnings), and thus we are unable to conclude much about students' labor supply from these results.

Table 3: RD-Estimated Effects of Failing GPA Performance Standard at End of Year 1, Aid Recipients Only

	Preferred Bandwidth (+/- 0.5)		Alternate BW (with covars.)		RD-DID (0.5 bw)
	With Cov.	No Cov.	RD-0.25	RD-1.0	with covars
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Outcome	Coef. (S.E)	Coef. (S.E)	Coef. (S.E)	Coef. (S.E)	Coef. (S.E)
Enrolled, Fall Year 2	-0.01 (0.02)	-0.01 (0.02)	-0.08 (0.05) *	0.00 (0.01)	-0.03 (0.03)
Term GPA, Fall Year 2	0.07 (0.04) *	0.08 (0.04) **	0.11 (0.09)	0.05 (0.03) **	0.04 (0.06)
Credits Attempted, Fall Y2	-0.36 (0.34)	-0.33 (0.34)	-0.98 (0.74)	-0.07 (0.23)	-0.76 (0.48)
Credits Earned, Fall Y2	-0.12 (0.29)	-0.09 (0.29)	-0.52 (0.64)	0.09 (0.19)	-0.25 (0.42)
Enrolled, Fall Year 3	-0.03 (0.02)	-0.03 (0.02)	-0.02 (0.05)	-0.04 (0.01) ***	-0.02 (0.03)
Term GPA, Fall Year 3	0.00 (0.03)	0.01 (0.03)	-0.04 (0.08)	0.02 (0.02)	-0.03 (0.05)
Total Credits Att., Y3	-1.47 (0.50) ***	-1.42 (0.51) ***	-1.36 (1.09)	-1.43 (0.33) ***	-1.63 (0.75) **
Total Credits Earned, Y3	-0.85 (0.42) **	-0.80 (0.42) *	-0.65 (0.91)	-0.86 (0.28) ***	-0.83 (0.62)
Still enrolled, Spring Y3	-0.08 (0.02) ***	-0.08 (0.02) ***	-0.07 (0.05)	-0.06 (0.01) ***	-0.08 (0.03) **
Cumulative GPA, end of Y3	0.00 (0.02)	0.01 (0.02)	0.02 (0.04)	0.03 (0.01) **	-0.02 (0.03)
Total Credits Att., Y2-Y3	-2.16 (0.91) **	-2.07 (0.92) **	-3.25 (1.97)	-1.66 (0.60) ***	-2.72 (1.32) **
Total Credits Earned, Y2-Y3	-1.35 (0.77) *	-1.25 (0.78)	-1.86 (1.70)	-0.85 (0.51) *	-1.34 (1.13)
Earned Certificate, by Y3	-0.02 (0.01) *	-0.02 (0.01)	-0.04 (0.02) *	-0.01 (0.01)	-0.03 (0.01) **
Earned AA/AS, by Y3	-0.03 (0.01) ***	-0.03 (0.01) **	-0.03 (0.03)	0.01 (0.01)	-0.02 (0.02)
Transferred to 4Yr, by Y3	0.00 (0.02)	0.00 (0.02)	0.00 (0.04)	0.02 (0.01) **	0.02 (0.02)
School-Year Earnings, Y2	$83 (141)	$1 (165)	-$38 (296)	-$49 (095)	-$188 (202)
Ln(earnings), Y2	0.00 (0.08)	-0.04 (0.09)	-0.13 (0.17)	-0.02 (0.06)	-0.16 (0.12)
School-Year Earnings, Y3	$203 (188)	$112 (200)	-$398 (406)	-$21 (129)	-$63 (276)
Ln(earnings), Y2	0.13 (0.09)	0.09 (0.09)	-0.05 (0.19)	-0.01 (0.06)	-0.02 (0.12)
Earnings, Y2-Y3	$282 (440)	$12 (498)	-$620 (944)	-$134 (297)	-$555 (640)
Ln(earnings), Y2	0.04 (0.08)	-0.01 (0.09)	-0.12 (0.17)	-0.02 (0.06)	0.01 (0.12)
Sample size	13,506	13,506	5,111	24,673	25,557

Note. Source is authors' calculations using restricted SCCS administrative data, 2004–2010 first time fall entrants who initially enrolled full-time. Robust standard errors in parentheses. All specifications use local linear regression with observations at precisely 2.0 GPA dropped. Control variables include all variables listed in Table 2: age, gender, race dummies, placement test scores if available, placement test flags, flag for predicted remedial need, flag for ever dual enrolled, first year credits attempted, first year employment status, and first year earnings. For term GPA estimates, term GPA is imputed to the last known cumulative GPA (this ensures that any impacts on this measure come only from students who reenroll without introducing attrition bias).

*** $p < .01$. ** $p < .05$. * $p < .1$

Difference-in-Difference Results

As discussed in section 3, a drawback of both the RD and the RD-DID is that effects are estimated only for students near the 2.0 threshold. Yet our model clearly predicts heterogeneous effects by ability. We expect encouragement effects to be strongest for students just below the threshold, while we expect discouragement effects to grow as we move further down the GPA distribution. Our DID specification enables us to capture the effects of SAP policy for a wider range of students affected. Our results are shown in Table 4, which varies the range of observations included below the threshold while keeping the bandwidth above the cutoff fixed at 0.5. Results for specifications with no covariates are shown in Appendix Table A1, and are virtually identical.

Indeed, the pattern of impacts on enrollment in fall of Year 2 suggests that discouragement effects are larger for students further below the cutoff. The estimated 5 percentage point decline in reenrollment for the 0.15 bandwidth is statistically significant but grows to an 8 percentage point decline for the 1.0 bandwidth. Conversely, the 0.09 GPA point improvement for the 0.15 bandwidth falls to an insignificant 0.03 points for the 1.0 bandwidth. This pattern still shows, though much more weakly, in fall of Year 3. On the other hand, credits attempted and completed do not appear to vary much by bandwidth, either in the short or longer term. This may be because students near the margin may reduce their course loads in order to improve their GPAs. Overall, our preferred bandwidth of 0.5 (preferred because it captures a much wider range than the RD but avoids possible contamination from other policies for students below 1.5) suggests a decrease of 3.1 credits attempted and 1.1 credits completed after three years, similar to the RD-DID results.

As in the RD and RD-DID, we find 2–3 percentage point reductions in certificate completion in the DID specifications. But in contrast to the RD and RD-DID, the DID suggests null or even positive effects on associates degrees and large positive effects on likelihood of transferring to a four-year institution. While in theory the DID examines an estimand of greater policy interest (measuring average effects for a greater range of students below the GPA threshold), it is surprising that any degree completion/transfer impacts would become *more* positive when we include students further below the threshold. Indeed, the impact on transfer is *largest* in the DID when the bandwidth is expanded to +/- 1.0 (see Table 4, column 4). The outcome graphs in Figure 7 provide an additional reason for concern: these degree and transfer outcomes appear to virtually bottom out at a GPA of 1.5 (a phenomenon *not* observed for our other outcomes).[15] Aid recipients tend to have lower levels of degree completion/transfer than

[15] While the rate of "transfer" is non-zero even for students with near-zero GPAs, we suspect this is because there is some baseline noise in the definition of the outcome, combined with true transfers bottoming out around 1.5. The noise may be due to students who co-enrolled at a four-year in their first year or even prior to their first year; or, because it is based on NSC data it is possible that some of these students have transferred to for-profit institutions. We are working to create a cleaner measure that would capture only transfer to a public/non-profit four-year institution after the first year.

non-recipients regardless of GPA, but the difference narrows as we move down the GPA scale towards 1.5. The DID will attribute this narrowing difference to SAP policy, when in fact it may simply be attributable to floor effects in the outcome. For this reason, we are hesitant to take the DID effects on degree completion/transfer at face value, even while we find the DID estimation credible for other outcomes.

Finally, for student earnings we continue to see generally negative but insignificant estimates of a magnitude similar to what we found in the RD-DID. The notable exception is in the narrow bandwidth sample, for which we find very large and statistically significant reductions in earnings. Given that these enormous reductions do not show up in any other specification, are not visually discernable in Figure 7, and are in contrast with the prediction that labor supply should increase when students drop out, we prefer not to over-interpret this. Nonetheless, it is worth noting that the earnings coefficients are, at least, quite consistent in their negative sign across all DID specifications as well as most of the RD specifications. This is suggestive of another channel for earnings effects besides the school–work time allocation tradeoff initially hypothesized.

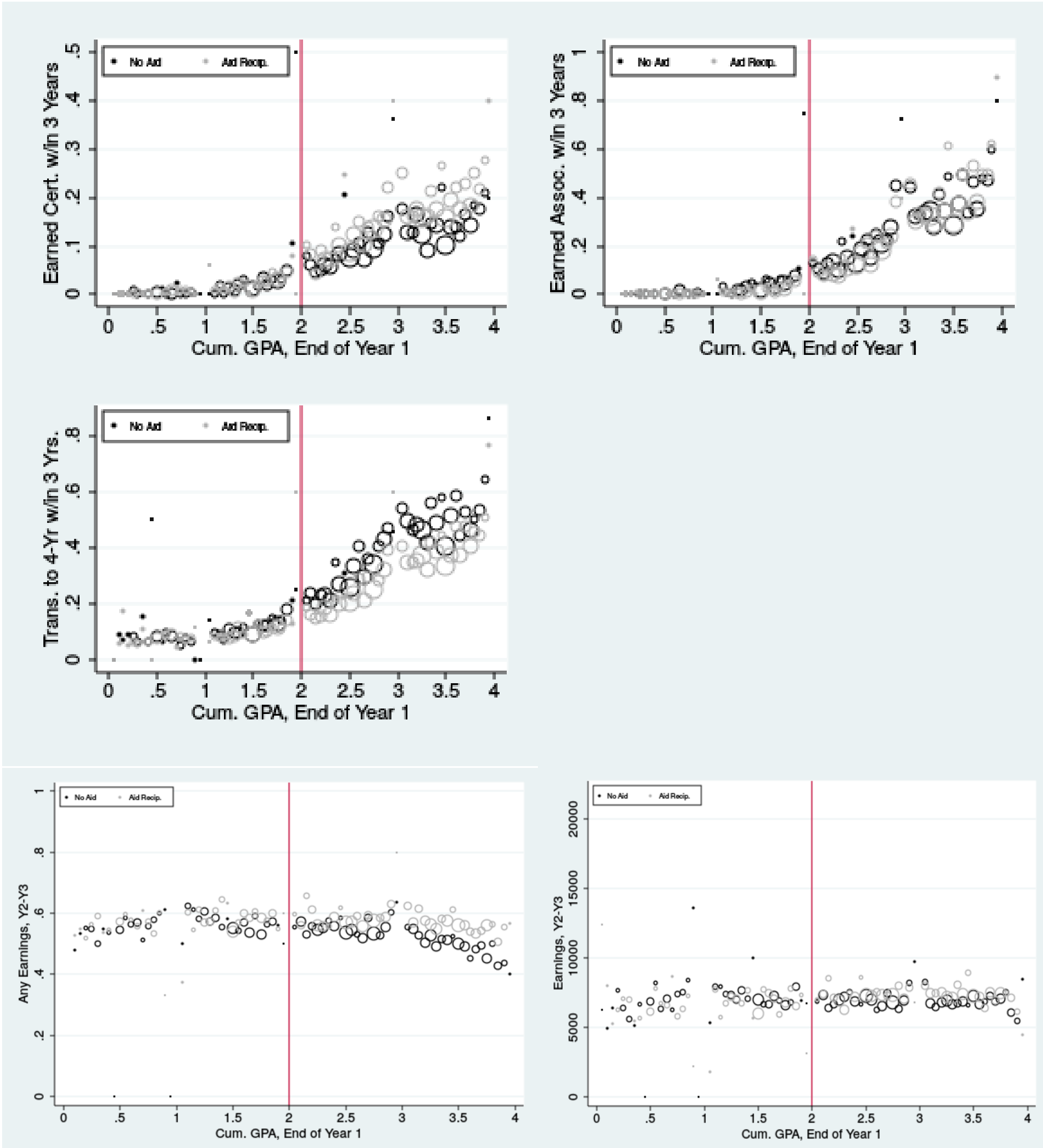
Table 4: DID Estimated Effects of Failing GPA Performance Standard at End of Year 1 (Above Versus Below for Aided Versus Unaided Students)

Outcome	Model 6 DID-0.15 Coef. (S.E)		Model 7 DID-0.25 Coef. (S.E)		Model 8 DID-0.5 Coef. (S.E)		Model 9 DID-1.0 Coef. (S.E)	
Enrolled, Fall Year 2	-0.05 (0.02)	**	-0.05 (0.02)	***	-0.06 (0.01)	***	-0.08 (0.01)	***
Term GPA, Fall Year 2	0.09 (0.05)	*	0.05 (0.03)		0.04 (0.02)	**	0.03 (0.02)	
Credits Attempted, Fall Y2	-1.12 (0.40)	***	-0.98 (0.25)	***	-0.91 (0.18)	***	-1.01 (0.16)	***
Credits Earned, Fall Y2	-0.28 (0.35)		-0.26 (0.21)		-0.16 (0.15)		-0.19 (0.14)	
School-Year Earnings, Y2	-$412 (162)	**	-$103 (104)		-$97 (077)		-$45 (069)	
Ln(earnings), Y2	*-0.23 (0.10)*	**	*-0.08 (0.06)*		*-0.08 (0.05)*	*	*-0.04 (0.04)*	
Enrolled, Fall Year 3	-0.03 (0.03)		-0.04 (0.02)	***	-0.05 (0.01)	***	-0.06 (0.01)	***
Term GPA, Fall Year 3	0.02 (0.04)		0.01 (0.03)		-0.01 (0.02)		-0.02 (0.02)	
Total Credits Attempted, Y3	-1.76 (0.61)	***	-1.79 (0.38)	***	-1.62 (0.28)	***	-1.58 (0.25)	***
Total Credits Earned, Y3	-0.68 (0.51)		-0.85 (0.32)	***	-0.82 (0.23)	***	-0.82 (0.21)	***
Still enrolled, end of Y3	-0.05 (0.03)	*	-0.06 (0.02)	***	-0.05 (0.01)	***	-0.05 (0.01)	***
Cumulative GPA, End of Y3	0.04 (0.02)	*	0.01 (0.01)		0.01 (0.01)		0.00 (0.01)	
Total Credits Attempted, Y2-Y3	-3.61 (1.08)	***	-3.37 (0.67)	***	-3.10 (0.49)	***	-3.17 (0.43)	***
Total Credits Earned, Y2-Y3	-1.34 (0.93)		-1.34 (0.57)	**	-1.13 (0.42)	***	-1.12 (0.37)	***
Earned Certificate, by Y3	-0.03 (0.01)	**	-0.02 (0.01)	***	-0.02 (0.01)	***	-0.02 (0.00)	***
Earned AA/AS, by Y3	0.00 (0.01)		0.00 (0.01)		0.01 (0.01)		0.01 (0.01)	**
Transferred to 4Yr, by Y3	0.02 (0.02)		0.04 (0.01)	***	0.04 (0.01)	***	0.05 (0.01)	***
School-Year Earnings, Y2	-$412 (162)	**	-$103 (104)		-$97 (077)		-$45 (069)	
Ln(earnings), Y2	*-0.23 (0.10)*	**	*-0.08 (0.06)*		*-0.08 (0.05)*	*	*-0.04 (0.04)*	
School-Year Earnings, Y3	-$515 (220)	**	-$186 (141)		-$127 (104)		-$167 (094)	*
Ln(earnings), Y3	*-0.16 (0.09)*	*	*-0.08 (0.06)*		*-0.01 (0.05)*		*-0.02 (0.04)*	
Earnings, Y2-Y3	-$1,332 (517)	***	-$394 (328)		-$329 (243)		-$332 (219)	
Ln(earnings), Y2-Y3	*-0.17 (0.09)*	*	*-0.06 (0.06)*		*-0.05 (0.04)*		*-0.02 (0.04)*	
Sample size	16,326		19,223		25,557		31,562	

Note. Source is authors' calculations using restricted SCCS administrative data, 2004–2010 first time fall entrants who initially enrolled full-time. Robust standard errors in parentheses. Coefficients shown are on the interaction term aided*below. Aid status is based on first year awards and includes all forms of aid. Range of data above the cutoff is held fixed across specifications at 0.5; range of data included below the threshold varies by model. All specifications include fixed effects for first-year GPA bin in increments of 0.05, with observations at precisely 2.0 GPA dropped. Control variables include all variables listed in Table 2: age, gender, race dummies, placement test scores if available, placement test flags, flag for predicted remedial need, flag for ever dual enrolled, first year credits attempted, first year employment status, and first year earnings. For term GPA estimates, term GPA is imputed to the last known cumulative GPA (this ensures that any impacts on this measure come only from students who reenroll without introducing attrition bias).

*** $p < .01$. ** $p < .05$. * $p < .1$

Figure 7. Degree Completion and Earnings Outcomes, End of Year 3



Note. Aid recipients are in gray; non-recipients are in black. "Earnings, Y2–Y3" includes zeros and covers six quarters beginning in the fall of Year 2 and ending in the spring of Year 3 (Q4–Q1–Q2–Q3–Q4–Q1).

7. Discussion

In this paper, we attempt to provide a conceptual framework for thinking about the role and consequences of imposing performance standards in the context of financial aid. The framework suggests that some minimum standard is desirable. Determining whether a standard is too high or too low will require weighing the value of encouragement effects for those who are motivated to work harder against the discouragement effects for those who are induced to dropout.

Consistent with the model and with prior research by Lindo, Sanders, and Oreopoulos (2010), we find behavioral effects in the expected directions that are particularly strong in the first term after a warning is issued, the fall of the second year. Also consistent with our theoretical model, student responses to performance standards appear to be larger for students receiving financial aid (seen by comparing the RD-DID specifications to the RD). Discouragement effects appear larger, and encouragement effects smaller, when we include students further below the GPA threshold (seen by comparing DID specifications of different bandwidths). In our preferred DID specification (Table 4, third column), aid recipients who fail the SAP grade standard are 6 percentage points less likely to reenroll in the second year, but second-year GPAs rise by 0.04 points compared with similar unaided students. These results are also consistent with Schudde and Scott-Clayton (2016), which applied a DID specification to examine SAP policy in a different state and found significant negative effects on reenrollment and positive (but small and insignificant) effects on GPAs in the second year.

Over the longer term, our results across all specifications suggest reductions of about three credits attempted and one credit earned after three years. This pattern suggests that students are discouraged from attempting credits they were unlikely to complete, and thus SAP policy may improve the efficiency of aid distributed. If we multiply the decline in credits attempted by an estimate of students' per-credit aid eligibility, the decline corresponds to a $440–$530 decline in estimated aid disbursed per student in the second and third years. For comparison, $440–$530 is four to five times the tuition cost of one credit, which was below $115 for this sample and timeframe. Moreover, this could be a conservative estimate of the cost savings since tuition itself is subsidized, and because even some of the students in our sample who reenroll after failing SAP may have done so without aid.

Still, our findings generate a number of dimensions for possible concern. First, we consistently find declines of about 2–3 percentage points on certificate completion for aided students who fail SAP. Because of the shorter length of these programs (often one year or even less), the second year may already be too late to recover if students have a below-standard GPA at the end of the first year. Recent estimates of the labor market payoff to certificates, using an individual fixed-effects approach, suggest an earnings gain of perhaps $1,400 annually for certificate completers (Di & Trimble, 2014; though it is not clear whether these gains might be bigger or smaller for students on the margin of failing SAP). If only 2–3 percent of the sample experiences this loss, it would take over a decade for the earnings losses to outweigh the financial aid savings on average. Still, some individuals are clearly worse off as a result: the discouragement effects of the policy mean that some students who could have earned a degree are dissuaded from reenrolling. The concentrated consequences they experience may outweigh the social benefit of reduced aid expenditures, which are dispersed across many. Moreover, it is possible that the students who are least likely to earn a degree are those that benefit the most from doing so (Brand & Xie, 2010).

More generally, it does not appear to be the case that students themselves receive much benefit from SAP policy. The short term improvements in GPA are not sustained over the long term; the isolated positive impact on transfer is sensitive to specification and follows a pattern that suggests it may be spurious. Regarding the negative effects on credits and enrollment, in theory, students with a low likelihood of completing the courses they attempt might benefit from leaving school sooner rather than later in order to devote more time to gaining experience in the labor market. However, we find little evidence of any positive effects on labor supply; indeed, most point estimates on earnings were negative.

Taken broadly, the pattern of effects here suggests that SAP policy is at least partly doing its job (at least from the perspective of a social planner who weights all students equally): minimizing unproductive reenrollments while providing some encouragement for students to perform better. This hardly implies that SAP policy is optimized, however. Our review of college catalogs, as well as anecdotal reports from college staff, suggests that many students may not learn about SAP until they lose aid. If true, this is a missed opportunity: if students are poorly informed it will mute the incentive effects of standards, and the longer it takes for students to realize they are failing, the harder it will be for them to get back above the GPA threshold.

From an equity stance, the implications of SAP policy are complex. Poor academic performance is widespread across student demographics. SAP policy targets undergraduates from America's most disadvantaged families (median family income among aid recipients in our sample is about $28,000). Students who are reliant on federal financial aid face the consequences of academic standards more quickly than students who can afford to pay for college out of pocket. A student with unlimited funds can, theoretically, continue to enroll in community college for as many iterations as necessary to attain the 2.0 cumulative GPA required for graduation. A student who relies on federal funding to cover tuition expenses ultimately receives fewer chances to "get it right." While SAP standards may help some students avoid overinvesting time, money, and energy into college schooling, it also may also prevent students from economically disadvantaged households from an equal chance at earning a diploma. Heterogeneous effects within the economically disadvantaged group may further exacerbate inequality: though we cannot examine it here, prior work by Barrow & Rouse (2013) suggests that students with children are less able than those without children to shift their time allocation toward academics in order to meet performance incentives.

Finally, an open question is how the effects of SAP may be different following a significant tightening of the standards in 2011 (too late for us to examine in our sample). Federal regulations now specify that SAP status must be evaluated at least once annually; only those institutions evaluating more than once per year can use a “warning” status (and then only for one term); and students who file a successful appeal may be placed on “probationary” status only for one term (Satisfactory Academic Progress, 2012; U.S. Department of Education, 2014). In effect, the new regulations mean that students who fail SAP cannot receive aid for more than one subsequent term without filing an appeal; even if the appeal is successful, students can only receive aid for one additional term unless they improve sufficiently to pass the SAP standard. Because of these changes, SAP policy is likely to affect more students more quickly than it has in the past.

These changes could be beneficial if students are encouraged to improve earlier in their college careers, but they could be detrimental if enforcement is so draconian that students do not have sufficient time to improve. Nor is it clear what would happen if standards were set at a higher level such as 2.5, above the GPA typically required for graduation, as was recently proposed by the Obama administration with respect to the President’s “free community college” proposal. What is certain is that SAP policy is not going away, and may affect even more students in future years—so the stakes are high to understand its impacts for both students and public coffers.

References

Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics, 1*(1), 136–63.

Altonji, J. G. (1993). The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics*, *11*(1), 48–83.

Bailey, M. J., & Dynarski, S. M. (2011). Inequality in postsecondary education. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life choices* (pp. 117–132). New York, NY: Russell Sage Foundation.

Barreca, A. I., Guldi, M., Lindo, J. M., & Waddell, G. R. (2011). Saving babies? Revisiting the effect of very low birth weight classification. *Quarterly Journal of Economics*, *126*(4), 2117–2123.

Barreca, A. I., Lindo, J. M., & Waddell, G. R. (2011). *Heaping-induced bias in regression-discontinuity designs* (NBER Working Paper No. 17408). Cambridge, MA: National Bureau of Economic Research.

Barrow, L., Richburg-Hayes, L., Rouse, C. E., & Brock, T. (2014). Paying for performance: The education impacts of a community college scholarship program for low-income adults. *Journal of Labor Economics, 32*(3), 563–599.

Barrow, L., & Rouse, C. E. (2013). *Financial incentives and educational investment: The impact of performance-based scholarships on student time use* (NBER Working Paper No. 19351). Cambridge, MA: National Bureau of Economic Research. [Revised version forthcoming in *Education Finance and Policy*].

Baum, S., & Payea, K. (with Kurose, C.). (2013). *Trends in student aid 2013*. Retrieved from the College Board website: http://trends.collegeboard.org/sites/default/files/student-aid-2013-full-report-140108.pdf

Belfield, C., Liu, V. Y. T., & Trimble, M. J. (2014). *The medium-term labor market returns to community college awards: Evidence from North Carolina* (A CAPSEE Working Paper). New York, NY: Center for the Analysis of Postsecondary Education and Employment.

Belley, P., & Lochner, L. (2007). The changing role of family income and ability in determining educational achievement. *Journal of Human Capital, 1*(1), 37–89.

Bénabou, R., & Tirole, J. (2000). *Self-confidence and social interactions* (NBER Working Paper No. 7585). Cambridge, MA: National Bureau of Economic Research.

Bénabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. *Quarterly Journal of Economic*s, *117*(3), 871–915.

Bennett, W., & Grothe, B. (1982). Implementation of an academic progress policy at a public urban university: A review after four years. *Journal of Student Financial Aid*, *12*(1), 33–39.

Brand, J. E., & Xie, Y. (2010). Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education. *American Sociological Review, 75*(2), 273–302.

Carruthers, C., & Özek, U. (2013). *Losing HOPE: Financial aid and the line between college and work* (Working Paper No. 91). Retrieved from Center for Analysis of Longitudinal Data in Education Research website: http://www.caldercenter.org/sites/default/files/wp91.pdf

Casey, M., Cline, J., Ost, B., & Qureshi, J. (2015, February). *Academic probation, student performance and strategic behavior*. Paper presented at the annual meeting for the Association for Education Finance and Policy, Washington, DC.

Cornwell, C., Mustard, D. B., & Sridhar, D. J. (2006). The enrollment effects of merit-based financial aid: Evidence from Georgia's HOPE program. *Journal of Labor Economics, 24*(4), 761–786.

Cornwell, C. M., Lee, K. H., & Mustard, D. B. (2005). Student responses to merit scholarship retention rules. *Journal of Human Resources*, *40*(4), 895–917.

Dynarski, S. M. (2008). Building the stock of college-educated labor. *Journal of Human Resources, XLIII*(3), 577–610.

Fischer, F. (1987). Graduation-contingent student aid. *Change* (Nov–Dec), 40–47.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635.

Jacobson, L., & Mokher, C. (2009). *Pathways to boosting the earnings of low-income students by increasing their educational attainment*. Washington, DC: Hudson Institute.

Jepsen, C., Troske, K., & Coomes, P. (2014). The labor-market returns to community college degrees, diplomas, and certificates. *Journal of Labor Economics, 32*(1), 95–121.

Lavecchia, A. M., Liu, H., & Oreopoulos, P. (2014). *Behavioral economics of education: Progress and possibilities* (NBER Working Paper No. 19306). Cambridge, MA: National Bureau of Economic Research.

Lindo, J. M., Sanders, N. J., & Oreopoulos, P. (2010). Ability, gender, and performance standards: Evidence from academic probation. *American Economic Journal: Applied Economics*, *2*(2), 95–117.

Manski, C. F. (1988, June). *Should we subsidize enrollment in or completion of postsecondary schooling?* Paper presented at the Conference on Outcome-Based Policy Options for Vocational Education, Washington, DC.

Manski, C. F. (1989). Schooling as experimentation: A reappraisal of the postsecondary dropout phenomenon. *Economics of Education Review*, *8*(4), 305–312.

Manski, C. F., & Wise, D. (1983). *College Choice in America*. Cambridge, MA: Harvard University Press.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698–714.

Patel, R., & Valenzuela, I. (with McDermott, D.). (2013). *Moving forward: Early findings from the Performance-Based Scholarship Demonstration in Arizona*. New York, NY: MDRC.

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, NY: MDRC.

Satisfactory Academic Progress, 34 C.F.R. § 668.34 (2012).

Schudde, L., & Scott-Clayton, J. (2016). Pell grants as performance-based scholarships? An examination of satisfactory academic progress requirements in the nation's largest need-based aid program. *Research in Higher Education*. doi:10.1007/s11162-016-9413-3

Scott-Clayton, J. (2011). On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources*, *46*(3), 614–646.

Scott-Clayton, J. (2013). Information constraints and financial aid policy. In D. E. Heller & C. Callender (Eds.), *Student financing of higher education: A comparative perspective* (pp. 75–97). New York, NY: Routledge.

Stinebrickner, T., & Stinebrickner, R. (2012). Learning about academic ability and the college dropout decision. *Journal of Labor Economics, 30*(4), 707–748.

U.S. Department of Education, Federal Student Aid. (2014). School-determined requirements. In *Student eligibility [2014-2015 federal student aid handbook]* (Vol. 1). Retrieved from https://ifap.ed.gov/fsahandbook/attachments/1415Vol1Ch1.pdf

Wiswall, M., & Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *Review of Economic Studies, 82*(2), 791–824.

Xu, D., & Trimble, M. J. (2014). *What about certificates? Evidence on the labor market returns to non-degree community college awards in two states.* New York, NY: Columbia University, Teachers College, Community College Research Center.

Appendix

Table A1. DID Estimated Effects of Failing GPA Performance Standard At End of Year 1, No Covariates

Outcome	DID-0.15 No Cov. Coef. (S.E)		DID-0.25 No Cov. Coef. (S.E)		DID-0.5 No Cov. Coef. (S.E)		DID-1.0 No Cov. Coef. (S.E)	
Enrolled, Fall Year 2	-0.05 (0.02)	**	-0.05 (0.02)	***	-0.06 (0.01)	***	-0.08 (0.01)	***
Term GPA, Fall Year 2	0.10 (0.05)	**	0.05 (0.03)	*	0.04 (0.02)	**	0.03 (0.02)	
Credits Attempted, Fall Y2	-1.15 (0.41)	***	-0.98 (0.25)	***	-0.89 (0.19)	***	-1.03 (0.16)	***
Credits Earned, Fall Y2	-0.27 (0.35)		-0.25 (0.21)		-0.15 (0.16)		-0.21 (0.14)	
Enrolled, Fall Year 3	-0.03 (0.03)		-0.04 (0.02)	***	-0.05 (0.01)	***	-0.06 (0.01)	***
Term GPA, Fall Year 3	0.03 (0.04)		0.02 (0.03)		-0.01 (0.02)		-0.02 (0.02)	
Total Credits Attempted, Y3	-1.73 (0.61)	***	-1.76 (0.38)	***	-1.63 (0.28)	***	-1.61 (0.25)	***
Total Credits Earned, Y3	-0.64 (0.51)		-0.82 (0.32)	**	-0.84 (0.23)	***	-0.85 (0.21)	***
Still enrolled, end of Y3	-0.05 (0.03)	*	-0.06 (0.02)	***	-0.05 (0.01)	***	-0.05 (0.01)	***
Cumulative GPA, End of Y3	0.04 (0.02)	**	0.02 (0.01)		0.01 (0.01)		0.00 (0.01)	
Total Credits Attempted, Y2-	-3.59 (1.10)	***	-3.31 (0.69)	***	-3.07 (0.50)	***	-3.24 (0.45)	***
Total Credits Earned, Y2-Y3	-1.25 (0.95)		-1.25 (0.58)	**	-1.12 (0.43)	***	-1.18 (0.38)	***
Earned Certificate, by Y3	-0.03 (0.01)	**	-0.02 (0.01)	***	-0.02 (0.01)	***	-0.02 (0.00)	***
Earned AA/AS, by Y3	0.00 (0.01)		0.00 (0.01)		0.01 (0.01)		0.01 (0.01)	*
Transferred to 4Yr, by Y3	0.03 (0.02)		0.04 (0.01)	***	0.05 (0.01)	***	0.05 (0.01)	***
School-Year Earnings, Y2	-$384 (195)	**	-$38 (121)		-$77 (091)		-$45 (082)	
Ln(earnings), Y2	*-0.18 (0.10)*	*	*-0.06 (0.07)*		*-0.06 (0.05)*		-0.03 (0.04)	
School-Year Earnings, Y3	-$505 (235)	**	-$145 (150)		-$122 (112)		-$178 (101)	*
Ln(earnings), Y3	*-0.13 (0.10)*		*-0.08 (0.06)*		*-0.01 (0.05)*		-0.02 (0.04)	
Earnings, Y2-Y3	-$1,272 (589)	**	-$236 (371)		-$292 (277)		-$351 (251)	
Ln(earnings), Y2-Y3	*-0.13 (0.10)*		*-0.04 (0.06)*		*-0.04 (0.05)*		-0.02 (0.04)	
Sample size	16,326		19,223		25,557		31,562	