

February – 2019

# On the Efficacy of Open Educational Resources: Parametric and Nonparametric Analyses of a University Calculus Class



Huimei Delgado<sup>1</sup>, Michael S. Delgado<sup>1</sup>, and John Hilton III<sup>2</sup>

<sup>1</sup>Purdue University, <sup>2</sup>Brigham Young University

## Abstract

Open educational resources (OER), which are free and openly licensed educational materials, have been a widely discussed topic in response to high textbook costs, the need for more pedagogical flexibility, and inequality in access to educational materials. In this study we examine the efficacy of OER through a quantitative analysis of the impact of OER on student final exam performance in a large calculus course. Our dataset affords us a relatively large sample size, allows us to classify students in both treatment and control groups, and includes a variety of covariates that allow us to control for multiple correlated factors. We estimate causal treatment effects using several econometric approaches. Our study adds the following insights into the research on OER efficacy: (i) OER materials do not, in general, lead to any significant change in student final exam performance; and (ii) OER materials have a significantly positive impact on both international students and Pell Grant eligible students.

*Keywords:* open education resources, teaching calculus, treatment effects, OER

## Introduction

Open Educational Resources (OER) are “teaching, learning, and research resources that reside in the public domain or have been released under an intellectual property license that permits their free use and repurposing by others” (Hewlett, 2017, para. 7), and recent years have seen a surge in interest in open education and the use of OER. Among many potential virtues that we will describe, OER have been touted as a solution to the problem of high textbook costs – the general hypothesis is that with access to educational materials at no cost, students will save money, and may perform better. For instance, a survey of 22,906 post-secondary students in Florida found that 67% of students reported that they did not purchase a required textbook because of its high cost. As well, this same survey found that because of the lack of access to learning materials, 37.6% had earned a poor grade and 19.8% had failed a course. Nearly half the students surveyed said they occasionally or frequently take fewer courses because of the cost of textbooks, and one-quarter of students drop courses for the same reason (Florida Virtual Campus, 2016). Moreover, if, as Florida Virtual Campus (2016) indicates, the cost of textbooks leads students to withdraw from courses, we might expect to see fewer withdrawals if costly learning materials are substituted with OER.

In this paper, we rigorously analyze the efficacy of OER materials in a large college calculus class, and provide insight into the different ways in which OER have a positive or negative impact on student learning outcomes. We begin in the following section with a review of the literature surrounding research on OER that examines its quality and efficacy.

## Review of Literature

In terms of trusted quality, those who have used OER generally believe that OER are high quality. Watson, Domizi, and Clouser (2017) surveyed 1,299 students at the University of Georgia who used the OpenStax biology textbook. Students were directly asked to rate the quality of the OpenStax textbook relative to other textbooks they had used. The majority of students (64%) reported that the OpenStax book had approximately the same quality as traditional books, while 22% said it had higher quality. Jhangiani, Pitt, Hendricks, Key, and Lalonde (2016) examined awareness, usage, outcomes, and perceptions of OER among a survey of 78 British Columbia post-secondary faculty. Of the respondents, 77% had used OER and most respondents rated OER quality as comparable or superior to that of traditional materials. The California OER Council (2016) surveyed 351 college students and found that of those who had used OER, 42% said that OER was better than traditional materials, 39% said they were about the same, 11% rated OER as worse, and 8% declined to answer. Hilton (2016) reviews an additional nine published studies of OER perceptions and found that across 2,366 students and 2,144 surveyed or interviewed, a strong majority perceive that OER have the same or higher quality as traditional resources.

Although a large number of faculty and students have been surveyed in a variety of contexts, perception of quality is not the same as proven efficacy. To date, only a handful of empirical studies have been published that examine the impact of OER adoption on student learning. Through his review, Hilton (2016) draws a general consensus across the literature that OER students typically perform just as well as students using commercial materials. At the same time he acknowledges that many of the published studies assessing OER efficacy are weak. For example, Lovett, Meyer, and Thille (2008) studied those who used OER across multiple semesters, but their total number of OER adopters was only  $n = 66$ , a significant limitation.

In some cases, there are serious methodological problems with OER efficacy studies. For example, Feldstein et al. (2012) compares courses that use OER with *different* courses that are not using OER. Given that the comparison courses are different, it is unclear whether any measured differences across OER and traditional courses can be attributed to the use of OER. Hilton and Laman (2012) has a different weakness; they compared the grades, withdrawal rates, and pass rates of students who took a psychology course in Spring 2011 (traditional textbooks) with those who took the same course in Fall 2011 (OER); yet, they did not account for any difference between the two student populations. Moreover, changes were made in the course learning outcomes and final exam during the time period of the study, significantly weakening any ability to link the change in student outcomes with the introduction of OER. Pawlyshyn, Braddlee, Casper, and Miller (2013) reported dramatic improvement when OER was adopted; however, their results are confounded by the fact that simultaneous with the new curriculum materials was a new form of instruction (flipped classrooms). Bowen, Chingos, Lack, and Nygren (2014) did control for student differences by randomly selected treatment and control groups and used multiple characteristics to determine that the two groups were equivalent. However, their design also introduced elements of blended learning for students who used OER, possibly confounding their results that both treatment and control performed equally well. Other recent studies such as Grewe and Davis (2017) and Ozdemir and Hendricks (2017) find similar or slightly better student outcomes when using OER, but lack rigorous controls.

Thus, a serious issue with OER research to date is that more than half of the efficacy studies conducted do not make any attempt to control for student variables that could be influencing the difference in their performance. This critique was raised by Gurung (2017), who argues that “The results highlight limitations of current attempts to assess learning in psychology and underline the need for robust comparisons of a wider variety of OER, with a focus on lower ability students” (p. 233).

Colvard, Watson, and Park (2018) examined course-level faculty adoption of OER at the University of Georgia, by evaluating eight undergraduate courses that switched from commercial textbooks to OpenStax (OER) textbooks. They only included sections of these courses where instructors had taught with both textbook versions to control for teacher effect. This was the first published study to examine the relationship between OER and Pell Grant recipients; researchers found a 6.90% GPA increase for non-Pell recipients and an 11.0% increase for Pell recipients. Furthermore, OER adoption resulted in a 2.1% reduction in DFW grades for non-Pell eligible students versus a 4.4% reduction for Pell-eligible students, indicating that the OER effect was stronger for these students with greater financial needs. While these results are promising for OER advocacy, one limitation of this study their approach was that results were only reported at an aggregate (not student) level. This level of reporting may have masked or created differences that would not have been present had results been disaggregated. Thus further research is needed in this area.

We examine the results of 1,488 students who took a calculus course using traditional learning materials with 1,521 students who took the same course but used OER. We account for specific gaps in the literature by (1) rigorously accounting for relevant student variables, and (2) examining whether there is an OER intervention effect that varies across student sub-groups. Our specific research questions are as follows:

1. Do students who use OER withdraw at the same rate as students who do not use OER?

2. Do students who use OER perform as well, worse, or better than their counterparts who use traditional learning materials on the final exam?
3. Do students who are Pell Grant eligible perform better on the final exam when using OER?
4. Do international students perform better on the final exam when using OER?

## Method

### Setting

Our analysis pertains to student performance in *MA16010: Applied Calculus I*, which is the first course in a two-semester applied calculus sequence for non-science and non-engineering majors at Purdue University. This course was selected because one of the authors of the present study is responsible for the curriculum of this course, and as such was in a position to implement this study. The course is offered all year round, but the student composition is substantially different among the fall, spring, and summer classes. To maintain the consistency of population composition, our analysis compares two fall semesters: Fall 2014 (traditional learning materials) and Fall 2015 (OER).

Within each semester, the entire course is centrally coordinated; students are divided into sections of approximately 40 for lectures that are taught primarily by graduate student teaching assistants and limited term lecturers. All course policies and exams are identical for all sections. For each section, lectures are given three times a week for 50 minutes each. In both Fall 2014 and Fall 2015 semesters, sections were offered throughout the day between 7:30am and 5:20pm with a 10-minute gap in between any two consecutive sections. Multiple sections (3-6 sections) were offered simultaneously at each time slot. In Fall 2015, four sections that were either flipped or online were offered; these sections are excluded from the present study so that the results are not biased by different instructional formats.

### Participants

The adoption of OER in this course was done systematically, allowing us to carefully define treatment and control groups of students (i.e., students are “treated” if they used OER, and are in the “control” group if they used commercial materials). The full sample of students includes 1,372 students from the Fall 2014 (control) semester and 1,437 students from the Fall 2015 (treatment) semester. This includes all students who were registered in a traditional section of the course and received a letter grade, and includes withdraws and incompletes. We exclude any student who took an alternate final. As an important covariate, we used the students’ SAT math scores (or SAT math scores converted from ACT math scores) as a measure of mathematical training prior to taking this course. These scores are not available for all students in each semester, and as a result we consider two samples: the full sample, excluding SAT scores, and a reduced sample that includes the SAT scores. Specifically, the reduced sample contains 1,245 students from Fall 2014 and 1,357 students from Fall 2015.

## Procedures

The dependent variable in this study is the students' scores on the end-of-semester final exam. The exact same final exam was used in both Fall 2014 and Fall 2015. Within each semester, all students in the entire course were given the same homework assignments and exams (though not quizzes). There were three one-hour midterms and one two-hour final, which all the students in the course took at the same time and in the same location. The exams were entirely multiple-choice questions and were machine graded, with final exams administered in both Fall 2014 and Fall 2015.

The course is highly centralized, which eliminates much of the variation that might otherwise exist across sections, such as difficulty of the exams, equity in grading, and consistency of policies. In addition, both courses were taught in the fall semester, allowing us to avoid any concern that fall and spring courses have nontrivial differences.

## Materials

In the Fall 2014 semester, all sections were taught using traditional course materials, which included a textbook and an online homework system. Every student was required to purchase a commercial online access code (\$121.00) that provided access to an online homework system and the online version of the commercial textbook that would be valid for two semesters of calculus courses. Students also had the option of purchasing a bundle of the access code and a hard copy textbook for \$168.50, but it was not an option to purchase only the hard copy. The most conservative estimate of cost to students assumes that each student only purchased the online access code.

In Fall 2015, students used OER, which consisted of an e-text, videos, and online homework problems, centrally hosted on an open online learning platform. The course coordinator extensively modified an open textbook called *Calculus: Early Transcendentals*, by David Guichard. The coordinator also recorded a series of lecture videos and online homework problems. All materials were free to students.

## Minor Differences Between Fall 2014 and 2015

There were a few minor differences between Fall 2014 and 2015. We note these differences because while we do not have any data regarding whether they would affect the results, we acknowledge the possibility that they could. First, the grading policy for the course changed from Fall 2014 to Fall 2015. In Fall 2014, final letter grades were given on a curve, whereas in Fall 2015 letter grades were assigned based on a pre-determined scale. Another difference between semesters concerns the final exam schedule, which is assigned by the Registrar's office. In Fall 2015, the final exam took place on the first day of finals (Monday, December 14) from 8:00-10:00 AM, and in the Fall 2014 semester the final was on the second day of finals week (Tuesday, December 16) from 7:00-9:00 PM. Finally, while both the questions and ordering of the questions on the exams were identical across semesters, the multiple-choice response ordering was different across versions. In Fall 2014 there were two versions, and in Fall 2015 there were 10 versions.

## Data Analysis

Table 1 provides a list of variables included in our statistical analyses.

Table 1

*List of Variables and Definitions*

Variable	Type	Definition
<b>Outcome</b>		
Final exam grade	Continuous	MA16010 final exam grade, measured as percentage of total points earned out of maximum points possible on the final.
<b>Control Variables</b>		
Gender	Binary	Male and female.
Ethnicity	Categorical	American Indian or Alaska Native, Asian, Black or African American, Hispanic/Latino, International, White, 2 or more races, Unknown.
College	Categorical	
Student classification	Categorical	Freshman (0-29 hours), Sophomore (30-60 hours), Junior (61-89 hours), Senior (90-105+ hours).
GPA	Continuous	Overall GPA; ranges from 0.0 to 4.0.
Class time	Categorical	7:30am, 8:30am, 9:30am, 10:30am, 11:30am, 12:30pm, 1:30pm, 2:30pm, 3:30pm, 4:30pm.
Pell grant eligibility	Binary	1 = eligible, 0 = ineligible.
Repeat course	Binary	1 = yes, 0 = no.
SAT mathematics	Continuous	Ranges from 400 to 800, and is computed from either SAT Math scores or ACT Math scores converted to SAT units.
Instructor	Categorical	Unordered indicator to control for unobservable instructor effects.

The dependent variable is the final exam grade, in percent, for each student enrolled in the course in Fall 2014 or Fall 2015. The control variables we include are indicator variables for gender and ethnicity, student academic classification, the college from which the student is a major, the student's overall GPA, the time of day in which each student took the class, indicators for whether or not the student is Pell Grant eligible or is repeating the course, the student's SAT mathematics score (or ACT math score converted to SAT units), and the instructor of the student's session. We use these variables to control for student socio-economic

characteristics, academic standing and background, and classroom environment for his/her section of calculus.

Our purpose in examining these covariates is to assess whether there are substantial differences in the distributions of the control variables between semesters that may, if not adjusted for, lead to biased estimates. Our assessment of these distributions is based on the normalized difference, log ratio of standard deviations, and the fraction of the distribution of each variable that lies in the tails of the distribution for the opposing semester. The normalized difference is a unit-free measure of difference in the central tendency of the variable distribution, and the other two measures provide indication of distributional overlap (Imbens & Rubin, 2015).

We examine the impact of OER in two ways. First, our benchmark model of an average overall effect of OER is a linear-in-parameters conditional mean regression model, in which we regress the student's final exam score on the OER indicator variable and controls. The statistical model is

$$Exam_i = \beta_0 + \beta_1 OER_i + X_i \gamma + \varepsilon_i \quad (1)$$

for  $i = 1, 2, \dots, n$  being an index of students and  $X_i$  being a vector of control variables. We are interested in  $\beta_1$  as the parameter that captures the impact of using OER on student final exam performance. We estimate this model using ordinary least squares. Consistent with our first hypothesis, we expect that  $\beta_1$  is statistically insignificant. A significantly positive  $\beta_1$  would imply that OER have a positive overall impact on final exam performance, whereas a significantly negative estimate would indicate that OER have an overall detrimental impact. A standard  $t$ -test for statistical significance of  $\beta_1$  provides a means of testing our first hypothesis.

As a secondary model, we consider a nonparametric specification that is analogous to Equation (1) except that we no longer impose strict functional form assumptions (i.e., linearity, additive separability, and parameter homogeneity). The nonparametric model is written

$$Exam_i = g(OER_i, X_i) + \varepsilon_i \quad (2)$$

where we assume that the function  $g(\cdot)$  is a twice differentiable conditional mean function, but otherwise do not place any restraints on the nature of (non)linearities or interactions between variables. In Equation (1), the OER effect is given by  $\beta_1$ , and is constant for each student in the course, regardless of his/her characteristics or background. In contrast, in the nonparametric specification, the OER effect is given by  $\mathbb{E}[Exam_i | OER_i = 1, X_i = x] - \mathbb{E}[Exam_i | OER_i = 0, X_i = x] = g(OER_i, X_i)_{OER_i=1} - g(OER_i, X_i)_{OER_i=0}$ , which is not constant across students and is allowed to vary generally with student characteristics or background. For instance, if OER have a different effect on domestic vs international students, or Pell Grant eligible vs ineligible students, then the constant  $\beta_1$  will not adequately capture these differences; the nonparametric OER effect, on the other hand, provides a way for us to assess these differences. Note that since the linear structure in (1) is a special, restricted case of the nonparametric structure in (2), if, in fact, the OER effect is not different for different students, the nonparametric effect will itself be constant. We estimate the nonparametric model using kernel regression methods (Li & Racine 2007).

## Results

First, we compare overall withdrawal rates between the Fall 2014 and Fall 2015 semesters (Table 2). These statistics are based on the total number of students that stayed in the course after the last day to drop without any record on the transcript had already passed. We find that the withdrawal rate in Fall 2015 is statistically significantly lower than the withdrawal rate in Fall 2014 ( $p=.012$ ).

Table 2

### *Enrollment and Withdrawal Rates*

	Fall 2014	Fall 2015
Number of students enrolled	1,488	1,521
Withdrawal (W)	7.80%	5.52%

We report descriptive statistics for all the variables in the model, separated by semester, in Table 3.

Table 3

### *Descriptive Statistics*

Variable	Fall 2014		Fall 2015					
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Final exam grade	65.312	25.199	0	100	63.903	24.577	0	100
Male	0.480	0.500	0	1	0.530	0.499	0	1
White	0.603	0.49	0	1	0.596	0.491	0	1
Asian	0.074	0.261	0	1	0.081	0.274	0	1
Hispanic	0.058	0.234	0	1	0.049	0.217	0	1
Black	0.037	0.189	0	1	0.042	0.200	0	1
International	0.184	0.388	0	1	0.175	0.380	0	1
Other	0.044	0.205	0	1	0.057	0.232	0	1
Freshman	0.521	0.500	0	1	0.515	0.500	0	1
Sophomore	0.391	0.488	0	1	0.390	0.488	0	1
Junior	0.063	0.242	0	1	0.070	0.255	0	1
Senior	0.026	0.158	0	1	0.025	0.156	0	1
GPA	2.980	0.675	0	4	3.019	0.690	0	4
SAT Math	613.257	75.629	400	800	612.085	77.415	330	800
Pell grant eligible	0.204	0.403	0	1	0.203	0.403	0	1
Repeated course	0.197	0.398	0	1	0.173	0.379	0	1



Across most of the variables, we do not see large differences in the sample means across the Fall 2014 and Fall 2015 groups, indicating no systematic differences in student characteristics across the two semesters. Several facts are noteworthy from Table 3. First, about 18% of the students enrolled in this course (in either semester) are international students, and above half are freshman with another 40% being sophomores. The average SAT Math score for students in this course is just under 615, and about 20% of students enrolled are Pell Grant eligible. Likewise, just under 20% of the students are repeating the course.

In Table 4, we report the balance and overlap measures for the full sample of observations (excluding SAT Math scores) and for the reduced sample that includes SAT Math scores. All of the normalized differences for both samples lie below 0.1 (in absolute value), and the overlap measures are all relatively small – the variables are well balanced and have sufficient overlap between semesters. Thus, there is no evidence of any systematic difference in student groups across the commercial and OER semesters. This implies that our regression analyses will be robust.

Table 4

*Pre-Match Balance and Overlap Assessment Between Fall 2014 and Fall 2015*

	Normalized difference	Log ratio	Tails treated	Tails control
<b>Full sample -- Excludes math SAT</b>				
Gender	0.099	-0.001	0.470	0.520
Ethnicity	0.016	0.028	0.629	0.604
College	0.017	-0.067	0.100	0.131
Student classification	0.030	-0.006	0.128	0.149
GPA	0.057	0.022	0.062	0.027
Class time	-0.037	0.087	0.127	0.101
Pell grant eligibility	-0.002	-0.002	0.797	0.796
Repeated course	-0.061	-0.049	0.827	0.803
<b>Reduced sample -- Includes math SAT</b>				
Gender	0.097	-0.002	0.468	0.516
Ethnicity	0.059	0.041	0.655	0.652
College	0.002	-0.066	0.103	0.133
Student classification	-0.002	0.017	0.117	0.113
GPA	0.045	0.036	0.071	0.027
Class time	-0.025	0.087	0.130	0.107
SAT math score	-0.015	0.023	0.066	0.036
Pell grant eligibility	-0.015	-0.010	0.789	0.783
Repeated course	-0.067	-0.055	0.829	0.803

\*Note. Full sample: 2,809 observations, 1,437 treated Fall 2015 students; reduced sample: 2,602 observations, 1,357 treated Fall 2015 students.

In Table 5 we report the parameter estimates from our parametric, linear-in-parameters model in which  $\beta_1$  is the effect of OER materials on student final exam performance. We consider three models: Model 1 is the baseline model that includes student and class-time control variables; Model 2 augments the specification to include instructor fixed effects; and Model 3 augments the Model 1 specification to include fixed effects for the college in which each student is a major.

Table 5

*Least Squares Regressions of the Final Exam Score of the OER Indicator and Controls*

	Dependent variable in the regression: Final exam grade in percentage		
	Model 1	Model 2	Model 3
OER indicator	-1.784** (0.768)	-3.332** (1.534)	-1.831** (0.763)
Male	-2.724*** (0.766)	-2.618*** (0.767)	-1.108 (0.824)
Asian	-1.505 (1.432)	-1.326 (1.435)	-1.660 (1.424)
Hispanic	0.206 (1.693)	0.299 (1.695)	-0.030 (1.682)
Black	-0.764 (1.965)	-0.497 (1.968)	-1.184 (1.952)
International	-0.637 (1.310)	-0.688 (1.319)	0.868 (1.364)
Other	1.569 (1.716)	1.419 (1.723)	1.671 (1.701)
Freshman	-2.914 (2.748)	-3.030 (2.760)	-4.451 (2.794)
Sophomore	1.430 (2.744)	1.179 (2.757)	0.048 (2.787)
Junior	-0.112 (3.078)	-0.315 (3.095)	-0.930 (3.101)
8:30am class	-0.443 (1.486)	-0.492 (1.502)	0.104 (1.479)
9:30am class	-3.589** (1.615)	-5.058* (2.772)	-2.879* (1.608)
10:30am class	-1.747 (1.695)	-3.743 (2.930)	-1.277 (1.689)
11:30am class	-0.684 (1.689)	0.692 (3.093)	0.218 (1.704)
12:30pm class	-1.853 (1.613)	0.089 (3.085)	-1.116 (1.609)
1:30pm class	-2.204 (1.543)	0.247 (3.392)	-1.573 (1.539)
2:30pm class	-4.931*** (1.781)	-2.650 (3.305)	-3.907** (1.795)
3:30pm class	-3.801** (1.729)	-4.897** (2.445)	-3.090* (1.729)
4:30pm class	-2.072 (1.673)	-3.148 (2.505)	-1.261 (1.668)
SAT math	0.136*** (0.006)	0.137*** (0.006)	0.131*** (0.006)
Pell grant eligible	0.340 (0.974)	0.127 (0.977)	0.311 (0.967)
Repeated course	-24.816*** (1.017)	-24.665*** (1.023)	-25.094*** (1.012)
Constant	-9.373** (4.358)	-9.585 (5.895)	-3.834 (4.437)
Instructor FEs	No	Yes	No
College FEs	No	No	Yes
Observations	2,602	2,602	2,602
R-Squared	0.403	0.415	0.418
Adj. R-Squared	0.398	0.402	0.41
Residual Std. Error	19.198	19.139	19.006
F Statistic	79.254***	32.786***	52.701***

\*Note. Statistical significance at the 1, 5, and 10 percent level is denoted by \*\*\*, \*\*, \* respectively.

We find that in each model specification, the effect of OER on final exam performance is negative and statistically significant, ranging from -3.332 to -1.784. These estimates mean that, on average, OER students scored about 2 or 3 percentage points lower on their final exam compared to non-OER students in Fall 2014. We also find that many of the student characteristic variables are not significant predictors of student final exam performance, but there is some evidence that students taking the course at different times of the day perform differently. In addition, our estimates show that male students tend to perform worse than females (in Models 1 and 2); students with higher SAT Math scores perform better; and students who are repeating the course perform substantially worse.

The nonparametric model of the effect of OER on final exam performance is student-specific, providing us with a distribution of marginal impacts of OER grade effects. Specifically, for each OER student we compute the difference between his/her expected exam score following Equation (2) and his/her counterfactual predicted exam score had the same student been without OER. To provide a summary of these effects, we report the kernel density of these impacts in Figure 1 for three nonparametric models that correspond to the three linear-in-parameters models reported in Table 5 – that is, the baseline model, and the models that add instructor and college fixed effects, respectively. In all models the distribution of OER effects spans both the negative and positive regions; the distributions of effects in Models 1 and 2 are nearly identical, and are centered around -2.00, whereas the distribution of OER effects in Model 3 is centered just below zero and has a much larger tail in the positive region. When controlling for the college from which the student is a major, the distribution of nonparametric OER effects changes shape, and shifts to the right.

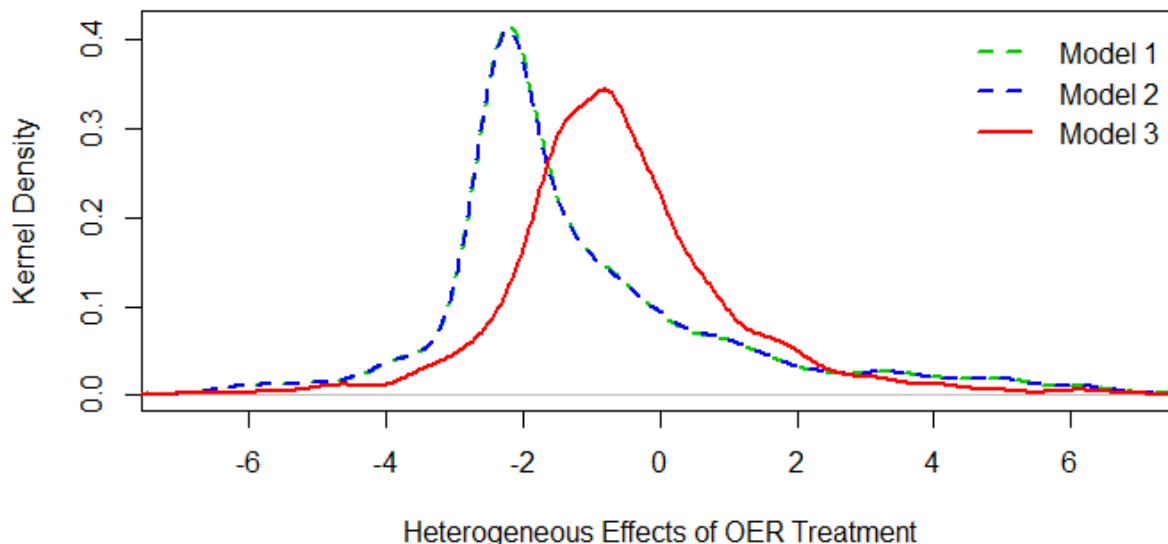


Figure 1. Kernel densities of the heterogeneous OER effects on student final exam scores from three nonparametric specifications.

To examine the grade effects of OER on various student sub-populations, we regress the student-specific nonparametric OER effects on the control variables to ascertain which variables explain the OER effect on final exam performance. We report these regression estimates in Table 6.

Table 6

*Regression of Nonparametric Gradient on Student and Class Characteristics*

Dependent variable: Nonparametric OER gradient			
	Model 1	Model 2	Model 3
Male	0.208* (0.112)	0.211* (0.113)	0.057 (0.087)
Asian	0.641*** (0.206)	0.648*** (0.207)	0.364** (0.160)
Hispanic	-0.060 (0.260)	-0.062 (0.262)	0.068 (0.202)
Black	-0.153 (0.282)	-0.160 (0.283)	-0.142 (0.219)
International	2.209*** (0.187)	2.238*** (0.188)	1.438*** (0.145)
Other	0.900*** (0.238)	0.916*** (0.239)	0.663*** (0.185)
Freshman	-3.512*** (0.401)	-3.524*** (0.403)	-2.039*** (0.312)
Sophomore	-1.847*** (0.401)	-1.856*** (0.402)	-1.158*** (0.311)
Junior	-0.714 (0.446)	-0.720 (0.448)	-0.500 (0.346)
8:30am class	0.302 (0.208)	0.300 (0.209)	0.247 (0.162)
9:30am class	0.499** (0.230)	0.496** (0.232)	0.241 (0.179)
10:30am class	0.524** (0.247)	0.519** (0.248)	0.435** (0.192)
11:30am class	0.024 (0.249)	0.015 (0.250)	0.222 (0.194)
12:30pm class	0.652*** (0.229)	0.653*** (0.230)	0.475*** (0.178)
1:30pm class	0.137 (0.252)	0.127 (0.253)	0.131 (0.196)
2:30pm class	0.011 (0.257)	0.007 (0.258)	-0.022 (0.200)
3:30pm class	0.349 (0.233)	0.343 (0.234)	-0.004 (0.181)
4:30pm class	0.374 (0.234)	0.371 (0.235)	0.134 (0.182)
SAT Math	-0.005*** (0.001)	-0.005*** (0.001)	-0.002*** (0.001)
Pell eligible	0.262* (0.145)	0.261* (0.146)	0.208* (0.113)
Repeated course	0.115 (0.153)	0.115 (0.154)	-0.018 (0.119)
Constant	3.570*** (0.620)	3.624*** (0.623)	1.675*** (0.482)
Observations	1,357	1,357	1,357
R-Squared	0.249	0.249	0.148
Adj. R-Squared	0.237	0.237	0.135
Residual std. error	2.012	2.03	1.57

\*Note. Statistical significance at the 1, 5, and 10 percent level is denoted by \*\*\*, \*\*, \* respectively.

Table 6 answers our third research question, by showing that there is a positive and statistically significant relationship between Pell Grant eligibility and the student-specific effects of OER on exam performance – in Models 1 and 2 the coefficient estimate is approximately 0.26. We find that for Asians, international

students, and “other” students, the effect of OER on final exam performance is significantly higher compared to that of whites and domestic students, with the impact being particularly large for international students. In addition, the effect of OER on final exam performance for freshman and sophomores is significantly lower compared to juniors and seniors. We find that students who took morning sections of the course tend to have higher OER gradients, and students with higher SAT Math scores had marginally lower OER effects.

Keeping with our preferred nonparametric Model 3, we next turn to whether these nonparametric estimates are statistically significant throughout the distribution of effects. We use a residual bootstrap to recover pointwise standard errors for each of the nonparametric point estimates (i.e., each student receives his/her own OER effect estimate, and each of these effect estimates has its own standard error). We summarize these effects in Figure 2, which allows us to simultaneously assess the magnitude as well as statistical significance of each of these point estimates. All the nonparametric OER estimates are placed on the 45-degree line, shown by the black dots. We then overlay the 95% confidence interval above and below each point; these confidence intervals are point-specific, and are shown by the red dots above and below each black point estimate. If the horizontal line at zero cuts between the confidence interval for any particular point, that point estimate is statistically insignificant, while if the horizontal line at zero runs outside of the confidence interval, then that point is statistically significant (at the 95% level).

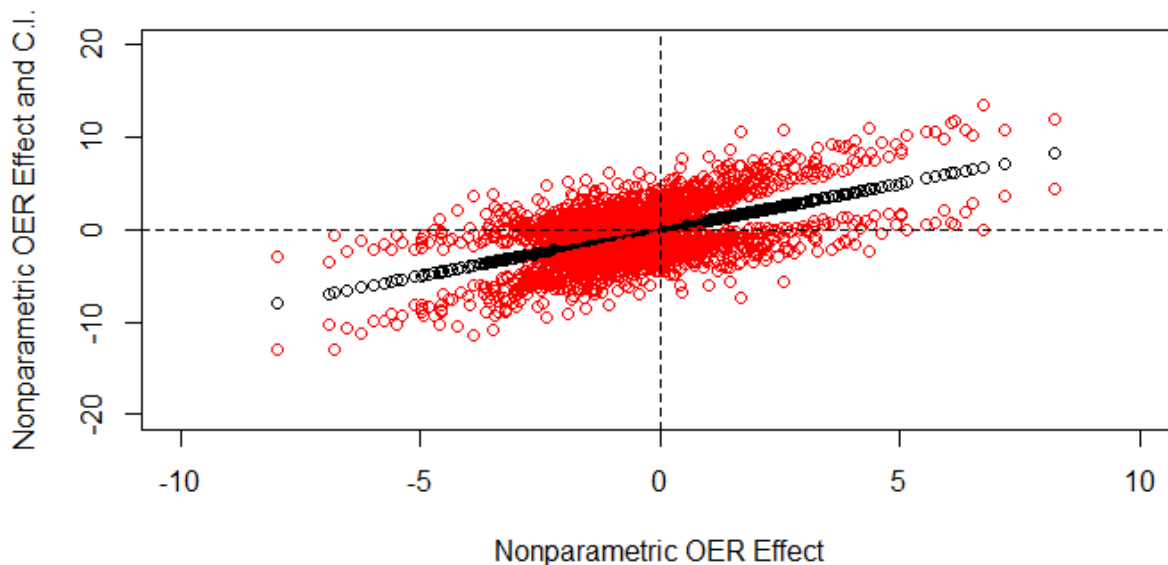


Figure 2. A 45 degree gradient plot of the OER effect from nonparametric Model 3.

We can clearly see that many of the point estimates are statistically insignificant; specifically, out of the 1,357 total point estimates, 1,035 are statistically insignificant, 281 are negative and significant, and 41 are positive and significant. To provide further description, in Figure 3 we plot the kernel densities of the OER effects by group: the group that is negative and significant, statistically insignificant, and positive and significant. It is clear that the group that is negative or positive and significant are the smallest and largest point estimates of the OER effects, whereas the insignificant effects are those that lie closest to zero. Further, the group with positive and significant OER effects average about 4 percentage points higher on the final exam when using OER, compared to their expected performance when using commercial textbook materials. Students who performed significantly worse averaged about 2.5 percentage points lower when using OER. Thus, for the majority of students, OER do not generally lead to a significant difference in student final exam performance, relative to commercial materials. When OER have a negative effect, the effect tends to be relatively small compared to when OER have a positive effect.

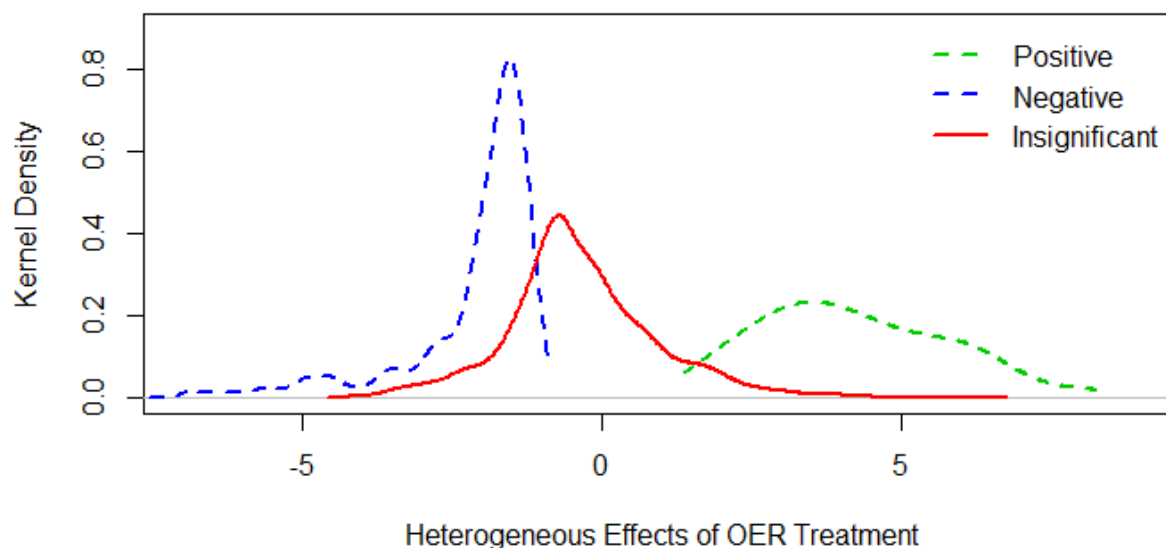


Figure 3. Distributions of heterogeneous OER effects from nonparametric Model 3.

As a final analysis, we would like to uncover exactly which types of students perform better, worse, or no different when using OER. To make this determination, we estimate a multinomial regression in order to predict the probability that each student falls into the negative, insignificant, or positive OER effects category. We have shown that the majority of students have an insignificant effect of OER on grade performance; however, we are interested in understanding if there are systematic differences between students who fall into these three categories of OER effects on final exam performance. In the multinomial model, we estimate parameters for each category (for the base category, negative, the coefficients are constrained to zero for identification), and assess statistical significance of these parameters via standard errors. However, since the multinomial logit is nonlinear, we compute average marginal effects for each

category for each student characteristic, in order to understand how characteristics influence the probability that the student ends up in one of the three groups of OER effects. We report these results in Table 7.

Table 7

*Multinomial Logit Regression Estimates and Implied Marginal Effects*

	<i>Dependent variable: Nature of OER effect</i>		
	Negative	Insignificant	Positive
Male	- [-0.003]	0.032 (0.152) [0.012]	-0.268 (0.365) [-0.009]
Asian	- [-0.225]	1.360*** (0.319) [0.209]	1.632** (0.718) [0.016]
Hispanic	- [-0.189]	1.173*** (0.388) [0.200]	0.553 (0.864) [-0.011]
Black	- [-0.108]	0.652* (0.382) [0.101]	0.755 (0.696) [0.007]
International	- [-0.546]	3.273*** (0.482) [0.484]	4.719*** (0.729) [0.063]
Other	- [-0.147]	0.909*** (0.320) [0.152]	0.545 (0.838) [-0.005]
Freshman	- [0.573]	-3.452*** (0.727) [-0.523]	-4.412*** (0.987) [-0.050]
Sophomore	- [0.418]	-2.578*** (0.729) [-0.428]	-1.682* (0.893) [0.010]
8:30am class	- [-0.021]	0.132 (0.269) [0.023]	0.024 (0.615) [-0.002]
9:30am class	- [0.062]	-0.374 (0.282) [-0.057]	-0.471 (0.706) [-0.005]
10:30am class	- [-0.067]	0.394 (0.333) [0.055]	0.708 (0.697) [0.012]
11:30am class	- [0.036]	-0.190 (0.310) [-0.014]	-0.908 (0.886) [-0.022]
12:30pm class	- [-0.178]	1.124*** (0.343) [0.201]	0.108 (0.892) [-0.023]
1:30pm class	- [-0.006]	0.002 (0.327) [-0.019]	0.822 (0.667) [0.024]
2:30pm class	- [0.079]	-0.456 (0.335) [-0.058]	-1.043 (0.903) [-0.020]
3:30pm class	- [-0.044]	0.298 (0.331) [0.067]	-0.565 (0.886) [-0.023]

On the Efficacy of Open Educational Resources: Parametric and Nonparametric Analyses of a University Calculus Class  
Delgado, Delgado, and Hilton III

4:30pm class	-	-0.361 (0.301)	-0.317 (0.715)
	[0.059]	[-0.058]	[-0.001]
SAT math	-	0.0001 (0.001)	-0.008*** (0.003)
	[0.000]	[0.000]	[0.000]
Pell eligible	-	-0.109 (0.186)	0.607 (0.414)
	[0.013]	[-0.034]	[0.020]
Repeated course	-	1.366*** (0.280)	1.877*** (0.452)
	[-0.227]	[0.204]	[0.024]
Constant	-	3.553*** (1.021)	4.442** (1.794)
Frequencies	0.207	0.763	0.03
McFadden R-Squared		0.172	
Log-Likelihood		-717.625	
Likelihood ratio test		297.386 (0.000)	

*\*Note.* Table reports coefficient estimates with standard errors in parentheses. The numbers in square brackets are implied average marginal effects. Statistical significance at the 1, 5, and 10 percent level is denoted by \*\*\*, \*\*, \* respectively. Note that for identification, the coefficients in the negative category are constrained to zero.

Looking first at the estimated parameters, shown in the first line for each variable in the table, we see that the most significant predictor variables are indicators for Asian, international, freshman, and sophomore students, as well as for those repeating the course. Many of the parameters on the other covariates are not statistically significant, which simply means that students of these other types are not systematically winding up in different OER effect groups. For example, the gender of the student, or the student's Pell Grant eligibility status, do not influence the likelihood of ending up in one of these three OER effect groups. To interpret these Pell Grant eligibility estimates in the context of the significantly positive effect of Pell Grant eligibility on the OER gradient in Table 6, we find that while Pell Grant eligible students have a larger OER gradient than non-eligible students (Table 6), Pell Grant eligibility does not lead to a systematic grouping of students into the positive, insignificant, or negative OER effect groups. Thus, we find that OER materials have a significantly positive effect on exam performance for Pell Grant eligible students, though not to the extent that Pell Grant eligibility leads to systematic differences in overall OER effects.

Asian students are 22.5% less likely to be in the negative OER group, compared to white students, and are 21% and 2% more likely to be in the insignificant and positive effect groups, respectively. Similarly, we find that international students are 55% less likely to be in the negative group, and are 48% and 6% more likely to end up in the insignificant or positive effect groups, respectively. Freshman and sophomores are 57% and 42% more likely to end up in the negative OER effect group. Finally, students who are repeating the course are 23% less likely to be in the negative effect group, and are 20% and 2% more likely to end up in the insignificant and positive groups, respectively.



## Discussion

During the semester that OER were implemented, students were significantly less likely to withdraw from the course (7.8% withdrew in 2014, 5.5% in 2015). This appears to be a potentially important benefit of OER; other studies (e.g., Hilton, Fischer, Wiley, & Williams, 2016) have found a similar benefit of lower withdrawal rates with OER.

When looking at an overall, average effect, we find that students using OER materials performed about 2% worse in the course compared to students who used commercial materials. This finding is unique in OER efficacy literature; to date, only Robinson (2015) has found that more students did worse using OER than better. To provide some perspective on the relative size of these effects, we see that the magnitude of the OER effect in Model 1 (-1.784) is less than half the magnitude of the coefficients on the 2:30 and 3:30 PM class time indicators (respectively, -4.931 and -3.801). This means that if OER have a detrimental effect on student course performance, the effect is substantially smaller than moving a student from one section to another. We also note that for 76% of students, OER had no effect. There were more students negatively affected than were positively affected; however, those who were positively affected had a larger overall effect. We speculate that the fact that fewer students withdrew from the course may also have had an influence on the average final exam scores.

We find that freshmen and sophomores did not perform as well when using OER. Pell Grant eligible students performed statistically better using OER materials compared to statistically identical students who used commercial materials. This effect, however, is small, and does not translate into systematic qualitative differences in exam performance for Pell Grant eligible versus ineligible students – i.e., while Pell grant eligible students, on average, scored about 0.26% higher on their final exam compared to non-eligible students, Pell grant eligibility does not help us predict whether a student will have a positive, insignificant, or negative OER effect. These results are different from those of Colvard et al. (2018), indicating that further work needs to be done in this area.

Evidence suggests that OER has a substantial, positive impact on exam performance for international students. Not only do international students using OER score higher on their final exam than domestic students using OER, they also consistently rank in either the insignificant or positive OER effect group. At a minimum, international students are not harmed by the OER we used, and at best are positively impacted. We conjecture that the online materials associated with this course were beneficial to students who are not native English speakers because there were videos and they were subtitled, providing them with an opportunity to repeat the materials as needed until mastery is attained.

This study only focuses on one discrete set of OER used in one context. This discussion cannot be construed to mean that all OER will have similar results. What this study does do is, in a rigorous way, examine the influence of one set of OER in one particular context. These results suggest that additional rigorous studies of OER should be done to confirm these results with other OER in different contexts.

## Conclusion

While cost-savings is not an explicit part of our study, we do feel that it is notable that the adoption of OER collectively saved students in Fall 2015 at least of \$101,700.50 (given the actual minimum per-student cost of \$60.50 costs in Fall 2014). This is a deliberate underestimate; depending on the number of students who would have purchased a hard copy and/or only took one semester of calculus, the savings could be much higher. This leads us to wonder about the relative cost/benefit ratio for OER materials in the context of heterogeneous, student-specific OER effects. Clearly for students who have positive or insignificant OER effects, it is advantageous to use an open textbook that is free and can be retained by the students for years to come. Yet, for students who have a negative OER effect, if in fact, students can save significant amounts of money through the use of OER with only a small decrease in their final exam score, should OER be utilized? How low would the savings need to be (or how severe the drop in exam score) before OER was not considered to be a potentially good choice? Is there a threshold in terms of the percentage of students in the course with a negative OER effect that would make it undesirable to adopt OER? These questions are difficult to answer, and we recognize that the optimal instructional approach may not be one-size-fits-all. And so, while we are reluctant to provide an answer these questions, the very existence of OER invites their consideration.

## References

- Bowen, W. G., Chingos, M. M., Lack, K. A., & Nygren, T. I. (2014). Interactive learning online at public universities: evidence from a six-campus randomized trial. *Journal of Policy Analysis and Management, 33*, 94-111.
- California OER Council. (2016). OER adoption study: Using open educational resources in the college classroom (White paper). *California Open Educational Resources Council*. Retrieved from <https://docs.google.com/document/d/1sHrLOWEiRs-fgzN1TZUlmjF36BLGnICNMbTZIP69WTA>
- Colvard, N. B., Watson, C. E., & Park, H. (2018). The Impact of Open Educational Resources on Various Student Success Metrics. *International Journal of Teaching and Learning in Higher Education, 30* (2), 262-276.
- Feldstein, A., Martin, M., Hudson, A., Warren, K., Hilton, J., & Wiley, D. (2012). Open textbooks and increased student access and outcomes. *European Journal of Open, Distance and E-Learning, 15*(2), 1-9. Retrieved from <https://eric.ed.gov/?id=EJ992490>
- Florida Virtual Campus. (2016, October 7). 2016 Florida student textbook and course materials survey. *The Orange Grove*. Retrieved from <https://florida.theorange Grove.org/og/items/3a65c507-2510-42d7-814c-ffdefd394b6c/1/>
- Grewe, K., & Davis, W. P. (2017). The impact of enrollment in an OER course on student learning outcomes. *The International Review of Research in Open and Distributed Learning, 18*(4). Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/2986/4211>
- Gurung, R. A. R. (2017). Predicting learning: Comparing an open education research and standard textbooks. *Scholarship of Teaching and Learning, 3*, 233-249.
- Hewlett (2017). Open educational resources (Blog post). Retrieved from <http://www.hewlett.org/strategy/open-educational-resources/>
- Hilton, J. (2016). Open educational resources and college textbook choices: A review of research on efficacy and perceptions. *Educational Technology Research and Development, 64*(4), 73-590. Retrieved from <https://link.springer.com/article/10.1007/s11423-016-9434-9>
- Hilton, J., Fischer, L., Wiley, D., & Williams, L. (2016). Maintaining momentum toward graduation: OER and the course throughput rate. *International Review of Research in Open and Distance Learning, 17*(6). Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/2686>
- Hilton, J., & Laman, C. (2012). One college's use of an open psychology textbook. *Open Learning: The Journal of Open, Distance and e-Learning, 27*(3), 265-272.
- Jhangiani, R. S., Pitt, R., Hendricks, C., Key, J., & Lalonde, C. (2016). *Exploring faculty use of open educational resources at British Columbia post-secondary institutions*. BCcampus Research

- Report. Victoria, BC: BCcampus. Retrieved from [https://bccampus.ca/files/2016/01/BCFacultyUseOfOER\\_final.pdf](https://bccampus.ca/files/2016/01/BCFacultyUseOfOER_final.pdf)
- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences*. New York: Cambridge University Press.
- Li, Q., & J. S. Racine (2007). *Nonparametric econometrics: Theory and practice*. New Jersey: Princeton University Press.
- Lovett, M., Meyer, O. & Thille, C. (2008). The open learning initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning. *Journal of Interactive Media in Education*, 1, 1-16. Retrieved from <https://jime.open.ac.uk/articles/10.5334/2008-14/>
- Ozdemir, O., & Hendricks, C. (2017). Instructor and student experiences with open textbooks, from the California open online library for education (Cool4Ed). *Journal of Computing in Higher Education*, 29(1), 98-113. Retrieved from <https://link.springer.com/article/10.1007/s12528-017-9138-0>
- Pawlyshyn, N., Braddlee, Casper, L., & Miller, H. (2013). Adopting OER: a case study of cross-institutional collaboration and innovation. *EDUCAUSE Review*. Retrieved from <https://er.educause.edu/articles/2013/11/adopting-oer-a-case-study-of-crossinstitutional-collaboration-and-innovation>
- Robinson, T. J. (2015). *The effects of open educational resource adoption on measures of post-secondary student success* (Doctoral dissertation). Available from BYU ScholarsArchive. (Accession No. 5815)
- Watson, C. E., Domizi, D. P., & Clouser, S. A. (2017). Student and faculty perceptions of OpenStax in high enrollment courses. *The International Review of Research in Open and Distributed Learning*, 18(5). Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/2462/4299>

