Efficacy of Open Textbook Adoption on Learning Performance and Course Withdrawal Rates: A Meta-Analysis

Virginia Clinton
University of North Dakota, virginia.clinton@und.edu

Shafiq Khan

Follow this and additional works at: https://commons.und.edu/ehb-fac

Part of the Education Commons

Recommended Citation
https://commons.und.edu/ehb-fac/42

This Article is brought to you for free and open access by the Department of Education, Health & Behavior Studies at UND Scholarly Commons. It has been accepted for inclusion in Education, Health & Behavior Studies Faculty Publications by an authorized administrator of UND Scholarly Commons. For more information, please contact zeinebyousif@library.und.edu.
Efficacy of Open Textbook Adoption on Learning Performance and Course Withdrawal Rates: A Meta-Analysis

Virginia Clinton
Shafiq Khan
University of North Dakota

Open textbooks have been developed in response to rising commercial textbook costs and copyright constraints. Numerous studies have been conducted to examine open textbooks with varied findings. The purpose of this study is to meta-analyze the findings of studies of postsecondary students comparing learning performance and course withdrawal rates between open and commercial textbooks. Based on a systematic search of research findings, there were no differences in learning efficacy between open textbooks and commercial textbooks ($k = 22$, $g = 0.01$, $p = .87$, $N = 100,012$). However, the withdrawal rate for postsecondary courses with open textbooks was significantly lower than that for commercial textbooks ($k = 11$, OR (odds ratio) = 0.71, $p = .005$, $N = 78,393$). No significant moderators were identified. Limitations and future directions for research, such as a need for more work in K–12 education, outside of North America, and that better examine student characteristics, are discussed.

Keywords: open textbooks, course withdrawal, open educational resources, meta-analysis

Textbooks are useful and ubiquitous teaching and learning resources in education (Behnke, 2018). Commercial textbooks have been frequently criticized for their high prices, which have increased substantially in the past few decades (Blomgren, 2018; Perry, 2015). These prices contribute to the financial cost of education that is shouldered by taxpayers and students. As an alternative to commercial textbooks, open textbooks, which are electronically accessible without charging the user, have been developed (Smith, 2009). Many studies have examined differences in learning performance as well as course withdrawal rates between open and commercial textbooks (see Hilton, 2016, 2018, for reviews). However, results have been mixed with some showing positive results (Colvard, Watson, & Park, 2018; Hilton, Fischer, Wiley, & Williams, 2016), some showing negative results (e.g., Gurung, 2017), and some showing no difference between commercial and open textbooks (Allen et al., 2015; Grissett & Huffman, 2019; Medley-Rath, 2018). The purpose of this article is to meta-analyze the research findings comparing learning performance and course withdrawal rates between open and commercial textbooks. The findings from this meta-analysis may be used to inform instructors and institutions with their choice of textbooks.

Literature Review

The development of open textbooks is part of a broader movement in open educational resources (OER). OER is an umbrella term for a variety of learning materials, including textbooks, videos, online modules, music, brief readings, and music (Butcher, 2015). According to the William and Flora Hewlett Foundation (2019), which has supported OER for decades, OER are

teaching, learning, and research resources that reside in the public domain or have been released under an intellectual property license that permits their free use and re-purposing by others. (para. 7)

Textbooks were the focus of this manuscript for three reasons. One is that textbooks are commonly used learning resources throughout postsecondary education (Illowsky, Hilton, Whiting, & Ackerman, 2016). Another reason is that open textbooks are a type of OER that has a notable body of research on its efficacy (Weller, de los Arcos, Farrow, Pitt, & McAndrew, 2015). Finally, by focusing the analyses on a single type of OER (in this case, textbooks), the comparisons between OER and commercial resources are clearer.

A survey of postsecondary faculty found that many faculty express frustrations with the high costs of commercial materials, and a majority of faculty agreed that the high cost of commercial materials was a problem (Seaman & Seaman, 2018). However, the same survey indicated that one of the major concerns that faculty have about adopting open textbooks is whether their quality is comparable with commercial textbooks (Seaman & Seaman, 2018). Concerns about quality have been noted as a barrier to open textbook adoption in other surveys of faculty as well (Belikov & Bodily, 2016; Jhangiani, Pitt, Hendricks, Key, & Lalonde, 2016).
These concerns are not unwarranted. Indeed, in two studies comparing performance on researcher-developed, objective learning measures, students enrolled in courses with commercial textbooks outperformed students enrolled in courses with open textbooks (Gurung, 2017). However, there have been numerous studies on open textbooks indicating no meaningful differences in learning compared with commercial textbooks (e.g., Clinton, 2018; Engler & Shedlosky-Shoemaker, 2019; Jhangiani, Dastur, Le Grand, & Penner, 2018; Medley-Rath, 2018) that need to be considered alongside Gurung’s (2017) findings. For this reason, a meta-analysis in which the overall efficacy of open textbooks across studies was summarized would be informative for instructors, administrators, and policymakers.

There are reasons to expect that students in courses with open textbooks would outperform those in courses with commercial textbooks. According to the access hypothesis, many students do not purchase the textbook because of the cost (Grimaldi, Basu Mallick, Waters, & Baraniuk, 2019). Therefore, the use of an open-source textbook would allow more students to have access to the textbook, which would logically improve student performance on learning measures dependent on textbook access (e.g., quiz performance on required readings). In addition, having access to a textbook also provides students with a resource to help them better understand the content covered in class. Following this hypothesis, students should do better in courses with open textbooks than commercial textbooks.

Examinations of studies on open textbook efficacy have been the topic of two systematic reviews (Hilton, 2016, 2018) as well as nonsystematic, narrative reviews (Clinton, 2019a; Wiley, Bliss, & McEwen, 2014). In these reviews, the overall conclusions based on qualitative analyses were that OER, including open textbooks, were as effective for student learning as commercial materials. These reviews provide thoughtful, in-depth thematic analyses of the research on the topic. However, there are no published meta-analyses on this topic, to our knowledge, and conducting meta-analyses would build on these reviews by combining the effects of multiple studies to get overall quantitative effect sizes.

Given that random assignment is generally not feasible for comparing textbooks, nearly all of the research has involved quasi-experiments with varying levels of quality in terms of controlling for confounders, such as comparing courses taught by different instructors. These methodological limitations constitute a significant limitation in research on the efficacy of open textbooks (see Griggs & Jackson, 2017; Gurung, 2017, for discussions). A meta-analysis could examine some of the potential effects of these confounders by including moderator analyses based on study quality variables. The use of the same instructor in courses with open textbooks compared with commercial textbooks is important given the variability in grading practices and pedagogical quality across individual instructors (de Vlieger, Jacob, & Stange, 2017). Another confounder to consider is whether student prior achievement or knowledge varied when comparing courses. This is important to consider given that prior academic performance is a clear predictor of future performance on learning measures (Cassidy, 2015). Furthermore, in one study, a course with an open textbook had higher average grades than the same course with the same instructor using a commercial textbook, but the students in the course with the open textbook had stronger academic backgrounds based on high school grade point averages (Clinton, 2018). In addition, many studies used different instruments to measure learning (e.g., different exams and quizzes that contributed to final grades). The use of different instruments is a clear confounder when comparing learning efficacy from open and commercial textbooks as different instruments would likely yield different scores due to measurement error.

The availability of an open textbook could potentially influence postsecondary students’ decisions on whether to complete a course or withdraw from it. One reason postsecondary students state for withdrawing from a course is that they cannot afford the textbook (Michalski, 2014). Indeed, approximately 20% of postsecondary students report that the expense of a textbook motivated their decision to withdraw from a course (Florida Virtual Campus, 2016). These course withdrawals lead to losses of time and tuition for students and increase the amount of time it takes to complete a degree, making the overall cost of a postsecondary education higher (Boldt, Kassis, & Smith, 2017; Nicholls & Gaede, 2014). Students have indicated that having an open textbook for the course was a reason they completed that course (Hardin et al., 2019). It is possible that postsecondary students who are behind in a course may be more likely to complete the course if there is an open-access textbook they can use to access the material rather than spend hundreds of dollars for a commercial textbook. This could potentially lead to courses with open textbooks having lower withdrawal rates than courses with commercial textbooks. However, it should be noted that postsecondary students withdraw from courses for a multitude of reasons that are unrelated to finances, such as stressors in their personal lives or dislike of the instructor (Hall, Smith, Boeckman, Ramachandran, & Jasin, 2003; Michalski, 2014).

Research Questions

Three research questions guided this meta-analytic research:

Research Question 1: What is the efficacy of open textbooks compared with commercial textbooks in terms of student learning?

Research Question 2: What are the differences in withdrawal rates for courses with open textbooks compared with commercial textbooks?

Research Question 3: Do differences in confounders in the methodology moderate the results for Research Questions 1 and 2?
Method

A search and analysis plan for this meta-analysis was pre-registered prior to beginning this project (Clinton, 2019b).

Inclusion Criteria

In order to be included in this meta-analysis, the report needed to meet certain criteria. First, the report needed to have findings comparing student learning performance and/or withdrawal rates between open and commercial textbooks. It needed to report sufficient information to be used in the meta-analysis (e.g., means and standard deviations, t tests, withdrawal rate, number of students for each textbook type), or the authors provided the necessary information on request (e.g., Grewe & Davis, 2017; Lawrence & Lester, 2018). The report needed to be in English. Postsecondary students needed to be the participants because they are expected to purchase their course materials and are more directly affected by cost. There were no restrictions regarding the type of dissemination (e.g., journal articles, dissertations, conference presentations), and unpublished findings that were available to the authors of this meta-analysis were eligible.

Search Procedure

A systematic search for studies comparing learning measures and/or withdrawal rates between students using open and commercial textbooks was conducted in multiple steps. First, scholarly dissemination was searched for in the following databases: Scopus, DOAJ (Directory of Open Access Journals), ProQuest, PSYCinfo, and ERIC using the search terms open source textbook* and OER AND textbook were conducted in August of 2018 (see Appendix A for search terms used). This yielded 578 citations, of which 117 duplicates were removed. The titles and abstracts of these 461 citations were screened by the first author, and 97 were determined to be relevant (the internet-based tool Abstrackr was used for screening; Wallace, Small, Brodley, Lau, & Trikalinos, 2012). The full texts of the remaining 97 reports were examined. Based on examinations of the full texts, 20 reports were determined to be relevant (however, further on in analyses, three reports from this list had to be removed because the necessary statistics were not available, but there were backward and forward searches of citations for these reports as described in the following paragraph).

A list of these relevant reports found at this stage was compiled and sent to the Community College Consortium for Open Educational Resources–Advisory listserv, the authors of these reports were emailed, and a tweet was posted on this manuscript’s first author’s Twitter page with the hashtags #oer and #opensource requesting any additional reports or unpublished data (this yielded one additional report). The research reports posted on “The Review Project” of the Open Education Group’s website (https://openedgroup.org/) were also examined, which yielded one more report. After these relevant reports were identified, a backward search of the reference lists of these reports was conducted (no new reports were identified in this way). Then, a forward search of the citations of the relevant reports in Google Scholar was conducted, which identified one more report. The titles and abstracts of all presentations at Open Education 2018 were examined, which led to the identification of one more report. The systematic reviews of OER conducted by Hilton (2016, 2018) were examined, but no additional relevant studies were found. This led to a total of 22 reports with 23 independent studies (Gurung, 2017, reported two independent studies). This process ceased in October 2018. Twenty-two of these studies were analyzed in the learning efficacy meta-analysis (Table 1). Eleven of these studies were analyzed in the course withdrawal meta-analysis (Table 2).

Coding of Studies

To provide descriptive information, assess study quality, and obtain information for moderator analyses, the studies were coded using the criteria found in Appendix B. Items pertaining to study quality were based on the Study Design and Implementation Assessment Device (Valentine & Cooper, 2008). The Study Design and Implementation Assessment Device provides researchers with questions to study quality regarding construct, internal, external, and statistical conclusion validity. The studies were coded and double-coded by the first author. In addition, research assistants coded 25% of the reviews (intrarater reliability was good, κ = 0.78; see Follmer, 2018, for a similar approach). The first author resolved disagreements.

Statistical Procedures

To summarize the findings across learning measures, Hedges’s g was used. Hedges’s g is a standardized mean difference metric that is bias-corrected based on sample sizes (Enzmann, 2015; Hedges, 1981). To calculate Hedges’s g for each learning measure, either descriptive statistics (e.g., means and standard deviations, number of students passing a course) or inferential statistics (e.g., t tests) in addition to sample sizes entered into Comprehensive Meta-Analysis software (Version 3; Biostat, Englewood, NJ; note that some effect sizes based on dichotomous dependent variables were converted from the odds ratio to standardized mean differences following Polanin and Snilsstveit, 2016). Corresponding authors contacted with requests for any missing statistical information that was necessary. If the relevant statistics could not be obtained, the study was not included (as shown in Figure 1, this was the case for one full text report). A positive Hedges’s g indicates that the mean values for open textbooks were greater than those for commercial textbooks.
<table>
<thead>
<tr>
<th>Author/s (Year)</th>
<th>Sample Size (Open and Commercial)</th>
<th>Learning Measure</th>
<th>Content Area(s)</th>
<th>Institution Type (Country)</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allen et al. (2015)</td>
<td>Open = 377, commercial = 348</td>
<td>Normalized learning gain on standardized assessment of chemistry knowledge</td>
<td>Hard science</td>
<td>4-year public university (United States)</td>
<td>There were substantial changes to the course beyond adopting an open-source textbook</td>
</tr>
<tr>
<td>Basu Mallick, Grimaldi, Whittle, Waters, and Baraniuk (2018)</td>
<td>Open = 4,247, commercial = 14,474</td>
<td>Course grades</td>
<td>Humanities</td>
<td>2-year community college (United States)</td>
<td>There were differences in prior academic achievement that could not be controlled for in the analyses</td>
</tr>
<tr>
<td>Chiorescu (2017)</td>
<td>Open = 159, commercial = 447</td>
<td>Course grades (C or better)</td>
<td>Math</td>
<td>4-year public college (United States)</td>
<td>No assessment of prior knowledge or achievement</td>
</tr>
<tr>
<td>Choi and Carpenter (2017)</td>
<td>Open = 175, commercial = 114</td>
<td>Course grades</td>
<td>Social science</td>
<td>4-year public university (United States)</td>
<td>No inferential statistics reported</td>
</tr>
<tr>
<td>Clinton (2018)</td>
<td>Open = 204, commercial = 316</td>
<td>Course grades</td>
<td>Social science</td>
<td>4-year public university (United States)</td>
<td>There were differences in prior academic achievement that could not be controlled for in the analyses</td>
</tr>
<tr>
<td>Colvard et al. (2018)</td>
<td>Open = 10,141, commercial = 11,681</td>
<td>Course grades</td>
<td>Varied</td>
<td>4-year public university (United States)</td>
<td>No assessment of prior knowledge or achievement</td>
</tr>
<tr>
<td>Engler and Shedlosky-Shoemaker (2019)</td>
<td>Open = 24, commercial = 25</td>
<td>Course exam performance</td>
<td>Social science</td>
<td>4-year private college (United States)</td>
<td>Small sample sizes</td>
</tr>
<tr>
<td>Feldstein et al. (2012)</td>
<td>Open = 1,394, commercial = 2,176</td>
<td>Course grade C or above</td>
<td>Business</td>
<td>4-year public university (United States)</td>
<td>No assessment of prior knowledge or achievement; different courses compared</td>
</tr>
<tr>
<td>Grewe and Davis (2017)</td>
<td>Open = 58, commercial = 60</td>
<td>Course grades</td>
<td>Humanities</td>
<td>2-year public college (United States)</td>
<td>Insufficient statistics reported to estimate effect size</td>
</tr>
<tr>
<td>Grissett and Huffman (2019)</td>
<td>Open = 31, commercial = 29</td>
<td>Course grades</td>
<td>Social science</td>
<td>4-year public university (United States)</td>
<td>Small sample sizes</td>
</tr>
<tr>
<td>Gurung (2017), Study 1</td>
<td>Open = 405, commercial = 398</td>
<td>Multiple-choice quiz</td>
<td>Social science</td>
<td>Multiple 4-year public colleges and universities (United States)</td>
<td>Open and commercial students from different populations; different instructors</td>
</tr>
<tr>
<td>Gurung (2017), Study 2</td>
<td>Open = 753, commercial = 584</td>
<td>Multiple-choice quiz</td>
<td>Social science</td>
<td>Multiple 2-year and 4-year public colleges and universities (United States)</td>
<td>Open and commercial students from different populations; different instructors</td>
</tr>
<tr>
<td>Author/s (Year)</td>
<td>Sample Size (Open and Commercial)</td>
<td>Learning Measure</td>
<td>Content Area(s)</td>
<td>Institution Type (Country)</td>
<td>Limitations</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------</td>
<td>---------------------------</td>
<td>-----------------</td>
<td>-----------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Hardin et al. (2019)</td>
<td>Open = 1,014, commercial = 2,237</td>
<td>Exams</td>
<td>Social science</td>
<td>4-year public university (United States)</td>
<td>Different instructors taught courses with OER and commercial textbooks</td>
</tr>
<tr>
<td>Hendricks, Reinsberg, and Rieger (2017)</td>
<td>Open = 755, commercial = 1,492</td>
<td>Final exam</td>
<td>Hard science</td>
<td>4-year public university (Canada)</td>
<td>Limited inferential statistics reported; no analysis of possible differences in prior knowledge</td>
</tr>
<tr>
<td>Hilton, Gaudet, Clark, Robinson, and Wiley (2013)</td>
<td>Open = 2,043, commercial = 4,164</td>
<td>Course grade of C or better</td>
<td>Math</td>
<td>2-year community college (United States)</td>
<td>Confound of changes in course placement policies; inferential statistics not reported</td>
</tr>
<tr>
<td>Hilton et al. (2016)</td>
<td>Open = 2,014, commercial = 43,223</td>
<td>Course grade</td>
<td>Varied</td>
<td>2-year community college (United States)</td>
<td>Different instructors taught courses with OER and commercial textbooks</td>
</tr>
<tr>
<td>Jhangiani et al. (2018)</td>
<td>Open = 94, commercial = 83</td>
<td>Exams</td>
<td>Social science</td>
<td>4-year public university (Canada)</td>
<td>Different instructors taught courses with OER and commercial textbooks; summer terms were compared with academic year, may be a different population</td>
</tr>
<tr>
<td>Lawrence and Lester (2018)</td>
<td>Open = 178, commercial = 200</td>
<td>Course grades</td>
<td>Social science</td>
<td>4-year public university (United States)</td>
<td>Limited reporting of descriptive and inferential statistics for grades; no assessment of prior knowledge or achievement</td>
</tr>
<tr>
<td>Medley-Rath (2018)</td>
<td>Open = 23, commercial = 24</td>
<td>Improvement on posttest course grade</td>
<td>Social science</td>
<td>2-year community college (United States)</td>
<td>Online courses were mentioned, but not included in analyses; small sample sizes</td>
</tr>
<tr>
<td>Robinson (2015)</td>
<td>Open = 1,908, commercial = 1,908</td>
<td>Course grade</td>
<td>Varied</td>
<td>Multiple: 2-year community colleges, 4-year private college, and 4-year public college (United States)</td>
<td>Different instructors; no assessment of prior knowledge or achievement</td>
</tr>
<tr>
<td>Westermann Juárez and Venegas Muggli (2017)</td>
<td>Open = 31, commercial = 30</td>
<td>Course grades</td>
<td>Mathematics</td>
<td>Postsecondary, details N/A (Chile)</td>
<td>Small sample sizes</td>
</tr>
<tr>
<td>Winitzky-Stephens and Pickavance (2017)</td>
<td>Open = 329, commercial = 1,120</td>
<td>Course grades</td>
<td>Varied</td>
<td>2-year community college (United States)</td>
<td>Different courses compared; no assessment of prior knowledge or achievement</td>
</tr>
</tbody>
</table>

Note: OER = open educational resources; N/A = not applicable.
For the learning measure, if course grades and learning measures included in the overall course grades were reported (e.g., exam scores), then only course grades were used in the meta-analysis. This was to avoid redundant effect sizes because course grades are inclusive of performance on learning measures within a course (note that Medley-Rath, 2018, had a posttest measure that was not included in the final grade). One exception is Allen and colleagues (2015), because the information to calculate effect sizes based on a standardized learning measure was available, but such information was not available for course grades. However, there were studies in which multiple measures were reported that had separate scores (e.g., Hardin et al., 2019; Jhangiani et al., 2018). These measures were not independent because they came from the same samples. To account for these multiple measures that were from the same study (but were separate measures), robust variance estimation (RVE) was used. RVE is a method of incorporating correlations of dependent effect sizes within studies. This approach is more accurate for estimating effect sizes than the traditional approach of aggregating multiple effect sizes within a study (Tanner-Smith, Tipton, & Polanin, 2016). Also, a small sample correction was used (Tipton, 2015; see Tipton & Pustejovsky, 2015). The package “robumeta” was used (Fisher & Tipton, 2014). There were 22 studies with 26 measures (and subsequently 26 effect sizes) assessing learning efficacy.

For the withdrawal information, the odds ratio was used to compare the odds of withdrawal in a course that uses commercial textbooks compared with one that uses open textbooks (see Freeman et al., 2014, for a similar approach). The

\[ \text{Odds Ratio} = \frac{\text{Probability of Withdrawal in Commercial Textbooks}}{\text{Probability of Withdrawal in Open Textbooks}} \]

TABLE 2

<table>
<thead>
<tr>
<th>Author/s (Year)</th>
<th>Sample Size (Open and Commercial)</th>
<th>Content Area(s)</th>
<th>Institution Type (Country)</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiorescu (2017)</td>
<td>Open = 159, commercial = 447</td>
<td>Math</td>
<td>4-year public college (United States)</td>
<td>No assessment of prior knowledge or achievement</td>
</tr>
<tr>
<td>Clinton (2018)</td>
<td>Open = 232, commercial = 435</td>
<td>Social science</td>
<td>4-year public university (United States)</td>
<td>No subgroup analyses</td>
</tr>
<tr>
<td>Colvard et al. (2018)</td>
<td>Open = 10,141, commercial = 11,691</td>
<td>Varied</td>
<td>4-year public university (United States)</td>
<td>No assessment of prior knowledge or achievement</td>
</tr>
<tr>
<td>Grewe and Davis (2017)</td>
<td>Open = 95, commercial = 95</td>
<td>Humanities</td>
<td>2-year public college (United States)</td>
<td>No subgroup analyses; different instructors</td>
</tr>
<tr>
<td>Grissett and Huffman (2019)</td>
<td>Open = 36, commercial = 34</td>
<td>Social science</td>
<td>4-year public university (United States)</td>
<td>Small sample sizes</td>
</tr>
<tr>
<td>Hendricks et al. (2017)</td>
<td>Open = 787, commercial = 1,583</td>
<td>Hard science</td>
<td>4-year public university (Canada)</td>
<td>No analysis of possible differences in scientific attitudes</td>
</tr>
<tr>
<td>Hilton and Laman (2012)</td>
<td>Open = 370, commercial = 370</td>
<td>Social science</td>
<td>2-year community college (United States)</td>
<td>Inferential statistics not reported; no assessment of prior knowledge or achievement; courses were redesigned along with adoption of an open-source textbook</td>
</tr>
<tr>
<td>Hilton et al. (2013)</td>
<td>Open = 2,043, commercial = 4,164</td>
<td>Math</td>
<td>2-year community college (United States)</td>
<td>Confound of changes in course placement policies; inferential statistics not reported</td>
</tr>
<tr>
<td>Hilton et al. (2016)</td>
<td>Open = 2,014, commercial = 43,223</td>
<td>Varied</td>
<td>2-year community college (United States)</td>
<td>Different instructors taught courses with OER and commercial textbooks</td>
</tr>
<tr>
<td>Jhangiani et al. (2018)</td>
<td>Open = 139, commercial = 105</td>
<td>Social science</td>
<td>4-year public university (Canada)</td>
<td>Summer terms were compared with academic year, may be a different population</td>
</tr>
<tr>
<td>Lawrence and Lester (2018)</td>
<td>Open = 193, commercial = 243</td>
<td>Social science</td>
<td>4-year public university (United States)</td>
<td>Limited reporting of descriptive and inferential statistics for grades; no assessment of prior knowledge or achievement</td>
</tr>
</tbody>
</table>

Note. Sample size may vary between withdrawal rate and learning performance measures for the same studies depending on who was included in the authors’ samples and differences in available data.
odds ratio was chosen because findings regarding withdrawal rates, unlike learning measures, were always reported as dichotomous variables. Because there was only one measure of withdrawal rate (and subsequently only one effect size) per the study, RVE was not used. The odds ratio indicates the relative odds of withdrawal rate associated with the type of textbook used. An odds ratio less than 1 indicates that the use of open textbooks in courses was associated with a lower withdrawal rate compared with commercial textbooks. The withdrawal rate was reported in five studies and provided by authors on request for an additional six studies (i.e., the withdrawal rate not stated in the dissemination, but six authors whose reports were included in the learning efficacy meta-analysis were able to locate and share the relevant numbers regarding course withdrawals).

The $I^2$ index was used to report the heterogeneity of effect sizes. The $I^2$ index ranges from 0 to 100 and indicates the percentage of heterogeneity across studies not due to chance or sampling error—higher levels indicating a greater degree of heterogeneity. Moderator analyses were appropriate if the $I^2$ index is more than 20% (Bloch, 2014).

**Results**

**Learning Efficacy**

See Appendix C for R code used in analyses and for data sets. A data dictionary for the data sets is also provided in Appendix C.

Outliers in the effect sizes were examined using a violin plot prior to synthesizing effect sizes in RVE (see Tanner-Smith et al., 2016). To construct the violin plot, the package vioplot was used (Adler & Kelly, 2018). A violin plot is a combination of a box plot, which displays a measure of central tendency (either mean or median) and interquartile

FIGURE 1. Flow diagram of the systematic review process.
range, and a kernel density plot, which has a smooth curve to indicate the probability density of a variable in a manner similar to a histogram (Hintze & Nelson, 1998). This allows for a visual display of the distribution of the data as well as the summary statistics. The violin plot, shown in Figure 2, has a white circle to indicate the mean and the thin black line to represent the interquartile range. Weights were not used, but the data displayed in the violin plot, Hedges’s $g$, were themselves biased correct for sample size. The violin plot, shown in Figure 2, is normally distributed and slightly left skewed. The effect sizes outside of the thin gray line indicate that there were outliers. However, this was approximately evenly distributed on both sides, and all effect sizes were included so as not to lose information (e.g., Kong, Seo, & Zhai, 2018).

The learning performance overall was first examined, and the $I^2$ was 98.17, which would indicate a substantial amount of heterogeneity. Based on the results of the RVE meta-analysis assuming correlated effects of .8, there were no reliable differences in learning performance between students in courses with open textbooks compared with students in courses with commercial textbooks, $g = 0.01$, SE (standard error) = 0.08, 95% CI (confidence interval) = [−0.16, 0.19], $p = .87$ (see Table 3 for findings by study).

To further examine the null effect found, two approaches were used. The first was a power analysis (Harrer & Ebert, 2018). Based on a small effect size of Cohen’s $d = 0.2$, the average number of participants in each study, the number of studies, and a high degree of heterogeneity, the power was 1—indicating a great likelihood that there was sufficient power to detect a small effect. However, there was substantial variability with the number of participants in each study. Given this, the analysis was redone with the same variables except that the median number of participants in each study was used. Again, the power was 1.

The second examination was an equivalence test using the package “toster” in R (Lakens, Scheel, & Isager, 2018). The equivalence test was significant, $Z = −2.25$, $p = .01$, assuming a small effect size of Cohen’s $d = 0.2$. An interpretation of this finding was that the main effect of learning performance in courses with open textbooks compared with those with commercial textbooks was statistically equivalent to zero.

Publication bias, which is the increased likelihood that statistically significant findings were reported, was assessed in two manners using aggregated effect sizes because there currently is not a validated measure of publication bias with RVE (Friese, Frankenbach, Job, & Loschelder, 2017). The first was the graphical technique of the funnel plot. A funnel plot shows studies graphed according to their size along the $y$-axis in which smaller studies were toward the bottom, and the effect sizes were along the $x$-axis. The line in the middle represents the mean effect. If there were an asymmetrical distribution on the sides of the middle line and/or if the distribution of studies was broader at the bottom, then publication bias was likely (Egger, Smith, Schneider, & Minder, 1997). Based on the funnel plot in Figure 3, there did not appear to be publication bias. Publication bias was also tested quantitatively using Egger’s test of the intercept, which did not differ significantly from zero, $\beta = −1.49$, 95% CI [−5.12, 2.13], $p = .40$, indicating unlikely publication bias (Cooper, 2015; see Follmer, 2018; Koponen, Georgiou, Salmi, Leskinen, & Aro, 2017, for similar approaches for testing publication bias).

**Moderator Analyses.** To test for potential moderators, a meta-regression model with each of the moderators as coefficients was estimated (see Tipton & Pustejovsky, 2015). Following Borenstein, Hedges, Higgins, and Rothstein (2009), a minimum of six effect sizes per potential moderator category was required to conduct moderator analyses (see Elleman, 2017, for a similar approach). Therefore, potential moderators such as the same course being compared and whether the learning measure was part of a course grade or only for a research study (e.g., for the studies in Gurung, 2017) were not examined.

The moderators examined were whether the instructor was the same for open and commercial conditions, whether prior knowledge or prior student achievement was considered or controlled in the analyses, and whether the learning measure was identical (e.g., the same exam was used) for
### TABLE 3

**Learning Efficacy Statistics for Each Study and Model Statistics (Positive Hedges’s g Indicates Better Performance With an Open-Source Textbook Compared With a Commercial Textbook)**

<table>
<thead>
<tr>
<th>Study Name</th>
<th>Same Instructor(s)</th>
<th>Prior Achievement</th>
<th>Identical Learning Measure(s)</th>
<th>Hedges’s g</th>
<th>SE</th>
<th>O</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allen et al. (2015)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.02</td>
<td>0.07</td>
<td>377</td>
<td>348</td>
<td>725</td>
</tr>
<tr>
<td>Basu Mallick et al. (2018)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>−0.14</td>
<td>0.02</td>
<td>4247</td>
<td>14474</td>
<td>18721</td>
</tr>
<tr>
<td>Chiorescu (2017)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>0.20</td>
<td>0.14</td>
<td>159</td>
<td>447</td>
<td>606</td>
</tr>
<tr>
<td>Choi and Carpenter (2017)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>−0.44</td>
<td>0.12</td>
<td>175</td>
<td>114</td>
<td>289</td>
</tr>
<tr>
<td>Clinton (2018)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>0.23</td>
<td>0.09</td>
<td>204</td>
<td>316</td>
<td>520</td>
</tr>
<tr>
<td>Colvard et al. (2018)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>0.21</td>
<td>0.01</td>
<td>10141</td>
<td>11681</td>
<td>12692</td>
</tr>
<tr>
<td>Engler and Shedlosky-Shoemaker (2019)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.00</td>
<td>0.28</td>
<td>24</td>
<td>25</td>
<td>49</td>
</tr>
<tr>
<td>Feldstein et al. (2012)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>0.20</td>
<td>0.05</td>
<td>1394</td>
<td>2176</td>
<td>3570</td>
</tr>
<tr>
<td>Grewe and Davis (2017)</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>1.104</td>
<td>0.20</td>
<td>58</td>
<td>60</td>
<td>118</td>
</tr>
<tr>
<td>Grissett and Huffman (2019)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>−0.16</td>
<td>0.26</td>
<td>31</td>
<td>29</td>
<td>60</td>
</tr>
<tr>
<td>Gurung (2017), Study 1</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>−0.87</td>
<td>0.07</td>
<td>405</td>
<td>398</td>
<td>803</td>
</tr>
<tr>
<td>Gurung (2017), Study 2</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>−0.48</td>
<td>0.07</td>
<td>823</td>
<td>584</td>
<td>1407</td>
</tr>
<tr>
<td>Hardin et al. (2019), a content knowledge</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>0.39</td>
<td>0.05</td>
<td>682</td>
<td>1122</td>
<td>1704</td>
</tr>
<tr>
<td>Critical thinking</td>
<td></td>
<td></td>
<td></td>
<td>0.06</td>
<td>0.06</td>
<td>332</td>
<td>1115</td>
<td>1447</td>
</tr>
<tr>
<td>Aggregated</td>
<td></td>
<td></td>
<td></td>
<td>0.22</td>
<td>0.06</td>
<td>507</td>
<td>1118.5</td>
<td>1575.5</td>
</tr>
<tr>
<td>Hendricks et al. (2017)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>0.12</td>
<td>0.05</td>
<td>755</td>
<td>750</td>
<td>1505</td>
</tr>
<tr>
<td>Hilton et al. (2013)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>−0.09</td>
<td>0.30</td>
<td>2043</td>
<td>4164</td>
<td>6207</td>
</tr>
<tr>
<td>Hilton et al. (2016)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>0.12</td>
<td>0.03</td>
<td>2014</td>
<td>43223</td>
<td>45237</td>
</tr>
<tr>
<td>Jhangiani et al. (2018), Exam 1</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>0.01</td>
<td>0.15</td>
<td>94</td>
<td>83</td>
<td>177</td>
</tr>
<tr>
<td>Exam 2</td>
<td></td>
<td></td>
<td></td>
<td>0.02</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam 3</td>
<td></td>
<td></td>
<td></td>
<td>0.52</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggregated</td>
<td></td>
<td></td>
<td></td>
<td>0.18</td>
<td>0.15</td>
<td>94</td>
<td>83</td>
<td>177</td>
</tr>
<tr>
<td>Lawrence and Lester (2018)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>−0.06</td>
<td>0.10</td>
<td>178</td>
<td>200</td>
<td>378</td>
</tr>
<tr>
<td>Medley-Rath (2018), posttest</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>−0.21</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course grade</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>0.36</td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggregated</td>
<td></td>
<td></td>
<td></td>
<td>0.08</td>
<td>0.29</td>
<td>23</td>
<td>24</td>
<td>47</td>
</tr>
<tr>
<td>Robinson (2015)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>−0.12</td>
<td>0.03</td>
<td>1908</td>
<td>1908</td>
<td>3816</td>
</tr>
<tr>
<td>Westermann Juárez and Venegas Muggli (2017)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>−0.10</td>
<td>0.04</td>
<td>31</td>
<td>30</td>
<td>61</td>
</tr>
<tr>
<td>Winitzy-Stephens and Pickavance (2017)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>−0.02</td>
<td>0.06</td>
<td>329</td>
<td>1120</td>
<td>1449</td>
</tr>
<tr>
<td>Random model (all $k = 22$)</td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
<td>.08</td>
<td>25920</td>
<td>83272.5</td>
<td>100012.5</td>
</tr>
</tbody>
</table>

**Note.** Prior achievement = study either incorporated measure of prior knowledge and/or achievement into findings or conducted analyses to determine no significant differences by condition in prior knowledge and/or achievement. Sample size O = open textbook, C = commercial textbook, and T = sum of students; SE = standard error.

Some students took both measures and some were randomly assigned to only take one. In the interests of being conservative, the aggregated sample size consists of the mean of the sample sizes for the two measures.
open and commercial conditions. Each of these potential moderators pertains to methodological quality. Note that Clinton (2018) was coded as “does not consider prior knowledge or prior academic performance” because the measure of student prior knowledge was significant between conditions and could not be covaried out of the effect size (data were only available at the course level, it was not possible to get student-level data). As can be noted in Table 4, none were significant. However, there were fewer than 4 degrees of freedom for the results with prior knowledge or prior student achievement as well as for whether the learning measure was identical. With fewer than 4 degrees of freedom, the results cannot be trusted, and therefore, it was inconclusive whether these two moderators have effects. In addition, it is not good practice to conduct a power analysis for moderators in a meta-regression (Pigott, 2012); therefore, it is uncertain whether the findings for instructor indicated that there is truly no difference or if there was simply a lack of power to detect a difference.

**Withdrawal Rate**

For the withdrawal rate, the heterogeneity of effect sizes was substantial, with an I² of 83.09, which indicated that the findings varied considerably. Based on the random-effects model using the package “meta” (Schwarzer, 2007), the withdrawal rate for students in courses with open textbooks was lower than that of students in courses with commercial textbooks; OR (odds ratio) = 0.71, k = 11, 95% CI [0.56, 0.90], p = .005 (see Table 5 for statistics by study). As with learning efficacy, a violin plot was made to examine outliers. As can be seen in Figure 4, the distribution of effect sizes was quite negatively skewed, indicating likely outliers. However, the outliers were opposite the direction of the

---

**TABLE 4**

**Meta-Regression Results for Learning Efficacy**

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>SE</th>
<th>t</th>
<th>df</th>
<th>p</th>
<th>95% CI Lower</th>
<th>95% CI Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.03</td>
<td>0.1</td>
<td>-0.26</td>
<td>13.27</td>
<td>.80</td>
<td>-0.24</td>
<td>0.19</td>
</tr>
<tr>
<td>Prior</td>
<td>0.22</td>
<td>0.52</td>
<td>0.43</td>
<td>3.54</td>
<td>.69</td>
<td>-1.30</td>
<td>1.74</td>
</tr>
<tr>
<td>Instructor</td>
<td>-0.02</td>
<td>0.15</td>
<td>-0.122</td>
<td>14.81</td>
<td>.91</td>
<td>-0.35</td>
<td>0.31</td>
</tr>
<tr>
<td>Measure</td>
<td>-0.39</td>
<td>0.46</td>
<td>-0.83</td>
<td>3.34</td>
<td>.46</td>
<td>-1.78</td>
<td>1.01</td>
</tr>
</tbody>
</table>

*Note: Prior = study either incorporated measure of prior knowledge and/or achievement into findings or conducted analyses to determine no significant differences by condition in prior knowledge and/or achievement. Instructor = whether or not the same instructor taught the courses with open or commercial textbooks. Measure = whether or not the measure for learning efficacy was identical for courses with open or commercial textbooks. SE = standard error; T = t-test value; df = degrees of freedom; 95% CI Lower = 95% confidence interval lower limit; 95% CI Upper = 95% confidence interval upper limit.*
main effect. Therefore, the overall finding (that the withdrawal rate is lower for students in courses with open textbooks compared with commercial textbooks) was not affected by the outliers. For that reason and to not lose information, the outliers were kept in the analyses.

For withdrawal rate, over half of the studies did not report this information (it was obtained from the authors). Therefore, examining publication bias in a manner similar to what was done with learning efficacy (i.e., funnel plot and Egger’s test of the intercept) would be inappropriate given that half of the findings were not published.

Discussion

The purposes of these meta-analyses were to summarize the efficacy of open textbooks compared with commercial textbooks on learning performance and withdrawal rate. Based on the findings of 22 independent studies, there appeared to be no effect on learning performance in courses with open textbooks compared with courses with commercial textbooks. This effect did not appear to vary between studies based on having the same instructors for the courses with open and commercial textbooks. However, it should be noted that the absence of a significant effect from a moderator may be due to a lack of power rather than a true null effect (Hempel et al., 2013). The moderator analyses for incorporating prior student performance into the analyses or using identical instruments for measuring learning for open and commercial textbook conditions were inconclusive. The withdrawal rate from courses with open textbooks, based on 11 independent studies, was reliably lower than that for courses with commercial textbooks.

The null results for learning performance from open compared with commercial textbooks supports the notion that open textbooks save students money without a detrimental effect on learning. This quantitative summary of the findings converges with conclusions based on qualitative systematic reviews comparing learning in courses with open and commercial textbooks (Hilton, 2016, 2018). Furthermore, this meta-analysis builds on these previous reviews by comparing effect sizes based on study quality through moderator analyses. These moderator analyses indicated that controlling for the confounder of whether or not the instructor was the same did not significantly vary the results. Therefore, although many studies on open textbooks have been justifiably critiqued for methodological quality (see Griggs & Jackson, 2017; Gurung, 2017, for critiques), this particular confounder did not appear to be skewing the findings on open textbook efficacy. However, this could be due to lack of power, and there was not sufficient power to trust the results for the other potential moderators examined (identical learning measure and whether or not student prior achievement was controlled).

Based on the access hypothesis, students in courses with open textbooks should be academically outperforming their peers in courses with commercial textbooks. This is because every student would have access to open textbooks, but some students would not be able to afford access, or at least reliable access, to a commercial textbook. However, the findings from this meta-analysis do not provide support for

<table>
<thead>
<tr>
<th>Study Name</th>
<th>Withdraw Published</th>
<th>Odds Ratio</th>
<th>95% CI</th>
<th>p</th>
<th>O</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiorescu (2017)</td>
<td>Yes</td>
<td>0.5</td>
<td>0.23, 1.08</td>
<td>.08</td>
<td>159</td>
<td>447</td>
<td>606</td>
</tr>
<tr>
<td>Clinton (2018)</td>
<td>Yes</td>
<td>0.36</td>
<td>0.23, 0.57</td>
<td>&lt;.001</td>
<td>232</td>
<td>435</td>
<td>667</td>
</tr>
<tr>
<td>Colvard et al. (2018)</td>
<td>No</td>
<td>0.82</td>
<td>0.73, 0.93</td>
<td>.002</td>
<td>10,141</td>
<td>11,681</td>
<td>21,822</td>
</tr>
<tr>
<td>Grewe and Davis (2017)</td>
<td>No</td>
<td>0.56</td>
<td>0.25, 1.26</td>
<td>.16</td>
<td>95</td>
<td>95</td>
<td>190</td>
</tr>
<tr>
<td>Grissett and Huffman (2019)</td>
<td>No</td>
<td>3.00</td>
<td>0.30, 3.30</td>
<td>.35</td>
<td>36</td>
<td>34</td>
<td>70</td>
</tr>
<tr>
<td>Hendricks et al. (2017)</td>
<td>No</td>
<td>0.75</td>
<td>0.48, 1.17</td>
<td>.21</td>
<td>787</td>
<td>1,583</td>
<td>2,370</td>
</tr>
<tr>
<td>Hilton and Laman (2012)</td>
<td>Yes</td>
<td>0.46</td>
<td>0.28, 0.76</td>
<td>.002</td>
<td>370</td>
<td>370</td>
<td>740</td>
</tr>
<tr>
<td>Hilton et al. (2013)</td>
<td>Yes</td>
<td>1.23</td>
<td>1.09, 1.38</td>
<td>.001</td>
<td>2,043</td>
<td>4,168</td>
<td>6,211</td>
</tr>
<tr>
<td>Hilton et al. (2016)</td>
<td>Yes</td>
<td>0.97</td>
<td>0.84, 1.12</td>
<td>.68</td>
<td>2,014</td>
<td>43,223</td>
<td>45,237</td>
</tr>
<tr>
<td>Jhangiani et al. (2018)</td>
<td>No</td>
<td>0.44</td>
<td>0.10, 1.89</td>
<td>.27</td>
<td>139</td>
<td>105</td>
<td>244</td>
</tr>
<tr>
<td>Lawrence and Lester (2018)</td>
<td>No</td>
<td>0.51</td>
<td>0.25, 1.03</td>
<td>.06</td>
<td>193</td>
<td>243</td>
<td>436</td>
</tr>
<tr>
<td>Random model (k = 11)</td>
<td>0.71</td>
<td>0.56, 0.90</td>
<td>0.05</td>
<td>16,209</td>
<td>62,384</td>
<td>78,593</td>
<td></td>
</tr>
</tbody>
</table>

Note: Withdrawal published is whether or not the withdrawal rate was reported in the publication (Yes) or the authors provided the withdrawal rate on request (No). Sample size O = open textbook, C = commercial textbook, and T = sum of students.

Withdrawal rate was published as part of a composite measure including grades of D and F (DFW rate), but specific withdrawal rate analyses were not reported.
this hypothesis. One reason could be that having free access to a textbook may only have a benefit for learning for a small number of students. For example, Colvard and colleagues (2018) found that students eligible for federal assistance grants based on low household income yielded greater learning benefits from open textbooks relative to their peers with higher household incomes. It could be that if the information on student socioeconomic status (SES) were available, SES would be a significant moderator in which students with lower SES backgrounds would have benefits not seen with students of higher SES who can more easily afford and access commercial textbooks. Alternatively, it is possible that the open and commercial textbooks compared were of similar quality and perhaps access was not necessarily an issue, so the learning findings were similar. Another possible reason, based on conjecture, is that postsecondary instructors may be aware that not all of their students have the commercial textbook and adjust their pedagogy accordingly. For example, instructors may be less likely to require reading or assignments requiring the use of the textbook if they are concerned about the financial impact on some of their students. Finally, it could also be possible that the type of textbook simply has little influence on student learning. For example, one experiment comparing commercial textbooks from different publishers found no differences between types in learning from reading them (Durwin & Sherman, 2008). This possibility is supported by findings from studies on open textbook efficacy (not eligible for this meta-analysis) in which all students had access to the commercial textbooks, but there were either no differences or a small benefit in learning performance with open textbooks (Clinton, Legerski, & Rhodes, 2019; Robinson, Fischer, Wiley, & Hilton, 2014).

Although there was no benefit of open textbooks on the performance of students who completed courses, there was a significant reduction in the likelihood of students withdrawing from the course. The most common reason students state for course withdrawal is that they anticipate failing the course or receiving a low grade (Wheland, Butler, Qammar, Katz, & Harris, 2012). It is possible that having access to a textbook may help a student who is behind to cover missed material and not withdraw from the class. In this way, students who are struggling in a course may be less likely to withdraw if they can access an open textbook for free as opposed to paying hundreds of dollars for a commercial textbook to succeed in the course. However, there should be caution with interpreting withdrawal findings because some studies compared courses with different instructors (e.g., Hilton et al., 2016; Wiley, Williams, DeMarte, & Hilton, 2016), and there were not enough studies in each cell to conduct moderator analyses.

Although not identified as an outlier in analyses, there was one study in which the withdrawal rate was significantly higher in a mathematics course with an open textbook than with a commercial textbook (Hilton et al., 2013). In the discussion of this finding, Hilton and colleagues (2013) describe how institutional policy changes for placement in mathematics courses coincided with the adoption of an open textbook for that course. This led to students with different levels of preparedness in mathematics in the compared courses, which creates a clear confounder for comparing withdrawal rates.

Limitations and Future Directions

There are a number of limitations in these meta-analyses as well as in research on open textbooks more generally that should be acknowledged. One limitation of this learning efficacy meta-analysis was the focus on postsecondary students. As with postsecondary education, the adoption of open textbooks in K–12 education can yield significant cost savings, even with the time involved with modifying curricula and pedagogy (Wiley, Hilton, Ellington, & Hall, 2012).
Furthermore, one study found that K–12 teachers rated open textbooks as better quality than commercial textbooks (Kimmons, 2015) and another found that K–12 teachers saw open textbooks as helpful when personalizing student learning (de los Arcos, Farrow, Pitt, Weller, & McAndrew, 2016). However, there is only one study that examined K–12 student learning that we know of that would have otherwise met the criteria to be included in this meta-analysis (Robinson et al., 2014). The increased focus on open textbooks in postsecondary education may be because the costs of commercial materials are more obvious to postsecondary students who typically have direct expenses for course materials, whereas the cost to public K–12 students are indirectly supported through taxpayers. That said, with more OER being developed for and used in K–12 schools (Pitt, 2015), more inquiry in their efficacy is needed (see Blomgren & McPherson, 2018, for a review of K–12 OER research on a variety of issues).

The geographic scope of this meta-analysis was not representative of OER use. The reports in this meta-analysis were based on studies set in the United States and Canada, with one exception of a study based in Chile (Westermann Juárez & Venegas Muggli, 2017). The advocacy and use of OER, including open textbooks, is international (Bliss & Smith, 2017). Furthermore, teaching and learning practices with textbook use vary considerably by geographic region, partly due to textbook availability (Milligan, Tikly, Williams, Vianney, & Uworwabayeho, 2017). It is unclear if open textbooks would ameliorate difficulties with textbook availability in regions in which access to the internet and electronic devices may be limited (Butcher, 2015). For these reasons, more inquiry into open textbook efficacy is needed outside the United States and Canada.

One limitation in the methodology of this meta-analysis is the use of a single screener for abstract screening. Logically, more than one screener would certainly reduce the likelihood a relevant citation was missed. Therefore, it is possible a relevant report was missed during abstract screening because of the use of a single screener. That said, an assessment of Abstractor, the abstract screening software used in this meta-analysis, used only one screener (Gates, Johnson, & Hartling, 2018). Moreover, other possibly relevant reports were searched for in alternative methods (e.g., backward and forward citations searches) that may have compensated for the use of a single screener.

There were no significant moderators identified in the meta-regression analyses. However, there was substantial variability in the findings that would indicate the possible existence of significant moderators. It is possible that as more research is conducted in OER, there would be sufficient power to find that the proposed moderators related to study methodological quality explained some of this variability. In addition, future research could consider possible moderators, such as student background and how the textbook is used by the instructor and students, to determine sources of variability in findings.

Another reason for the variability could be that the studies themselves were quite different. Although all studies compared students enrolled with courses with open textbooks with those with commercial textbooks, there was a broad range of content areas and sample sizes. Moreover, grading criteria vary considerably among instructors and institutions. Regarding course withdrawals, each institution in this study had an initial drop period without a withdrawal on the transcript at the beginning of the term and a deadline at some point in the term to receive a W and not an F. However, the specific deadlines and policies varied.

The noted differences in the course withdrawal rates for open and commercial textbooks may have subsequently led to the characteristics of the students who completed the courses to be different. Although it is possible that students opted not to withdraw because they had access to an open textbook that allowed them to perceive a better chance of success in the course, in general, students who withdraw tend to have weaker academic backgrounds than those who persist in courses (McKinney, Novak, Hagedorn, & Luna-Torres, 2019). It is possible that this could lead to student differences in the courses that cannot be accounted for in the data available for these meta-analyses.

A potential issue is the medium (paper or screen) from which students read their course textbook. In Gurung’s (2017) second study, the performance between students in courses using open textbooks was lower than students in courses using commercial textbooks. However, the difference was much smaller when only examining students who accessed their textbooks only electronically (thus reading from screens). Because open textbooks are free to access electronically but involve costs for hard copies, it may be that students may be more likely to read their open textbooks electronically than commercial textbooks (that involve costs in any medium; Clinton, 2018). This issue of reading medium used may be worth considering given that meta-analyses have been published indicating a small benefit in learning performance when reading from paper compared with reading from screens (Clinton, 2019c; Kong et al., 2018).

Conclusion

Open-source textbooks have been developed primarily in response to the rising costs of commercial materials. Concerns over quality and effects on learning have prompted numerous studies in this area. Based on the meta-analytic findings here, there are no meaningful differences in learning efficacy between students using open textbooks and students using commercial textbooks. However, students in courses with open textbooks appear to be less likely to withdraw. There are several limitations in research on open textbooks that indicate future research should consider K–12 students, the needs of students outside of the United States and Canada, and the potential moderating factors of student characteristics.
Appendix A
Search Terms Used for Each Database

Scopus
“open source textbook*”
“OER” + textbook*

ERIC
“open source textbook*”
“OER” + textbook*

DOAJ
“open source textbook*”
“OER” + textbook*

ProQuest
“open source textbook*”
“OER” textbook*

PSYCinfo
“open source textbook*”
“OER” textbook*

Appendix B
Coding of Studies

If the information was not available in the report, the corresponding author was contacted with a request for the information. If the information was not or could not be obtained, then the code “N/A” for “not available” was used.

Descriptive Information
See Tables 1 and 2

1. Bibliographic information: Authors, year of publication, and title
2. Dissemination type: Peer-reviewed journal article, master’s thesis or dissertation, conference presentation or proceedings, or unpublished data
3. Year of publication: The year it was published or the necessary statistical information was provided
4. Sample size: Number of participants in each condition
5. Type of learning measure: Exam scores, course grade, researcher development measure, other
6. Content area: Social sciences, mathematics, hard sciences, humanities, vocational/business education, multiple, other
7. Institution type: K–12, public 2-year postsecondary, private 2-year postsecondary, public 4-year postsecondary, private 4-year postsecondary
8. Country: The country in which the study took place

Study Quality
See Table B1

1. Were the participants in the open textbook condition comparable with the participants in the commercial textbook condition? If yes, then there was a pretest or prior academic measure that was found to be similar across conditions or statistically controlled for in the analyses.
2. Were the participants in each condition receiving similar instruction? Specifically, the following three questions were asked:
   a. Were the courses in the open and commercial conditions the same?
   b. Were the instructors in the open and commercial conditions the same?
   c. Were the course modalities the same (online, face to face, blended)? Whenever possible, only courses of the same modality were included (e.g., Hendricks et al., 2017, involved one online course section with an open textbook, but all other courses were face to face so only courses of the same modality, face to face, were included).
3. To what extent was the intervention tested for effectiveness within important subgroups of participants? Specifically, were participant subgroups, such as variation in socioeconomic status or prior academic achievement examined?
4. Were the outcomes measured in a way that is consistent with the proposed effects of the intervention? If yes, the outcome measure (grades, exam scores, withdrawal rate) was relevant to the course(s) being compared in the study.
5. Did the description of the study give any other indication of the strong plausibility of other intervention contaminants? If yes, there was evidence of potential contaminants such as substantial course redesign or changes in institutional policies.
6. Were the sample sizes adequate to provide sufficiently precise estimates of effect sizes? If yes, there were at least 50 participants per condition.
7. To what extent were sample sizes reported (or estimable) from statistical information presented? If yes, sample sizes for each condition were explicitly reported.
8. To what extent could the direction of effects be identified for important measured outcomes? Specifically, were the direction of effects explicitly reported?
9. Were statistical tests adequately reported? If yes, the inferential statistics (e.g., F statistic, t tests, p values) were explicitly stated in the report.
10. Could estimates of effect sizes be computed using a standard formula (or algebraic equivalent)? If yes, the necessary information was reported. If not, the corresponding author was contacted for the information. If the information for necessary effect sizes could not be obtained, the study could not be included in the meta-analysis.

11. Was the assessment measure for learning the same for students with open textbooks and those with commercial textbooks? If yes, the learning measure was identical.
Appendix C

Data Dictionary

### Learning Data Set

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>studynum</td>
<td>Study number</td>
</tr>
<tr>
<td>ES.ID</td>
<td>Effect size ID number</td>
</tr>
<tr>
<td>Author</td>
<td>Study author</td>
</tr>
<tr>
<td>Measure</td>
<td>Effect size description</td>
</tr>
<tr>
<td>effect.size</td>
<td>Hedges’s $g$ effect size</td>
</tr>
<tr>
<td>var</td>
<td>Variance of effect size</td>
</tr>
<tr>
<td>prior</td>
<td>Prior knowledge measured (0 for no, 1 for yes)</td>
</tr>
<tr>
<td>Instructor</td>
<td>Instructor was the same for open textbook and commercial textbook (0 for no, 1 for yes)</td>
</tr>
<tr>
<td>measure</td>
<td>Measure was the same for open textbook and commercial textbook (0 for no, 1 for yes)</td>
</tr>
</tbody>
</table>

### Withdrawal Data Set

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>studynum</td>
<td>Study number</td>
</tr>
<tr>
<td>esnumber</td>
<td>Effect size ID number</td>
</tr>
<tr>
<td>Author</td>
<td>Study author</td>
</tr>
<tr>
<td>Ee</td>
<td>Number of withdrawals for commercial textbook</td>
</tr>
<tr>
<td>Ne</td>
<td>Total participants for commercial textbook</td>
</tr>
<tr>
<td>Ec</td>
<td>Number of withdrawals for open textbook</td>
</tr>
<tr>
<td>Nc</td>
<td>Total participants for open textbook</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>log</td>
<td>Log odds ratio</td>
</tr>
<tr>
<td>SE</td>
<td>Standard error</td>
</tr>
<tr>
<td>var</td>
<td>Variance of effect size</td>
</tr>
</tbody>
</table>

---

R Code for Analyses

```r
install.packages("robumeta")
install.packages("devtools")
install.packages("vioplot")
install.packages("grid")
install.packages("meta")
install.packages("vioplot")
install.packages("TOSTER")
library(robumeta)
library(devtools)
library(vioplot)
library(grid)
library(meta)
library(vioplot)
library(TOSTER)

data <- read.csv("learning.csv") #reading in the data

names(data) #reads off the variable names in your dataset
levels(data$Follow.Up)

is.numeric(data$Follow.Up)
summary(data$Follow.Up)

is.numeric(data$prior)
summary(data$prior)

is.numeric(data$instructor)
summary(data$instructor)

is.numeric(data$measure)
levels(data$measure)

#RVE analyses for main effect testing of learning efficacy

learning <- robu(formula = effect.size ~ 1, data = data, studynum = studynum, var.eff.size = var, rho = .8, model.weights = "CORR", small = TRUE)

print(learning)

#POWER ANALYSIS


power.analysis.random<-function(d,k,n1,n2,p,heterogeneity){
  n1<-n1
  n2<-n2
  d<-d
  k<-k
  p<-p
  heterogeneity<-heterogeneity
  if(heterogeneity=="low"){
    v.d<-(n1+n2)/(n1*n2)+((d*d)/(2*(n1+n2)))
    v.m<v.d/k
    v.m<-1.33*v.m
    lambda<-(d/sqrt(v.m))
    plevel<-(1-(p/2))
    zval<-qnorm(p=plevel, 0,1)
    power<-(1-(pnorm(zval-lambda))+(pnorm(-zval-lambda))
    return(power)
  }
  if(heterogeneity=="moderate"){
    v.d<-(n1+n2)/(n1*n2)+((d*d)/(2*(n1+n2)))
  }
```
v.m<-v.d/k  
v.m<-1.67*v.m  
lambda<-(d/sqrt(v.m))  
plevel<-1-(p/2)  
zval<-qnorm(p=plevel, 0,1)  
power<-1-(pnorm(zval-lambda))+(pnorm(-zval-lambda))  
return(power)  
}

if(heterogeneity==“high”){
v.d<-(n1+n2)/(n1*n2)+((d*d)/(2*(n1+n2)))  
v.m<v.d/k  
v.m<-2*v.m  
lambda<-(d/sqrt(v.m))  
plevel<-1-(p/2)  
zval<-qnorm(p=plevel, 0,1)  
power<-1-(pnorm(zval-lambda))+(pnorm(-zval-lambda))  
return(power)  
}

power.analysis.random(d=0.20,k=22,n1=1178,n2=3785,p=0.05,  
heterogeneity = “high”)

power.analysis.random(d=0.20,k=22,n1=353,n2=423,p=0.05,  
heterogeneity = “high”)

####non equivalence test

TOSTmeta(ES = 0.0136, se = 0.0827, low_eqbound_d=-0.1,  
high_eqbound_d=0.1, alpha=0.05)

print(TOSTmeta)

TOSTmeta(ES = 0.0136, se = 0.0827, low_eqbound_d=-0.2,  
high_eqbound_d=0.2, alpha=0.05)

####META REGRESSION

data$prior_c <- data$prior - mean(data$prior)  
data$instructor_c <- data$instructor - mean(data$instructor)  
data$measure_c <- data$measure - mean(data$measure)

res_4<-robu(formula = effect.size ~ prior_c + instructor_c +  
measure_c, var.eff.size=var, studynum =  
studynum, modelweights = “CORR”, rho = 0.8, small=TRUE,  
data=data)

print(res_4)

Wald_test(res_4, constraints = 2:3, vcov=“CR2”)
Clinton and Khan


Perry, M. (2015, July 16). *The new era of the $400 college textbook, which is part of the unsustainable higher education bubble* (American Enterprise Institute Ideas) [Web log post].
Clinton and Khan


Authors

VIRGINIA CLINTON, PhD, is an assistant professor of Educational Foundations and Research at the University of North Dakota. Her research interests are open educational resources and text comprehensiveness.

SHAFIQ KHAN is a doctoral student in Educational Foundations and Research at the University of North Dakota. His research interests are open educational resources and instructional design.