How High School, College, and Online Students Differentially Engage with an Interactive Digital Textbook

Jeremy Warner[1], John Doorenbos[2], Bradley N. Miller[2], Philip J. Guo[1]

[1] University of Rochester, Rochester, New York, USA
[2] Luther College, Decorah, Iowa, USA

ABSTRACT

Digital textbooks have been growing popular as a lower-cost and more interactive alternative to paper books. Despite the recent rise in adoption, little is known about how people use these resources. Prior studies have investigated student perceptions of digital textbooks in the classroom via interviews and surveys but have not quantified actual usage patterns. We present, to our knowledge, the first large-scale quantitative study of digital textbook usage. We mined 6.8 million log events from over 43,000 people interacting with *How To Think Like a Computer Scientist*, one of the most widely-used Web-based textbooks for learning computer programming. We compared engagement patterns among three populations: high school students, college students, and online website viewers. We discovered that people made extensive use of interactive components such as executing code and answering multiple-choice questions, engaged for longer when taking high school or college courses, and frequently viewed textbook sections out of order.

Keywords

Digital textbooks; student engagement; server log data mining

Categories and Subject Descriptors

H.5.1. [**Information Interfaces and Presentation (e.g. HCI)**]: Multimedia Information Systems

1. INTRODUCTION

Digital textbooks have grown popular in the past decade as more students gain access to laptop computers, tablet devices, and broadband Internet. Some of their claimed benefits over paper textbooks include lower cost, lighter physical weight, full-text search, electronic note-taking, and better accessibility for sight-impaired students via text-to-speech [4]. As the costs of paper textbooks continue to rise, university professors are adopting digital alternatives to save money for their students [13]. Governments are pushing for widespread adoption of digital textbooks at the K-12 level as well. For instance, in his 2011 State of the Union address, U.S. President Barack Obama challenged all K-12 schools to adopt digital textbooks by 2016, and the FCC Chairman and Secretary of Education

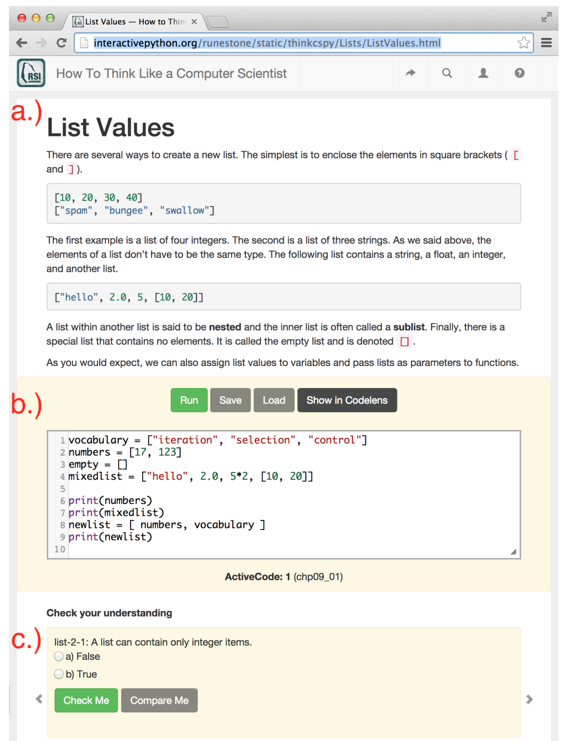


Figure 1: *How To Think Like a Computer Scientist* [8] is a Web-based interactive digital textbook for learning computer programming. A user can: a.) read text, b.) edit and run Python code to see outputs, and c.) answer multiple-choice questions.

followed up with a plan to implement this vision [12]. The publishing industry has responded to recent events by converting many of their paper textbooks into digital formats. By some estimates, digital textbook sales will be a $1.5 billion business and account for over 25% of all new textbook sales by 2016 [11]. In parallel, universities [1], non-profits, and independent volunteers [8] are developing freely-available digital textbooks.

Aside from classroom use, online digital textbooks are a form of educational technology similar to MOOCs. Anyone with a computer and Internet connection can learn topics ranging from computer programming [8, 10] to math using digital textbooks. In recent years, many researchers have studied how students use MOOCs [3, 6], but to our knowledge, there has never been an analogous large-scale study of digital textbook usage. Given the growing prominence of digital textbooks, it is important to understand how students use them in a variety of educational settings, and how that could inform the design of the next generation of digital textbooks.

This paper contributes, to our knowledge, the first large-scale study of how students use an interactive digital textbook. We studied *How To Think Like a Computer Scientist* [8], a Web-based digital textbook for learning computer programming (Figure 1). We analyzed two years of server logs containing 6.8 million events from 43,416 students. This data is far larger, more diverse, more precise, and finer-grained than prior digital textbook studies that relied on questionnaires sent on university campuses [2, 9, 13].

Specifically, we quantified how students navigated through the textbook and engaged with interactive components such as live code and multiple-choice questions. We segmented students into three populations: those taking a high school course, a college course, and those visiting the public textbook website. These comprise the three main populations of textbook readers. We investigated three sets of research questions: 1.) How much does each population engage with interactive components of the textbook? 2.) When do people in each population access the textbook, and for how long do they persist before quitting? 3.) How do readers navigate non-linearly and skip around when accessing textbook contents?

The first generation of digital textbooks were simply paper books converted into electronic formats such as PDF. The current generation features interactive topic-specific widgets (Figure 1) but does not take advantage of the scale afforded by tens of thousands of online readers. This study is one step toward providing data to inform the design of the next generation of digital textbooks, which can leverage such data to assist students, instructors, and book authors.

2. RELATED WORK

Researchers have studied student attitudes toward digital textbooks in the classroom, with mixed findings. Questionnaire studies of 446 students in the University of Cape Town in South Africa [13] and of 5,000 business school students across 127 U.K. universities [9] found high self-reported enthusiasm for adopting digital textbooks. In contrast, a survey of 662 students across five California State University campuses found that only 1/3 were satisfied with digital textbooks and only 1/2 felt they were easy to use [2]. Prior studies were all done on non-interactive digital textbooks, comparing them to nearly-identical paper counterparts. And they all relied on questionnaires and exam results but did not analyze log data on actual textbook usage. To our knowledge, we are the first to study an interactive digital textbook in-the-wild in a large-scale online setting. Our sample contains 43,416 students from around the world, which is one to two orders of magnitude more students than prior studies.

3. METHODOLOGY

We studied usage patterns of *How To Think Like a Computer Scientist* [8], a widely-used Web-based digital textbook for learning introductory computer programming. This textbook is viewable online for free at **`http://interactivepython.org/`**. Figure 1 shows how it intersperses textual content, snippets of editable and runnable Python code, and multiple-choice questions. This digital textbook shares similarities with computer programming MOOCs. Both feature multiple-choice questions and runnable Python code as interactive components. However, unlike a MOOC, the main pedagogical modality here is text rather than video. Also, registration is not mandatory. Readers can register with a free account to save their code and track personal analytics, but this is an open resource that anyone can access on the Web. Finally, there is no notion of a fixed course schedule with, say, weekly releases of new materials like there is in some MOOCs. All textbook materials are always present, which supports self-paced learning.

We mined the server logs from June 2012 to 2014, fetching 6,834,244 events from 43,416 students. Each event has the following fields:

- **Timestamp** – server time in the U.S. Central Time Zone
- **Student type** – High School, College, Open (public website)
- **Student ID** – either a registered username or an IP address
- **Event type** – Page load, Run code, Code error, Viz interaction (Python code visualization), or Multiple-choice attempt
- **Textbook location** – the chapter and sub-chapter to which this event belongs (e.g., chapter 5, sub-chapter 3).

Event types: The *Event type* field has one of the following values:

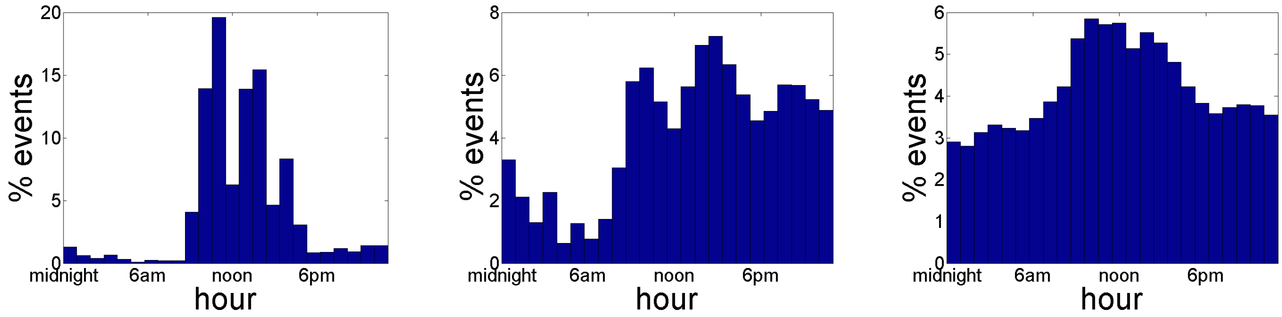
- **Page load** – Load a webpage, which displays the content for a specific sub-chapter of the textbook
- **Run code** – Press the “Run" button to run a piece of Python code, and the code executes successfully (Figure 1b)
- **Code error** – Press the “Run" button to run a piece of Python code, but the code has a syntax or runtime error
- **Viz interaction** – Interact with a Python code visualization widget by taking one step forward or backward in the embedded visual single-step debugger tool [5]
- **Multiple-choice attempt** – Attempt to answer a multiple-choice question within a webpage (Figure 1c)

Non-Linear Navigation: We define a ***backjump*** as any consecutive pair of events for one student where the first occurred in chapter n and the second in chapter m, where $n > m$. A ***sub-backjump*** is either a regular backjump, or a pair of events in the same chapter that went from sub-chapter n to sub-chapter m, where $n > m$. We define skip and sub-skip similarly. A ***skip*** is any consecutive pair of events where a student jumped from chapter n to chapter m, where $m > n{+}1$. Note that we use $n{+}1$ because simply going to the next chapter is ordinary sequential navigation, not a skip. A ***sub-skip*** is either a regular skip, or a pair of events in the same chapter that went from sub-chapter n to sub-chapter m, where $m > n + 1$. The intuition behind these metrics is that if a student navigated through the textbook in a perfectly sequential fashion, starting with chapter 1, sub-chapter 1, and ending with the final sub-chapter of chapter 15, then they would have zero backjumps or skips. Thus, backjumps and skips indicate non-linear navigation.

4. FINDINGS

4.1 Engagement with Interactive Components

Most students actively engaged with the interactive components rather than just passively reading. Figure 2 shows that page loads accounted for only around 10% of total events. If students had simply been using this textbook as a static reference, then *all* events would have been page loads. By far the most common event type was attempting to run Python code. *Run code* and *Code error* events comprise around three quarters of total events. Recall that pieces of Python code are embedded throughout the textbook (Figure 1b.). Some are complete working examples that can be run verbatim without triggering errors, while others are incomplete snippets that students must complete as an exercise. For all three populations, attempting multiple-choice problems and interacting with code visualizations were about as common as page loads, which again indicates that students did not just passively read the book.



(a) High School (b) College (c) Open

Figure 3: Distributions of events throughout the day, recorded as server time in the U.S. Central Time Zone.

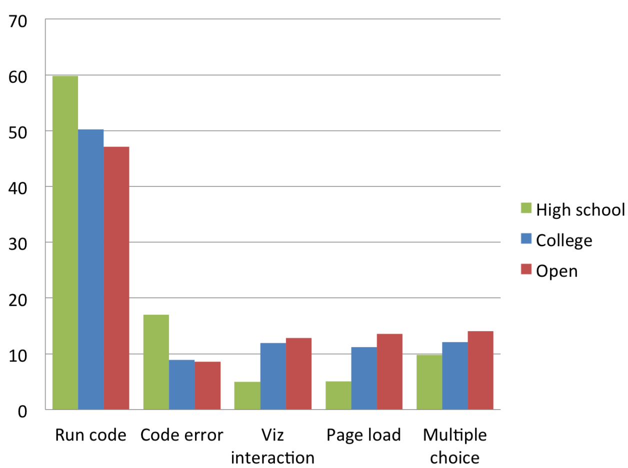


Figure 2: Percentages of total events by type.

4.2 Writing, Running, and Debugging Code

Figure 2 shows that high school students ran the most code, with ∼10% more *Run code* events and twice as many *Code error* events than college and open. Also, for high school students, 22% of total code run attempts resulted in an error, versus only 15% for college and open. High school students made, on average, 112 errors per student, versus 35 errors per student for college and 12 for open.

One interpretation is that high school students made more errors because they were less experienced at coding, but we do not have the data to support this claim. Since this is an introductory textbook, presumably the college and open students also did not have much prior coding experience. A more likely interpretation is that the high school students used this textbook in a more structured and instructor-guided manner than college and open. We have anecdotal evidence from high school teachers who sent emails to the textbook creators requesting technical support that many intended to use this strictly within their classrooms. A typical use case is a teacher directing students to spend the class period reading through a sub-chapter and attempting to do all of the code-related exercises. The teacher would then walk around the classroom and help students debug their faulty code. Thus, high school students ran more code and persisted in debugging, fixing their errors, and re-running possibly because an instructor was present in the classroom.

In contrast, college and open students are usually less supervised. College instructors typically assign readings from a textbook but do not monitor students as closely as high school teachers do. Since running code and attempting multiple-choice problems are ungraded formative exercises, students can work on them at their leisure. Open students might be self-directed learners with little to no supervision. Thus, they make fewer code errors (12 per student) not necessarily because they are better at coding, but simply because they might give up after seeing an error and not persist in fixing it.

4.3 Activity Levels by Time of Day

Visualizing activity levels by time of day confirms that high school students mostly use this textbook in class during school hours, while college and open students use it throughout the day. Figure 3 shows the distribution of event times. The majority of high school activity occurs between school hours of 9am to 4pm, with a sharp dip at noontime for lunch. This pattern indicates in-class usage, supervised by a teacher. In contrast, college activity occurs evenly throughout most waking hours from 8am to midnight.

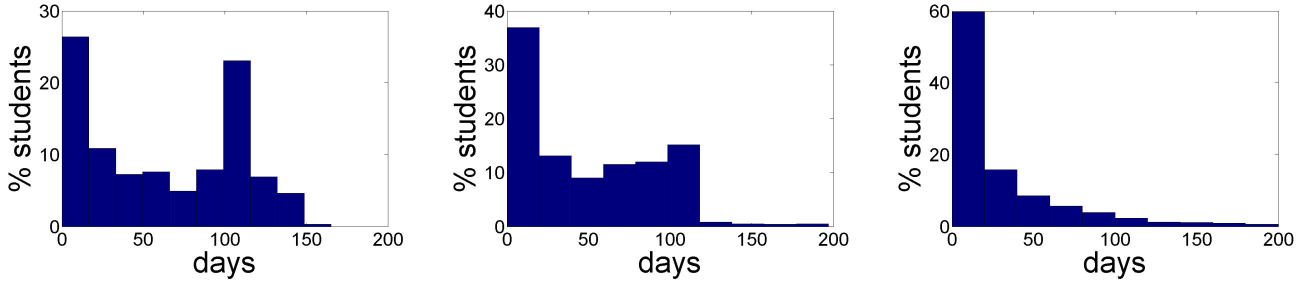
Note that the event timestamp is the server's time (U.S. Central Time Zone), so it does not take the student's local time zone into account. However, by geolocating IP addresses of high school and college students, we found that the majority with a geolocatable IP were from the U.S. and Canada (89% of high school and 94% of college students), so the true time for those students lies within a few hours of the U.S. Central Time Zone.

Whereas high school and college students came mostly from the U.S. and Canada, the open student population was much more international. Only 57% of open students were from the U.S. or Canada, and many came from countries such as Australia, New Zealand, the U.K., and India. Unsurprisingly, those are all English-speaking countries, since this textbook is in English. The presence of many international students explains the relatively even levels of activity throughout the day and night in Figure 3c, although there is still a spike during mid-day in the U.S. and Canada.

4.4 Engagement Duration

For how long does each student engage with the textbook before quitting? We quantified engagement duration by calculating the difference between the first and last event times for each student. Figure 4 plots the distributions for all three student types. High school and college students engaged for up to a semester (∼ 105 days) because they used the textbook as part of a course. The high school spike at around 105 days is much more pronounced than the college one, which could be a result of greater teacher supervision.

In contrast, the open population engagement drops off sharply in a long-tail-like distribution, which mirrors the high initial dropout



(a) High School (mean=62, median=62) (b) College (mean=48, median=39) (c) Open (mean=28, median=13)

Figure 4: Distributions of how many days each student was actively engaging with the textbook, split by student type.

Student Type	Backjumps mean	Backjumps median	Sub-Backjumps mean	Sub-Backjumps median
High School	41.0	2	58.4	11
College	13.2	2	21.8	4
Open	3.8	0	6.1	1

	Skips mean	Skips median	Sub-Skips mean	Sub-Skips median
High School	38.5	4	67.3	13
College	13.1	1	27.4	7
Open	4.3	1	9.2	3

Table 1: Non-linear navigation statistics for all student types.

rates in MOOCs [3, 7]. Half of the open students used the textbook for less than two weeks. However, unlike many MOOCs, which incrementally release new course materials on a weekly basis, all of the material in this textbook is always available. Thus, it is possible for self-directed learners in the open population to engage for a week or two, learn what they want, and then leave. Thus, semester-long engagement is simply an artifact of formal course schedules.

4.5 Non-Linear Navigation

How frequently did students jump backward to earlier textbook locations or skip forward to latter ones out of sequence? Table 1 summarizes the levels of backjump and skip activity by student type. For all four measures we defined (backjump, sub-backjump, skip, sub-skip), high school students exhibited the most non-linear navigation, followed by college, then open. Even controlling for differing levels of activity per student, high school students perform twice the number of backjumps and skips as college and open students. For instance, 6.2% of all high school events involved backjumps, versus only 3.4% of college and 2.7% of open events.

Non-linear navigation indicates engagement, since it takes more active effort to jump around rather than following the default sequential ordering of the textbook by simply clicking the "Next page" link at the bottom of each page. One explanation for the high numbers of backjumps and skips for high school students is that they are using the textbook in the classroom, so their teacher can proactively direct them to other parts of the textbook as they are trying to solve coding problems. Without other people present in-person to guide or direct one's learning, it is easier to default back to the more passive style of reading through the textbook in a linear way.

Another interpretation is that high school and college students navigate non-linearly to review materials when studying for exams. A related study of non-linear navigation in MOOCs showed that students often backjumped from exam pages back to earlier lecture pages [6]. In contrast, open students might be self-studying without taking a graded course, so they do not need to review as much.

5. REFERENCES

[1] Open SUNY Textbooks – `http://opensuny.org/`.

[2] Baek, E.-O., and Monaghan, J. Journey to textbook affordability: An investigation of students' use of eTextbooks at multiple campuses. *The International Review of Research in Open and Distance Learning 14*, 3 (2013).

[3] Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., and Seaton, D. T. Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research and Practice in Assessment 8* (Summer 2013).

[4] Courduff, J. Digital Textbooks and Students with Special Needs – `http://goo.gl/JGTaA9`, Accessed: Oct, 2014.

[5] Guo, P. J. Online Python Tutor: Embeddable Web-based Program Visualization for CS Education. SIGCSE '13, ACM (2013), 579–584.

[6] Guo, P. J., and Reinecke, K. Demographic differences in how students navigate through MOOCs. L@S '14, ACM (New York, NY, USA, 2014), 21–30.

[7] Kizilcec, R. F., Piech, C., and Schneider, E. Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. LAK '13 (2013), 170–179.

[8] Miller, B. N., and Ranum, D. L. Beyond PDF and ePub: Toward an interactive textbook. In *Proceedings of ITiCSE* (2012), 150–155.

[9] Nicholas, D., Rowlands, I., and Jamali, H. R. E-textbook use, information seeking behaviour and its impact: Case study business and management. *Journal of Information Science 36*, 2 (Apr. 2010), 263–280.

[10] Pritchard, D., and Vasiga, T. CS Circles: An In-browser Python Course for Beginners. SIGCSE '13, ACM (New York, NY, USA, 2013), 591–596.

[11] Reynolds, R. Trends influencing the growth of digital textbooks in us higher education. *Publishing Research Quarterly 27*, 2 (2011), 178–187.

[12] Usdan, J., and Gottheimer, J. FCC Chairman: Digital Textbooks to All Students in Five Years. `http://goo.gl/VJ9NA0`. Accessed: Oct, 2014.

[13] van Heerden, M., Ophoff, J., and Van Belle, J.-P. Are university students ready to dump their textbooks? A survey on student attitudes towards e-readers and tablet computers. *Int'l Jour. Cyber Ethics in Education 2*, 3 (2012), 15–44.