Student Voice in Textbook Evaluation: Comparing Open and Restricted Textbooks

Scott Woodward, PhD, Adam Lloyd, and Royce Kimmons, PhD
Brigham Young University

Abstract

Advocates for student voice in higher education believe students should have the right and power to engage in much of the decision-making traditionally dominated by instructors or administrators. This qualitative study examines the role of student voice in the evaluation of textbook quality. Evaluators included two graduate students enrolled in a project management course at a university in the western U.S. and their instructor. Evaluators used their own student-created metric to analyze the comparative quality of eight graduate-level project management textbooks, two of which were open and six copyright-restricted. The purposes of this study were to assess the comparative quality of low-cost open textbooks and traditional copyright-restricted textbooks and to identify key student-generated criteria wherein all textbooks may be improved to better serve student needs. The analysis revealed that textbooks can be rigorously and meaningfully evaluated by students, that open textbooks can compete with restricted textbooks in terms of quality, and that polyphonic approaches to textbook evaluation can be valuable for learning. We discuss the implications of open textbooks as viable, high-quality classroom options, and the importance of valuing both student voice and instructor perspectives to ensure the highest quality textbook selection for courses in higher education.

Keywords: open educational resources, open pedagogy, open textbooks
Introduction

*Student voice* refers to the “collective contribution of diverse students’ presence, participation, and power ... [in] educational planning, research, and reform” (Cook-Sather, 2014, p. 132), and its advocates in higher education believe that students should have the right and power to engage in much of the decision-making traditionally dominated by instructors or administrators. In part, this has been exemplified in a shift from research on students to research with students (Cook-Sather, 2006; Fine, Torre, Bums, & Payne, 2007), and few areas of practice are as isolated from student voice, as are the processes surrounding textbook evaluation and adoption. Due to advancements in digital technologies, which allow content to be openly licensed and freely distributed, textbooks themselves are in a state of flux (cf. Denoyelles, Raible, & Seilhamer, 2015), and open textbooks have emerged as an alternative to traditional copyright-restricted textbooks. Open textbooks differ from traditional textbooks in that they: (1) can be distributed with no licensing costs, and (2) can be edited by instructors without permission from the original authors. As one example of an open educational resource (OER), open textbooks may include other OER within them (e.g., images) and also may be broken up themselves for partial redistribution as smaller, focused resources (e.g., individual chapters).

Open textbooks are of great interest to students and instructors in large part because high textbook costs have a discernible impact in the classroom. One study found that the persistence probability for poor and working-class students in higher education decreases by 16-19% for every 1,000 USD tuition increase (Paulsen & St. John, 2002), which was less than the average cost of textbooks in a single year (The College Board, 2015). Similarly, one university system found that 56% of its students “worried very often” about family financial circumstances, and 81% spent less money on food to meet college expenses (UCR VCSA, 2011). Another study found that many students decided against buying an assigned textbook due to cost, and almost all of these were concerned that doing so would hurt their grades and believed that access to a free textbook would help them do better in their courses (Senack, 2014). Such cost and academic concerns are of importance to online and distance learning programs as well as traditional face-to-face programs, because many online courses are developed around costly textbooks, which drives up associated costs, and also face additional enrollment challenges.

Introduction of open textbooks requires new evaluation mechanisms that are faster (i.e., responsive to a massive amount of possible resources), more inclusive (i.e., recognizing value from multiple sources), more contextually appropriate (i.e., allowing for adaptation for localized needs), and more sensitive to student needs (i.e., cost, learning needs). In response to this special issue’s call for proposals, this study explores the third trajectory of student voice in research methodology by placing enrolled students as the primary evaluators of potential classroom textbooks for a graduate-level course for the purpose of: a) conducting valuable research on the comparative quality of low-cost open textbooks vs. traditional copyright-restricted textbooks, b) guiding departmental adoptions of high-quality textbooks, and c) identifying key criteria wherein all textbooks may be improved to better serve student needs. The role of students in this study is that of “active agents,” rather than “silent witnesses” (cf. Smyth & McInerney, 2012), as students conducted research on their own learning experience while providing valuable findings on textbook quality and polyphonic evaluation. We will proceed by briefly exploring literature surrounding open textbooks and evaluation, highlighting our methods for student-based evaluations, providing results, and discussing implications of the findings.
Literature Review

Digital technologies allow content to be freely shared and distributed at an unprecedented scale that may challenge traditional thinking about textbooks and publishing. OER are textbooks, syllabi, course packets, readings, lesson plans, images, videos, audio, and courses “that are in the public domain or introduced with an open license” (UNESCO, n.d., para. 1). These materials may be freely reused, remixed, revised, and redistributed (cf. Bissell, 2009) without additional permission or licensing fees. Such freedom makes these resources especially useful for online and distance learning scenarios because online instructors and course developers can freely share, remix, and employ direct-linking to resources rather than operating within the strict limitations of fair use (Belikov & Kimmons, in press). Research into open textbooks has yielded neutral-positive results for learning outcomes, meaning that students using open textbooks generally perform as well as or better than students using restricted textbooks, and have also yielded positive results for cost savings (Bliss, Hilton, Wiley, & Thanos, 2013; Fischer, Hilton, Robinson, Wiley, 2015; Hilton & Laman, 2012; Mtebe & Raisamo, 2014; Prasad & Usagawa, 2014; Robinson, Fischer, Wiley, & Hilton, 2014; Walker, 2007), but there is need to expand the literature in the area of textbook quality and to enact a paradigmatic shift in how we think about educational resource quality assurance. This is especially pertinent for professionals in ODL as a shift to open textbooks will need to be supported by studies on resource quality to help bolster the perceived overall quality of online learning experiences.

Open textbook stakeholders recognize quality control to be an essential ongoing priority (Atenas, Havemann, & Priego, 2014). Limited research into open textbook quality in K-12 has found that open textbooks may be of higher quality than their copyright-restricted (and easily outdated) counterparts and that because open textbooks can be adapted, they can continuously improve in quality as educators edit and adapt them to their needs (Kimmons, 2015). In general, open textbook “quality control is in the hands of learners and teachers ... instead of instructional [or content] experts” (Dinevski, 2008, p. 121), because open textbooks can be adopted at any time in a semester, can be utilized in a partial or scattershot manner, and can be adapted by instructors for ongoing improvement, thereby circumventing traditional vetting and approval mechanisms. Thus, new methods are needed for evaluating the quality of open textbooks and including student voice in this process. Research in this area is fledgling but should include more sophisticated means of evaluation than those offered by traditional peer review by content experts (Clements & Pawlowski, 2011; Kimmons, 2015; Kimmons, 2016). Meaningful evaluation of textbooks requires a deep understanding of contextualized educator and student needs, not just content knowledge (Carey & Hanley, 2007). Yet, textbook evaluations are generally instructor- or department-driven and altogether ignore student voice (cf. Mahmood, 2011; Stein, Stuen, Carnine, & Long, 2001). This means that textbook evaluations are currently monophonically shaped by the prejudices, attitudes, and beliefs of instructors.

From a hermeneutic perspective, such prejudices are expected and problematic when considered alone (Kinsella, 2006). Polyphony, or a range of voices, is altogether missing (Bakhtin, 1981), and since textbooks are selected for students, it seems reasonable that students should have voice in this decision. Widdershoven (2001) explains that “only if the evaluator tries to understand the values endorsed in a practice and is prepared to listen to the stories of all parties involved, can evaluation become a practice to which all participants contribute” (p. 253). In the current context, students can only contribute to meaningful selection of textbooks if they play a major role in the evaluation process, rather than being
viewed as lacking “information or the capacity to take part in the process of deliberation” (Widdershoven, 2001, p. 260).

In this study, we seek to answer the following research questions:

1. How can student voice guide the evaluation and selection of a course textbook?

2. What can this teach us about potential comparative quality of open textbooks?

**Methods**

This study applies a hermeneutic, student voice-driven lens to the evaluation of competing course textbooks and took place in a graduate-level project management course at a large, private, high research activity university in the U.S. Students in the course were full-time graduate students in the Instructional Psychology and Technology department at the institution, either at the Masters or Doctoral level, and enrolled students generally prepare for careers in instructional design, educational technology, or educational research. The first two authors of this study were enrolled students in the course, and the third author was the instructor. The instructor served as a facilitator to encourage students to thoughtfully shift between the roles of learner and evaluator and to mentor students through the evaluation process. Prior to this study, the student authors did not have training in course content or textbook evaluation, and the course was structured so that students would learn the subject matter while simultaneously conducting the research project. It was anticipated that this approach would: a) give students an authentic experience in project management, b) expose students to polyphonic perspectives on project management, c) mentor students in research and evaluation practices, d) empower students to exhibit voice and agency in their own learning and in the course structure, and e) provide valuable results on student voice, OER, and textbook evaluation.

The instructor began the class by introducing the course project and providing students with access to eight, de-identified electronic textbooks that were either commonly used in courses or by professionals in the field. The instructor selected textbooks by: a) consulting colleagues and syllabi from various peer institutions, and b) consulting Amazon sales ranks of books in the topic area. Of these, six textbooks were under copyright restrictions while two textbooks were openly released under a Creative Commons license. Restricted textbooks were made available to the students by licensing agreements through the institutional library.

Structure of the course beyond the first class period was directed by the student-directed project team as participants determined how to best approach the task of textbook evaluation and took turns serving as the project manager for each week. The team was given eight weeks to complete the project, and students elected to initially peruse all texts and to rank them according to their initial perceptions while memoing the process. Students then revisited the texts and articulated the pros and cons associated with their two “most liked” and two “least liked” textbooks. This information was then used to determine what criteria should be included in a more formal evaluation process, drawing upon their own initial reactions and using the existing literature for thematic comparison and common terminology. Recognizing that it would
be unfeasible to deeply read all eight textbooks in a single course, students decided to conduct evaluations at two levels: broad and deep.

**Broad Evaluation**

Broad evaluation consisted in the establishment of readily identifiable items that were valuable to the students and were applied to all eight textbooks at the book level. These broad evaluations included 24 questions representing four student-determined categories: **Content**, **Design**, **Pedagogy**, and **Cost**. **Content** included trustworthiness of the author and alignment with standards (i.e., PMBOK or project management body of knowledge); **design** included logical structure, visual layout, and aesthetics (e.g., font size, graphics); **pedagogy** included audience-appropriate word usage and inclusion of scaffolding and learning tools (e.g., glossaries, multimedia, summaries, and objectives); and **cost** included the general cost of the text. The included items were intentionally designed to consider only the criteria that students felt competent to evaluate, thereby excluding some possible but important criteria (e.g., content accuracy). Most items were evaluated on a 5-point Likert scale (ranging from Strongly Disagree to Strongly Agree with a neutral midpoint) by each student and then normalized to a 0.0 - 1.0 scale. Categories included both information- and value-based items, raw scores for value-based items were determined by averaging ratings together, and overall book scores were determined by weighting raw scores for each item according to each item’s importance, determined by student consensus. Categories were also weighted to capture subjective importance to students: Content (25%), Design (33%), Pedagogy (33%), and Cost (8%). Weightings were agreed upon by students through a process of consensus regarding what they believed to be most important when selecting a quality textbook. A complete list of evaluation items along with weightings are available in the online supplement (Kimmons, n.d.). Results of the broad evaluation allowed students to then select the two highest-rated textbooks for deeper evaluation, one of which happened to be open and the other restricted.

**Deep Evaluation**

Deep evaluation criteria were then constructed including the categories of **Content** (40%), **Design** (20%), and **Pedagogy** (40%) and were applied at the chapter level. Students created a reading schedule and evaluated each chapter by completing a 5-point Likert scale evaluation for each of the nine created questions. Overall deep evaluations were calculated by averaging weighted chapter-level evaluations. Deep evaluation criteria differed from broad evaluation items, because they required nuanced reading and immersion in the text. For instance, when considering visual design elements and the presence of illustrations, a broad item was worded as follows “the graphics look high-quality and seem helpful (rather than intimidating),” while a deep item was worded as follows “illustration/graphics/multimedia are effectively used to add to the instructional value of the material.” Evaluating visual design elements in this manner allowed the evaluators both to account for initial impressions of visual elements on the reader and their potential value for learning.

**Ensuring Qualitative Rigor**

Many considerations were made to help ensure qualitative rigor. First, evaluation team debriefs were conducted for 1.5+ hours, semi-weekly, in which student member checking occurred (Lincoln & Guba, 1985). Since students served as both learners and evaluators in this project, they had to regularly take time to thoughtfully switch roles, reflect on the process, and consider how results reflected their actual
perceptions. Second, ongoing memoing occurred during team meetings, evaluations, and readings, and these memos were used as a data source for reflection and analysis (Birks, Chapman, & Francis, 2008). And third, triangulation was employed both in the use of multiple data sources (broad evaluations, deep evaluations, and memos) as well as in the use of multiple evaluators (two students and one instructor), and results represent consensus from these parallel triangulation processes (Creswell, 2002; cf. Figure 1).

![Figure 1. Parallel triangulation of data sources and researchers to support rigor.](image)

Furthermore, evaluators conducted deep evaluations at the chapter level to allow for the generation of highly-focused data points, thereby allowing for greater analytical rigor. When doing these deep evaluations, it was not possible for students to read and evaluate every chapter in the studied textbooks within the available timeframe, but saturation in these evaluations was reached once evaluators felt confident that new chapters would not influence overall perspectives (Creswell, 2002). In addition, subsequent, deep criteria chapter evaluations were regularly compared to the emerging average evaluation for each text and provided a guide for identifying reduced variations over time. For example, in its penultimate deep evaluation, the Amado et al. (2015) chapter differed only .02 points from its mean, and the last deep evaluation differed less than .01 point from the mean. We also took this as a sign of saturation.

**Results**

We proceed by providing results for the broad, deep, and summative evaluations. Numeric results are provided in decimal form or as percentage points (i.e., .75 = 75%), and .50 represents neutrality on the Likert scale items.

**Broad Evaluation**

Results of the broad evaluation indicated a wide range of overall scores: 35-72%, $M = 51\%$, $SD = 14\%$. The average score of the two open textbooks was higher than restricted textbooks (60% vs. 49%), and the open
textbooks were ranked first and fifth in the list of eight. Category scores revealed that open and restricted textbooks were rated similarly on content and design, with a difference of 7-9%, but that open textbooks performed better on pedagogy by 29% (cf. Table 1).

Table 1

Broad Evaluation Results for All Eight Textbooks

<table>
<thead>
<tr>
<th>Rank</th>
<th>Type</th>
<th>Content</th>
<th>Design</th>
<th>Pedagogy</th>
<th>Cost</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Open</td>
<td>.79</td>
<td>.35</td>
<td>.99</td>
<td>1.00</td>
<td>.72</td>
</tr>
<tr>
<td>2</td>
<td>Restricted</td>
<td>1.00</td>
<td>.70</td>
<td>.51</td>
<td>.00</td>
<td>.65</td>
</tr>
<tr>
<td>3</td>
<td>Restricted</td>
<td>.89</td>
<td>.43</td>
<td>.77</td>
<td>.26</td>
<td>.64</td>
</tr>
<tr>
<td>4</td>
<td>Restricted</td>
<td>.36</td>
<td>.66</td>
<td>.54</td>
<td>.26</td>
<td>.51</td>
</tr>
<tr>
<td>5</td>
<td>Open</td>
<td>.38</td>
<td>.54</td>
<td>.34</td>
<td>1.00</td>
<td>.47</td>
</tr>
<tr>
<td>6</td>
<td>Restricted</td>
<td>.89</td>
<td>.44</td>
<td>.09</td>
<td>.00</td>
<td>.40</td>
</tr>
<tr>
<td>7</td>
<td>Restricted</td>
<td>.29</td>
<td>.48</td>
<td>.32</td>
<td>.26</td>
<td>.36</td>
</tr>
<tr>
<td>8</td>
<td>Restricted</td>
<td>.66</td>
<td>.42</td>
<td>.06</td>
<td>.26</td>
<td>.35</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>.66</td>
<td>.50</td>
<td>.45</td>
<td>.38</td>
<td>.51</td>
</tr>
</tbody>
</table>

Deep Evaluation

Results of the deep evaluation indicated a difference between the top open and top restricted textbook, favoring the open textbook by 5%. The open textbook performed better on content (15%) and pedagogy (12%), while the restricted textbook performed better on design (13%; cf. Table 2). These deep evaluation results corroborated results of the broad evaluation. Parallel reading schedules were developed between the texts so that student evaluators could read the same content as it was addressed in the separate texts. Differences in numbers of pages and chapters evaluated were a result of the writing style, layout, and authoring choices of each text, because authors chunked content differently or gave some topics more attention and because editors used different font sizes, graphical elements, and so forth that influenced how much text might appear on a page.

Table 2

Deep Evaluation Results for the Top Two Textbooks

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Chapters evaluated</td>
</tr>
<tr>
<td>Open</td>
<td>5</td>
</tr>
<tr>
<td>Restricted</td>
<td>13</td>
</tr>
</tbody>
</table>
**Summative Evaluation**

Following triangulation design analysis (Creswell, 2002), evaluators reread all memos and found that most thoughts expressed therein were adequately captured in the broad and deep results. However, a few differences and nuances were noted that deserved additional attention. Overall, student evaluators felt that the two textbooks selected for deep evaluation were superior to the rest but that there were some nuanced strengths and weaknesses of each that should be more fully reported.

The open textbook by Amado et al. (2015), stood apart from the rest by providing many useful details and anecdotal illustrations, reducing the need for the instructor to clarify difficult concepts. However, some of its design decisions were distracting from the text. Anecdotal illustrations and case studies, for instance, were often poorly situated or seemed abrupt or out of place, and graphics typically used unhelpful stock photos rather than illustrative diagrams.

The restricted textbook by Kuster et al. (2015), exhibited a writing style that was tight and conceptually concise with useful checklists that served as quick reference guides for practitioners. However, the theories covered were not grounded in clear examples and illustrations from the field, which meant that students relied heavily on the instructor to explain difficult principles and to situate concepts in real-life scenarios. Illustrations also exhibited varying degrees of usefulness, as some expertly clarified concepts while others were overly complex or distracting.

**Discussion**

Students evaluated textbook quality based upon the applicability of the content, the instructional value of the design, and the overall effectiveness of the pedagogical scaffolding in each chapter. Student perceptions of quality increased as more of these elements were present in a textbook, but students also recognized that they could not meaningfully evaluate all aspects of quality, such as content accuracy. Thus, though student voice revealed some important considerations for textbook quality, it was recognized that polyphony with an expert is necessary to ensure a fully-informed evaluation.

**Open vs. Restricted**

Interestingly, the two best textbooks identified as highest quality by students included an open textbook and the most expensive restricted textbook. Since cost was not a barrier in this study, cost-savings was not a factor in deep or summative analyses. While affordability is often the chief virtue lauded by advocates of open textbooks, this finding compellingly shows that when cost is neutralized in a comparative evaluation, open textbooks can compete with restricted textbook quality. However, each open textbook did not perform equally well, suggesting that future research should explore more generalizable patterns of comparative quality in these resources. One of the instructor’s goals for this project, which he did not disclose to students until after analysis was complete, was to determine comparative open textbook quality when cost was not a factor. He will use results to guide textbook selection in future course sections.
Multi-Authored vs. Single-Authored

Another finding is that of the eight textbooks analyzed in the broad evaluation, the top two selected were the only multi-authored options. This inadvertent selection may suggest the value of polyphony as an authoring decision as well as the value of openly licensed resources that allow for on-going revision, remixing, and collaboration. Future research should explore the effects of collaborative authorship on textbook quality at a larger scale.

Instructor Presence and Textbook Selection

As the evaluators reflected on their processes, they concluded that the best textbook is not based solely on the course to be taught, but ultimately upon the learning context and the relative presence of the instructor in the learning process. A textbook like Kuster et al. (2015), for example, which contains very few anecdotal or real-life illustrations, may actually be a better textbook for an instructor who intends to provide this illustrative scaffolding in class through personal anecdotes or real-life illustrations. If, however, instructor presence is weaker (such as in an online course), a textbook like Amado et al. (2015), with its extensive anecdotal illustrations to scaffold conceptual learning, may be of superior quality. In this regard, open textbooks could also provide value by allowing instructors to modify and personalize materials to fit their intended level of instructor presence. This is especially important in online and distributed learning scenarios where instructors may be more limited in their access to students or they might have to rely upon increased mediation via online tools. If online instructors cannot, for instance, provide on-the-fly, real-world examples to students as they would in traditional classroom discussions, then a textbook that provides many such examples (i.e., more scaffolding) might conceivably be more valuable for the student learning at a distance than for the face-to-face student.

Student Takeaways

Others may find this study’s approach useful for: a) choosing a textbook, b) understanding student textbook perceptions, and c) guiding approaches to engaging students in analysis and evaluation. On this third point, it should be said that while the students believed the traditional approach of reading an instructor-assigned textbook to be a valuable way to learn the concepts of a field, they felt that engaging in a project to critically analyze multiple textbooks by their own standards fostered a much greater depth of learning, as it helped them simultaneously learn about content, research, writing, and learning. Additionally, as students would alternate managing the project, they were able to immediately transfer what they were learning from the textbooks to what they were doing on the project. The student authors believe this process allowed them to deepen their learning as they reflected together on a semi-weekly basis, both on their readings and their application of project management principles. Thus students felt that a deep critical analysis of multiple textbooks combined with the application of project management principles and regular group-reflection, created a rich and fertile context for deep learning to occur. Conversely, the instructor author believes that having students involved in this way provided insights on the strengths and limitations of different textbooks and also shed light on student expectations for learning within the context of the course and subject area. This allowed for a deep degree of reflective practice on the part of the instructor to consider what resources and teaching strategies are most suitable for achieving course learning outcomes with similar students in the future.
Limitations
This study has a number of limitations and delimitations that should be addressed. First, as novices in the content area, students could not be expected to evaluate the accuracy of textbook content. This is a limitation of student voice itself and was minimized by focusing on items that students could reasonably be expected to evaluate. Second, this study was limited to an evaluation team of two graduate students who had to strategically approach project scope in a manner that was suitable for a formal classroom setting. This made evaluation results authentic but potentially incomplete or non-representative of students generally. Readers should counterbalance potential student biases with reported methodological rigor to determine result transferability.

Conclusion
Leveraging the potentials of digital technologies to explore the transformational value of open textbooks upon the traditional copyright-restricted paradigm, this study has sought to establish the possibility of meaningfully including student voice in the cloistered processes of textbook evaluation and adoption. Results indicate that textbook quality can encompass a variety of dimensions for students and that not all textbooks perform equally in addressing matters of content, pedagogy, and design. Comparative findings on open textbooks reveal that these alternatives can compete with their restricted counterparts, even when cost advantages are ignored. From this study, we hope that instructors and institutions will explore processes for including student voice in textbook evaluation by exploring open alternatives to traditional textbooks, identifying student textbook quality considerations, and polyphonically empowering students to participate in textbook adoption and adaptation.

References


Walker, E. (2007). Evaluating the results of open education. In T. Iiyoshi, & M. S. V. Kumar (Eds.), *Opening up education: The collective advancement of education through open technology, open...*
content, and open knowledge, 77-88. Retrieved from