Educational Researcher

The Impact of Open Textbooks on Secondary Science Learning Outcomes
T. Jared Robinson, Lane Fischer, David Wiley and John Hilton III
EDUCATIONAL RESEARCHER 2014 43: 341 originally published online 15 September 2014
DOI: 10.3102/0013189X14550275

The online version of this article can be found at:
http://edr.sagepub.com/content/43/7/341

Published on behalf of

AERA AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

American Educational Research Association

and
SAGE
http://www.sagepublications.com

Additional services and information for *Educational Researcher* can be found at:

Open Access: Immediate free access via SAGE Choice

Email Alerts: http://er.aera.net/alerts

Subscriptions: http://er.aera.net/subscriptions

Reprints: http://www.aera.net/reprints

Permissions: http://www.aera.net/permissions

>> Version of Record - Oct 7, 2014

OnlineFirst Version of Record - Sep 15, 2014

What is This?

The Impact of Open Textbooks on Secondary Science Learning Outcomes

T. Jared Robinson[1], Lane Fischer[1], David Wiley[1], and John Hilton III[1]

Given the increasing costs associated with commercial textbooks and decreasing financial support of public schools, it is important to better understand the impacts of open educational resources on student outcomes. The purpose of this quantitative study is to analyze whether the adoption of open science textbooks significantly affects science learning outcomes for secondary students in earth systems, chemistry, and physics.

This study uses a quantitative quasi-experimental design with propensity score matched groups and multiple regression to examine whether student learning was influenced by the adoption of open textbooks instead of traditional publisher-produced textbooks. Students who used open textbooks scored .65 points higher on end-of-year state standardized science tests than students using traditional textbooks when controlling for the effects of 10 student and teacher covariates. Further analysis revealed statistically significant positive gains for students using the open chemistry textbooks, with no significant difference in student scores for earth systems of physics courses. Although the effect size of the gains were relatively small, and not consistent across all textbooks, the finding that open textbooks can be as effective or even slightly more effective than their traditional counterparts has important considerations in terms of school district policy in a climate of finite educational funding.

Keywords: open educational resources; open textbooks; science education; secondary education; propensity score matching

Introduction

For better or for worse, the textbook is the single most predominant curriculum delivery vehicle in schools in the United States (Jobrack, 2011). The textbook's role, however, extends beyond the dissemination of information. Textbooks play an important role in mediating the politics of what is taught, and even what methods are used to teach students.

Textbook spending data help to paint a picture of U.S. textbook use. According to the Federal Communications Commission (FCC), the United States spends more than $7 billion each year on textbooks for K–12 public schools (Usdan & Gottheimer, 2012). The FCC further points out that in many instances, this significant expenditure is insufficient to prevent outdated materials from being used in U.S. classrooms, where it is common to find textbooks that are 7–10 years old and contain outdated information (FCC, 2012).

The open educational resources (OER) movement seeks to address the cost issues associated with textbooks and the ability to flexibly present current, relevant content that is suited to how students learn. The William and Flora Hewlett Foundation, an early leader of the OER movement, defines open educational resources as "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 re-purposing by others. Open educational resources include full courses, course materials, modules, textbooks, streaming videos, tests, software, and any other tools, materials, or techniques used to support access to knowledge" (Hewlett, 2013). Thus, the "open" in open educational resources refers to the fact that these educational materials use copyright licenses that allow anyone to freely "reuse, revise, remix, and redistribute" the materials (Hilton, Wiley, Stein, & Johnson, 2010).

Much of the research and literature concerning OER adoption has focused on postsecondary education (see, e.g., Baraniuk & Burrus, 2008; Carson, 2006; Johansen & Wiley, 2010). However, OER is increasingly being adopted in K–12 settings as well. A recent report from the International Association for

[1]Brigham Young University, Provo, UT

K–12 Online Learning detailed OER-friendly legislation or policies in California, Florida, Maine, Maryland, Oregon, Texas, Utah, and Virginia (Bliss & Patrick, 2013). Particularly notable in this list are California and Texas, both of which have recently adopted legislative or policy initiatives facilitating and encouraging the adoption of OER in secondary education.

The number of secondary institutions adopting OER in place of traditional textbooks is increasing. For example, in 2009 the Open High School of Utah (now Mountain Heights Academy) became the first secondary school in the United States to adopt OER exclusively across its curriculum (Tonks, Weston, Wiley, & Barbour, 2013). From 2010 to 2012, science teachers in a Utah school district piloted the adoption of open science textbooks in biology, chemistry, and earth systems courses, with thousands of students using the open replacements for traditional science textbooks (Hilton, Wiley, Ellington, & Hall, 2012). Although open textbooks could be delivered electronically and include multimedia content not normally available in traditional textbooks, this pilot used printed versions of open textbooks that were similar in form and function to traditional textbooks. Even with open textbooks not being used digitally, but printed for distribution, the cost per textbook was approximately $5. The state of Utah is currently in the process of expanding this pilot statewide.

Although financial reasons to adopt OER might be persuasive to educational stakeholders, today's accountability-focused climate demands attention to the educational utility of OER, specifically as they relate to supporting student learning. Research addressing questions of OER effectiveness are noticeably absent from the literature. This may be, in part, due to the relative newness of OER and the limited instances of systematic adoption of OER. The purpose of this study is to analyze whether the use of open science textbooks significantly affected science learning outcomes in a group of 3,780 secondary students.

Review of Literature

Stern and Roseman (2004) pose the question, "Will better curriculum materials necessarily make a difference in student learning?" (p. 557). Intuitively, it seems that good curriculum materials should make a difference in student learning outcomes. Specifically, Chambliss and Calfee (1989) state, "Theory and practice both suggest that well-designed science textbooks can enhance student understanding" (p. 307).

The research into OER is still in a very early stage. As such, there are relatively few studies of the actual effectiveness of OER as a textbook replacement in educational settings. The work that has been done can be categorized into (a) frameworks for OER evaluation and (b) empirical research and evaluation of OER.

Frameworks for OER Evaluation

Two separate frameworks for OER evaluation were developed during the Open, Transferable, Technology-enabled Educational Resources (OTTER) project at University of Leicester, United Kingdom. Nikoi, Rowlett, Armellini, and Witthaus (2011) proposed the CORRE (content, openness, reuse, repurpose, and evidence) framework for the purpose of evaluating OER materials

or materials that could potentially be adopted into OER. The article suggests that switching from traditional materials to OER can be daunting and provides a workflow framework aiming to help teachers evaluate OER and create new, high-quality OER. Although the content and openness portions of CORRE specifically refer to the process of transforming materials into legally licensed OER, the reuse/repurpose and evidence elements of the framework present the authors' thinking on ways institutions can be involved in evaluating the quality and effectiveness of OER materials.

From the same OTTER Project, Nikoi and Armellini (2012) also developed and proposed the "OER mix framework" that examines adopters' purpose, process, product, and policy (the 4 Ps). The framework deals with the creation of OER and what variables influence the OER product that is shared with others. The authors suggested that different stances in regards to the four Ps can reflect fundamentally different stakeholder values and produce products with different strengths and weaknesses, and should, therefore, influence how OER is evaluated.

Clements and Pawlowski (2012) focused on teachers' perspective of the quality of OER. The study is relevant to the current study in two ways. First, Clements and Pawlowski identified a perceived lack of quality as one of the key barriers to broader adoption of OER. The authors draw from the quality literature in other fields to contend that quality directly relates to perceptions, and that there are different approaches to ascertaining the quality of OER.

The second contribution by Clements and Pawlowski was the creation and administration of a teacher survey aimed at measuring what teachers perceive as key to OER quality. Results from their survey indicated that most teachers want OER to employ multimedia, be accurate in terms of content, meet preestablished curricular guidelines, work well with their learning management system, and come from a reputable source. Although this study did not ask teachers to evaluate materials, the authors indicated that many teachers surveyed would be willing to serve on review boards for materials. Abeywardena, Raviraja, and Tham (2012) problematized peer review of OER, however, arguing that peer review is infeasible when resources are proliferated as quickly as OER and can be legally revised or remixed by any user.

Clements and Pawlowski (2012) examined quality from the perspective of teachers, highlighting what issues teachers found key for quality OER without providing a way to compare resources or even measure the quality of resources. Similarly, neither the Nikoi and Armellini (2012) OER mix framework nor the Nikoi et al. (2011) CORRE framework offers means or even justification for comparing curricular resources. In these frameworks, openness is itself the measure of a resource's desirability—that resources should be preferred as a function of their openness, with little regard to their broader quality compared to non-OER resources, such as traditional textbooks.

Empirical Research and Evaluation of OER

There has been some limited work done, though, in comparing OER replacements for textbooks to the non-OER materials they replaced. For example, Petrides and Jimes (2008) conducted survey research and case study analysis on the use of the South

African Free High School Science Texts. Although the study did not form broad conclusions about the quality of OER compared to the textbooks they replaced, they did suggest that comparing the resources being developed to prior curriculum was an important factor in increasing textbook quality.

Additionally, some research has examined teacher and student perceptions of the quality of open textbooks used in the classroom. Bliss, Hilton, Wiley, and Thanos (2013) surveyed 125 students and 11 faculty members who were involved in a pilot of OER as textbook replacement in eight separate courses at seven U.S. colleges. Students in the survey were asked to rate the quality of textbooks in the class compared to traditional textbooks. Three percent felt that the open textbooks were of significantly lower quality than traditional textbooks, 67% responded that the quality was about the same, and 41% reported that the open textbooks were significantly better than traditional textbooks. In open-ended responses, students reported ease of understanding, organizational features, the online nature of the books, and visual appeal of the books as reasons to prefer the open textbooks. Among faculty adopters, 5 out of the 11 faculty members were actively involved in creating the texts. The six faculty members who did not create the texts were asked to compare the quality of the open textbook to traditional textbooks. All six responded that the quality was about the same. Although these results seem to suggest that the quality of OER may be comparable to the textbooks they replace, it is worth questioning whether perceptions of quality are reliably correlated with student learning.

Hilton and Laman (2012) compared the performance of 690 students using an open textbook in an introductory psychology class to the performance of 370 students who used a traditional textbook in a previous semester. They concluded that students who used the open textbook achieved better grades in the course, had a lower withdrawal rate, and scored better on the final examination. The researchers acknowledge that the design elements of the study are insufficient to qualify the study as an experiment and so the research is discussed as a case study, and suggestions are made for more rigorous causal research in the future.

This research by Hilton and Laman does, however, suggest a fundamentally distinct way to think about evaluating textbook quality. By treating student learning outcomes as a dependent variable, researchers can design experiments and quasi-experiments to attempt to quantify differences in textbook quality, not in terms of perceptions of quality, but in terms of outcomes. Although there are not many studies of this sort in the emerging OER literature, examples of textbook evaluation using variations of this logic exist in the broader textbook literature (e.g., Chamberlin & Powers, 2007; Farragher & Yore, 1997; McCrudden, Schraw, Hartley, & Kiewra, 2004; Pyne, 2007). These studies highlight the need for creative and conscientious research design, accounting for competing explanations of results, and eliminating other potential threats to internal validity.

Method

Participants

The initial data set consisted of 4,183 students and 43 teachers from the Nebo School District in Utah who were enrolled in or taught science courses in 2012. Courses included regular sections (not AP) of earth systems, biology, and chemistry. Many of the teachers used traditionally published textbooks. A separate group of teachers created their own textbooks by revising and remixing OER originally published by the CK-12 Foundation. Before the 2010–2011 school year, a subgroup of six teachers had met in the summer to select content for inclusion, sequence that content, and add new sections where necessary. A grant from the William and Flora Hewlett Foundation paid for print copies of these materials for each student in their sections. Students were given the printed copies of the open textbooks to keep, which meant that students could highlight and take notes in the books if they desired. Before the 2011–2012 school year, 18 teachers in the district followed the same process of consultation, selection, printing, and delivery to students. In the 2011–2012 school year, approximately 43% of all students in earth systems, biology, or chemistry courses in the district used OER textbooks throughout the course of the school year.

University IRB and the school district approved the release of a deidentified, existing data set regarding these students and teachers. Student data included which Criterion Referenced Test (the end-of-year, state standardized test) was taken in 2012 (earth systems, biology, or chemistry), the item response theory (IRT) scaled score on that test, and a proficiency score on a scale of 1–4, with students scoring 3 or higher labeled as proficient. In addition, each student was identified by the type of textbook used, traditional or openly licensed. The data set also included student covariate data that could theoretically affect science achievement. These covariates included 2011 GPA, 2011 science test taken, and 2011 science criterion-referenced test (CRT) scores as estimates of students' general academic ability, motivation, and science ability. Also included were student age, gender, race, English proficiency, year in school, special-education status, eligibility for free- and reduced-price lunch, and teacher.

The data set also included the previous test performance for each teacher's students. These variables included the percentage of each teacher's students that were previously rated as proficient and their average IRT scaled scores for 2012, 2011, 2010, and 2009.

Because none of the sections taught using OER textbooks were designated special education sections, we omitted four special education sections with 47 total students from the control group in order to balance groups across treatment condition. We maintained students in the data set that had a special-education designation but who took the class mainstreamed with other students. The data set also had 357 students who were missing data for their 2011 GPA and science CRT and its accompanying score. We removed these students from the data set as well, which resulted in a final data set of 3,780 students. Student demographic data are depicted in Table 1.

Propensity Score Matching

Because students were not randomly assigned to textbook condition, any perceived differences in student achievement across textbook condition may have been due to systematic differences in groups rather than any specific feature of textbooks. When random assignment is not feasible, propensity score matching

Table 1
Participant Demographic Data Before and After Propensity Score Matching

Demographic characteristic	Unmatched Control	Unmatched Treatment	Unmatched Total	Matched Control	Matched Treatment	Matched Total
Number of students	2,168	1,612	3,780	1,274	1,274	2,548
Age (M)	15.69	16.09	15.99	15.99	15.98	15.99
Female	50.12	50.46	50.32	50.08	49.69	49.88
Male	49.88	49.54	49.68	49.92	50.31	50.12
ESL	0.78	1.30	1.01	0.94	1.02	0.98
Asian	0.55	0.19	0.40	0.31	0.23	0.27
Black	0.32	0.31	0.32	0.47	0.31	0.39
White	89.21	86.60	88.10	88.30	87.52	87.91
Hispanic	7.66	10.36	8.81	8.30	9.50	8.90
Native American	0.60	0.74	0.66	0.55	0.63	0.59
Pacific Islander	0.69	0.93	0.79	0.78	0.86	0.82
Multirace	0.97	0.87	0.93	1.20	0.94	1.10
9th-graders	62.41	31.76	49.34	39.08	39.80	39.44
10th-graders	27.40	50.43	37.22	44.82	44.98	44.90
11th-graders	9.87	16.87	12.86	15.54	14.68	15.11
12th-graders	0.32	0.93	0.58	0.55	0.55	0.55
Special education	6.55	7.51	6.96	6.99	7.22	7.10
Mean GPA	3.48	3.35	3.43	3.44	3.40	3.42
Students in biology	37.87	44.91	40.87	38.14	39.17	38.66
Students in chemistry	21.96	32.57	26.48	34.30	32.65	33.48
Students in earth systems	40.18	22.52	32.65	27.55	28.18	27.86

Note. Values are percentages unless otherwise noted. ESL = English as second language; GPA = grade point average.

can be used as a way to create matched groups that are equal in selection-expectation. A propensity score is defined as the conditional probability that an individual would be assigned to the treatment condition, given a set of relevant covariates. In other words, given what is known about students based on the covariate data, would they reasonably be found in either the treatment or control groups. To the degree that the covariates collected reflect potential differences across groups that might affect the outcome variable, propensity score matching approximates the design effects of random assignment (Guo & Fraser, 2010). Propensity score matching is a popular method of approximating random assignment in educational settings where such assignments are logistically difficult to achieve (e.g., Henson, Hull, & Williams, 2010; Riegle-Crumb & King, 2011; Stuart, 2007).

We used logistic regression to create propensity scores, regressing treatment condition on the covariates age, race, English language proficiency, year in school, special-education status, free and reduced-price lunch, 2011 GPA, 2011 CRT scores, the 2011 test taken, and the 2012 test taken. Using the R package *MatchIt*, we created a matched data set using nearest-neighbor matching within calipers (Guo & Fraser, 2010). In accordance with the recommendations of Rosenbaum and Rubin (1985), we used the formula $\varepsilon \leq .25\sigma\rho$, where ε is the caliper and $\sigma\rho$ indicates the standard deviation of the propensity scores of the sample. This resulted in a caliper of .04 for this study. The resulting data set consisted of 2,548 students, equally distributed across treatment condition. In other words, each student in the

treatment condition (open textbooks) was matched with the most statistically similar student from the pool of all students in the control condition (traditional textbooks) based on the available set of covariates listed above. There were 1,274 students in each condition, treatment and control. Propensity score matching led to a 96.99% balance improvement over the original sample (see Figure 1).

In order to facilitate comparative analysis of each specific subject group, we also used propensity score matching to create three additional smaller matched data sets—earth systems (*n* = 664), biology (*n* = 960), and chemistry (*n* = 784). In each case, we used the formula $\varepsilon \leq .25\sigma\rho$ to calculate calipers for each group.

Teacher Effect

In addition to controlling for student characteristics that might compete with treatment to explain student performance, controlling for teacher effects was essential to estimating any effect that might be due to textbook. This was particularly important considering that in this sample, every teacher used either a traditional textbook or an open textbook for all of their classes. Without controlling for teacher effect, any observed differences in student performance across textbook selection would be perfectly confounded with potential effects due to the effect of individual teachers. This was especially problematic considering that teachers independently chose whether or not to use the open textbooks. It was possible that teachers choosing to use the open
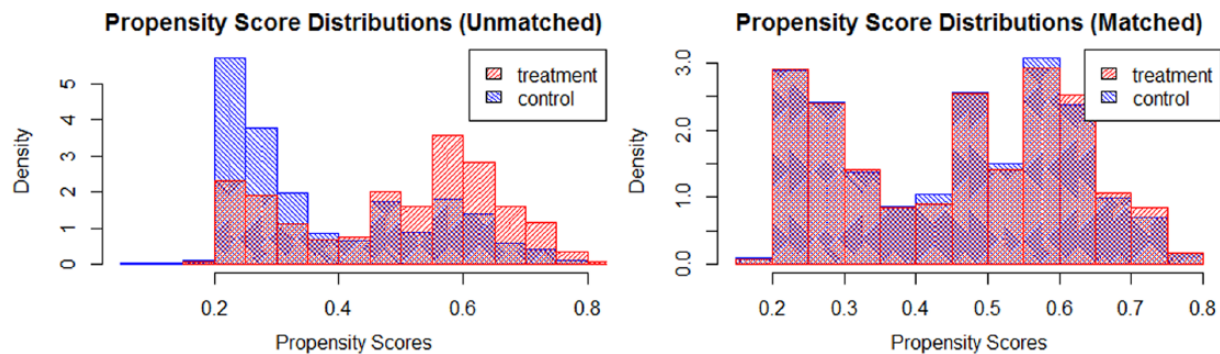
FIGURE 1. *Treatment and control propensity score distributions for the unmatched and the matched data sets*

textbook differed in systematic ways from teachers who chose to use traditional science textbooks.

We estimated the teacher effect initially based on the percentage of teachers' students labeled as proficient (scoring a three or better) for each teacher in each subject area for 2012, 2011, 2010, and 2009. We then converted these percentages to z scores. For teachers who used the traditional textbook, we used the 2012 z-score for percent proficient as an indicator of teacher effectiveness. For teachers using the open textbooks, we used the z-score for the most recent year that the teacher taught using a traditional textbook (either 2011 or 2010). For the two teachers who adopted the OER textbook in the same year that they began teaching in the school district, we imputed the mean of the distribution, which in the case of z scores, was zero. In order to evaluate the stability of this estimate of teacher effect, we correlated scores from 2010, 2011, with scores from 2012. The 2011 percent proficient scores were correlated with 2012 percent proficient scores at $r = .88$, whereas the 2010 scores correlated with 2012 scores at $r = .77$.

Data Analysis

After using propensity score matching to create the two groups and estimating teacher effects, we used ordinary least squares regression to examine the effect of textbook condition on student science achievement in the presence of multiple covariates. This analytical strategy was chosen because of its flexibility in estimating the covariance of both categorical and continuous scaled variables. We regressed IRT scaled science CRT scores on textbook condition and 10 covariates that might also have had an effect on outcome. We included the following 10 covariates: 2011 GPA (which might indicate general academic ability), 2012 science CRT (some of which might simply be harder than others), 2011 science CRT (which might indicate experience with a more difficult test), 2011 CRT scaled score (which might indicate aptitude or interest in science), gender (which might show differences in acculturation to science), age (which might show general cognitive maturity), English language proficiency (which might impact reading comprehension on the tests), special education status (which might include any number of cognitive challenges), free- and reduced-price lunch as a surrogate for family income (which might indicate parental education and access to opportunities for enriched learning outside of school), and teacher effect (which might indicate quality of teaching/

pedagogy). We believed that any textbook effect that emerged as significant in the presence of ten such covariates would be reasonably un-confounded and trustworthy.

We then repeated this analysis separately for earth systems, biology, and chemistry to explore whether the omnibus pattern held across all courses. Where necessary, we removed covariates from the regression analyses at the course level where there was no variation in that group. For example, all earth systems students took the exact same 2011 CRT, so this was removed from the analysis.

Results

As shown in Table 2, most of the covariates were significant in the presence of each other. Gender was not a significant predictor nor was free- and reduced-price lunch or the 2011 chemistry CRT. Students' previous science CRT scores and teacher effect accounted for the greatest proportion in variance in the 2012 science CRT scores, and all other covariates were significant predictors of 2012 CRT scaled scores. Even after accounting for all other covariates' influence on science performance, there was a significant difference in the treatment and control groups. Students that used open textbooks scored significantly higher than students that used traditional texts.

We analyzed two additional models to examine whether OER resources equally met the needs of low-income students and of male/female students. Because we found no significant moderation of treatment effect by income or gender by textbook, we rejected these models in favor of the more parsimonious model presented in Table 2.

Our research question dealt with whether the choice of open textbooks had a significant impact on student science learning. These results indicate that students who used open textbooks scored .65 points higher on the science CRTs than they would have scored if they had used traditional textbooks, even controlling for the effects of teacher, gender, socioeconomic status, science ability, prior academic achievement, prior science training, and student age. This difference was significant at $\alpha = .05$, $p = .008$. This increase in CRT scores is relatively small, however, when examined as an effect size $\beta = .03$, where β represents the standardized beta weight for the textbook variable in the regression analysis. It is interesting to note that variance due to textbook was still statistically significant, even in the presence of eight other significant predictor variables. The overall model

Table 2
Simultaneous Regression of 2012 Scaled Scores

Predictor	Estimate (B)	SE	β	t
Intercept	57.01	5.12		11.13***
Treatment	0.65	0.24	.03	2.67**
2011 GPA	1.64	0.23	.11	7.25***
2012 CRT[a]				
Chemistry	−6.59	1.01	−.57	−6.53***
Earth science	3.94	0.53	.11	7.44***
2011 CRT[a]				
Chemistry	3.91	2.51	.25	1.56
Earth science	−2.88	1.02	−.23	−2.83**
Physics	5.1	0.89	.25	5.75***
8th-grade science	−5.96	1.17	−.31	−5.08***
2011 scaled score	0.68	0.15	.06	44.42***
Gender[b]	1.16	0.24	.01	4.84***
Age	−0.49	0.26	−.03	−1.91
English language proficiency	−3.86	1.18	−.69	−3.27**
Special education[c]	−1.73	0.48	−.09	−3.62***
Free and reduced-price lunch[d]	−0.31	0.26	−.03	−1.19
Teacher effect	2.26	0.19	.21	11.45***

Note. $n = 2548$, $R^2 = .627$. CRT = criterion-referenced test; GPA = grade point average.
[a]The dummy-coded reference group for both the 2012 and 2011 CRTs is those who took the biology test.
[b]The dummy-coded reference group for gender is female.
[c]The dummy-coded reference group for special education is not special education.
[d]The dummy-coded reference group for free and reduced-price lunch is not free and reduced-price lunch.
*$p < .05$, **$p < .01$, ***$p < .001$.

R^2 value of .6273 indicates that this model accounts for approximately 63% of the variance in student science achievement on the science CRT scores.

These results are bolstered by the propensity score matching, which leads us to believe that the observed difference is not attributable to systematic student differences across treatment condition, and to conclude that students using open textbooks generally scored modestly higher on the state CRTs than they would have using traditional textbooks.

In order to obtain a more nuanced understanding of these results, we repeated the process separately for each course group: earth systems, biology, and chemistry. For each course, we used propensity score matching to balance covariates across treatment and control groups. We then used ordinary least squares regression to examine whether science CRT scores for students who used open textbooks differed from science CRT scores of students using traditional textbooks, controlling for the same sets of covariates as in the omnibus analysis. Results for earth systems are listed in Table 3, results for biology in Table 4, and results for chemistry in Table 5. Interestingly, textbook type was a nonsignificant predictor of student success for both earth systems and biology. In chemistry, however, students who used open textbooks performed significantly better than the control group, controlling for teacher, GPA, special-education status, gender, English language proficiency, the previous year's test and score, age, and socioeconomic status, $t = 2.49$, $p = .013$. In other words, students in the treatment group scored, on average, 1.23 points higher on the chemistry CRT, all other covariates held equal. As

in the omnibus test, the effect size of the difference in scores for chemistry students is relatively small as measured by the standardized beta weight, $\beta = .06$.

Discussion

Even though the increases in student CRT scores associated with open textbooks in the omnibus and chemistry analyses were statistically significant, they may have limited educational significance. The effect sizes as measured by standardized beta weights were relatively small compared to other predictors in the models. For example, in the omnibus analysis, the standardized beta weight of .03 was proportionally much smaller than standardized beta weights for other predictors such as prior GPA, $\beta = .11$ or teacher effect, $\beta = .21$ (see Table 2).

In chemistry, the treatment effect has a modest effect size of .06 compared to prior GPA, $\beta = .18$, or teacher effect, $\beta = .29$ (see Table 5). These findings conformed with our theoretical belief that teacher efficacy and prior ability would play a much more important role in educational achievement than textbook selection. In the model, students using open textbooks were predicted to score 1.23 points higher on the chemistry CRT. For chemistry students in the PSM matched data set, the mean IRT scaled score was 163.4 with a standard deviation of 9.36, where scores ranged from 130 to 192. So an expected 1.23-point increase was relatively small compared to the overall spread of chemistry CRT scores as expressed by the standard deviation.

Table 3
Simultaneous Regression of 2012 Earth Science CRT Scaled Scores

Predictor	Estimate (B)	SE	β	t
Intercept	46.45	10.67		4.35***
Treatment	−0.27	0.53	−.01	−0.5
2011 GPA	0.66	0.38	.05	1.73***
2011 scaled score	0.76	0.03	.75	26.41***
Gender[a]	1.43	0.48	.07	3**
Age	−0.58	0.66	−.02	−0.88
English language proficiency	−2.19	4.08	−.01	−0.54
Special education[b]	−1.96	0.72	−.07	−2.71**
Free and reduced-price lunch[c]	0.28	0.48	.01	0.58*
Teacher effect	−0.98	0.72	−.04	−1.36**

Note. $n = 664$, $R^2 = .652$. CRT = criterion-referenced test; GPA = grade point average.
[a]The dummy-coded reference group for gender is female.
[b]The dummy-coded reference group for special education is not special education.
[c]The dummy-coded reference group for free and reduced-price lunch is not free and reduced-price lunch.
*$p < .05$, **$p < .01$, ***$p < .001$.

Table 4
Simultaneous Regression of 2012 Biology CRT Scaled Scores

Predictor	Estimate (B)	SE	β	t
Intercept	54.46	10.01		5.44***
Treatment	0.22	0.35	.01	0.62
2011 GPA	1.47	0.37	.10	4.00***
2011 CRT[a]				
Chemistry	3.35	4.01	.34	0.84
Earth science	−4.83	3.25	−.25	−1.49
Physics	8.05	3.98	.52	2.02*
8th-grade science	−5.65	3.52	−.29	−1.61
2011 scaled score	0.65	0.02	.04	27.08***
Gender[b]	0.55	0.37	.03	1.48
Age	0.22	0.51	.02	0.44
English language proficiency	−2.31	1.28	−.04	−1.79
Special education[c]	−1.11	0.69	−.03	−1.61
Free and reduced-price lunch[d]	−0.93	0.40	−.06	−2.31*
Teacher effect	1.43	0.46	.12	3.09**

Note. $n = 960$, $R^2 = .668$. CRT = criterion-referenced test; GPA = grade point average.
[a]The dummy-coded reference group for the 2011 CRTs is those who retook the biology test.
[b]The dummy-coded reference group for Gender is Female.
[c]The dummy-coded reference group for Special Education is Not Special Education.
[d]The dummy-coded reference group for Free and Reduced Lunch is Not Free and Reduced Lunch.
*$p < .05$, **$p < .01$, ***$p < .001$.

It may provide some interpretive context to note that for the 2012 CRTs, the state of Utah categorized students into four categories based on IRT scaled CRT scores: "minimal proficiency," "partial proficiency," "sufficient proficiency," and "substantial proficiency." For chemistry, students above the cut score 156 were considered "proficient" for adequate yearly progress (AYP) evaluation purposes. More specifically, minimally proficient students scored below 151, partially proficient students scored from 152 to 159, sufficiently proficient students scored from 160 to 168, and substantially proficient students scored more than 169. Of the 784 students in the data set, 162 (20.66%) scored either at one of these cut scores or one score below. So although the increase in chemistry CRT scores associated with open textbooks might have limited educational significance, with textbooks being a relatively small component in student learning outcomes, there may be political significance for administrators or teachers concerned on a broader level with AYP. Additionally, the result that students using open textbooks demonstrated similar or slightly better learning outcomes may be important for science educators given budget constraints in secondary education.

Although the results from the omnibus analysis suggest that students who used open textbooks perform statistically significantly

Table 5
Table 5
Simultaneous Regression of 2012 Chemistry CRT Scaled Scores

Predictor	Estimate (B)	SE	β	t
Intercept	35.34	8.79		4.02***
Treatment	1.23	0.50	.06	2.49*
2011 GPA	3.82	0.60	.18	6.35***
2011 CRT[a]				
Earth science	3.21	1.91	.13	1.68
Physics	4.17	1.18	.22	3.54***
2011 Scaled Score	0.67	0.03	.04	22***
Gender[b]	1.55	0.47	.08	3.31***
Age	0.05	0.40	.00	0.12
English language proficiency	−10.22	4.41	−.06	−2.32*
Special education[c]	−10.89	3.64	−.02	−3.00**
Free and reduced-price lunch[d]	0.37	0.58	.02	0.64
Teacher effect	2.95	0.27	.29	10.82***

Note. $n = 784$, $R^2 = .561$. CRT = criterion-referenced test; GPA = grade point average.
[a]The dummy-coded reference group for the 2011 CRTs is those who took the biology test.
[b]The dummy-coded reference group for gender is female.
[c]The dummy-coded reference group for special education is not special education.
[d]The dummy-coded reference group for free and reduced-price lunch is not free and reduced-price lunch.
*$p < .05$, **$p < .01$, ***$p < .001$.

better on state science CRTs than students using traditional patterns, the three course-level analyses reveal that those results are likely a product of student gains in chemistry; students in earth systems and biology showed no significant difference in scores. Without speculating as to possible explanations for this pattern, it seems reasonable to suggest that not all open textbooks are created equal. It would be premature to suggest that every teacher who remixes their own open textbooks would expect to see uniform improvements in student performance.

The successful open textbook pilot in the Nebo school district, in which open textbooks, in some cases, improved student learning outcomes while providing a template for dramatically lowering the cost of providing access to core instructional materials for all students (the textbooks were free online or approximately $5 per printed book), has many implications. Below we discuss three: implications for access and equity, implications for teacher deskilling, and implications for the affordability of the transition from print to digital curriculum materials.

As budgetary pressures have increased, many districts now wait 7–10 years between textbook purchases. This prolonged delay between purchases results in textbooks that are out of date and badly damaged. Because these books must be handed down from student to student, students are not permitted to highlight or take notes in the books. Some districts have even moved to a "classroom set" model where the textbook to student ratio is 1:6 or 1:8. These changes in textbook acquisition behavior have created a crisis of access to curriculum material for students.

The open textbook model piloted in the Nebo school district, in which textbooks are available for $5 in print form and are freely available in digital form, provides a ready remedy to this crisis of access. At $5 per printed book, it becomes much more affordable to provide every child with their own copy of the core

instructional materials necessary to support learning. From this perspective, the open textbooks model has much to contribute to the broader dialog on equity and access.

The open textbooks model can also contribute meaningfully to the discourse on teacher deskilling. Deskilling is the separation of conception from execution (Apple, 1986, 1995). During the mid- and late twentieth century, the growing view that teachers lacked sufficient skills and content knowledge to successfully facilitate learning on their own led to attempts to create instructional materials that were "teacher proof," effectively cutting teachers out of the design process and relegating them to the role of implementers. After being forced into this role change, instruction can become "a managerial concern, not an educative one" for teachers (Shannon, 1989, p. 92).

In order to maximize profits, the creators of commercial curriculum materials actively pursue and protect their copyrights in these materials. This copyright protection prevents teachers from engaging in redesign and improvement activities with their textbooks. Although teachers have historically exercised a significant amount of autonomy once the classroom door was closed, the potentially illegal nature of redesigning curriculum (even in the context of Fair Use or TEACH Act claims) prevents these efforts from being widely viewed and valued. In this way, the adoption of copyrighted textbooks contributes directly to the deskilling of teachers and their sense that the curriculum is beyond their control. As Sizer (1984) wrote,

> Teaching often lacks a sense of ownership, a sense among the teachers working together that the school is theirs, and that its future and their reputation are indistinguishable. Hired hands own nothing, are told what to do, and have little stake in their enterprises. Teachers are often treated like hired hands. Not surprisingly, they often act like hired hands. (p. 184)

Adopting open educational resources and open textbooks puts ownership of curriculum directly back into the hands of teachers, both encouraging them to reflect on how the materials might be redesigned and improved and empowering them to make these improvements directly. These redesigns and improvements can then be broadly and legally shared. In these ways, adopting open textbooks combats the deskilling of teachers and reinstates them as skilled experts in both content and pedagogy.

Finally, the open textbooks model makes a significant contribution to our understanding of how to afford the transition from print to digital devices, sometimes called one-to-one computing. When schools and districts struggle to purchase new textbooks once per decade, the transition from print to digital can be daunting. In addition to continuing to pay for textbooks (now in electronic form), schools and districts must also find additional funding to purchase, maintain, and update digital devices (e.g., tablets or laptops) on which students will use the electronic textbooks. Although the prices of electronic textbooks appear to be lower than printed textbooks, most electronic textbooks are defective by design and contain code that causes them to self-delete or otherwise cease to function at the end of each school year so that they must be repurchased annually. Although a $20 electronic textbook sounds like a bargain compared to a $100 printed book, over a 7-year adoption cycle the electronic book costs more because it must be purchased seven times.

However, when a school or district makes the move to printed open textbooks they can achieve significant savings over traditional textbook adoptions, allowing them to begin saving for the transition to devices. This initial period during which teachers use open textbooks in print can provide teachers with the opportunity to become familiar with the new textbooks in a familiar format and begin the process of revising and improving the books.

Once a complete set of open textbooks for core subjects has been aggregated and deployed successfully in print, the next year's textbook budget can be used to acquire devices rather than textbooks. Savings from the period of print use can be used to support one-time transition costs. Instead of paying for content, schools and districts can now pay to acquire, maintain, and update devices. Content is free in the form of digital open textbooks, which teachers can continue to revise and improve as part of their professional development activities. In this way, the adoption of open textbooks can provide the financial support necessary to successfully transition from print to digital.

Limitations

Although this study's quasi-experimental design, use of propensity score matching, and covariate selection all facilitate more valid causal inference, it still has limitations. This study was not designed in a way to allow for conclusions about what specific aspects of the open textbooks caused student gains. One teacher in the study noted that because teachers only included content that was isomorphic with their curriculum and pedagogy, instruction seemed more seamless and rational to both teachers and students. Observed gains may be caused by things other than substantive content differences between textbooks. Students who used open textbooks each received their own copy of the book, whereas other students may have had limited access to classroom sets. Relatedly, students with open textbooks had the ability to annotate their books, which may, in turn, increase comprehension (Simpson & Nist, 1990). Alternatively, gains could be explained by the effect that the process of open textbook adoption may have had on teachers. Teachers who feel increased autonomy or investment in open textbooks might teach better, even if the quality of the texts is roughly equal. Although examining mediating influences of the effect of open textbooks on learning should be studied, the actual increase in student learning is important in its own right.

One important aspect of this study was controlling for the effect of individual teachers on student learning outcomes. Lacking some accounting for teachers, any result would be perfectly confounded with teacher self-selection into the treatment group. Although the correlation coefficients between previous years' scores and 2012 scores was shown to be high for control teachers, such a method would be controversial, at best, in any attempt to measure the value that any teacher brings to the classroom. We recognize that there many factors outside of teacher control that lead to correlated patterns of student scores across school years. So, although we do not advocate our method of quantifying teacher effect for value-added models or teacher evaluation, we believe that the attempt to account for these patterns is essential for the internal validity of our results.

Finally, we should exercise extreme caution in generalizing results beyond the population sampled or assuming this pattern of gains would happen in other locations. 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.

Conclusion and Future Research

The open textbooks pilot in the Nebo school district demonstrates that open textbooks can be used to improve learning outcomes while simultaneously dramatically lowering the cost of providing access to core instructional materials for all students. Partly because of the success of this pilot, the Utah State Office of Education (USOE) has begun a program to support teacher aggregation, alignment, and development of open science textbooks for statewide use in Grades 7 through 12. The most recent version of these six books passed the state's instructional materials review process in Winter 2013 and are now available for adoption statewide beginning in Fall 2013. The USOE has also undertaken an open textbooks initiative in secondary mathematics. This appears to be a generally fruitful collaboration between research and state-level policy, as opposed to the more cynical relationship described by Glass (1987).

Although we will research the statewide roll out of these texts, similar experiments must be conducted in other states in order to substantiate or repudiate our findings. Research regarding lobbying and other publisher responses to the move to open textbooks should be conducted in parallel.

Additionally, this study's finding of a significant effect for students in chemistry, with no significant effect in biology or earth

systems deserves further inquiry. Our research questions and design did not allow us to speculate on possible reasons for these differences. Future research, however, could explore how open content may interact with different subject matters, what text characteristics make some open textbooks better or worse than others, and what pedagogical decisions best coordinate with open textbook adoption.

New research is also needed to understand how the legal permissions associated with open textbooks may enable changes in pedagogy, assessment, and student engagement. When teachers feel genuine ownership of their materials, what changes in the way they teach? When students can become coauthors of those materials, what changes in the way they engage? Because open textbooks permit, in a legal sense, a wide range of revise and remix activities that have not been possible in the classroom historically, it will take time for teachers, students, administrators, and the public to understand their full potential to enable transformations in teaching and learning (Reich, Murnane, & Willett, 2012).

REFERENCES

Abeywardena, I. S., Raviraja, S., & Tham, C. Y. (2012). Conceptual framework for parametrically measuring the desirability of open educational resources using D-index. *International Review of Research in Open and Distance Learning*, *13*(2), 59–76.

Apple, M. W. (1986). *Teachers and texts: A political economy of class and gender relations in education*. New York: Routledge & Kegan Paul.

Apple, M. W. (1995). *Education and power*. New York: Routledge.

Baraniuk, R. G., & Burrus, C. S. (2008). Viewpoint: Global warming toward open educational resources. *Communications of the ACM*, *51*, 30–32.

Bliss, T. J., Hilton III, J., Wiley, D., & Thanos, K. (2013). The cost and quality of online open textbooks: Perceptions of community college faculty and students. *First Monday*, *18*, 1–7.

Bliss, T., & Patrick, S. (2013). OER state policy in K–12 education. Retrieved from http://www.inacol.org/resources/publications/inacol-reports/

Carson, S. (2006). *2005 program evaluation findings report*. Cambridge, MA: Massachusetts Institute of Technology.

Chamberlin, M. T., & Powers, R. A. (2007). Selecting from three curricula for a preservice elementary teacher geometry course. *Issues in the Undergraduate Mathematics Preparation of School Teachers, 4*.

Chambliss, M. J., & Calfee, R. C. (1989). Designing science textbooks to enhance student understanding. *Educational Psychologist*, *24*(3), 307–322.

Clements, K. I., & Pawlowski, J. M. (2012). User-oriented quality for OER: understanding teachers' views on re-use, quality, and trust. *Journal of Computer Assisted Learning*, *28*(1), 4–14.

Farragher, P., & Yore, L. D. (1997). The effects of embedded monitoring and regulating devices on the achievement of high school students learning science from text. *School Science and Mathematics*, *97*(2), 87–95.

Federal Communications Commission. (2012). FCC chairman Genachowski joins Secretary of Education Duncan to unveil new "digital textbook playbook," a roadmap for educators to accelerate the transition to digital textbooks. Retrieved from http://transition.fcc.gov/Daily_Releases/Daily_Business/2012/db0201/DOC-312244A1.pdf

Glass, G. V. (1987). What works: Politics and research. *Educational Researcher*, *16*(3), 5–10.

Guo, S., & Fraser, M. W. (2010). *Propensity score analysis. Statistical methods and applications*. Los Angeles: Sage.

Henson, R. K., Hull, D. M., & Williams, C. S. (2010). Methodology in our education research culture toward a stronger collective quantitative proficiency. *Educational Researcher*, *39*(3), 229–240.

Hewlett. (2013). Open educational resources. http://www.hewlett.org/programs/education-program/open-educational-resources

Hilton, J., & Laman, C. (2012). One college's use of an open psychology textbook. *Open Learning*, *27*(3), 265–272.

Hilton, J., Wiley, D., Ellington, S., & Hall, T. (2012). A preliminary examination of the cost savings and learning impacts of using open textbooks in high school science classes. *International Review of Research in Open and Distance Learning*, *13*(3), 262–276. Retrieved from http://hdl.handle.net/10515/sy5707x40

Hilton, J., Wiley, D., Stein, J., & Johnson, A. (2010). The four "R's" of openness and ALMS analysis: frameworks for open educational resources. *Open Learning*, *25*(1), 37–44.

Jobrack, B. (2011). *Tyranny of the textbook: An insider exposes how educational materials undermine reforms*. Lanham, MD: Rowman & Littlefield.

Johansen, J., & Wiley, D. (2010). A sustainable model for OpenCourseWare development. *Educational Technology Research and Development*, *59*, 369–382. doi: 10.1007/s11423-010-9160-7

McCrudden, M., Schraw, G., Hartley, K., & Kenneth, A. K. (2004). The influence of presentation, organization, and example context on text learning. *Journal of Experimental Education*, *72*(4), 289–306.

Nikoi, S., & Armellini, A. (2012). The OER mix in higher education: purpose, process, product, and policy. *Distance Education*, *33*(2), 165–184.

Nikoi, S. K., Rowlett, T., Armellini, A., & Witthaus, G. (2011). CORRE: a framework for evaluating and transforming teaching materials into open educational resources. *Open Learning*, *26*(3), 191–207.

Petrides, L., & Jimes, C. (2008). Building open educational resources from the ground up: South Africa's free high school science texts. *Journal of Interactive Media in Education*, *2008*(1).

Pyne, D. (2007). Does the choice of introductory microeconomics textbook matter? *Journal of Economic Education*, *38*(3), 279–296.

Reich, J., Murnane, M., & Willett, J. (2012). The state of wiki usage in U.S. K–12 schools: Leveraging Web 2.0 data warehouses to assess quality and equity in online learning environments. *Educational Researcher*, *41*(1), 7–15. doi: 10.3102/0013189X11427083

Riegle-Crumb, C., & King, B. (2010). Questioning a white male advantage in STEM: Examining disparities in college major by gender and race/ethnicity. *Educational Researcher*, *39*(9), 656–664.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33–38.

Shannon, P. (1989). *Broken promises: Reading instruction in twentieth century America*. Granby, MA: Bergin & Garvey.

Simpson, M. L., & Nist, S. L. (1990). Textbook annotation: An effective and efficient study strategy for college students. *Journal of Reading*, *34*(2), 122–129.

Sizer, T. R. (1984). *Horace's compromise: The dilemma of the American high school*. Boston: Houghton Mifflin.

Stern, L., & Roseman, J. E. (2004). Can middle-school science textbooks help students learn important ideas? Findings from project 2061's curriculum evaluation study: Life science. *Journal of Research on Science Teaching*, *41*(6), 538–568.

Stuart, E. A. (2007). Estimating causal effects using school-level data sets. *Educational Researcher*, *36*(4), 187–198.

Tonks, D., Weston, S., Wiley, D., & Barbour, M. (2013). "Opening" a new kind of school: The story of the open high school of Utah. *International Review of Research in Open and Distance Learning*, *14*(1). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1345/2419

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

AUTHORS

T. JARED ROBINSON is a doctoral candidate at Brigham Young University, 150-P MCKB, Provo, UT 84602; *t.jared.robinson@gmail.com*. His research focuses on open educational resource adoption in secondary and postsecondary education.

LANE FISCHER is an associate professor of counseling psychology and special education at Brigham Young University, 340-E MCKB, Provo, UT 84602; *lane_fischer@byu.edu*. His research has focused on psychometric issues in the assessment of sexual deviance and the prevention of educator sexual misconduct.

DAVID WILEY is currently a Shuttleworth fellow, an education fellow at Creative Commons, and adjunct faculty in the graduate program in instructional psychology and technology at Brigham Young University, 150 MCKB, Provo, UT 84602; *david.wiley@gmail.com*. He is also chief academic officer of Lumen Learning, an organization dedicated to increasing student success and improving the affordability of education through the adoption of open educational resources by middle schools, high schools, community and state colleges, and universities.

JOHN HILTON III is an assistant professor at Brigham Young University, 270 M, JSB, Provo, UT 84602; *johnhiltoniii@byu.edu*. He is the author of over thirty peer-reviewed articles and eight books. He has a variety of research interests including the processes of learning and teaching, and the effect of open educational resources.