

The Social Lives of Books: Reading Victorian Literature on *Goodreads*

Karen Bourrier, Mike Thelwall

^aUniversity of Calgary, Alberta ^bUniversity of Wolverhampton, England

ARTICLE INFO

Peer-Reviewed By: Katherine Bode, Leah Price

Article DOI: 10.22148/001c.12049

Dataverse DOI: 10.7910/DVN/ZMVANC

Journal ISSN: 2371-4549

ABSTRACT

This paper compares social media traces from *Goodreads* to data from the *MLA International Bibliography* and the *Open Syllabus Project*, in order to better understand the preferences of readers of Victorian literature from different but overlapping communities. We find that the majority of works of Victorian literature that are indicated as being read on *Goodreads* occur about as often as they are taught or written about in the academy, although books aimed at an adult audience are written about more frequently in peer-reviewed venues. Interestingly, those works that are statistical outliers in terms of their greater popularity with a general audience than an academic audience tend to feature women authors, children's literature, and works with a strong female protagonist. Turning to an analysis of the written reviews on *Goodreads* of three outliers that were more popular with a general audience--*A Tale of Two Cities*, *Jane Eyre*, and *The Secret Garden*--we find that readers tend to comment on plot (especially in Dickens), feminist themes (in *Jane Eyre*), and the importance of characters (in all three works). In conclusion, we suggest ways in which postsecondary teachers might draw on these results to inform their syllabi and formulate strategies for teaching Victorian literature. We argue that in terms of outliers, popular taste in Victorian literature among *Goodreads* users reflects more general reading preferences among this user group, as readers turn to the Victorian era to read children's literature and books featuring strong female characters.

Much of what we know about how twenty-first century readers engage with Victorian literature outside of the academy tends to come from personal essays and memoirs. Recently, we might think of Rebecca Mead's *My Life in Middlemarch* (2014) or Nell Stevens's *Mrs. Gaskell and Me* (published in the U.S. as *The Romantic and the Victorian* 2018) as examples of memoirs that interweave the experience of reading (and rereading) George Eliot or Elizabeth Gaskell alongside intimate events from the authors' lives, showing what these nineteenth-century authors continue to mean for twenty-first century readers.¹ Scaling up from the level of the individual in order to understand the larger patterns that govern contemporary reading preferences has proven more difficult. Yet, as scholars turn to the digital sphere, and to online platforms including but not limited to Amazon book reviews

and *Goodreads*, recovering these reading patterns is becoming more possible as digitization makes these reading traces increasingly accessible. As Simone Murray points out, perhaps the most significant aspect of these online reviewing platforms is “its greatly broadened base of participation.”² The large scale of online reviews also requires new methods. As Tatlock et al have argued in an article analysing data on library patrons in Muncie, Indiana, from 1891 to 1902, computational methods may be especially suited to “investigations of reader agency”; such “quantitative analyses of reader behavior” may allow us “to enhance our understanding of how meaning is co-constructed.”³

In this paper, following scholarship in the reception history of literature, we compare social media traces from *Goodreads* to data from the *MLA International Bibliography* and the *Open Syllabus Project*, in order to better understand the preferences of readers of Victorian literature from different but overlapping communities. We find that the majority of works of Victorian literature that are indicated as being read on *Goodreads* occur about as often as they are taught or written about in the academy, although books aimed at an adult audience are written about more frequently in peer-reviewed venues. Interestingly, those works that are statistical outliers in terms of their greater popularity with a general audience than an academic audience are dominated by women authors, children’s literature, and works with a strong female protagonist. We argue that in terms of outliers, popular taste in Victorian literature among *Goodreads* users reflects more general reading preferences among this user group, as readers turn to the Victorian era to read children’s literature and books featuring strong female characters. This could be the case because 76% of *Goodreads* users are women, and that the books that they are likely to have read are in the majority by women authors, from Jane Austen to J. K. Rowling, many of whom are writing in the young adult genre.⁴ In contrast to the outliers on *Goodreads*, syllabi in English literature courses and works in the *MLA bibliography* continue to focus on male authors.⁵

In the second part of the paper, we move to an analysis of the written reviews on *Goodreads* of three books—*A Tale of Two Cities*, *Jane Eyre*, and *The Secret Garden*—which are all outliers in terms of being more popular with a general audience than we would predict given how often they are taught and written about in the academy. Character was the most commonly commented-upon category in the

written reviews of all the novels, but readers of *A Tale of Two Cities* were likely to frame the characters using E. M. Forester's terms of flatness versus roundness, to comment on the character development of the children in *The Secret Garden* (without reference to flatness or roundness), and to sympathize with Jane Eyre's feminism, without focusing on her childhood. This list of outliers is augmented by works that would traditionally be considered minor by authors like Charlotte Brontë and Oscar Wilde, who would traditionally be considered major authors, which suggests that readers pick their next book by author.

Data set

In quantifying the twenty-first century reader response to Victorian literature, we follow the example of works like Janice Radway's *Reading the Romance*, which, as Ted Underwood observes, relies on an "experimental method drawn from the social sciences," and relies largely on "questionnaires, interviews and numbers" to analyze the type of romances that readers found satisfying.⁶ We too rely on written reviews and numbers, drawing on publicly available information on *Goodreads*, the MLA bibliography and the *Open Syllabus Project* to analyze reader preferences. Although part of our methodology is computational, we would stress that this is not a distant reading of literature, but rather a type that Alison Booth has characterised as "midrange reading," in which we use computational data to shed new light on a medium-sized corpus of works of literature we know well as scholars of Victorian literature.⁷

To begin our analysis, we compiled a list of every author who published a book during Queen Victoria's reign (1837 to 1901) that was included in the Chadwyck-Healey database of nineteenth-century fiction.⁸ The Chadwyck-Healey database casts a wide net, including canonical as well as popular authors, from Charlotte Brontë to Charlotte Yonge. Searching by author stretched our results beyond the boundaries of the Victorian period, since many authors, including Frances Hodgson Burnett and H. G. Wells, continued publishing after Queen Victoria's death in 1901. It also stretched our results beyond fiction, since many Victorian authors wrote in multiple genres. For example, Oscar Wilde's plays and children's literature as well

as his fiction for adults continue to be widely read. We then scraped the rating count and the average rating for all of the books on *Goodreads* associated with these individual authors, as well as a few canonical Victorian sage writers and Victorian poets for comparison.⁹ This resulted in 203 books, when conditioning on only those books with more than 1,000 ratings on *Goodreads* at the time of analysis. 79.8% of the books in our study of Victorian works of literature were written by a male author and 20.2% were written by a female author, 63% had a male protagonist and 31% had a female protagonist (in 6% of cases the gender of the protagonist was undetermined or not applicable). 87% of books in our study were aimed at an adult audience (we counted a book as a work of children's literature if one of the top ten user-defined tags on *Goodreads* was for children's literature). For each of these books, we added the number of peer-reviewed articles in the MLA bibliography, and the number of times the book has been taught in the syllabi aggregated on the *Open Syllabus Project*, to our data set. Our final step was to calculate the statistical outliers using linear regression in order to find out which books are more often read by *Goodreads* users than we would predict given how often they are taught or written about in the academy, and which books are less frequently read by *Goodreads* users than we would predict given how often they are taught or written about. We also considered whether authors and main protagonists were male or female, and whether the main audience for the book was children or adults, as a dimension of the analysis.¹⁰

Some background on the sources of our data may be helpful to contextualize our findings. Most familiar to literary scholars will be the MLA bibliography, which aggregates information on works published in literary studies from 1926 to the present. In order to determine how many articles were published on a given book, we searched the MLA bibliography with the text in question listed as the "subject work" (e.g. *Jane Eyre*), and filtered the results by those articles and books marked as peer-reviewed. Our second source of data, focusing on which works we teach, is the *Open Syllabus Project*, an outgrowth of Dan Cohen's Million Syllabus Project, which scrapes syllabi from the web (though users can also contribute their syllabi directly) and aggregates the data to show the number of times different works are taught. The current database contains approximately 1.1 million syllabi from disciplines including history, English and biology.¹¹ It is possible to filter the syllabi by discipline (i.e. just those books taught on English courses), but for the purposes

of this paper we counted books as taught at the university level regardless of discipline; in other words, we counted *Jane Eyre* regardless of whether it appeared on a history syllabus or an English literature syllabus.¹² This data set is limited in terms of geographic and temporal scope; the vast majority of syllabi are from universities in the U.S., U.K., Canada, and Australia in the past fifteen years. The data set also favours those syllabi that have been posted online for public consumption.

Our final source of data, *Goodreads*, currently the 99th most trafficked website in the US (Quantcast Sept 28, 2018), is a social cataloguing site which allows readers to list books that they “want-to-read” are “currently-reading” and have “read,” to review these books and rank them on a five-star scale, and to share what they are reading with their friends and followers.¹³ Launched in 2007 and acquired by Amazon in 2013, *Goodreads* is by far the most popular social media site devoted to books, with 65 million users and counting. Research in both English literature and computer science has found in *Goodreads* a rich source of knowledge about the way that people read now. In Lisa Nakamura’s words, *Goodreads* offers an embarrassment of riches for scholars looking to track reading habits “in the wild,”¹⁴ although, of course, like any source of information about readers, there are demographic biases inherent in *Goodreads* that limit how far we can generalize about all readers from *Goodreads* data (see below). Scholarship on *Goodreads* so far has investigated topics ranging from which genres men and women readers tend to favour (Thelwall), the differences between written reviews on *Goodreads* and Amazon (Dimitrov et al), and Flannery O’Connor’s reception amongst twenty-first century readers (Moran).¹⁵ There is significant overlap between *Goodreads*, *Open Syllabus*, and the MLA bibliography. For example, Victorianists may teach some of the same things that they publish on, and some members of *Goodreads* are academics.¹⁶ Because there is overlap between these domains, there is potential for the lack of differences between these groups to be explained in part because the groups are populated by some of the same people, who have similar reading preferences inside and outside the classroom. Nonetheless, the similarities and differences among what we teach and write about in the university, and what *Goodreads* users report reading, deserve further exploration.

Some demographic information on *Goodreads* users sheds light on who holds the reading preferences we explore in the rest of the paper. Approximately 76% of *Goodreads* users are women.¹⁷ Women read almost twice as many books as men, though they are more willing to read books by authors of either gender.¹⁸ *Goodreads* users are educated: 47% have some college, and 26% have been to graduate school. In terms of race, 79% of *Goodreads* users are white, 9% Hispanic, 7% African American, 4% Asian, and 1% other. In terms of age, an estimated 88% are under age 54.¹⁹ These readers participate in a variety of book-based and social activities within the site, which allows users to form book clubs.²⁰

In order to take a closer look at the habits of those *Goodreads* users whose lists revealed a preference for Victorian novels, we scraped the virtual bookshelves of readers who belonged to two popular groups mainly dedicated to nineteenth-century literature: *The Readers Review: Literature from 1714 to 1910*, and *Victorians! Members of Victorians!* who had read *Jane Eyre* were 89% female (as opposed to the 76% female users of *Goodreads* overall), which suggests that women may be particularly interested in Victorian literature. (This claim may not surprise those who have taught *Jane Eyre* recently and observed the warm reception many female students give the novel, but it is worthwhile to have data beyond anecdotal evidence to back up the claim.) This finding suggests that women readers on *Goodreads* may prefer works by women and works with female protagonists. However, since we sampled the reading habits of book club members, it may also be that a preponderance of women are likely to join virtual book clubs and that the men who read *Jane Eyre* are less likely to be members of such a group.

Looking at the general preferences of readers who joined the *Victorians!* book club can help us better understand their cultural context. The top five books read by this group, which were not exclusively Victorian, were: *Jane Eyre*, *Pride and Prejudice*, *The Great Gatsby*, *To Kill a Mockingbird* and *Harry Potter and the Philosopher's Stone*. As this list indicates, in general, members of these two groups read a combination of classic literature and contemporary bestsellers. In a separate study of the fifty books read by the fifty most popular English-language reading groups on *Goodreads*, we found that while 91% of people who were members of one of the fifty largest reading groups on *Goodreads* had read Suzanne Collins's popular YA dystopian novel, *The Hunger Games*. This novel was less popular among those who

self-selected into a group focusing on nineteenth-century literature. Only 49% of the members of *The Reader's Review* had read *The Hunger Games*, which indicates those readers who join an online forum dedicated to nineteenth-century literature may be less likely to keep up with contemporary young adult fiction than the average book club member on *Goodreads*. To give another example, of the fifty books most commonly read by book club members, *Pride and Prejudice* was the most commonly read book in *The Reader's Review*, while *Insurgent* was the least commonly read; Jane Austen was the most commonly read author, while young adult author Rainbow Rowell was the least read.²¹

Findings

Top Victorian authors on *Goodreads*

In absolute terms, the top ten most-read works by Victorian authors on *Goodreads* at the time of writing are *Jane Eyre*, *Wuthering Heights*, *Dracula*, *The Secret Garden*, *The Picture of Dorian Gray*, *A Tale of Two Cities*, *Alice in Wonderland*, *Great Expectations*, *A Christmas Carol* and *Treasure Island*. We extrapolate this information from the number of individual ratings for each book (on a one to five-star scale). We rely on user rating for a book having been read at least in part. (*Goodreads* allows readers to catalogue books they would like to read with a “to-read” tag). On the page for each author, *Goodreads* keeps an updated list of that author’s average rating, total ratings, and total number of written reviews. A bar chart shows us the twenty-two most commonly read Victorian authors, colour-coded by overall rating (see figure three). The most-read Victorian author is Charles Dickens, whose books have been rated 2,661,330 times at the time of data collection (May 2018). Following Dickens, Arthur Conan Doyle, Charlotte Brontë, Oscar Wilde, Emily Brontë, Lewis Carroll, and Frances Hodgson Burnett all boast more than a million ratings on *Goodreads*. Victorian writers of adventure, horror and fantasy, R. L. Stevenson, H. G. Wells and Bram Stoker make up the rest of the top ten most-read Victorian authors. It is significant that with the exception of Dickens, Wilde, and the Brontës, the top ten writers are primarily known as either children’s authors (Carroll and Burnett) or as what we would now call genre authors, working in horror or fantasy. The most highly rated Victorian authors included in this study

on *Goodreads*' five-star scale are Arthur Conan Doyle (4.21), John Henry Newman (4.16, not shown), and Frances Hodgson Burnett (4.13).

Top Victorian Authors

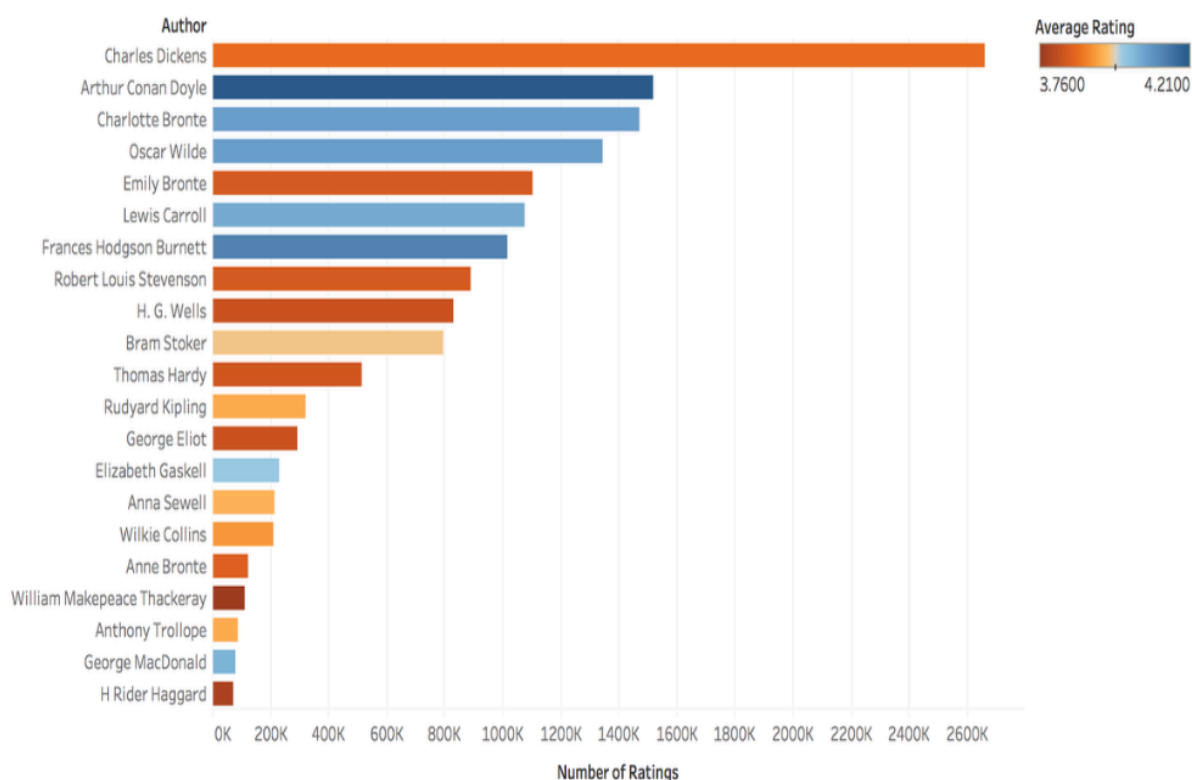


Figure 1. The twenty-one most read Victorian authors included in our study, as of May 18, 2018, colour-coded according to average rating, with the most highly rated authors in dark blue and the least highly rated authors in dark orange

Reading patterns on *Goodreads* vs Open Syllabus vs the MLA bibliography

While the raw numbers of ratings of Victorian literature on *Goodreads* are interesting in themselves, comparing what is read by a general audience to what we teach in the college classroom and write about in peer-reviewed journals can give us a more nuanced picture of reader preferences. In order to compare data from *Goodreads*, the MLA bibliography, and the *Open Syllabus Project* we used multiple linear regression.²² In the case of comparing what works of Victorian literature are

read by a general audience, taught in the college classroom, or written about in peer-reviewed venues by academics, we would predict that the more often university professors teach a book, or the more they write about it, the more social media users would report reading it on *Goodreads*. In some cases, as we will see below in the cases of William Morris and George Meredith, this is wishful thinking. However, for the most part, the books readers report reading on *Goodreads* are the same books we teach and publish on; only 27 out of 203 titles, or 13% of the books included in our study were outliers in the statistical sense of having standardised residuals with magnitude above 1.96 or below -1.96. The gender of the author or the protagonist was not a factor overall in determining which books were read, taught or studied (see Appendix for regression results). In most cases, regardless of the gender of the author or the book's protagonist, a book was read about as often as it was taught or studied. However, the main audience of the book did have some influence, with books with a mainly adult audience being written about more often in peer-reviewed venues. Simple multiple linear regression identifies which books are outliers in terms of being read much more often than we would predict for how often they are taught, or read much more often than we would predict given how often they are the subject of peer-reviewed articles (as an approximate guide, the points representing these books on the scatterplot fall far outside the line, although they were detected from their scores as residuals from the regression equations). It also allows us to determine which books are taught or written about in the academy more often than we would predict given how often they are read by general readers.

In order to determine the relationship between what academics write about in peer-reviewed articles and what a more general audience reads, we conducted an ordinary least squares regression of (log) *Goodreads* readers against (log) MLA subject tags, with audience (adult/children), main character (female/male) and author (female/male) as additional independent variables. In this particular dataset the gender of the authors was relatively straightforward. We marked a book as a work of children's literature if one of the top ten user-determined tags on *Goodreads* put the work in this category. The gender of the protagonist was more ambiguous, and we omitted works in which it was unclear whether a male or female character was the protagonist. The purpose of this regression was to assess whether these factors systematically influenced the relationship between the number of *Goodreads* readers

and the MLA citations. The residuals were close to normal, there was evidence of only minor heteroscedasticity and negligible collinearity, so the results are reasonably statistically robust, except that some of the books are related (they have the same author) and their residuals may therefore not be independent. Our results showed that books had relatively few *Goodreads* readers given the number of articles published on them in the MLA if the audience was adult and the author was female, but more *Goodreads* readers if the main character was female. Only the first (adult audience) achieved statistical significance ($p=0.05$), however, so it is reasonably likely that the character ($p=0.91$) and author ($p=0.89$) gender associations in the data are due to chance factors.

In order to determine the relationship between what we teach and what a general audience reads, we also conducted an ordinary least squares multiple linear regression conducted of (log) *Goodreads* readers against (log) *Open Syllabus* citations, with audience (adult/children), main character (female/male) and author (female/male) as additional independent variables. Books when one of these factors was unclear (e.g., multi-gender main characters or none) were omitted. The purpose of this regression was to assess whether these factors systematically influenced the relationship between the number of *Goodreads* readers and the *Open Syllabus* mentions. The residuals were reasonably close to normal, there was evidence of very minor heteroscedasticity and negligible collinearity, so the results are reasonably statistically robust, except that some of the books are related (same author) and their residuals may therefore not be independent. None of the gender and audience variables came close to achieving statistical significance ($p>0.2$ in all cases) and so there may well not be a general trend for any of these factors to lead to relatively many or few *Open Syllabus* mentions compared to *Goodreads* readers.

The chart below compares the number of times the books in our study have been read on *Goodreads* (y-axis, running vertically), to the number of times they have been taught (x-axis, running horizontally). Those books that are read corresponding to how much we would predict given how often they are taught follow the line through the centre of the graph. Good examples of books that follow the predicted trajectory, being read commensurate to how often they are taught include Edward

Lear’s “The Owl and the Pussycat” and Trollope’s *Can You Forgive Her?* To give an intuitive visual impression of the outliers, those that are read less often than we would predict given how often they are taught tend to occur towards the top left of the graph, and those that are taught less often than we would predict given how often they are read tend to occur towards the bottom right of the graph, although they were identified as outliers from their regression residuals rather than their position on the graph.

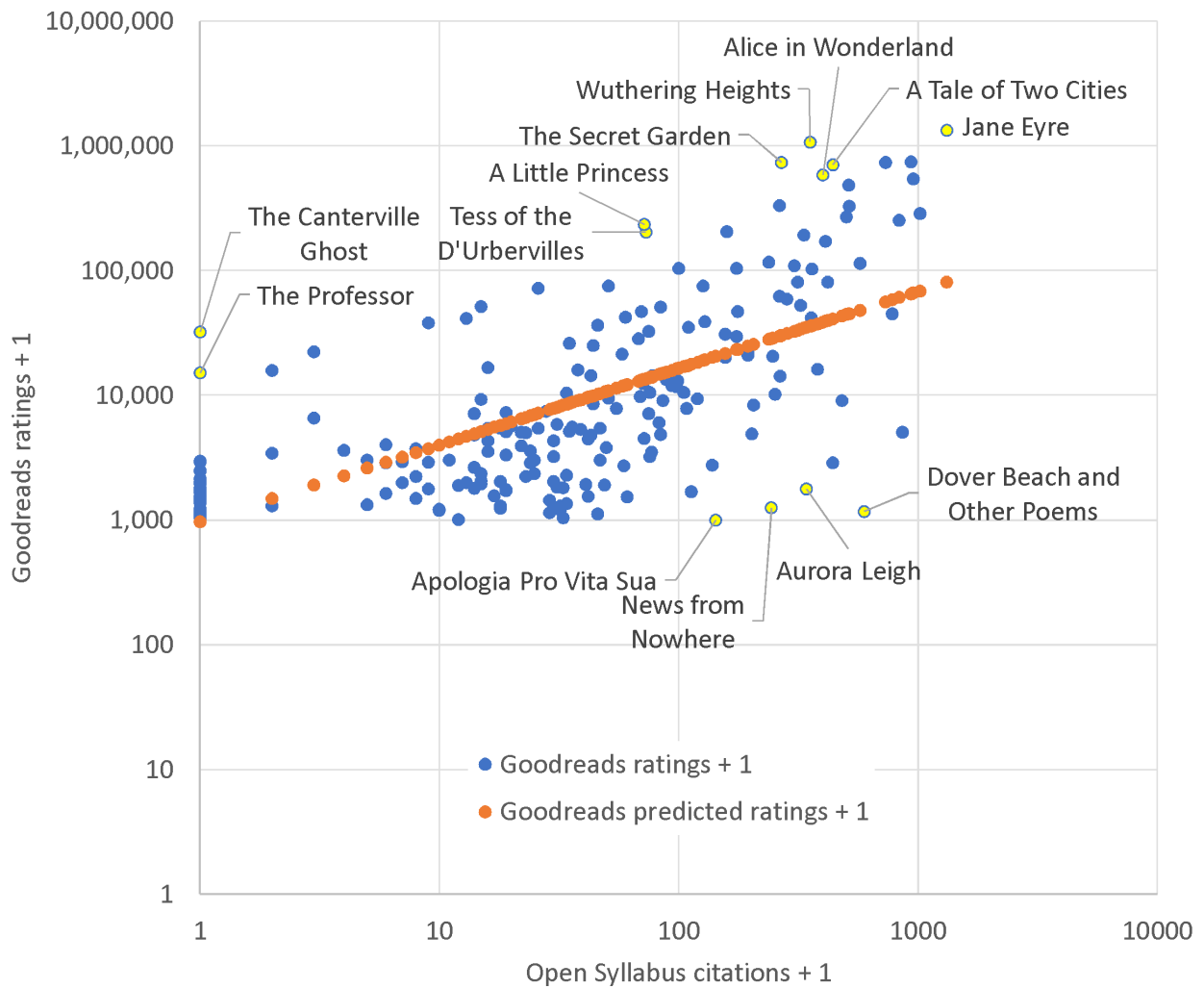


Figure 2. Works of Victorian literature that are read by a general audience (y-axis) versus those that are taught (x-axis). Log-log analysis. Graphs converted to a log-log scale, with works mentioned in the text of the article labelled. See Appendix for regression statistics. The orange regression line predictions ignore audience age, author gender and main character gender. Outliers are in yellow. Those outliers that are read less often than we would predict given how often they are taught appear towards the top left of the graph, and those that are taught less often than we would predict given how often they are read appear towards the bottom right of the graph.

Although, overall, the works of literature are read by a general audience about as often as they are taught, a few key patterns emerge from the top outliers, which are read more by a popular audience. The works that are outliers are represented by yellow points in the graph: Oscar Wilde, “The Canterville Ghost,” *Wuthering Heights*, *The Secret Garden*, *A Little Princess*, *Alice in Wonderland*, *A Tale of Two Cities*, *Jane Eyre*, *The Professor*, and *Tess of the D’Urbervilles*. This small sample is used here to draw qualitative insights and the conclusions are limited in generalisability from a quantitative perspective as a result.

First, all of these works, with the exception of “The Canterville Ghost,” are novels. Second, children’s literature (three out of nine, compared to 12% of the overall list) and novels with a strong female protagonist are strong presences on the list (six out of nine, 66.7%, compared to 31% of the overall list), and five out of nine books (55.6%) are by women writers (55.6% compared to 20% of the overall list). Third, the inclusion of “The Canterville Ghost” and *The Professor*, which academics would traditionally consider minor works by Charlotte Brontë and Oscar Wilde, suggests that readers may be beginning with well-known works by these authors (for example, *Jane Eyre* and *The Picture of Dorian Gray*) and working their way through an author’s corpus. In the initial stages of this work, we had hoped to be able to test this hypothesis using reading patterns among *Goodreads* users who have joined a group devoted to Victorian literature, but not enough readers included the date read for each book marked as read for us to be able to determine whether they started with *Jane Eyre* and then moved on to *Villette* and *The Professor*, for example. This may change in the future as *Goodreads* accrues more data on reading habits. However, the sheer numbers of readers attracted by certain authors are suggestive. Indeed, the 203 Victorian works with more than 1,000 ratings on *Goodreads* included 19 works by Dickens, 16 by Wilde, 15 by Trollope, 14 by Arthur Conan Doyle, 14 by Hardy, 13 by H. G. Wells, and 10 by George MacDonald, suggesting that readers went deep into the catalogues of certain authors. Taken together, these seven (male) authors wrote 49% of books in our study.

In contrast to the books which are outliers with a general audience, which are dominated by children’s literature and novels by women writers or with strong female protagonists, those that are statistical outliers in terms of being taught more often than they are read include poetry and non-fiction prose. These works are:

Matthew Arnold, “Dover Beach”, William Morris, *News from Nowhere*, John Henry Newman, *Apologia Pro Vita Sua*, Elizabeth Barrett Browning, *Aurora Leigh*. Looking at the scatterplot (figure 2), although they are not technically statistical outliers, *Hard Times*, *The Strange Case of Dr. Jekyll and Mr. Hyde*, and *The Importance of Being Earnest* are also favourite works inside the classroom but less so with a general audience given how often they are taught. We might suspect that relatively short length of these works—*Hard Times* is Dickens’s shortest novel—combined with important themes in Victorian studies, including the Industrial Revolution, degeneration theory, and decadence, accounts for their popularity in the classroom.

A second chart compares how many general readers a book attracts on *Goodreads* (y-axis running vertically) to how often these works are written about in the MLA bibliography (x-axis running horizontally). This model fits slightly better than the previous one, explaining 47.8% of the variance in the data compared to 46.9% for the syllabus mode (see Appendix A), although the difference is too small to draw conclusions from. Here again, we would emphasize that most works are read in proportion with how often they are written about; good examples of books that closely follow the regression line include Anthony Trollope’s *Doctor Thorne* and Thomas Hardy’s *Under the Greenwood Tree*. There is a strong overlap between those works of literature that are outliers in terms of being written about in peer-reviewed venues less than they are read by a general audience and those that are taught but not read, with the exception that George Meredith’s *The Egoist* replaces Cardinal Newman’s *Apologia Pro Vita Sua* on the list of what academics write about.

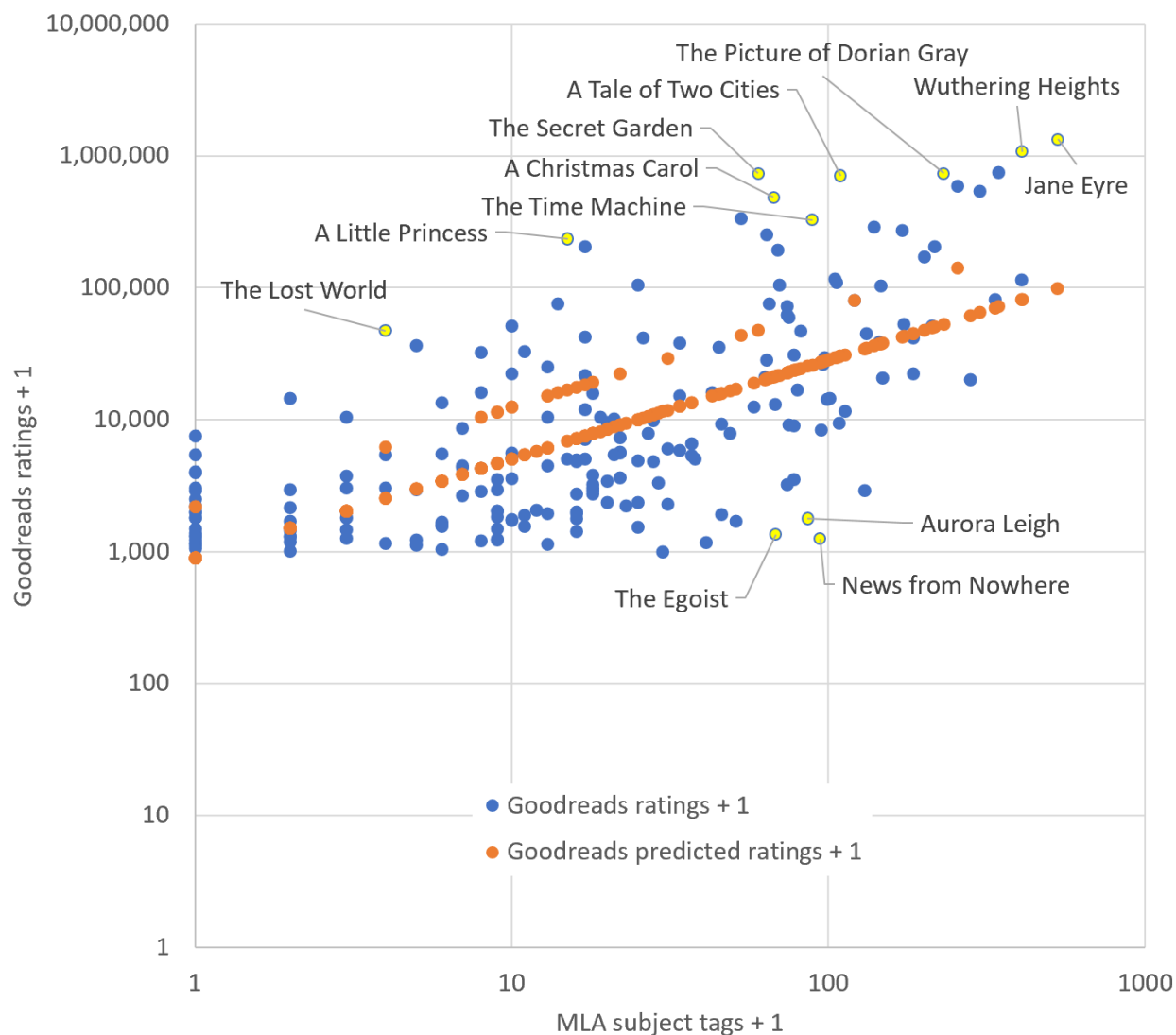


Figure 3. Books that are read by a general audience according to Goodreads (y-axis) versus those that are written about in peer-reviewed work indexed by the MLA bibliography (x-axis). Graph converted to a log-log scale with works mentioned in the text of the article labelled. See Appendix for regression statistics. The orange regression line predictions account for audience age (statistically significant) but ignores author gender and main character gender. Outliers are in yellow. Those outliers that are read more often than we would predict given how often they are written about in the MLA bibliography appear towards the top left of the graph, and those that are read less often than we would predict given how often they are written about in the MLA bibliography appear towards the bottom right of the graph.

It is worth noting, as well, that although they are not statistical outliers in our model, Tennyson's *In Memoriam*, Olive Schreiner's *Story of an African Farm*, and William Makepeace Thackeray's *Barry Lyndon* all appear in the top ten books that are not frequently read by a general audience given how often they are studied or taught. A Venn diagram demonstrates the strong overlap between these two lists of the top ten works that are written about by academics and the top ten works that are taught more

in the college classroom often than we would predict given how much they are read. On the right are the books that are taught more than read, on the left are books that are written about more than read, and in the middle are those that appear on both lists.

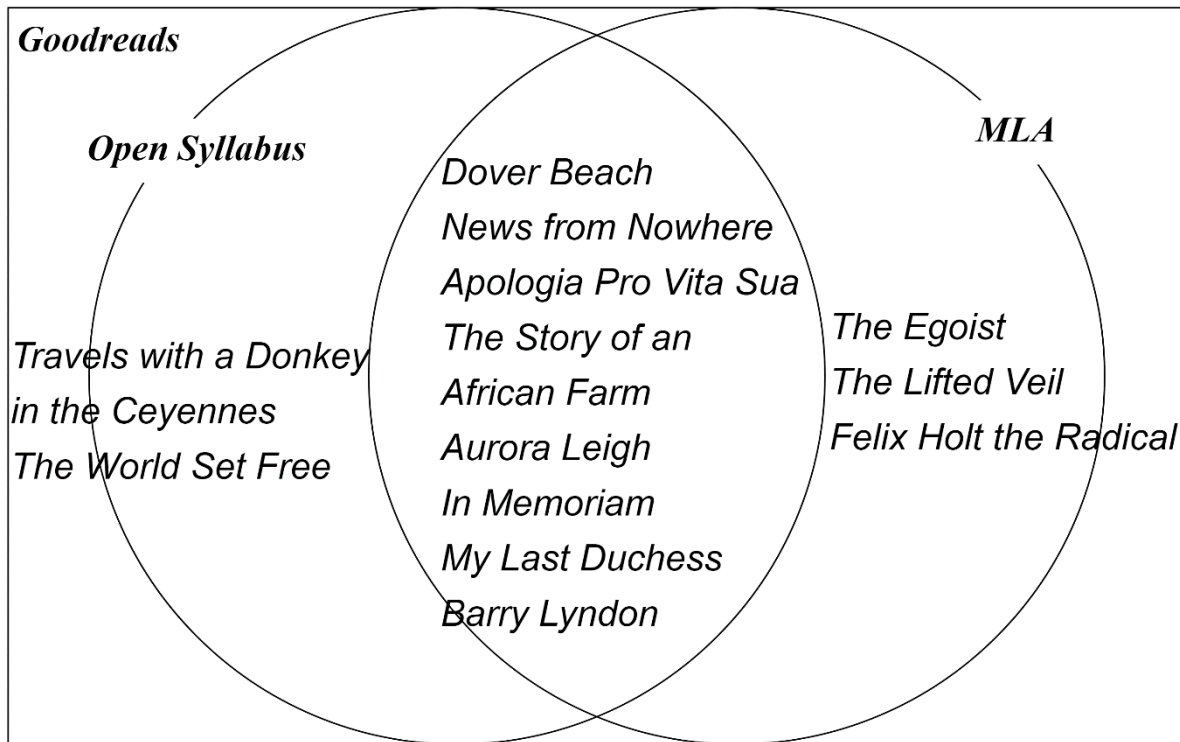


Figure 4. Books that are more popular in the academy. The left-hand circle shows those books that are read less often than we would predict given how often than they are taught, the right-hand circle shows those books that are read less often than we would predict given how often they are the subject of peer-reviewed work. The overlap between the two circles shows those books that fall on both lists.

Looking above the regression line on figure 3 (MLA subject tags versus *Goodreads* ratings) tells us which works of Victorian literature are read more often than we would predict given how often academics write about them in peer-reviewed journals. These works are: *The Secret Garden*, *A Little Princess*, *Black Beauty*, *A Tale of Two Cities*, *A Christmas Carol*, *Treasure Island*, *The Lost World*. These outliers are mainly works of children’s literature or works that have a strong history of being adapted for children and young adults. The popularity of children’s literature with a general audience is the flipside of our finding that an academic audience is more likely to write about adult literature. Indeed, for many Victorian authors, their work for children is the most popular on *Goodreads*; Charles

Kingsley's most read work is *The Water-Babies*, John Ruskin's is *The King of the Golden River*, and Ouida's is *A Dog of Flanders*.

Just as there is a strong overlap between which books are popular inside the academy (those that are more often written about and taught than read); there is a strong overlap between the top ten outliers of books that are read more often by a general audience than we would predict given how often they are taught and written about. A Venn diagram shows the overlap between the top ten works that are more popular with general readers than they are in the college classroom and the top ten works that are more popular with general readers than they are in peer-reviewed scholarship.

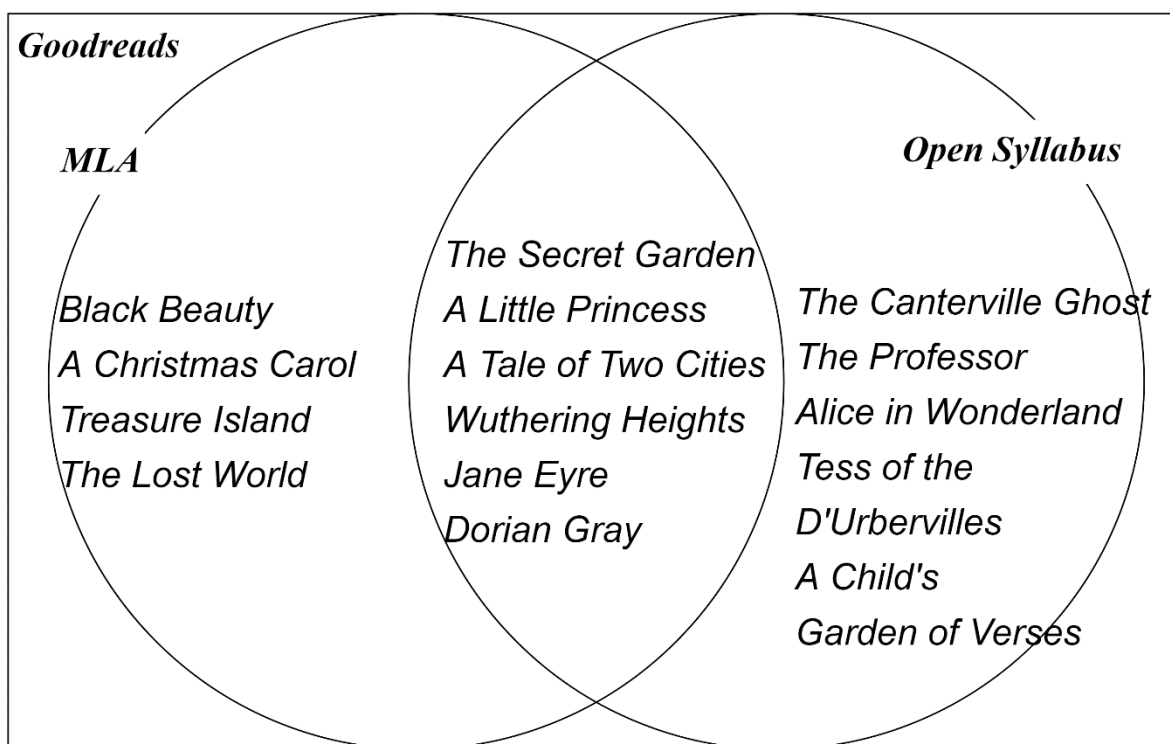


Figure 5. Books that are more popular with a general audience. The left hand circle shows those books that are read more often than we would predict given how often they are the subject of peer-reviewed work, the right hand circle shows those books that are read more often than we would predict given how often they are taught, the center shows the overlap between the two (works that are more popular on Goodreads than they are in the classroom or in peer-reviewed publications).

Three patterns emerge from these two lists of books that are more popular outside of the academy than they are inside it: first, Victorian children's literature remains popular; second, books with a strong female protagonist are popular; and third,

Goodreads readers seem to choose their reading according to author to some extent, with even minor works by Oscar Wilde and Charlotte Brontë being read more often than we would predict given how often they are taught or written about. *The Secret Garden*, *A Little Princess*, *Alice in Wonderland*, *Black Beauty* and *Treasure Island* are all readily classified as works from the Golden Age of children's literature, which roughly coincides with the Victorian era. The two works by Dickens, in particular *A Christmas Carol*, have a strong history of adaptation for children; in her work on crossover fiction, Sandra Beckett suggests that Charles Dickens's most famous novels "were written for adults, but were popular with readers of all ages."²³ Beckett also suggests that *Wuthering Heights* and *Jane Eyre* have "long been among the first adult novels to be read by adolescents."²⁴ This long history of novels by Dickens and the Brontë sisters as works for children and young adults bolsters our sense that children's literature is an important category for works of Victorian literature that continue to be popular.

Written reviews on *Goodreads*

In the second part of this essay, we turn to an analysis of the top 100 written reviews of *Jane Eyre*, *A Tale of Two Cities*, and *The Secret Garden*—all works that were top outliers in terms of being more popular than we would predict with a general audience given how often they are taught or written about—in order to glean further insight into what continues to attract general readers to these books.²⁵ Overall, we find that characters are the most important attraction for *Goodreads* users in any book; we also find that pre-existing expectations about a book's genre may be important in determining reader responses, with readers commenting extensively on the romance plot in *Jane Eyre* (but not the protagonist's childhood) and on the children's moral growth in *The Secret Garden*.

We used the qualitative analysis software NVivo, which offers a free educational license, to facilitate our reading of the reviews. NVivo offered us two main advantages over pen and paper: first, it allowed us to automatically code for word frequency, which was useful, for example, when we wanted to see how many readers of *The Professor* mentioned *Jane Eyre* or *Villette* in their reviews, in order to determine whether the decision to read Brontë's least-known novel was influenced

by having read her more well-known works.²⁶ Second, when we read through and coded by hand for concepts and themes that are not easily captured by word frequency, for example, the idea that *The Professor* was a practice novel for *Villette* or *Jane Eyre*, NVivo kept an automatic tally for us of the number of times that we encoded this concept and allowed us to pull up these quotations again instantly, which helped us to ensure consistency in the kinds of quotations we coded under different themes.²⁷ That said, despite our reliance on software and numbers, this close reading is still an interpretative act based on the model we have constructed: not everyone will agree, for example, that a reference to Dickens's "masterful storytelling" in *A Tale of Two Cities* should be encoded as a positive reference to the way the novel is plotted, to take one of our more ambiguous examples. We readily acknowledge that written reviews of literature can be ambiguous, but methodologically speaking, we hope that attempting to quantify mentions of certain themes across 100 reviews can help us move beyond an impressionistic reading to a more systematized one. For example, one theme that we had expected (or hoped?) would come up in written reviews was the continued relevance of Victorian literature in the twenty-first century. This theme did crop up, but only in 17/85 reviews of *A Secret Garden*, 12/92 in *Jane Eyre*, and 1/80 in *A Tale of Two Cities*.²⁸ Had we not been keeping tally, we might have been tempted to overemphasize the significance of these responses.

As the word frequency across written reviews indicated, "character" emerged as the central way that readers engaged with Victorian literature.²⁹ This result dovetails with Deidre Lynch's argument that "In the late twentieth-century, after all, it is (still) the time that we spend with characters that matters the most to many readers."³⁰ Data from *Goodreads* indicates that this continues to be the case in the twenty-first century. However, readers used different frameworks to interpret characters depending on the author.

Readers of *A Tale of Two Cities* (1859) framed their experience of Dickens's historical novel in terms of how much they liked the characters and how real or well-rounded the characters felt to the reader (40/80 mentions), the dark themes raised by the historical setting of the French revolution (36/80 mentions), and the novel's plot (30/80 mentions). Although many reviews mentioned the names of various characters in Dickens's novel, we only encoded a reference to the importance of

Dickens's characters when the reviewer included a meta-reflection on the characters in general, e.g. The novel has "a cast of quirky characters only Dickens could create."³¹ Comments on *Goodreads* reveal the persistent influence of E. M. Forster's distinction between flat characters, or those "constructed round a single idea or quality," and round characters, who "cannot be summed up in a single phrase" and are capable of surprising us.³² While Forster argues that flat characters are useful, for most contemporary readers the term flat character is negative and the term round character is positive. Readers who commented on Dickens's characters for the most part enjoyed them (25/40 comments were positive). The positive comments referred to characters as "memorable," "exceptional," "vivid," and "amazingly life-like"; these readers noted that they came to care deeply for the characters and that they felt real.³³ The influence of Forster becomes especially evident in the negative assessments of Dickens's characters, which reviewers refer to as "not fully developed," lacking in "depth" or "roundness" "two-dimensional" "one-dimensional" or "superficially-drawn." More mixed or neutral reviewers noted that they didn't feel an "emotional tug" toward any character until the end, or that "resplendent" female characters like Madame Defarge made up for "insipid" ones like Lucie Manette.³⁴

Readers' pleasure in the roundedness, or mixed nature of the characters was echoed in their taking pleasure in the mixed nature of the themes that Dickens deals with in *A Tale of Two Cities*. Thirty-six of eighty reviews in English mentioned the dark themes that Dickens explores in the novel; more than a third of the reviewers (16/36) who mentioned the novel's dark themes mentioned that Dickens juxtaposes these dark themes with uplifting themes in the same breath. As one reviewer put it: Dickens "crafts a tale of sacrifice and redemption set against the bleak background of the French Revolution"; another rather pithily wrote: "It's got love, sacrifice, revenge, revolt and other exciting verbs!"

Finally, readers of *A Tale of Two Cities* were likely to comment on the plot of the novel (30/80 mentions).³⁵ While many reviewers offered some plot summary as part of their review, we only encoded a review as mentioning plot if it was referenced on a meta-level, e.g. the book was "tightly-plotted." Reviewers were mainly positive about the plot of *A Tale of Two Cities*, with 83% (25/30) of those who mentioned it commending Dickens's storyline. As one reviewer put it: "One thing I love is

[Dickens's] ability to create a perfect storyline. Everything in this book fits together in the end like a perfect, intricate puzzle." Of all the books we analyzed written reviews of in detail, reviewers of *A Tale of Two Cities* were most likely to comment on whether they found the book challenging to read; as one reviewer commented, "It was as if the book was a thick piece of fabric, and I was a needle that was trying to break through to the other side."³⁶ At this time, no concrete data is available on how often *A Tale of Two Cities* is assigned in the high school classroom, but we might suspect that how often it is taught in secondary school accounts for the novel's popularity with a general audience. Indeed, *Goodreads* users were likely to tag it as required-reading for school, with "school" as the thirteenth most popular tag for the book (608 tags).³⁷

Character continued to be a salient theme for readers of *Jane Eyre*, although instead of thinking of Brontë's characters in terms of flat and round characters, the top 100 written reviews emphasize Jane's love story and its attendant passionate emotions (47/92 reviews written in English), as well as her role as a strong female heroine (40/92 reviews). These were usually but not always positive elements of the novel for readers. Typical positive comments about the courtship plot include: "I will return to this book if I ever become doubtful of true romantic love" and "I ended up being a sucker for the romantic subplot in this book, too, even though I can see how many terrible, wrong, bad choices the love interest made"; a more negative reviewer noted: "I never bought the romance between Jane and Mr. Rochester."

A much more universally appreciated theme than the romance plot was Jane Eyre's role as a strong female heroine. As one reviewer put it: "Once you get to make the acquaintance of courageous, zealous, outspoken, energetic, intelligent, principled, respectable Jane, you are bound to remember her forever." For almost all of the forty reviewers who mentioned Jane as a strong female heroine, the protagonist was a proto-feminist, though two readers expressed some reservation at this idea. One reviewer questioned: "What is it about Jane Eyre that seems to be an educated female rite of passage? I was somewhat looking forward to this book as it's an example of a strong woman who knows herself, but no."³⁸

More than a quarter of reviews (24/92) mentioned that *Jane Eyre* was a novel they had or would reread. One element that readers did not focus on was Jane's childhood,

which takes up the whole first volume of the three-volume novel, but received only 12 mentions in 92 reviews. To put this another way, “Rochester” was mentioned more than ten times as often in written reviews (429 times in 300 reviews) as “Reed” (42 times in 300 reviews). It is interesting to note that childhood was not a theme readers chose to comment on given the general popularity of children’s literature on *Goodreads*.

Reviewers of *A Secret Garden* were also most likely to mention characters as a major element in their reading of the novel, but they framed this discussion in terms of character development and childhood rather than love, feminism, or flatness and roundness. *Goodreads* reviewers were just as likely to comment on the character development of the two child protagonists, Mary and Colin (36 mentions), as they were on whether or not they had read the novel in their own childhood. For readers, this character development was tied to the theme of nature in the book (25 mentions), with many reviewers (though not all) explicitly connecting the growth of Mary and Collin to the growth of the secret garden. As one reviewer commented, “In contrast to the traditional Victorian literary trope of angelic children, the two main protagonists in *The Secret Garden* are extremely unlikable; yet despite, or even because of their flaws, they are able to heal others--and themselves”. Some reviewers expressed skepticism about the transformative power of the garden; as one reviewer noted: “If you are ugly, sick, bad-tempered, and nasty, you can become beautiful, healthy, happy, and nice, and all it takes is the fresh clean air of the Yorkshire moors and the companionship of people of an inferior class (as long as they are white and very, very clean)”.

Twenty-first century readers were not enamoured of Frances Hodgson Burnett’s racism, with nine reviewers mentioning it explicitly. Some readers were forgiving. As one reviewer from India put it: “Except for the persistent India bashing, I loved this book. In fact Mistress Mary, I loved the ending so much that I forgive your English superiority complex. Next time you visit here though, allow me to take you on the ride across India, I hope your impression will change.” Others were less forgiving, as one reviewer noted of her poor impression of the book: “the casual racism didn't make things much better. like I GET IT this is an old book but that doesn't mean I have to like it.”³⁹

The popularity of *The Secret Garden* with a general audience suggests the continued importance of the Golden Age of children's literature in determining which works of long nineteenth-century literature readers turn to. One of the most remarked-on themes in *Goodreads* reviews of *The Secret Garden* was whether or not the user had read the book as a child, with 36/87 reviews alluding to the book's status as beloved childhood reading, whether they had read it during their childhood or not. Indeed, of the reviewers that remarked on the book's status as childhood reading, almost a third (11/36) explicitly stated that they had not read the book in their childhood; as one reviewer noted: "I seem to be the only woman I know who didn't read and cherish this book as a child. So I decided to see what all the fuss was about." More than one reviewer mentioned owning and rereading the same copy since childhood; one reviewer mentioned his delight at regaining a copy that he had given away to his cousins in Singapore, another wrote of her childhood copy "I read the book to bits (I still have a copy held together with brown tape)". Some readers who had not read *The Secret Garden* as children mentioned being familiar with the story through having seen the film as children. For the most part, these reviewers seemed to be referencing the 1993 film directed by Agnieszka Holland.

Indeed, the two film versions of books by Frances Hodgson Burnett released in the 1990s, Holland's *The Secret Garden* and *A Little Princess* (1995, directed by Alfonso Cuarón) may be a large part of what is attracting general readers to her work. Frances Hodgson Burnett is not alone in being adapted for film and television. With the exception of *The Professor*, all of the works of literature that are more popular with a general audience than they are in the classroom or in peer-reviewed articles have been adapted for a visual medium. But, perhaps surprisingly, in a search for the words "movie," "film," "DVD," "TV," and "television" across 300 written reviews for the twelve outliers that are popular with a general audience, these words showed up most frequently in *A Little Princess* (140 times). *A Little Princess* outpaced even *A Christmas Carol* (126 mentions) for mentions of words related to adaptation. *Treasure Island* (113 mentions) also had significant mentions of these words; after *Treasure Island* there is a steep drop-off to *The Lost World* (71 mentions) and *Jane Eyre* (55 mentions). *Black Beauty* (19 mentions) and "The Canterville Ghost" (15 mentions) have the fewest mentions of film or television of those works that have been adapted.⁴⁰ Frances Hodgson Burnett's continued

popularity in particular seems traceable to film adaptations from the 1990s rather than the classroom.⁴¹

Limitations

Perhaps the largest limitation of this study is that data from *Goodreads* cannot tell us much about the way that people read Victorian poetry, essays, short stories or other commonly anthologized pieces now. We have not excluded sage writing, plays, and poems from our data as the results may still be of interest; nor did including them change which novels were statistical outliers. However, we would need a different model to study these works in-depth. The affordances of *Goodreads*—including the ability to add books to the database by barcode and ISBN—encourage users to rate single works by a single author that fall between two covers. Leah Price’s work on the novel and the anthology suggests that general readers know most of their Victorian poetry through anthologies. Price writes that in Britain “anthologies count among the only volumes of poetry that even stand a chance at mass-market success” while in North America “the economics of college survey courses have made ‘poem’ nearly synonymous with ‘anthology-piece.’”⁴² Furthermore, it is difficult to compare numbers on poems across our data sources. For example, the anthology *Love Poems: A Collection of Heart-Felt Verses* (68,647 ratings on *Goodreads*), which includes poetry by Tennyson as well as Byron, Shelley, Shakespeare and Blake, is popular, but the total number of ratings on *Goodreads* does not tell us how many people were reading Tennyson’s “Mariana” in particular. Similarly, Christina Rossetti’s *Complete Poems* is the furthest outlier in terms of books that are taught but not written about, likely because we teach by anthology. Browning’s “My Last Duchess” is the sixteenth most taught work on English syllabi on *Open Syllabus*, but the 297 ratings for the individual poem on *Goodreads* likely underestimates the total number of general readers. A different model—perhaps looking at the number of hits that a poem gets on Poetry.org—could give us a much better idea of which Victorian poems continue to be read today. As well, given that we only scraped data on a handful of Victorian poets and sage writers for this study and mainly focused on authors of fiction in Chadwyck-Healey (some of whom, like Charlotte and Emily Brontë also wrote poetry), a more extensive study

of these writers would take a different starting point for authors considered—perhaps those poets commonly anthologized.

A second limitation of this study is that although the authors that we scraped data on were all from the Chadwyck-Healey database of Victorian fiction, the results do not exclusively focus on what is being taught in the Victorian studies classroom or written about by Victorianists. Because we scraped data by author (as opposed to date), the works we collected include Edwardian works by those whose lives and careers spanned the early twentieth century, including Frances Hodgson Burnett and H. G. Wells. Thus, some works studied more properly belong to the long nineteenth century, though the Victorian era was our starting point. *Open Syllabus* and the MLA bibliography do not parse their data by subfield. In other words, while we can filter the results from *Open Syllabus* to show only works taught on English literature syllabi, we cannot filter to what is being taught on Victorian studies syllabi. Edwin A. Abbot's *Flatland: A Romance of Many Dimensions* (1884) has 27 citations on the MLA bibliography and 42,000 ratings on *Goodreads*. It is viewed as an early masterpiece of speculative fiction, but it is also a Wildean satire of Victorian society. At least some, if not the majority, of the 59 results for *Flatland* on *Open Syllabus* are likely to be from speculative fiction classes, but at this point we cannot determine how many. Similarly, we looked at all peer-reviewed works in our study, not just those in Victorian studies journals. Eight of 12 peer-reviewed articles on George MacDonald's best-known work, *The Princess and the Goblin* appeared in the George MacDonald journal, *The Northwind*, three appeared in children's literature journals, and one in *The Journal of English Language and Literature*. No peer-reviewed articles listing *The Princess and the Goblin* as a keyword subject appeared in a general Victorian studies journal. Although George MacDonald (1824-1905) is a Victorian author, and one who is still widely read, he is not in the mainstream of Victorian studies.

A third limitation of this study is that although we may have suspicions about why Victorianists write about authors like William Morris more than the general public reads them, or why books like *Aurora Leigh* and *Apologia Pro Vita Sua* are more often taught than read, this particular data set tells us little about why academics favour the books they do. (To answer that question, we might perform a text analysis of articles written by academics.) While, in order to further explore the reasons

behind the preferences of general readers, we were able to look at written reviews of books on *Goodreads*, to analyze the preferences of academic readers, we would need to undertake a different strategy, such as analyzing co-taught works on syllabi or surveying Victorianists. In other words, although our model does offer us glimpses of the specialist in Victorian literature and those works she writes about and teaches, this particular study does not offer us concrete data as to why certain works are favoured.⁴³

A fourth and final limitation of our dataset is that it cannot tell us how reading, teaching, and writing about Victorian literature have changed over time. While dates of publication are available for works catalogued in the MLA bibliography, the dates that books were taught are not available on the *Open Syllabus Project*, and, as discussed above, data on the dates that social media users read books is currently too incomplete on *Goodreads* to make meaningful conclusions.⁴⁴ While we may have a hunch about why academics cherish works like *News from Nowhere* and *Apologia Pro Vita Sua*, which fail to catch on with a general audience, we do not at this point have any concrete data that could tell us why this is so.

Conclusion

While there has been influential work on the Victorian common reader (Altick, Flint), there has been surprisingly little work on the preferences of the late twentieth and twenty-first century common reader who continues to enjoy nineteenth-century literature.⁴⁵ For the most part, the academic studies that venture explanations for the continued popularity of certain Victorian novels are on heritage film adaptations of novels by the Brontës, Dickens, and Austen, rather than on how contemporary readers consume the books themselves.⁴⁶ This study offers a data-rich analysis of reader preferences inside and outside of the academy. In an era of declining enrollments in historical English courses, it is important for those of us who teach and research these subjects to understand the way we read Victorian literature now.⁴⁷

The foremost finding of our study is that there is a strong correlation between what works of Victorian literature we teach and write about in the academy and what works are still read by a popular audience. We might find this correlation worrying,

suggesting as it does that a relatively small number of Victorian authors and books are read at all. In his work on canon formation, John Guillory argues that the "social function and institutional protocols of the school" helps us to understand how works of literature "are preserved, reproduced, and disseminated over successive generations and centuries."⁴⁸ While our data set only shows that there is a correlation between reader preferences inside and outside of the academy, and not that the academy determines reader preferences, we might take the 203 books that were widely read, taught and written about to be a contemporary canon. Looking at Romantic and World literature, David Damrosch suggests that there is a hypercanon, (those authors like William Wordsworth who have been popular since literary study was established as a discipline and by the numbers are only getting more so), a countercanon (authors like Felicia Hemans who have been brought in to diversify the white, male hypercanon), and a shadow canon (authors like William Hazlitt who were once considered "minor" authors and are increasingly fading from view).⁴⁹ The strong correlation between what we read, teach, and write about suggests that such a hypercanon, half of which is populated by seven male authors on *Goodreads*, may also define Victorian literature across three different spheres.

If our goal in researching and teaching Victorian literature and culture is to gain and impart a broad understanding of the era and its continued relevance, the hypercanon, which focuses our attention on a select few authors and texts, is certainly limiting. However, looking at those works which were outliers in terms of being read more by a general audience, which tended to be works featuring a strong female protagonist and works of children's literature, as well as "minor" works by major authors, may offer us a way of diversifying our syllabi and attracting more students. For example, we might capitalize on the continued popularity of *Jane Eyre* by offering a course that compares Brontë's novel to other countercanonical novels with strong female heroines and love plots, such as Margaret Oliphant's *Phoebe Junior* (1876) or Dinah Craik's *Olive* (1850). Our results also suggest that Victorian children's literature has an outsized popularity with a general audience, and that we might incorporate more children's literature into standard Victorian studies syllabi both in order to draw students and to enrich our understanding of the time period. There is no reason that children's literature needs to be relegated to special courses on that topic: reading *Alice's Adventures in Wonderland* alongside *Oliver Twist* or Elizabeth Barrett Browning's "The Cry of the Children" would certainly help

students gain a broader appreciation of Victorian childhood than reading only books meant for “grownups.” Like single-author dissertations, single-author courses are not seen as cutting edge, but general readers’ appetites for works like *The Professor* or “The Canterville Ghost” might lead us to believe that there would be an audience for these minor works by major authors, which could be taught alongside or instead of *Jane Eyre* or *The Picture of Dorian Gray*.

Taking the written reviews of popular novels on *Goodreads* seriously may also lead us to different teaching strategies once the syllabus is set. Contemporary reviewers on *Goodreads* have much in common with Merve Emre’s “bad readers,” that is, postwar American readers “socialized into the practices of readerly identification, emotion, action, and interaction.”⁵⁰ Readerly identification, reading for character, and reading for plot have all been dismissed as unacademic forms of reading (see Brooks, Green, and Lynch), but as our analysis of the written reviews shows, these forms of reading clearly persist among general readers.⁵¹ When we assign long novels, we might consider having students follow a minor character as a way of formalizing the investment that many readers already have in them, as Joyce Huff does when she asks students to create a digital commonplace book for one of the characters in *David Copperfield*.⁵² It might be particularly important for us focus on plot in teaching Dickens, who was the author *Goodreads* reviewers most appreciated for his ingenious storylines. We might also consider allowing space for a discussion of whether students identify with a character like Jane Eyre, contextualizing this discussion with theoretical work on the history of readerly identification (Green) and the psychology of reading (Auyoung).

Finally, our study suggests that the taste of twenty-first century readers in Victorian literature may broadly reflect taste in literature more generally, as readers (who read almost exclusively novels) are attracted to literature for young adults and children. In another study, we found that novels written by women authors in the YA and classic genre dominated the books that the members of the fifty most popular book clubs on *Goodreads* were likely to have read. The top fourteen authors—in order of popularity: JK Rowling, Suzanne Collins, Stephanie Meyer, George Orwell, Harper Lee, Stephen King, John Green, JRR Tolkien, Jane Austen, Dan Brown, F Scott Fitzgerald, Shakespeare, Neil Gaiman, and Veronica Roth—included eight women. While this is still only 57% of authors, it is a world away from university English

literature syllabi where only 27% of writers assigned are women.⁵³ Of the top nineteen novels—all seven books in the *Harry Potter* series, the three books in *The Hunger Games* series, *To Kill a Mockingbird*, *The Great Gatsby*, *The Fault in Our Stars*, *Pride and Prejudice*, *1984*, *The Hobbit*, *Animal Farm*, and *The Catcher in The Rye*—thirteen were written by women and another thirteen have a young adult or child protagonist, in large part due to the popularity of series of young adult fiction written by women. In this study of the habits of book club readers, the only popular work of literature that was not a novel was *The Diary of Anne Frank*. With its focus on a teenage girl in Nazi-occupied Holland, Frank's diary has thematic similarities with the dystopia of *The Hunger Games*, despite its very different origins. By the same token, although members of groups on *Goodreads* dedicated to reading nineteenth-century literature were less likely to have read *The Hunger Games* than the average reader, we can certainly trace thematic similarities between dystopian young adult fiction and *Tess of the D'Urbervilles* or *Wuthering Heights*. It may also be that the rise of young adult fiction is part of the same zeitgeist that draws general readers to *Treasure Island*, *Black Beauty*, or *A Little Princess*. As we move consider the place of Victorian literature in the twenty-first century, looking at the habits of general readers may lead us to reconsider the place of these popular works in our syllabi and our research.

Acknowledgments

Thanks to John Brosz, Kelly Hager, Sonia Jarmula, Christopher Keep, Paul Pival, and the editor and peer reviewers at *Cultural Analytics*.

Appendix: Regression results

Regression 1

An ordinary least squares multiple linear regression was conducted (in SPSS) of (log) *Goodreads* readers against (log) MLA subject tags with audience (adults/children), main character (female/male) and author (female/male) as additional independent variables. Books in which one of these factors was unclear (e.g., multi-gender main characters or none) were omitted. This regression assessed whether these factors systematically influenced the relationship between the number of *Goodreads* readers and the MLA citations. The residuals were close to normal, there was evidence of only minor heteroscedasticity and negligible collinearity, so the results are reasonably statistically robust, except that some of the books are related (same author) and their residuals may therefore not be independent. From the results, books attracted relatively few *Goodreads* readers for their MLA subject tags if the audience was adult or the author was female, but relatively many if the main character was female. Only the first (adult audience) achieved statistical significance ($p=0.05$), however, so it is reasonably likely that the character ($p=0.91$) and author ($p=0.89$) gender associations in the data are due to chance factors.

Table A1

Regression 1 results (1=adult, 0=children; or 1=Female, 0=male). Dependent variable: Ln(Goodreads readers +1).

Effect	Estimate	SE	95% CI		p	Collinearity statistics	
			LL	UL		Tolerance	VIF
Intercept	7.573	.327	6.932	8.214	.000		
Ln(MLA tags +1)	.763	.062	0.641	0.885	.000	.916	1.092
Audience age	-.781	.304	-1.377	-0.185	.011	.994	1.006
Main character gender	.026	.232	-0.429	0.481	.912	.756	1.323
Author gender	-.037	.269	-0.564	0.490	.890	.779	1.284

Note. Number of studies = 203. CI = confidence interval; LL = lower limit; UL = upper limit. Fit statistic: $R^2=0.478$.

Regression 2

An ordinary least squares multiple linear regression was conducted of (log) *Goodreads* readers against (log) *Open Syllabus* mentions, with audience (adults/children), main character (female/male) and author (female/male) as additional independent variables. Books in which one of these factors was unclear (e.g., multi-gender main characters or none) were omitted. The regression assessed whether these factors systematically influenced the relationship between the number of *Goodreads* readers and the *Open Syllabus* mentions. As for regression 1, the residuals were reasonably close to normal, there was evidence of very minor heteroscedasticity and negligible collinearity, so the results are reasonably statistically robust, except that some of the books are related (same author) and their residuals may therefore not be independent. None of the gender and audience variables came close to achieving statistical significance ($p > 0.2$ in all cases) and so there may well not be a general trend for any of these factors to lead to relatively many or few *Open Syllabus* mentions compared to *Goodreads* readers.

Table A2

Regression 2 results (1=adult, 0=children; or 1=female, 0=male). Dependent variable: Ln(Goodreads readers +1).

Effect	Estimate	SE	95% CI		p	Collinearity statistics	
			LL	UL		Tolerance	VIF
Intercept	7.048	.353	6.356	7.740	.000		
Ln(Syllabus citations +1)	.639	.053	0.535	0.743	.000	.974	1.026
Audience age	-.386	.307	-0.988	0.216	.210	.989	1.011
Main character gender	.280	.231	-0.173	0.733	.227	.776	1.289
Author gender	.198	.270	-0.331	0.727	.463	.787	1.270

Note. Number of studies = 203. CI = confidence interval; LL = lower limit; UL = upper limit. Fit statistic: $R^2=0.469$.

Notes

¹ See Rebecca Mead, *My Life in Middlemarch* (Toronto: Random House Canada, 2018) and Nell Stevens, *Mrs. Gaskell and Me* (London: Pan Macmillan UK, 2018).

² Simone Murray, *The Digital Literary Sphere: Reading, Writing and Selling Books in the Internet Era* (Baltimore: Johns Hopkins University Press, 2018), 118.

³ Lynne Tatlock, Matt Erlin, Stephen Pentecost, and Douglas Knox, "Crossing Over: Gendered Reading Formations at the Muncie Public Library, 1891-1902," *Journal of Cultural Analytics*. March 22, 2018. DOI: 10.31235/osf.io/k29bw

⁴ For information on the habits of *Goodreads* book club members, see Mike Thelwall and Karen Bourrier, "The Reading Background of *Goodreads* Book Club Members: A Female Fiction Canon?" *Journal of Documentation* 75, no. 5 (2019): 1139-1161.

⁵ For example, a study recently found that at McGill, English literature syllabi were 73% male authors and 27% female. See "Investigating Topic Bias and Gender Representation in Syllabi," Txlabs.org, ed. Andrew Piper. Last modified January 12, 2018. <https://txlab.org/2018/01/investigating-topic-bias-and-gender-representation-in-syllabi/>. Piper also finds that 33 authors are the topic of more than twenty percent of articles and book chapters listed in the MLA bibliography. Only four of these authors are women, and only one (W.E.B DuBois) was not white. See Piper, "Think Small: On Literary Modeling," *PMLA* 132, no. 3 (May 2017): 651.

⁶ Ted Underwood, "A Genealogy of Distant Reading," *Digital Humanities Quarterly* 11, no. 2 (June 2017): para. 13,15.

⁷ Alison Booth, "Mid-Range Reading: Not a Manifesto." *PMLA* 132, no. 3 (May 2017): 620–27.

⁸ "Nineteenth-Century Fiction," Chadwyck-Healey. Accessed March 22, 2018. <http://collections.chadwyck.com>.

⁹ We extracted this data in May 2018 using the free Webometric Analyst (lexiurl.wlv.ac.uk) software crawling *Goodreads*, with a maximum of one page per second to avoid overloading the servers. Authors in addition to Chadwyck-Healey were: Matthew Arnold, Elizabeth Barrett Browning, Robert Browning, Thomas Carlyle, Dinah Craik, Christina Rossetti and John Ruskin. We excluded multi-author anthologies from our results, for example, collections of Christmas Tales or ghost stories by several Victorian authors.

¹⁰ In this particular dataset the gender of the authors was relatively straightforward. The gender of the protagonist was more ambiguous, but did not turn out to have a statistically significant effect overall.

¹¹ *Open Syllabus Project*. The American Assembly at Columbia University, accessed June 3, 2018, <http://opensyllabusproject.org/>.

¹² These numbers were current as of June 3, 2018, when we manually looked up each individual work of literature in these two databases. In doing so we combined titles of different editions, so that *Vanity Fair* and *Vanity Fair: A Novel without a Hero*, were both counted as the same novel.

¹³ "Goodreads Traffic Card," Quantcast, accessed September 28, 2018, <https://www.quantcast.com/goodreads.com/#trafficCard>.

¹⁴ Lisa Nakamura, "'Words with Friends': Socially Networked Reading on *Goodreads*," *PMLA* 128, no. 1 (January 2013): 241.

¹⁵ See Stefan Dimitrov, Faiyaz Zamal, Andrew Piper, and Derek Ruths. “Goodreads vs Amazon: The Effect of Decoupling Book Reviewing and Book Selling,” *Proceedings of the Ninth International AAAI Conference on Web and Social Media* (2015): 602-605, Daniel Moran, “O’Connor and the Common (Online) Reader,” *Creating Flannery O’Connor: Her Critics, Her Publishers, Her Readers* (Athens: The University of Georgia Press, 2016), Mike Thelwall, “Reader and Author Gender and Genre in Goodreads,” *Journal of Librarianship and Information Science* (2017): 1-28, and Mike Thelwall and Kayvan Kousha. “Goodreads: A Social Network Site for Book Readers.” *Journal of the Association for Information Science and Technology* 68, no. 4 (2017): 972-83.

¹⁶ For a history of the lay reader and the scholarly reader, see John Guillory, “How Scholars Read,” *ADE Bulletin* 146 (Fall 2008): 8-17.

¹⁷ Mike Thelwall, “Reader and Author Gender and Genre in Goodreads,” *Journal of Librarianship and Information Science* (2017): 5.

¹⁸ Elizabeth Khuri Chandler, “Sex and Reading: A Look at Who’s Reading Whom,” *Goodreads*, last modified November 19, 2014, <https://www.goodreads.com/blog/show/475-sex-and-reading-a-look-at-who-s-reading-whom>.

¹⁹ Quantcast, “Goodreads Traffic Card.”

²⁰ Thelwall, “Reader and Author Gender,” 5.

²¹ See Thelwall and Bourrier, “The Reading Background of *Goodreads* Book Club Members: A Female Fiction Canon?” *Insurgent* (2012) is the second book in Veronica Roth’s enormously popular young adult science fiction trilogy, *Divergent*. Rainbow Rowell is the author of the popular young adult novel, *Eleanor and Park* (2013).

²² We used a log-transformation first because the data was highly skewed.

²³ Sandra Beckett, *Crossover Fictions: Global and Historical Perspectives* (New York: Routledge, 2008), 18.

²⁴ Beckett, *Crossover Fictions*, 18.

²⁵ *Goodreads* uses a proprietary algorithm to determine the top written reviews of each work of literature; it seems to be based on the number likes each review receives, with the most liked reviews appearing at the top of the page.

²⁶ The answer: in 300 reviews, *Jane Eyre* was mentioned 390 times, and *Villette* 97 times.

²⁷ Thirty-four of sixty-two reviews in English mentioned the notion that *The Professor*, Brontë’s first novel and one that remained unpublished in her lifetime, is a minor work, with many noting that it seemed to serve as a form of practice for Charlotte Brontë’s greater novels, *Jane Eyre* and *Villette*.

²⁸ Readers of *A Tale of Two Cities* were more likely to comment on the novel’s engagement with the French Revolution as a form of social criticism of that time (18/80 mentions), rather than our own time.

²⁹ In a corpus of 3,600 reviews (the top three hundred reviews for *Black Beauty*, “The Canterville Ghost,” *A Christmas Carol*, *Jane Eyre*, *A Little Princess*, *The Lost World*, *The Professor*, *A Tale of Two Cities*, *Tess of the D’Urbervilles*, *Treasure Island*, *The Secret Garden*, and *Wuthering Heights*) the most frequent recurring words, excluding common stop words in English, were: “story” (854), “book” (753), and “character” (734).

³⁰ Deidre Shauna Lynch, *The Economy of Character: Novels, Market Culture, and the Business of Inner Meaning* (Chicago: University of Chicago Press), 1.

³¹ Following advice from the University of Calgary ethics committee, although *Goodreads* is a public site and users post their reviews for public consumption, we have anonymized quotations from reviews.

³² E.M. Forster, *Aspects of the Novel* (New York: Harcourt, 1927), 103-4, 106. We would note that no reviewer mentioned Forster directly, and it is likely that the terms “flat” and “round” have entered vernacular literary

criticism without being tied to the critic directly. Nonetheless, the sheer frequency with which general readers describe Dickens's characters as flat or round in their reviews seems a testament to Forster's continuing influence.

³³ For a recuperation of this response to fiction, which has often been dismissed as naïve, see Elaine Auyoung, *When Fiction Feels Real: Representation and the Reading Mind* (New York: Oxford, 2018).

³⁴ Across 300 reviews, the characters that received the most mention were "Carton" (128), "Lucie" (116), "DeFarge" (105, could refer to Citizen or Madame) "Manette" (79, could refer to Lucy or her father), "Madame DeFarge" (59), "Pross" (33) and "Cruncher" (20).

³⁵ Of all twelve works that we scraped 300 reviews for, the word "plot" showed up by far the most often in reviews of *A Tale of Two Cities* (103 mentions, followed by 61 in *Tess of the D'Urbervilles* and 59 in *The Professor* and 53 in *Jane Eyre*, with only 18 mentions in *A Christmas Carol*. "The Canterville Ghost" had the fewest mentions of "plot" at 11 across 300 reviews.)

³⁶ Of the top 100 reviews, 80 were written in English, with an additional six in Arabic, four in Persian, three in Turkish two each in Portuguese, Spanish and Greek, and one in Vietnamese. (Two of the reviews were in both Persian and English and Arabic and English.) It is tempting to see the reviews in Arabic, Persian and Turkish as evidence of a renewed interest in Dickens's tale of democracy and revolution in light of the Arab Spring; indeed, BBC Radio 4 recently adapted Dickens's novel into a three-episode show entitled *A Tale of Two Cities: Aleppo and London*. Further research into *Goodreads* reviews of Dickens written in Arabic and Persian, conducted by those with expertise in those languages, could yield new insight into the kinds of political conditions that make Dickens relevant in the twenty-first century to readers beyond English speaking countries.

³⁷ *Goodreads* allows users to invent their own tags, and similar tags for *A Tale of Two Cities* included "high-school" (287), "read-for-school" (285), "school-books" (232), "for-school" (232), "read-in-school" (98).

³⁸ Following research that demonstrates a dip in the proportion of fiction written by women and representing female characters from 1850 to 1950, we might speculate that part of the continued popularity of works like *Jane Eyre* is that female characters were allocated less space as the nineteenth century wore on. See Ted Underwood, David Bamman, and Sabrina Lee, "The Transformation of Gender in English-Language Fiction," *Journal of Cultural Analytics*, Feb. 13, 2018. DOI: 10.31235/osf.io/fr9bk

³⁹ The experience of reading *The Professor* was also marred by xenophobia for many readers: another quarter of reviewers (17/62) mentioned that Brontë's belief in English racial superiority and Anglican religious superiority detracted from the novel for them.

⁴⁰ *The Professor* has the fewest mentions of words related to adaptation of all (2); one reader found the depiction of a respectful romance more satisfying than those of contemporary books or films, and another hoped to see the novel adapted for film.

⁴¹ Space prevents us from exploring the issue of how adaptation for film and television influences readership here, but as more data continues to accrue on *Goodreads*, it would be interesting to see the extent to which adaptations provoke a spike in readership.

⁴² Leah Price, *The Anthology and the Rise of the Novel* (Cambridge: Cambridge University Press, 2000), 2.

⁴³ The works in our study which were statistical outliers in terms of being taught more often than we would predict given how often they are studied (*Open Syllabus* versus MLA) were: 1) Christina Rossetti, *Complete Poems*, 2) Edward Lear, "The Owl and the Pussycat," 3) Frederick Marryat, *The Children of the New Forest*, 4) Anthony Trollope, *Can You Forgive Her?*, 5) Robert Louis Stevenson, *The Black Arrow*, 6) George and Weedon Grossmith, *The Diary of a Nobody*, 7) H. G. Wells, "The Food of the Gods" and 8) Thomas Hardy, *Wessex Tales*. Those works that are studied more often than we would predict given how often they are taught (MLA versus *Open Syllabus*): 1) Charlotte Brontë, *The Professor*, 2) George MacDonald, *Phantastes* 3) Charles Dickens, *The Signalman*.

⁴⁴ We would also note that the MLA bibliography covers a much longer time span, from 1926 to 2018 at the time of writing, than *Goodreads* or *Open Syllabus* which focus on the last ten to fifteen years. Due to the slowness of academic publishing, we included the fullest possible dataset from the MLA bibliography, but acknowledge that this does not take into account how we what have written about in peer-reviewed venues has changed from the twentieth to the twenty-first centuries.

⁴⁵ See Richard Altick, *The English Common Reader: A Social History of the Mass Reading Public, 1800-1900*, second edition. (Columbus, OH: Ohio State University Press, 1998 [1957]) and Kate Flint, *The Woman Reader: 1837 to 1914* (New York: Oxford University Press, 1995).

⁴⁶ See John Glavin, ed, *Dickens on Screen* (New York: Cambridge University Press, 2003) and Diane Sadoff, *Victorian Vogue: British Novels on Screen* (Minneapolis: University of Minnesota Press, 2009).

⁴⁷ “A Changing Major: The Report of the 2016–17 ADE Ad Hoc Committee on the English Major” New York: The Association of Departments of English, 2018. <https://www.ade.mla.org/content/download/98513/2276619/A-Changing-Major.pdf>

⁴⁸ John Guillory, *Cultural Capital: The Problem of Literary Canon Formation*. (Chicago: University of Chicago Press, 1993), vii.

⁴⁹ David Damrosch, “World Literature in a Postcanonical, Hypercanonical Age,” in Haun Saussy, ed., *Comparative Literature in an Age of Globalization* (Baltimore: Johns Hopkins University Press., 2006), 45.

⁵⁰ Merve Emre, *Paraliterary: The Making of Bad Readers in Postwar America*. (Chicago, University of Chicago Press, 2017): 3.

⁵¹ See Peter Brooks, *Reading for the Plot: Design and Intention in Narrative*. (New York: Alfred A. Knopf, 1984), Laura Green, *Literary Identification: from Charlotte Brontë to Tsitsi Dangarembga*. (Columbus: Ohio State University Press, 2012), and Deidre Shauna Lynch, *Loving Literature: A Cultural History* (Chicago: University of Chicago Press, 2015).

⁵² Joyce Huff. “Dickens Remediated: The Victorian Character Commonplace Project and the Teaching of Victorian Literature.” (Conference Presentation, North American Victorian Studies Association, Columbus, Ohio, October 18, 2019).

⁵³ Txlalab.org. “Investigating Topic Bias and Gender Representation in Syllabi.”