Substituting open educational resources for commercial curriculum materials: effects on student mathematics achievement in elementary schools

John Hilton III, Ross Larsen, David Wiley and Lane Fischer
Brigham Young University

ABSTRACT

Open Educational Resources (OER) have the potential to replace commercial learning materials in education. An empirical examination of this potential was conducted, comparing the end-of-year mathematics test results of 12,110 elementary school students clustered within 95 schools from five school districts in the state of Washington in the United States of America. Of this group, 6796 students used open learning materials, and 5314 used commercial educational resources. When three years of test scores were considered, there were no statistically significant differences in the exam scores of students who used open versus commercial curriculum materials. The lack of statistical significance may have practical significance, demonstrating that OER can replace conventional materials without impacting student performance, while potentially reducing costs and allowing for local modification.

ARTICLE HISTORY

Received 20 September 2017
Accepted 3 January 2019

KEYWORDS

Textbooks; policy; mathematics education; curriculum materials; elementary schools; cost of education; open educational resources; OER

Introduction

Successful mathematics teaching has several important components. For example, Charalambous and Hill (2012) note that complex interactions between students, teachers and content all impact how students learn. These interactions are influenced by a variety of factors such as the knowledge and beliefs of both students and teachers, as well as how both parties utilise available resources (including physical facilities, available time, and so forth). A key aspect of content is the choice of curriculum materials. While there are various definitions of curriculum materials, in this paper we use this term to "refer to artifacts such as student textbooks, teacher guides, lesson plans, and instructional materials kits that, by communicating ideas and practices, can shape classroom activity" (Charalambous & Hill, 2012, p. 444).

Few people doubt that curriculum materials are important; they play a key role in scope, sequence and many other facets of instruction (Crawford & Snider, 2000). However, determining whether one set of curriculum materials is more effective than another is often difficult. For example, the National Research Council (2004) reviewed
698 peer-reviewed studies of nineteen different mathematics curriculum materials at the K-12 level (ages 5–18). They found that due to several limitations in and across the studies they were unable to state which programmes were most effective.

In some respects, such a conclusion is not astonishing given the complexity of educational research. Yet given the large amount spent on curriculum materials, it is surprising that in many instances no evidence exists to corroborate its value. The Federal Communications Commission (FCC) states that in the United States, more than $7 billion is spent each year on textbooks at K–12 public schools (Usdan & Gottheimer, 2012). While the costs of various programmes vary, Oviett (2014) calculated the total expenditures by public schools in the state of Utah over seven years and found expenditures were approximately 25 million dollars per year.

The disconnect between the costs of learning materials and the lack of their proven efficacy has become more apparent as relevant Open Educational Resources (OER) have increased in quality and availability. The term “Open Educational Resources” was coined at the 2002 UNESCO Forum on the Impact of Open Courseware for Higher Education in Developing Countries. At this forum, OER was defined as “The open provision of educational resources, enabled by information and communication technologies, for consultation, use and adaptation by a community of users for non-commercial purposes” (UNESCO, 2002, p. 24). Thus OER are educational materials (including courses, textbooks, videos, and test items) that are licensed in such a way (typically through Creative Commons) so as to allow for reuse and revision to meet the needs of teachers and students (Bissell, 2009). Wiley, Bliss, & McEwen (2014) provide an overview of the history, models and challenges of OER. In addition to providing a history of both OER and its definition, they describe various models of both creating and sharing OER. They also highlight specific benefits and challenges pertaining to OER (one of which is discussed in greater detail below).

Some have suggested that as high-quality OER proliferate, the money spent on curriculum materials could be better allocated to other aspects of pedagogy. For example, Randy Dorn, the Superintendent of Instruction for the state of Washington stated that the result of Washington’s use of OER is “that quality classroom materials are getting to students at a lower cost. That gives districts more money to spend in critical areas like professional learning and technology infrastructure” (OPSI, 2016).

The argument that money could be allocated from curriculum to professional development is intriguing. Slavin and Lake’s (2008) synthesis of 87 mathematics curriculum studies found that instructional improvement had a larger impact on student performance than the choice of curriculum. This may be in part because of the size of teacher influence relative to curriculum. Baumert et al. (2010) show that teachers’ pedagogical content knowledge played an important role in student mathematics success. Some school districts are innovating in this manner; Song (2018) reports on how a district in Pennsylvania recognised they were spending significant amounts of their budget on learning materials. They adopted OER and allocated 31% of the curriculum savings to pay teachers for editing and (where needed) authoring OER, and spent the remaining 69% of saved funds to creating professional learning experiences. Song also notes that other districts are engaging in similar practices.

While OER are free to users, like all educational materials there are costs (in both time and money) to create, adopt and adapt OER. In some instances, foundations (such as the Bill and Melinda Gates Foundation) sponsor the creation of OER. In other cases,
government institutions (such as a state department of education) provide funding for OER. The underlying theory of OER is that once it has been paid for it can be freely reused, thus allowing for significantly lower per-student costs. As with many curriculum materials, there is a wide variety in the quality of OER. When considering adopting OER (as with traditional curriculum materials) teachers and other stakeholders should carefully examine them and compare them with competing materials according to pre-determined rubrics. If OER are of comparable quality, they are particularly attractive, not only because of the low cost (free if they can be used digitally and tablets or other devices are already available), but because they allow teachers, schools, and districts the flexibility of modifying the OER to meet local needs.

In 2015, the United States Department of Education launched #GoOpen, which is a coordinated effort to encourage educators to use OER (U.S. Department of Education, 2015). As of September, 2016, multiple school districts across seventeen states have engaged with this initiative and are actively working to adopt OER. In some instances, these efforts have been in place for several years; for example, in January of 2012, the Utah State Office of Education (2012) announced it would support the creation of open textbooks and encourage their adoption throughout the state.

While the idea of OER seems attractive, and theoretically whether a product is published as OER or a commercial product should have no bearing on its efficacy, some believe that OER cannot be as efficacious as their more expensive commercial counterparts. Wiley et al. (2014) state that a major obstacle for increasing OER adoption is “the pervasive perception that, because they are free, OER are necessarily of inferior quality” (p. 785). Allen and Seaman (2014) did a nationally representative survey of 2,144 higher-education faculty members in the United States and found that while faculty who were aware of OER rated it of a similar quality as traditional resources, a significant majority of faculty were not sufficiently aware of OER to make a judgement.

With respect to this issue, Butcher (2015) describes an entrenched notion that publishers are responsible for guaranteeing the quality of specific educational materials, raising a concern that non-traditional materials such as OER may be inferior. Such perceptions have been prevalent for several years. For example, Harley (2008) notes concerns regarding the quality of OER. Bossu and Tynan (2011) state that a major barrier for OER adoption is that the free and open nature of OER raises suspicions that they are necessarily lower quality than commercial resources. Wiley and Gurrell (2009) likewise state that there is a common intuition that OER are poor quality. The purpose of this present study is to empirically examine the issue of OER quality in an elementary school setting by comparing the end of year mathematics test results of 12,110 elementary school students in the state of Washington. Of this group, 6796 students used OER and 5314 used traditional curriculum materials.

**Review of the literature**

The literature regarding the use of curriculum materials is extensive and complex. Crawford and Snider (2000) argue that curriculum materials are a vital part of the educational enterprise, pointing to studies indicating that up to 90% of classroom instruction is centred on textbooks (Tyson & Woodward, 1989; Woodward & Elliot, 1990). Specifically within the domain of mathematics, textbooks (a key component of curriculum materials)
have been termed “the most important tools in guiding teachers’ teaching” (Van den Heuvel-Panhuizen, 2000, p. 10). More recently, Banilower et al. (2013) surveyed 7752 mathematics and science teachers in the United States of grades K-12 (ages 5–18) and found that textbooks were a key component of mathematics curriculum materials. They write, “Textbooks appear to exert substantial influence on instruction, from the amount of class time spent using the textbook (especially in mathematics) to the ways teachers use them to plan for and organise instruction” (p. 108).

And yet as Remillard (2005) points out, both researchers and practitioners have wide views of what it means to study and use curriculum materials. She states that there are “various ways that teachers draw on their own resources and capacities to read, make meaning of, evaluate, adopt, adapt, and replace the offerings of the curriculum” (p. 234). Thus caution must be exercised when making claims about the influence of a particular set of curriculum materials. Moreover, as Hill and Charalambous (2012) demonstrate, two teachers may take the same curriculum materials and utilise them in very different ways, making it difficult to point to curriculum materials as the driver of educational change.

Given the complexity of measuring the usefulness of curriculum materials it is not surprising that decisions regarding adopting curriculum materials are often not research based (Crawford & Snider, 2000). Part of this lack of research-based adoption is likely the paucity of definitive research demonstrating the superiority of specific curriculum materials. For example, Tarr et al. (2008) examined the test results of 2533 students to learn whether there were differences in student performance based on whether the mathematics curriculum was developed with funding from the National Science Foundation (NSF) or was a publisher-developed textbook. They found no significant differences in student test scores based on the curriculum materials used. This difficulty is not unique to mathematics; indeed, as Taylor et al. (2015) note, “although numerous federal agencies have funded the development of curriculum materials over the past 30 years, the field of science education still lacks evidence regarding what programmes (or types of programmes) have noteworthy effects” (p. 985).

The What Works Clearinghouse (WWC) was established to help stakeholders identify curriculum materials whose efficacy has been rigorously established through research. Unfortunately, results from the WWC indicate that while much research has been performed it often does not meet the criteria that the WWC has set for rigorous research on the effectiveness of curriculum materials. For example, while seventeen studies of “Singapore Math” have been published between 1983 and 2014, none of these met WWC standards (What Works Clearinghouse, 2015a, p. 1). Similarly, across eleven studies of the “Dreambox Learning” supplemental online mathematics curriculum materials, only one met the evidence standards of the What Works Clearinghouse (What Works Clearinghouse, 2013); for the “Everyday Mathematics Curriculum” only one out of ninety-two studies met the specified criteria (What Works Clearinghouse, 2015b).

Thus, while curriculum materials are clearly deemed to be important, and set the direction for classroom instruction, identifying their relative value has proved in many cases to be elusive. Given the many other factors that go into student learning, not the least of which is the effective pedagogy of the teacher (Wenglinsky, 2002) the efficacy of a specific set of curriculum materials may not be the most important factor in increasing student learning.

Crawford and Snider (2000) point out that selections of curriculum materials “are more likely to be guided by political and economic factors than by qualities that are known to
benefit students” (p. 123). The cost of curriculum can be a significant issue, amounting to state expenditures measured in the tens of millions of dollars (Oviett, 2014). Charles Tack, a spokesperson for the Arizona Department of Education (which is part of the #GoOpen initiative) stated,

In tight budgetary times that we currently find ourselves in and that districts have to live with, every little bit can help. It [OER] will give them additional access to materials that can be used to help students and soften the blow to their already strained finance. (Jung, 2016)

In addition to the financial costs, the high price of curriculum materials can lead to schools reusing it for several years in order to maximise the value of the expenditure (Armstrong & Bray, 1986), which can lead to students studying outdated materials. Another negative side effect of the high cost of curriculum materials is that in some settings schools only purchase desk copies, thus preventing students from studying the materials at home.

The lack of proven efficacy of traditional materials, coupled with their high cost, has caused some schools to transition to OER. OER adoption can lead to significant cost-savings for states, districts, and schools who adopt them. While OER are often digital, they do not have to be in order to provide cost savings. Wiley, Hilton, Ellington, and Hall (2012) found that paperback versions of open science textbooks could be produced for less than half the cost of traditional textbooks, even though their model provided new textbooks for students each year and the cost of traditional textbooks was amortised over seven years. Moreover, their model accounted for the costs of the teachers’ time spent in revising pre-existing OER so that the resources would be suitable for their context. If the open textbooks had the same shelf life as traditional textbooks (e.g. the same book is used for several years) then the cost-savings would be much higher.

Kimmons (2015) studied the perceptions of thirty primary and high school teachers who compared textbooks built from OER with traditional textbooks and found that teachers felt that the open textbooks were of higher quality. Similarly, de los Arcos, Farrow, Pitt, Weller, and McAndrew (2016) surveyed 323 primary and high school teachers located primarily in the United States, United Kingdom, and South Africa regarding how they perceived OER that they had utilised. They found that teachers were generally positive about the OER that they had used. A synthesis of nine studies by Hilton (2016) found similar perception results at the post-secondary level.

Much of the literature surrounding the efficacy of OER material has been done at the post-secondary level. For example, Allen, Guzman-Alvarez, Molinaro, and Larsen (2015), tested the efficacy of an OER called ChemWiki in a general chemistry class at the University of California, Davis. Students in one section \(n = 478\) used ChemWiki as its main learning resource, while students in another section \(n = 448\) used a commercial textbook. Both classes were taught by the same instructor at back-to-back hours and employed the same teaching assistants, exams, and grading rubrics were utilised. Researchers administered a pretest that indicated there were no significant prior knowledge differences between the two groups. Researchers found no significant differences between the two groups when examining their midterm and final exam scores.

Another OER efficacy study in higher education is Fischer, Hilton, Robinson, and Wiley (2015). They analyzed the academic results of students in fifteen courses across seven different institutions. Students used either OER \(n = 1087\) or traditional textbooks \(n = 9264\). The researchers found that in two of the fifteen classes, students using OER...
were significantly more likely to complete the course (there were no differences in the remaining thirteen). In five of the OER classes, students were significantly more likely to receive a C- or better. In one study control students were more likely to receive a C- or better. Similarly, in terms of the overall course grade, students in four of the OER classes received higher grades, ten of the classes had no significant differences, and students in one control class received higher grades than the corresponding treatment class. They also found that students who enrolled in courses that utilised OER took on average two credit hours more than those in the control group, even after controlling for demographic covariates and hypothesised that perhaps the lower cost of textbooks led to an increased financial ability to take more classes. This study is limited however in that it does not control for previous academic student success or teacher variables.

Wiley, Williams, DeMarte, and Hilton (2016) studied a pilot adoption of OER at Tidewater Community College. Although their study has significant limitations given relatively few students who used OER and lack of control for important covariates, the researchers did find that across two semesters students who enrolled in classes utilising OER were significantly less likely to drop the class than peers enrolled in courses using traditional textbooks. Hilton (2016) analyzed seven additional OER efficacy studies in higher education settings, and found that while many of the studies had weak research designs, collectively, they indicate that students do as well or better when using OER as opposed to commercial textbooks. More recent studies (e.g. Colvard, Watson, & Park, 2018) have similar results.

To date, only two studies have focused on K12 OER efficacy. Wiley et al. (2012) examined the standardised test scores of students using these open curriculum materials in secondary science classes in three different school districts. Approximately 1200 students used open curriculum materials during this study; their results were compared with students in the same district who used commercial materials. These researchers examined their end-of-year standardised test results and found no apparent differences between the results of students who used traditional and open curriculum materials. There were, however, several methodological problems with this study. Principle among these was a lack of control for many potential confounding variables, both at the teacher and student levels.

This limitation was addressed by a later study (Robinson, Fischer, Wiley, & Hilton, 2014). They examined the use of open science curriculum materials in three secondary science subjects across several schools in a suburban school district. This rigorous study used propensity score matching to account for teacher effect, socioeconomic status, and eight other potentially confounding variables. There were 1274 students in each condition (treatment and control). In examining the results of the end-of-year state standardised test there was a very small, but statistically significant, difference between the two groups, favouring those who utilised open curriculum materials. However, Robinson et al. (2014) noted,

Because students were only sampled from one school district with a distinct demographic footprint, it would be problematic to claim that other students would experience similar results. However, this result does provide a rationale for other systematic evaluations of the effects of open textbook in other locations, grade levels, and subjects. (p. 349)

In the present study, we seek to do just that by examining the adoption of an OER known as EngageNY Math, specifically the Eureka Math version. These curriculum materials were
developed using funds from the state of New York and have a Creative Commons license, allowing others to freely reuse them. We focus on two years following the adoption of EngageNY Math by two school districts in the state of Washington and compare the end-of-year results on the state standardised tests between students who used these open curriculum materials with those in comparable districts who continued using traditional learning materials. This is the first efficacy research study of OER materials at the primary school level. Our specific research question is as follows: are there detectable differences in the end-of-year test results of students who used OER versus students who used a commercial product as their core mathematics curriculum materials?

**Method**

**Participants**

The initial data set of students in schools using OER consisted of students in two large school districts in the state of Washington who were in 3rd grade (age 8–9) during the 2013–2014 school year, and follows them in 4th and 5th grades (ages 11–12). During their 3rd grade year, these students were taught with commercial curriculum; in 4th and 5th grade they were taught using the EngageNY Math OER. Information regarding their use of OER was provided by Washington State’s Office of Superintendent of Public Instruction (OSPI). Realising there was the possibility of a natural experiment (some districts adopting OER while others did not), we used public data supplied by OSPI to determine which districts would best serve as controls using information such as demographics, socio-economic status, geography, and 2013 end of year exam scores. We identified eleven potential comparison school districts. Because of local choice in determining curriculum materials, OSPI contacted representatives at these districts in order to determine whether schools in their district had used EngageNY Math. Eight of these districts had at least some schools that had used EngageNY Math as core or supplemental materials, which left us with three comparison districts. In total, our resulting sample included 6796 students who used OER and 5314 who used commercial materials. These students were clustered within 95 schools. Student data included mathematics test scores for 2016, 2015, and 2014. The dataset also included student covariate data such as race, gender, eligibility for free-and-reduced lunch, migrant and special education statuses.

**Curriculum materials**

Three sets of curriculum materials were utilised in the present study. The OER used by the two OER districts was EngageNY Math. Two of the control districts utilised Math Expressions published by Houghton Mifflin Harcourt, and the third control district used Math Connects published by Macmillan/McGraw-Hill. Each of these curriculum materials were the core instructional materials for mathematics in their respective settings and consisted of instructional components and supports for students as well as teacher materials (including pedagogical instructions, assessments, and rubrics).

Washington State has adopted the Common Core State Standards as their state learning standards; their end-of-year exam is aligned with these standards. The Louisiana Department of Education (2016) has posted an online review of several curriculum materials that examine the extent to which they align with Common Core standards. They rate the
*Eureka Math* version of *EngageNY Math* as tier 1, meaning that it has superior quality in terms of meeting the content standards. It was ranked as “strong” on seven out of seven criteria. They rate *Math Expressions* as tier 2, indicating that it meets all the key criteria and has some indicators of superior quality. It was ranked as “strong” on five out of seven criteria. They did not rate *Math Connects* as it has not been updated to align with the Common Core standards. Aligning curriculum materials with the Common Core has been challenging, as indicated by Polikoff (2015). In order to test whether the relative alignment with Common Core standards was influential in end-of-year exam scores we compared the results of students who used *Math Expressions* and *Math Connects* in addition to other analyses to test for measurable differences.

**Data analysis**

Our independent outcomes in this study were the student scores on the 2015 and 2016 state end-of-year mathematics assessment known as the Smarter Balanced Assessment (SBA). The SBA test was new in 2015; it replaced the Measurements of Student Progress (MSP) test. Among the important covariates were student test scores in 2014 (prior to the implementation of OER). One challenge we faced was that a large number of students (81.5%) did not have 2014 data because they were part of a pilot test in 2014 for the SBA and thus did not report scores on the MSP. The Pilot SBA scores were also not made available. To account for this issue, all analyses were run with and without the covariate MSP 2014.

In order to answer the question of whether the OER programme affects outcomes as measured by the state test scores in 2015, two linear regression models were run in MPLUS 8.0. Multiple regression was chosen as the method to isolate the effect of the programme over other methods (such as propensity score matching) to maximise statistical power (Bloom, Michalopoulos, & Hill, 2005).

The assumptions of multiple regression are: (a) linearity, (b) independence, (c) normality, (d) equality of variance, (e) lack of multicollinearity of independent variables, and (f) that the missing data is dealt with in an appropriate way. The data were examined through histograms, scatterplots, variance inflation factors, and residual plots, and it was found that the assumptions of linearity, normality, equality of variance, and lack of multicollinearity were met. Independence was not met as students are nested within schools. This limitation is handled by allowing the data to be clustered by using the TYPE = COMPLEX option in MPLUS with schools that the student attended in 2015 as the clustering variable. The model will have the general form:

\[ Y_i = X\beta + e_i \]

Where \( Y_i \) is the response for student \( i \), \( X \) is the matrix of covariate values or dichotomous scores for dummy variables where appropriate, \( \beta \) is the vector of betas to be estimated, and \( e_i \) are the errors associated with the predictions of \( Y_i \). Unlike other hierarchal linear modelling approaches (HLM), separate equations at different levels are not defined, rather the standard errors of the parameter estimates are adjusted. For more information readers are referred to Muthén and Satorra (1995).

While the data is also nested within classrooms, this information was not included in the dataset and so was not clustered for. The data was assumed to be missing at random and was handled by the Full Information Maximum Likelihood (FIML) method in
MPLUS which has been found to outperform other methods, such as listwise deletion (Little & Rubin, 2014). MPLUS was chosen as the programme as it can do FIML in the presence of clustered data while other programmes do not (Muthén & Muthén, 1998–2017). FIML does not impute data, instead it uses all possible information to inform the parameter estimates. The likelihood thus incorporates the information provided by those individuals with missing data rather than throwing their information out as in listwise deletion. FIML and Multiple Imputations are asymptotically equivalent, meaning that as the sample size increases both methods will provide the same results. If the data are Missing Completely at Random (MCAR) or Missing at Random (MAR) the resulting estimates and standard errors will be unbiased. For a fuller treatment of the FIML, readers are referred to Little & Rubin, 2014. The missing data mechanism was examined in the data.

The two regressions have SBA 2015 and SBA 2016 as outcomes and relevant controls (ethnicity, special education status, migrant status, and free and reduced lunch status). For the SBA 2016 regression both the SBA 2015 and the MSP 2014 were included as covariates while the SBA 2015 regression only had the MSP 2014 as a covariate. These results were a result of convenience as they were collected and available. Nevertheless, the researchers felt confident in the results especially with the same standardised test being given across time allows for a more longitudinal approach which allows greater statistical power.

Results

Table 1 shows the correlations of the variables used in the models.

The negative statistically significant result of OER status with SBA (2015) scores is deceptive, as the clustering of the data (students nested within schools) is not accounted for by SPSS in the calculation of the correlation. This result was corrected by subsequent analyses done in MPLUS 8.0.

Table 2 shows the descriptives of the variables used in the model broken down according to OER use or traditional curriculum.

The missingness is quite high for the outcome variables, especially SBA (2016) and the MSP (2014) scores. Nevertheless, there are no significant differences in the test scores between groups. The selected demographics are different between the groups with more Asians in the control group (11%) than the OER group (3%), more Latinos in the control group (27%) than the OER (12%) and less Whites in the control group (8%) than the OER group (12%). Special Ed percentages are equal between the groups, while migrant status is higher in the control group (4%) than the OER group (1%). Finally, Free or Reduced lunch status is slightly lower in the control group (52%) than the OER group (58%). As mentioned in the data analysis strategy section, missingness is handled via the FIML method. While the missingness is high on one of the outcomes of interest (SBA 2016) nevertheless the percent of missing is nearly identical across the groups (56.3% control, 56.2%) which shows that differential attrition is not a problem. When the missingness for SBA (2016) was regressed on the other variables in the model in a logistic regression, weak relations were observed between the control variables SBA (2015) and the MSP (2014) scores, nevertheless the model fit was fairly low with Cox and Snell R-square of .043. This indicates that the requirements for FIML are met and the results will be trustworthy. Cohen’s $d$ was calculated for the two main outcomes of interest (SBA (2016) and SBA (2015)) with a confidence interval performed (Cumming,
Table 1. Correlations of variables used in models.

|        | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1. SBA (2016) Score | 1.000 |       |     |     |     |     |     |     |     |     |     |
| 2. SBA (2015) Score | 0.730** | 1.000 |     |     |     |     |     |     |     |     |     |
| 3. MSP (2014) Score | 0.573** | 0.759** | 1.000 |     |     |     |     |     |     |     |     |
| 4. OER Status | -0.017 | -0.044** | 0.027 | 1.000 |     |     |     |     |     |     |     |
| 5. Asian | 0.047** | 0.133** | 0.037 | -0.152** | 1.000 |     |     |     |     |     |     |
| 6. Black | -0.040** | -0.071** | -0.056** | 0.000 | -0.062** | 1.000 |     |     |     |     |     |
| 7. Latino | -0.089** | -0.171** | -0.173** | -0.184** | -0.145** | -0.111** | 1.000 |     |     |     |     |
| 8. White | -0.004 | -0.020* | 0.002 | 0.071** | -0.094** | -0.072** | -0.169** | 1.000 |     |     |     |
| 9. Special Education Status | -0.357** | -0.347** | -0.247** | 0.001 | -0.056** | -0.001 | 0.040** | 0.005 | 1.000 |     |     |
| 10. Migrant Status | -0.032* | -0.086** | -0.145** | -0.135** | -0.036** | -0.034** | 0.233** | -0.052** | -0.024** | 1.000 |     |
| 11. Free or Reduced Lunch | -0.126** | -0.313** | -0.290** | 0.061** | -0.080** | 0.072** | 0.255** | -0.006 | 0.094** | 0.143** | 1.000 |

*p < .05, **p < .01.
Table 2. Descriptive statistics, and sample size and differences.

|                | OER (n = 5314) |                  |                  |                  |                  | Control (n = 6796) |                  |                  |                  |
|----------------|----------------|------------------|------------------|------------------|------------------|-------------------|------------------|------------------|------------------|
|                | Mean           | S.D.             | Min              | Max              | N                | Mean              | S.D.             | Min              | Max              | N                | % miss | Sig       | Mean            | S.D.          | Min              | Max              | N                | % miss | Sig       |
| 1. SBA (2016)  | 2482.75        | 279.25           | 100              | 2801             | 2325             | 56%               | 2492.66          | 285.75           | 100              | 2780             | 2971             | 56%               |       |         |
| 2. SBA (2015)  | 2492.65        | 270.47           | 100              | 2700             | 5275             | 7%                | 2471.03          | 296.01           | 100              | 2700             | 6625             | 3%                 |       |         |
| 3. MSP (2014)  | 363.35         | 118.22           | 0                | 575              | 647              | 87%               | 373.34           | 97.76            | 0                | 575              | 1704             | 75%               |       |         |
| 5. Asian       | 0.03           | NA               | 0                | 1                | 5314             | 0%                | 0.11             | NA               | 0                | 1                | 6796             | 0%                 | **    |      |
| 6. Black       | 0.05           | NA               | 0                | 1                | 5314             | 0%                | 0.05             | NA               | 0                | 1                | 6796             | 0%                 | **    |      |
| 7. Latino      | 0.12           | NA               | 0                | 1                | 5314             | 0%                | 0.27             | NA               | 0                | 1                | 6796             | 0%                 | **    |      |
| 8. White       | 0.12           | NA               | 0                | 1                | 5314             | 0%                | 0.08             | NA               | 0                | 1                | 6796             | 0%                 | **    |      |
| 9. Special Education Status | 0.15     | NA               | 0                | 1                | 5314             | 0%                | 0.15             | NA               | 0                | 1                | 6796             | 0%                 | **    |      |
| 10. Migrant Status | 0.01 | NA               | 0                | 1                | 5314             | 0%                | 0.04             | NA               | 0                | 1                | 6796             | 0%                 | **    |      |
| 11. Free or Reduced Lunch | 0.58 | NA               | 0                | 1                | 5314             | 0%                | 0.52             | NA               | 0                | 1                | 6796             | 0%                 | **    |      |

Sig shows differences in means (t-test) or proportions (Chi-Square) between groups. ** p < .10.
The results for SBA (2016) was a $d$ of $−0.03$ with a confidence interval of $(-0.08, 0.02)$ showing null effect. The results for SBA (2015) was a $d$ of $−0.004$ with a confidence interval of $(-0.04, 0.03)$ again showing a null effect.

The multiple regressions of SBA (2015) and SBA (2016) are shown in Table 3. In both of those models OER was not a significant predictor of the SBA test scores ($p > .05$). Nevertheless, the models were rerun without the MSP (2014) as a covariate given the scarcity of the data. This scarcity occurred because the data produced included all students who were registered for the year 2015; however, many of these students were missing data due to their participation in the 2014 SBA pilot test. In this analysis the OER effect becomes negative and statistically significant ($B = −12.021, p = 0.011, \beta = −0.076$) for the SBA (2015). This change does not occur with the SBA (2016) results ($B = 13.599, p = 0.086, \beta = 0.025$) as the results are still not statistically significant whether MSP (2014) is included as covariate or not. The results where the MSP (2014) is included as a covariate (Table 3) should be more accurate as it includes more information on these students.

In order to examine whether the relative alignment with Common Core standards influenced end-of-year exam scores we compared the results of students who used Math Expressions and Math Connects. Both of these groups were combined together as a control group in the previous analyses. We found no significant differences in test scores between the students who used these two different curricula. Furthermore, we compared the two districts that used OER with each other and likewise found no significant differences.

### Discussion

After examining student test scores across two years of using OER, we found that there were no statistically significant differences in the exam scores of students who used EngageNY’s open curriculum materials versus those who used commercial products when 2015 and 2016 scores were included (whether or not 2014 was included as a covariate). A key question concerns whether or not such non-significant results are important and bear reporting. From a theoretical standpoint, several researchers have discussed the vital importance of reporting non-significant findings. For example, Polanin, Tanner-Smith,
and Hennessy (2015) reviewed the work of several researchers who have argued that non-significant results need to be reported. They write,

The proliferation of dissemination biases has the potential to affect not only the validity of meta-analytic results, but also those of primary research. Dissemination biases create the illusion of theory conformation, potentially leading to the continuation of programs or policies that are ineffective, or worse, harmful. Moreover, continuation of funds to ineffective programs inhibits the growth of potentially new and important research. (p. 208)

Hence, reporting these statistically non-significant results is important.

While the results are not statistically significant, they may carry modest practical significance. If openly licensed resources like the EngageNY Math curriculum material can be substituted for traditional commercial resources without impacting student outcomes, then school districts may stand to save substantial amounts of money by using OER. This money could then be redirected to other aspects of pedagogy, potentially leading to larger changes in educational outcomes. Moreover, because OER can, by their nature, be adapted and modified, the potential exists for the materials to be improved in future iterations.

In the present study, we do not account for any potential cost savings. Depending on how OER are adapted and whether they are printed (and in what manner), adopting OER can provide extensive or no cost savings (Wiley et al., 2012). Nevertheless, the results of this study lead us to propose that much more research needs to be done in terms comparing the efficacy of a specific piece of curriculum materials with their cost. For example, if Curriculum A and Curriculum B cost the same amount of money, but Curriculum A leads to better outcomes, Curriculum A should be used. If Curriculum A costs four times more than Curriculum B and they both lead to the same educational outcomes, Curriculum B may be the obvious choice. Further questions such as “How much better does Curriculum A need to be than Curriculum B in order to justify a higher cost of X amount?” are also relevant. The emergence of high-quality OER should prompt further examination of questions regarding the learning return on curriculum investment.

**Conclusion**

The present study is the first efficacy study of OER materials in primary school (grades 3–6, approximately ages 9–12). Similar to Robinson et al. (2014) we found that there were no significant differences in standardised test scores when comparing students who use OER versus commercial materials. This finding is also in line with Slavin and Lake (2008) who synthesised several mathematics curriculum studies undertaken at the primary school level. Significantly, they found that the choice of curriculum has a smaller impact on student performance than instructional improvement (see also Slavin, Lake, & Groff, 2009). To the extent this is the case, there is merit to the possibility that substituting commercial curriculum for free OER could improve student learning if the resources allocated for curriculum were instead used in ways that would effectively change daily teaching practices.

We acknowledge that as a non-experimental study, this cannot be used to claim causality. Moreover, the present study only focuses on one collection of OER, namely the EngageNY curriculum materials; these results cannot be generalised to all OER. One methodological limitation of this study is that according to the Louisiana State Department of Education (2016) the three textbooks that we examined have different levels of alignment with the
Common Core Standards. While the SBA is not aligned to any particular set of curriculum materials, SBA does align to the Common Core Standards. Thus it may be that the difference in alignment was a significant factor in the present study. We attempted to address this limitation by comparing the results of the districts that used commercial curriculum materials that had different levels of alignment with Common Core standards. The finding that there was no difference in the test results of students in these districts may indicate a relatively small level of educational impact of these specific curriculum materials and/or the lack of alignment in these specific curriculum materials for this particular assessment.

Another limitation of the present study is that it does not attempt to examine the significant impact that teachers have on the use of curriculum materials. As Remillard (2005) states, the process of using a mathematics curriculum guide is complex and dynamic and is mediated by teachers’ knowledge, beliefs, and dispositions suggests that the decision to adopt a single curriculum in a school or district will not alone result in uniform mathematics instruction. (p. 239)

Baumert et al. (2010) show that teachers’ pedagogical content knowledge influenced the cognitive level of tasks with significant impact on pupil attainment in mathematics – this is something that the present study does not account for. Because we analyzed data that took part as part of a natural experiment (some districts adopting OER while others did not), it was not possible with our design to examine teacher differences and related important issues such as implementation fidelity or the quality of professional development teachers received. This lack of control admittedly makes it so this study does not meet the WWC criteria, a limitation of many other mathematics curriculum studies as discussed previously and noted by Slavin and Lake (2008).

While this is a major limitation and cannot be discounted, we believe that given the large amount of student data this study contains, the manner in which individual teachers used the three sets of curriculum materials may be less relevant than the overall results. Because there is no reason to suspect that there were systematic differences in how the curriculum materials were used or ignored, the key issue at hand is that students who used the open curriculum materials performed just as well as students who used the commercial materials. This may be because the materials are of equivalent quality, or because the materials have a relatively small overall effect, or for other reasons. Future studies could more carefully examine these issues by more carefully accounting for teacher effect by looking at student and teacher performance over longer periods of time. Another possibility for future research would be qualitative studies that examine and compare the actual use of OER and commercial materials in the classroom similar to Lloyd (2008).

Another aspect of OER adoption that bears more analysis in further studies concerns the potential influence of open licensing on OER adoption. In the present study, EngageNY Math was adopted wholesale by faculty, who generally did not take advantage of the additional legal permissions allowed by OER. It is possible that had teachers actively engaged in revising or remixing the OER, making them more directly relevant for their students, results would have been different. The extent to which teachers are willing to expend the time necessary to adapt OER remains unknown. Future research should more carefully examine this aspect of OER adoption.

The present study showed that across our sample of 12,110 students, clustered within 95 schools, there were no differences in the standardised end-of-year mathematics test scores
across two years when an open curriculum was substituted for a commercial curriculum. This study challenges the perception that OER are lower quality because the materials are available for free. While this is only one study, if multiple future studies replicate these non-significant findings, administrators could confidently consider implementing OER and identifying ways to more effectively reallocate funds designated for curriculum materials.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This work was supported by William and Flora Hewlett Foundation.

ORCID
John Hilton III http://orcid.org/0000-0002-4455-1125

References
Allen, G., Guzman-Alvarez, A., Molinaro, M., & Larsen, D. (2015, January). Assessing the impact and efficacy of the open-access ChemWiki textbook project. Educause Learning Initiative Brief. Retrieved from https://net.educause.edu/ir/library/pdf/elib1501.pdf
Allen, I., & Seaman, J. (2014). Opening the curriculum: Open educational resources in U.S. higher education. Retrieved from http://www.onlinelearningsurvey.com/reports/openingthecurriculum2014.pdf
Armstrong, J., & Bray, J. (1986). How can we improve textbooks? Publications, ECS Distribution Center, CO. Retrieved from http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED292208
Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). Report of the 2012 national survey of science and mathematics education. Chapel Hill, NC: Horizon Research, Inc.
Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., … Tsai, Y. (2010). Teachers’ mathematical knowledge, cognitive activation in the classroom and student progress. American Educational Research Journal, 47(1), 133–180.
Bissell, A. (2009). Permission granted: Open licensing for educational resources. Open Learning, The Journal of Open and Distance Learning, 24, 97–106.
Bloom, H. S., Michalopoulos, C., & Hill, C. J. (2005). Using experiments to assess nonexperimental comparison-group methods for measuring program effects. In H. S. Bloom (Ed.), Learning more from social experiments (pp. 173–235). New York: Russell Sage Foundation.
Bossu, C., & Tynan, B. (2011). OERs: New media on the learning landscape. On the Horizon, 19(4), 259–267.
Butcher, N. (2015). A basic guide to open educational resources (OER). Commonwealth of learning. Paris: United Nations Educational, Scientific and Cultural Organization.
Charalambous, C. Y., & Hill, H. C. (2012). Teacher knowledge, curriculum materials, and quality of instruction: Unpacking a complex relationship. Journal of Curriculum Studies, 44(4), 443–466.
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.
Crawford, D. B., & Snider, V. E. (2000). Effective mathematics instruction the importance of curriculum. Education and Treatment of Children, 23(2), 122–142.
Cumming, G. (2012). *Understanding the new statistics*. New York: Routledge. Retrieved from [http://proquest.safaribooksonline.com/9780415879675](http://proquest.safaribooksonline.com/9780415879675)

de los Arcos, B., Farrow, R., Pitt, R., Weller, M., & McAndrew, P. (2016). Personalising learning through adaptation: Evidence from a global survey of K-12 teachers’ perceptions of their use of open educational resources. *Journal of Online Learning Research*, 2(1), 23–40.

Fischer, L., Hilton, III, J., Robinson, T. J., & Wiley, D. A. (2015). A multi-institutional study of the impact of open textbook adoption on the learning outcomes of post-secondary students. *Journal of Computing in Higher Education*, 27(3), 159–172.

Harley, D. (2008). Why understanding the use and users of open education matters. In T. Iiyoshi, & M. S. Vijay Kumar (Eds.), *Opening up education: The collective advancement of education through open technology, open content, and open knowledge* (pp. 197–211). Cambridge, MA: MIT Press.

Hill, H. C., & Charalambous, C. Y. (2012). Teaching (un) connected mathematics: Two teachers’ enactment of the Pizza problem. *Journal of Curriculum Studies*, 44(4), 467–487.

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), 573–590.

Jung, Y. (2016). Some Sunnyside schools to drop textbooks for open-source material. Retrieved from [http://tucson.com/news/local/education/some-sunnyside-schools-to-drop-textbooks-for-open-source-material/article_40ec6f4d-e9e7-54b1-9a2f-b61c5046c604.html](http://tucson.com/news/local/education/some-sunnyside-schools-to-drop-textbooks-for-open-source-material/article_40ec6f4d-e9e7-54b1-9a2f-b61c5046c604.html)

Kimmons, R. (2015). OER quality and adaptation in K-12: Comparing teacher evaluations of copyright-restricted, open, and open/adapted textbooks. *The International Review of Research in Open and Distributed Learning*, 16(5). doi:10.19173/irrodl.v16i5.2341

Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Lloyd, G. M. (2008). Teaching mathematics with a new curriculum: Changes to classroom organization and interactions. *Mathematical Thinking and Learning*, 10(2), 163–195.

Louisiana Department of Education. (2016). Curricular resources annotated reviews. Retrieved from [http://www.louisianabelieves.com/academics/ONLINE-INSTRUCTIONAL-MATERIALS-REVIEWS/curricular-resources-annotated-reviews](http://www.louisianabelieves.com/academics/ONLINE-INSTRUCTIONAL-MATERIALS-REVIEWS/curricular-resources-annotated-reviews)

Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. Marsden (Ed.), *Sociological Methodology* (pp. 216–316). Oxford: Blackwell.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user’s guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

National Research Council. (2004). On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations. (J. Confrey & V. Stohl Eds.). Washington, DC: National Academies Press.

OPSI. (2016). Washington joins nationwide open resource initiative. Office of Superintendent of Public Instruction of the state of Washington News Release, 9/12/2016. Retrieved from [https://content.govdelivery.com/accounts/WAOSPI/bulletins/1638bec](https://content.govdelivery.com/accounts/WAOSPI/bulletins/1638bec)

Oviett, D. (2014). An analysis of textbook spending in K-12 public education in the State of Utah: A consideration of the effects of adopting open textbooks (Unpublished Masters Project).

Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2015). Estimating the difference between published and unpublished effect sizes a meta-review. *Review of Educational Research*, 86(1), 207–236.

Polikoff, M. S. (2015, December). How well aligned are textbooks to the common core standards in mathematics? *American Educational Research Journal*, 52, 1185–1211.

Remillard, J. T. (2005). Examining key concepts in research on teachers’ use of mathematics curricula. *Review of Educational Research*, 75(2), 211–246.

Robinson, T. J., Fischer, L., Wiley, D. A., & Hilton, J. (2014). The impact of open textbooks on secondary science learning outcomes. *Educational Researcher*, 43(7), 341–351.

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research*, 78(3), 427–515.

Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research*, 79(2), 839–911.
Song, J. S. (2018). District spotlight: Garnet valley school district in Joseph South and Kristina Ishmael, Creating Systems of Sustainability: Four Focus Areas for the Future of PK-12 Open Educational Resources. Retrieved from https://www.iste.org/docs/pdfs/new-america_oer_sustainability-research-paper_9-2018_v3_color_web.pdf?sfvrsn=2

Tarr, J. E., Reys, R. E., Reys, B. J., Chavez, O., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. Journal for Research in Mathematics Education, 39, 247–280.

Taylor, J. A., Getty, S. R., Kowalski, S. M., Wilson, C. D., Carlson, J., & Van Scotter, P. (2015). An efficacy trial of research-based curriculum materials with curriculum-based professional development. American Educational Research Journal, 52(5), 984–1017.

Tyson, H., & Woodward, A. (1989). Why students aren’t learning very much from textbooks. Educational Leadership, 47, 14–17.

UNESCO. (2002). Forum on the impact of open courseware for higher education in developing countries: Final report. Retrieved from www.unesco.org/iiep/eng/focus/opensrc/PDF/OERForumFinalReport.pdf

U.S. Department of Education. (2015). U.S. Department of Education launches campaign to encourage schools to #GoOpen with educational resources. Retrieved from http://www.ed.gov/news/press-releases/us-department-education-launches-campaign-encourage-schools-goopen-educational-resources

Usdan, J., & Gottheimer, J. (2012, February 3). FCC chairman: Digital textbooks to all students in five years [Blog post]. Retrieved from http://www.fcc.gov/blog/fcc-chairman-digital-textbooks-all-students-five-years

Utah State Office of Education (USOE, 2012). Utah state office of education to create open textbooks [Press Release]. Retrieved from http://www.schools.utah.gov/main/INFORMATION/Onlinenewsroom/DOCS/01252012OpenTextbook.aspx

Van den Heuvel-Panhuizen, M. (2000). Mathematics education in the Netherlands: A guided tour. Freudenthal Institute CD-rom for ICME9, 1–32. Retrieved from https://www.tandfonline.com/doi/pdf/10.1080/07370000802177235

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. Education Policy Analysis Archives, 10(12). Retrieved from http://epaa.asu.edu/epaa/v10n12/.

What Works Clearinghouse. (2013). WWC Intervention Report: Dreambox Learning. Retrieved from http://ies.ed.gov/ncee/wcc/pdf/intervention_reports/wcc_dreambox_121013.pdf

What Works Clearinghouse. (2015a). WWC Intervention Report: Singapore Math. Retrieved from http://ies.ed.gov/ncee/wcc/pdf/intervention_reports/wcc_singaporemath_122215.pdf

What Works Clearinghouse. (2015b). WWC Intervention Report: Everyday Mathematics. Retrieved from http://ies.ed.gov/ncee/wcc/pdf/intervention_reports/wcc_everydaymath_111015.pdf

Wiley, D., Bliss, T. J., & McEwen, M. (2014). Open educational resources: A review of the literature. In Handbook of research on educational communications and technology (pp. 781–789). New York: Springer.

Wiley, D., & Gurrell, S. (2009). A decade of development … Open Learning, 24(1), 11–21. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/02680510802627746

Wiley, D., Hilton, III, J., Ellington, S., & Hall, T. (2012). A preliminary examination of the cost savings and learning impacts of using open textbooks in middle and high school science classes. International Review of Research in Open and Distance Learning, 13(3), 262–276. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1153/2256

Wiley, D., Williams, L., DeMarte, D., & Hilton, J. (2016). The Tidewater Z-Degree and the INTRO model for sustaining OER adoption. Education Policy Analysis Archives, 23(41), doi:10.14507/epaa.v23.1828.

Woodward, A., & Elliot, D. L. (1990). Textbook use and teacher professionalism. In D. L. Elliott, & A. Woodward (Eds.), Textbooks and schooling in the United States (89th Yearbook of the National Society for the Study of Education, Part 1, pp. 178–193). Chicago: National Society for the Study of Education.