

The institutional repertoire in the hands of ordinary readers: a quantitative analysis of the impact of works from the French and Portuguese school canons on Goodreads

Lucía Hernández-Heras^{1,*}, Sara Duarte Reis Da Silva², Rosa Tabernero-Sala¹

¹Department of Specific Didactics, University of Zaragoza, Huesca 22003, Spain

²Instituto de Educação, Departamento de Ciências Sociais da Educação, Centro de Investigação em Estudos da Criança (CIEC), University of Minho, Braga 4710-057, Portugal

*Corresponding author. Department of Specific Didactics, University of Zaragoza, Zaragoza 50009, Spain. E-mail: lhernandezheras@unizar.es

Abstract

A growing body of research demonstrates how social networks and other digital platforms are redefining the concept of the literary canon, leading to a disconnect between institutional legitimization and public reception. Framed within systemic theories, this quantitative, non-experimental study seeks to measure the impact of institutional repertoires on ordinary readers by analysing social media platforms for readers, which serve as valuable spaces for examining society's reading preferences. To this end, the school canons of France and Portugal were selected as institutionally legitimized corpora, as they uniquely prescribe reading lists, unlike those of other countries. Goodreads was chosen as the primary environment for this investigation. A database was subsequently created, incorporating variables related to the works themselves—such as literary genre and internationality—as well as variables related to their impact on Goodreads. Descriptive and inferential statistical analyses were applied to this dataset. The results reveal that the scholarly articles in both countries are well received by ordinary readers and that readers combine traditional labelling methods associated with criticism with other more personalized methods.

Keywords: social media; network analysis; Goodreads; school canon; readers.

1. Introduction

The canon is a fundamental concept in Western culture, historically establishing repertoires of exemplary or imitable artists and values in art, education, science, sociology, and culture. However, in recent decades, the authority of the canon has diminished compared to the prestige it enjoyed during the 19th century, “when the ideas of authority, authenticity, genre value and nationalism began to come together closely” (Eagleton 1999: 61). In the 1970s, a series of cultural and political movements—such as deconstructionist theories, feminist critique, and Cultural Studies—began to challenge the selection and exclusion methods that had dominated literary theory up to that point, as they perpetuated certain values related to race, gender, sexual orientation, and ideology (Benton 2000). Criticized for its alleged ties to power structures (Eagleton 1999), the canon became a point of contention between its detractors and

defenders. The detractors, led by the New Criticism movement, condemned the selection process, viewing it as a mechanism serving the status quo. In contrast, the defenders advocated for an aesthetic view of literature, as exemplified in the most emblematic work of this school of thought: *The Western Canon. The Books and School of the Ages* by Harold Bloom (1995).

Systemic theories emerged as a conciliatory alternative, revisiting ideas about literature as a social and historical phenomenon that had been advanced by the positivist tradition of literary history, most notably represented by Gustave Lanson. Consequently, the systemic approaches of the late 20th century can be regarded as a reformulation of this holistic perspective rather than its inauguration. In any event, its conception of culture as a global system represented a significant turning point in the ongoing discourse surrounding the canon.

In accordance with this framework, the truly universal element lies not in aesthetic criteria, but in the

tensions and dynamics arising from this interaction. This system is fundamentally structured around a centre, which holds the works that have been canonized, and a periphery, which includes works aspiring to attain such status. Accordingly, the texts in the centre, having gained legitimacy, reproduce structures and models that serve as inclusion and exclusion criteria, thereby shaping the interpretive frameworks of communities (Lotman 1996). In this context, Even-Zohar (1997) introduced the concept of repertoires—sets of rules and principles that encompass readers' tastes and knowledge, cultural codes, and prerequisite systems.

Repertoires are shaped through the dynamic negotiation between various agents and resources (Even-Zohar 1997). Among these, institutional agents have traditionally wielded the most power in preserving a repertoire across generations, as their influence on culture is mediated by entities such as the Church and schools. Since every canon functions as a *paideia*, institutions control and dictate what occupies the centre through the curriculum and the “school canon” (Benton 2000). Education, therefore, acts as a mechanism for the social reproduction of exclusive cultural capital (Fokkema 1993), using it to mythologize the knowledge that becomes an unquestionably prestigious part of the repertoire (Bourdieu 1998). As a result, those who lack this knowledge or competency are excluded from the core of the system; in fact, they can only gain access to what is popular, which stands in opposition to what is considered high culture (Lotman 1996).

Another fundamental agent in repertoire formation is the producer or, in the case of the canon, the authors themselves, who introduce a series of characteristics that may result in a new model (Even-Zohar 1997). In this context, the market also plays a crucial role, mediating between the producer's attempt to create a product and the likelihood of a favoured reception, which depends on the availability and dissemination of the texts. Whether in the power of institutions, authors, or the market, the final components in establishing the system are metatexts—discursive practices that function as theoretical treatises and critical articles for organizing the literary system (Lotman 1996). This self-organizing mechanism determines both the inclusion and exclusion of texts, as well as their hierarchical classification.

While the primary canonizing agents up to the 20th century were institutions, the market, authors, and metatexts, the emergence of the aesthetics of reception and the fight for civil rights led by historically marginalized groups significantly expanded the role of receivers. Jauss and Iser's ideas led to the legitimization of the reader's role in the interpretive processes of works, which had previously been dominated by the author and the text alone. As a result, the “table of values” was fractured (Bourdieu 1998), leading to the

decentralization of the hegemonic canon. In its place, multiple “vocational canons” emerged (Mignolo 1991), shaped around the values and identities of their members.

This questioning of the established canon initiated by the social movements previously referenced has reached a new stage with the advent of the internet, as several empirical studies have shown. Unlike other groups, social media users do not feel excluded by traditional gatekeepers—they outright disregard their authority (Salgado 2022). This contrast between what institutions endorse and what actually unfolds within the cultural system has caused a certain disconnect between elites and the prevailing dynamics of society, as several authors have noted (Torre-Espinosa 2021; Salgado 2022).

Social media platforms for readers—such as LibraryThing, Lectoralia, and the most used and globally recognized Goodreads—are among the key players shaping this new media ecology. These alternatives offer all the standard features of social media—an inbox, notifications, status updates, a personal profile—but are all organized around one topic: reading. Users can review and rate the works they have read, create book lists around topics of their choice, participate in reading groups, catalogue titles using different tags, promote events, or develop their own digital libraries. Reading becomes a performative and social act, breaking away from the individuality that characterized traditional literacy models (García-Roca 2016). The experience of individual reading generates a public narrative that connects readers with one another (Driscoll and Rehberg 2019) in an environment driven by collaboration and reciprocity. By assuming roles as senders, receivers, experts, and opinion leaders, users undermine hegemonic canonizing frameworks and open the door to a repertoire that does not necessarily follow the traditional model (García-Roca 2016).

For all this, the internet's impact on the canon extends beyond mere selection, it disrupts traditional definitions in all senses: interpretation, selection systems, language, and format. By appropriating works and challenging canonical concepts of authorship and criticism, internet users invert the central paradigm. Rather than adhering to a vertical hierarchy—author-critic-receiver—user-generated content operates through a horizontal, multidirectional structure that contributes to the democratization of literary expressions (Achugar 2019).

Thus, several authors argue that this participatory culture (Jenkins, Ito, and Boyd 2015) has prompted a form of lay criticism in which internet users counterbalance professional criticism published in specialized magazines and journals (Salgado 2022). In this new cultural system, everyone becomes “amateur critics”

(Walsh and Antoniak 2021), taking on the role traditionally assigned to metatexts, although following their own criteria. This is particularly evident through the labelling process, known as folksonomy, which reflects the dynamics of collaborative creation, redefining hierarchical genre categories through more personalized observations (Walsh and Antoniak 2021).

While works are primarily evaluated based on reader satisfaction or dissatisfaction, their social prestige is measured through likes and other quantitative indicators: the greater the number, the better the work is perceived to be, reflecting a fusion of symbolic and economic capital (Salgado 2022). As a result, the hegemonic canon loses authority in favour of other popular and commercial canons, whose function is no longer to preserve works or authors, but rather to reduce them to objects of consumption (Lluch 2023). This explains the conclusions of several studies, which show a correlation between the titles most mentioned in internet conversations and the bestsellers (Francés-Díez 2021).

However, some authors, such as Sánchez-García (2020), debate the impact that these dynamics will have on the hegemonic canon, pointing out the differences between synchronic and diachronic canons. Consequently, what is currently considered popular on social media reading networks (synchronic canon) may or may not form part of the diachronic canon—the body of works that endure over time and are recognized as valuable by critics and readers. Furthermore, regarding the effects of Goodreads on the system, we must also consider what Pianzola (2025) points out: the phenomenon of social reading and the emergence of reader opinions have persisted throughout history. However, a distinguishing aspect of the contemporary era is the accessibility of reviews and raw data, which enables the discernment of readers' responses or consumption patterns. This affords them 'ecological validity' (Kuijpers 2023: 114), as they are not influenced by intermediaries.

In any event, the authors agree that the network is unleashing a social appropriation of knowledge, which is why the education is grappling with new challenges influenced by students' digital profiles. This shift has prompted a new conceptualization of the school canon, which was previously dominated by classic and national texts. Today, however, this selection has expanded to include paraliterary alternatives (Lluch 2023), such as comics, oral and popular heritage, and children's and young adult literature, as a means to increase student motivation.

As we move towards a more open model of literary education that better aligns with students' interests, few Western countries prescribe specific reading lists for primary education, although they do provide general guidelines in their national curricula.

These guidelines may be reinterpreted at the regional level—as in Belgium or Spain—or even at the school level—as in Italy or Denmark. Portugal and France are two exceptions to this trend, although their approaches differ significantly; Portugal maintains a more restricted catalogue established by the *Plano Nacional de Leitura* [National Reading Plan] (PNL2027) for reading in the field of Literary Education, while France offers an extensive repertoire, allowing teachers to select texts according to their needs. Another difference is that Portuguese lists are designed to be read in Portuguese classes. In contrast, the French lists are designed to educate students in literary culture. Although it is true that French teachers are the most frequent users of these recommendations, any other teacher of any subject can use these suggestions as a reference for developing the content of their discipline.

Although academic literature has proposed various theoretical frameworks for structuring school canons (Santoli and Elaine 2004), Lluch (2023) and Francés-Díez and Lluch (2024) caution against the lack of systematization and verifiable data in canon-related research within the humanities. In fact, Cordón-García (2018) addresses the complexity of understanding the relationship between canonical recommendations and their reception. Building on studies that explore the connection between literature and its reception on Goodreads (Sánchez-García *et al.*, 2021; Walsh and Antoniak 2021; Zheng and Fan 2023), this research seeks to 1) determine the characteristics of works included in the school canons of Portugal and France to understand how recent recommendations are structured; 2) measure the impact of these school canons on a social environment such as Goodreads by cross-tabulating variables related to the works with variables related to their reception on this social network; and 3) analyse the interaction between the variables related to the impact of the school canon on Goodreads to interpret the dynamics governing the behaviour of ordinary readers (Ceballos 2019) online. As the processes of canon legitimization depend on the reputation, visibility, and recognition conferred by both the literary system and society (Cordón-García 2018), this approach will shed light on how popular tastes engage with and respond to the legitimized repertoire.

2. Research design

This study adopts a non-experimental, quantitative methodology. By examining user interactions within a specific digital space, it also falls within the scope of e-research (Sánchez-García *et al.*, 2021). The first step consisted of creating a database of works from the primary school canons of both countries, sourced from the websites of their respective Ministries of

Education. In Portugal, primary education (Educação Básica) cover ages six to fifteen; however, only works from the first two stages—up to age twelve—were selected. This age range aligns with the end of primary education in France and other countries, such as Spain or England. Including the third stage of primary education in Portugal could have introduced bias, as it would have incorporated young adult literature. In total, the database includes 876 entries.

Each work was classified by literary genre, authorship (national or international), and inclusion in either the Portuguese or French canon. Additionally, given the growing importance of the study of non-fiction books in the field of literature education (Author), the works were also categorized as fiction or non-fiction. The original publication year of each work was recorded as well as any awards or nominations it had received. This approach seeks to determine whether a relationship exists between awards—which function as canon-shaping elements (Gómez-Díaz and García-Rodríguez 2024) due to their institutional ties (Lluch 2023; Lluch 2025)—and reader reception. These factors, therefore, represent variables linked to the intrinsic characteristics of the works.

To assess the impact of the works on the social network, several variables were analysed for each title. These included the average user rating, the number of ratings, the number of reviews, the number of users who marked the work as “want to read,” the number of users who marked it as “currently reading,” and the number of users who added the work to their bookshelves. In addition to these quantitative variables, the study incorporated the tags (genres) assigned by readers to the works, which are displayed on their main menu. It also examined whether the titles appeared in any user-generated lists, providing insights into the metatextual activity carried out by users. Notably, Goodreads consolidates data from multiple editions and translations of the same work into a single entry.

After organizing the database in Excel, it was coded and analysed using SPSS. This software facilitated the extraction of descriptive statistics and the application of inferential procedures. Since the central limit theorem establishes that large sample sizes (>40) tend to approximate a normal distribution, parametric tests were applied (Ghasemi and Zahediasl 2012). Before selecting the appropriate statistical tests, the homogeneity of variance was assessed using Levene’s test. For comparisons involving dichotomous nominal variables and quantitative variables, an independent samples t-test was conducted. In cases involving polytomous nominal variables alongside quantitative variables, a one-way ANOVA, or its Brown-Forsythe variant, was employed. Relationships between quantitative variables were examined using Pearson’s correlation

coefficient. In addition, a regression model was employed to determine which variables predict the number of reviews, used as an indicator to measure active reader participation, since writing a review requires greater effort and commitment than simply marking a work as read (Alghamdi and Ihshaish 2021). Following the guidelines proposed by López-Martin and Ardura-Martínez (2023) the effect size for each test was determined. Qualitative data related to the tags associated with the works were processed using the tidytext package in R. All code has been made publicly available in a data repository 2.

3. Results

3.1 Characteristics of both corpora

The French corpus contains 806 works. As shown in Table 1, the distribution between works by French authors and those by international authors is relatively balanced. In terms of author diversity, 23% ($n = 182$) of the entries correspond to authors who appear more than once. Table 2 shows that the most frequently represented authors are the Brothers Grimm, whose works appear six times. They are followed by Hans Christian Andersen and Charles Perrault, each with five works. This clearly indicates that the most recurring authors fall into the category of popular storytelling. Additionally, there are twelve authors represented in the corpus with three works each, namely Kitty Crowther, Roald Dahl, and Emily Gravett. Finally, sixty-four authors appear twice.

Regarding the genres of the works, Table 3 reveals a marked predominance of picture books, followed by narrative, poetry, comics, theatre, and, finally, wordless picture books. Additionally, while non-fiction works do not play a leading role, they still account for less than 10% of the total. As for the original publication dates, the oldest work in the corpus is Aesop’s *Fables*, dating back to the 1st century BC, while the most recent is from the 21st century. In fact, as illustrated in Table 1, the majority of the works (60.4%) are from the 21st century, followed by the 20th century (34.6%), with works from the 19th century also represented. The 16th and 15th centuries have only a marginal presence. This distribution suggests a clear preference for more contemporary works. Finally, 11.2% of the works in the corpus have either received an award or been nominated for one.

The Portuguese canon consists of seventy entries. As shown in Table 3, unlike the French canon, the Portuguese canon exhibits a strong preference for national works, which account for 80% of its selection. Additionally, this canon displays less creative diversity compared to the French corpus, with 57% ($n = 40$) of the works authored by writers who appear more

Table 1. Statistics for the school canon in France.

Variable	Descriptors	<i>n</i>	Frequency (%)
Origin of the author	National	377	46.8
	International	429	53.2
Literary genre	Picture book	403	50
	Narrative	240	29.8
	Wordless picture book	18	2.2
	Poetry	58	7.2
	Theatre or dramatic text	32	4
	Comic	55	6.8
Fictionality	Fiction book	752	93.3
	Non-fiction book	54	6.7
Century	1st century BC	1	0.10
	15th	1	0.10
	16th	4	0.5
	19th	22	2.7
	20th	279	34.6
	21st	487	60.4
Awards or nominations	Yes	90	11.2
	No	716	88.8

Table 2. Most frequent authors in the French canon.

Author	<i>n</i>
Jacob y Wilhelm Grimm	7
Hans Christian Andersen	5
Charles Perrault	5
Tomi Ungerer	3
Nadine Brun-Cosme	3
Philippe Corentin	3
Kitty Crowther	3
Roald Dahl	3
Emily Gravett	3
Alain Serres	3
Anne Herbauts	3
Pef	3
Claude Ponti	3
Jacques Prévert	3
Claude Roy	3

than once. The most frequently referenced creator is António Torrado, with five different works, followed by Alice Vieira, Luísa Ducla Soares, and Matilde Rosa Araújo, each with four entries. Three authors are represented by three titles each, while seven have two works included in the repertoire. An analysis of the most frequently referenced creators (Table 4) reveals that the most determining factor is not the genre practiced by the authors, but rather their Portuguese nationality, a characteristic that aligns with the overall structure of the corpus, as previously noted.

The most represented genre is narrative, accounting for 53.5% of the corpus, followed by poetry, picture books, and theatre. Notably, wordless picture books and comics are absent, suggesting that the Portuguese canon allocates less space to aesthetic formats,

although it does include several picture books. As for publication dates, the oldest work again dates back to 564 BC, since the Portuguese canon also includes Aesop's *Fables*. Similar to the French canon, the contemporary period is the most represented; however, here the 20th century dominates, followed by the 19th, 18th, and 17th centuries. Of all the works included in this canon, 33.8% have either won awards or received nominations. Thus, by prioritizing national works, canonical genres, and award-winning titles, the Portuguese canon seems to follow a more traditional approach to canonization compared to the French canon.

3.2 Rankings of the most highly rated and most reviewed works

Starting with the general trends of the French school canon (Table 5), it is evident that this corpus is well-received among ordinary readers, with an average rating of 3.8. Among the platform's indicators, the number of reviews has the lowest figures, likely because it requires more effort from users compared to other metrics. In contrast, "added to shelves" is the variable with the highest average. Users' bookshelves play a key role in constructing and displaying reader identity.

The Portuguese canon exhibits lower maximum values across all variables, despite having a higher average. This, along with a lower standard deviation across all indicators, likely results from the greater homogeneity of the Portuguese canon compared to the French one, as well as its more limited selection. While the French canon features a broader range of contemporary works, including both lesser-known and classic titles, the Portuguese canon leans towards more traditional works, particularly from its own national tradition, complemented by some international classics.

Table 3. Statistics for the Portuguese school canon.

Variable	Descriptors	<i>n</i>	Frequency (%)
Origin of the author	National	56	78.9
	International	14	19.7
Literary genre	Picture book	12	16.9
	Narrative	38	53.5
	Wordless picture book	0	0
	Poetry	17	23.9
	Theatre or dramatic text	3	4.2
	Comic	0	0
Fictionality	Fiction book	100	100
	Non-fiction book	0	0
Century	1st century BC	1	2
	17th	2	3
	18th	1	1
	19th	8	11
	20th	34	49
	21st	24	34
Awards or nominations	Yes	24	33.8
	No	46	64.8

Table 4. Most frequently mentioned authors in the Portuguese canon.

Author	<i>n</i>
Matilde Rosa Araújo	4
Luísa Ducla Soares	4
António Torrado	4
Alice Vieira	4
Manuel António Pina	3
Luísa Dacosta	3
Alves Redol	3
Sophia de M. B. Andresen	2
Sidónio Muralha	2
Oscar Wilde	2
Mia Couto	2
Maria Alberta Menéres	2
José Eduardo Agualusa	2
Álvaro Magalhães	2

Table 5. Impact of works from the French canon according to Goodreads indicators.

	Maximum	Average	Dev.
Rating	4.74	3.8182	0.54193
No. of ratings	2211528	10912.25	101570.2
Reviews	73134	361.47	3002.333
Want to read	826022	4215.51	38347.29
Currently reading	76908	497.59	4882.635
Added to shelves	3554754	17631.36	163574.2

However, [Table 6](#) challenges the perceived popularity of the Portuguese corpus.

The next step involves identifying the highest-rated works from each corpus on Goodreads. To ensure representativeness, only entries with over 100 ratings were considered, as some works had received very few

Table 6. Impact of the works from the Portuguese canon according to Goodreads indicators.

	Maximum	Average	Dev.
Rating	4.37	3.9201	0.39401
No. of ratings	314917	15891.84	56521.02
Reviews	20312	582.11	2576.766
Want to read	193696	8873.36	32588.02
Currently reading	44200	1033.97	5519.853
Added to shelves	581649	30311.9	109040.4

ratings—in some cases, just one. Including such works could distort the results and lead to unrepresentative conclusions.

The highest-rated work in the French canon is *Le livre de l'été* by Susanne Berner Rotraut ([Table 7](#)). Overall, the ranking of the top ten highest-rated works stands out for its international scope and a users' clear preference for aesthetic formats, including titles such as *Moi, si j'étais grand* or *Un livre*. The list also features non-fiction texts like *Zoo logique*, as well as works with material and interactive elements, such as *La maison hantée* or *Un livre*, which invite readers to participate and construct meaning through physical interaction with the book as an object. Thus, it is clear that, among the works in the French canon, ordinary readers particularly appreciate the hybridization and fragmentation of new discourses. This trend is further reflected in the absence of non-fiction readings in the Portuguese canon and their limited presence in the French one. Although wordless picture books are not strongly represented in the French selection, two works from the top ten belong to this genre, with *Le livre de l'été* occupying the top position. In line with the overall structure of the French canon, the titles in

Table 7. Highest-rated works of the French school canon.

Title in French	Author	Average rating	No. of ratings
<i>Le livre de l'été</i>	Susanne Berner Rotraut	4.74	248
<i>Moi, si j'étais grand</i>	Eva Janikovsky	4.59	264
<i>Zoo logique</i>	Joëlle Jolivet	4.51	156
<i>Gruffalo</i>	Julia Donaldson	4.46	43,605
<i>Le roi de la bibliothèque</i>	Michelle Knudsen	4.43	11,023
<i>Tu sais siffler, Johanna?</i>	Ulf Stark	4.43	588
<i>La maison hantée</i>	Jan Pienkowski	4.43	812
<i>Un livre</i>	Hervé Tullet	4.42	24,560
<i>L'ours et le chat sauvage</i>	Kasumi Yumoto	4.42	209
<i>La course au gâteau</i>	The Tjong-Khing	4.35	275

Table 8. Highest-rated works of the Portuguese school canon.

Title in Portuguese	Author	Average rating	No. of ratings
<i>O Beijo da Palavrinha</i>	Mia Couto	4.37	156
<i>A Ovelhinha Preta</i>	Elizabeth Shaw	4.32	109
<i>Ou isto ou aquilo</i>	Cecília Meireles	4.28	384
<i>Contos de Grimm</i>	Jacob and Wilhelm Grimm	4.24	214,001
<i>A História do Pedrito Coelho</i>	Beatrix Potter	4.22	260,899
<i>O Príncipe Feliz</i>	Oscar Wilde	4.2	35,389
<i>O Gato e o Escuro</i>	Mia Couto	4.15	181
<i>Contos de Perrault</i>	Charles Perrault	4.1	14,725
<i>A Maior Flor do Mundo</i>	José Saramago	4.09	1,122
<i>Contos de Andersen</i>	Hans Christian Andersen	4.08	12,762

this ranking were published in the 20th or 21st centuries.

The top ten in the Portuguese canon (Table 8) is led by Mia Couto with her work *O Beijo da Palavrinha*. Reflecting the structure of the corpus, this ranking shows greater homogeneity, with narrative works predominating, although other genres such as poetry are also present. The dominance of more traditional genres over others, like picture books, is expected given the composition of the Portuguese school canon, which includes older works compared to those in the French canon's top ten. While ordinary readers of the French canon valued new formats above all else, ordinary readers of the French canon particularly appreciate those works that have been incorporated into the canon of children's and young adult literature.

The classification of most reviewed works within the French corpus (Table 9) shows a strong presence of international authors. However, despite this trend, the most reviewed work in the French canon remains a classic of French literature: *Le petit prince*. The ranking also reaffirms the success of picture books—with titles such as *Max et les Maximonstres* and *L'arbre généreux*—as well as narrative works. A particularly notable finding is that the works most likely to engage users are those regarded as classics of children's and young adult literature. These titles also enjoy strong

reader approval, as reflected in their consistently positive average ratings, which exceed the overall average for the French corpus.

As in the French canon, the most reviewed works in the Portuguese corpus—those that prompt readers to share their opinions with the broader reading community—are classics (Table 10), such as *Robinson Crusoe*, *Contos de Grimm*, and *A História do Pedrito Coelho*. However, unlike the French corpus, where the most-reviewed titles are primarily from the 20th century, the Portuguese list includes works from the 20th century, the 19th century, the 17th century, and even as far back as the 1st century BC. Although the Portuguese school canon consists mainly of national titles, the most reviewed works are largely of international origin. Interestingly, four of the titles in this classification also appear among the highest-rated works, suggesting a correlation between these two metrics. Although nearly all titles in this list have received positive average ratings, the two exceptions still rank among the top three most-reviewed, with average ratings close to four.

3.3 Lists and tags

A total of 7.2% of the works in the French canon appear in seventy-seven different lists, with many titles recurring across multiple lists. As shown in Table 11,

reflect canon formation criteria traditionally shaped by structuralist and historical approaches. Similarly, lists such as “Best of Maurice Sendak” and “Best of Nathaniel Hawthorne” demonstrate the legitimization of an author’s brand. Some lists advocate for causes championed by specific groups, particularly since the sudden emergence of cultural studies and certain social movements. Examples include titles like “20 Children’s Books with Strong Female Characters,” “Children’s Books for the Resistance,” “Picture Books with Strong Heroines,” and “Books on the Israel-Palestine Conflict.”

There is a notable presence of lists that focus on recurring themes in children’s and young adult literature (“Pre-Potter,” “Magic Schools,” and “Fairy Tales in All Their Ways”), as well as those that reflect the moralizing intent behind the selection of such works by the hidden adult (“Build a Child” and “Best Educational Literature”). Finally, lists such as “Best Chapter Books to Read Out Loud” and “Best Read Alouds for Young Children,” draw attention to the importance of reading aloud and how ordinary readers experience the act of reading.

Table 12 displays the twenty-two lists that include 34.4% of the works from the Portuguese school canon. The characteristics of the lists in which users have included these canonical titles show a parallel between traditional methods of classifying classics in literary theory and criticism and the approaches used by Goodreads users. Specifically, there is a recurring use

of quality-based categories (“Best Books Ever”), generic or formal classifications (“Livros com rimas e Aliterações”), and chronological groupings (“Best Books of the 18th Century”). Similar to the French canon, lists that emphasize reading as a transformative experience—such as “Books You Will Keep Forever” and “Books That Everyone Should Read at Least Once”—are of particular interest. In this regard, all the lists, whether focused on French or Portuguese works, associate positive connotations with the titles of the school canon. However, in contrast to the French classification, there is a striking number of lists not only compiled in Portuguese but also explicitly referencing Portuguese literature (“Livros Portugueses”, “Portuguese Lyrical Writings”).

Figure 1 presents the thirty most frequently used tags assigned to the works in the French corpus. While the total number of tags assigned to these works amounts to 2,123, only 181 are unique, indicating a degree of homogeneity due to the marked prevalence of certain tags. Among these, “Children’s” appears nearly 300 times, followed by “Fiction” and “Picture Books.” This demonstrates the connection between the composition of the French corpus and these statistical trends. As with user-generated lists, the tags reflect terminology commonly used in literary criticism and theory, incorporating both geographical (“French Literature,” “France”) and generic descriptors (“Picture Books,” “Short Stories,” “Poetry”). Perhaps the tag most closely linked to this literary tradition is “Classics,” which appears nearly 100 times, in contrast to the concept of canon. Additionally, many tags reference themes traditionally associated with children’s and young adult literature, including “Humour,” “Animals,” “Fantasy,” or “Family.” Finally, there is a noticeable abundance of tags related to the school canon, such as “Middle Grade” or “Bande Dessinée.”

The thirty genres most commonly associated with the works in the Portuguese canon (Fig. 2) reveal notable similarities to the French corpus. Once again, the two most common tags are “Children’s” and “Fiction.” However, the third most frequently used genre points to a dominant trend in this classification: the reliance on geographical descriptors, which strengthen the connection between national culture and literature. Historicist tendencies are also evident in chronological descriptors, as well as in the canonizing efforts reflected in tags like “Classic Literature.” Mirroring the French corpus, there is an abundance of tags referring to the target audience of children’s and young adult literature (“Children’s,” “Young Adult”), recurring themes or characters (“Fantasy,” “Animals”), and traditional genres (“Fairy Tales”). This classification also reveals the notable presence of educational institutions (“Middle Grade,” “Read for School”). Due to the inherent structure of

Table 12. Lists from the Portuguese school canon.

Author	n
Livros portuguesas-O que já li ... Já li! - Portugal	7
Livros com rimas e Aliterações	3
Favorite books of my childhood	3
Books set in Portugal	2
Leituras de 2015-Portugal	2
Leituras de 2014-Portugal	2
Leituras de 2011-Portugal	2
Best books ever	2
Best Books of the 17th Century	1
Estrela do mar	1
Portuguese Lyrical Writings	1
Livros portugueses	1
Leituras de 2013-Portugal	1
Best Literature in Portuguese	1
South, Central and Latin American Women Authors	1
Leituras de 2012-Portugal	1
Books you will keep forever	1
Leituras de 2021-Portugal	1
Leituras de 2018-Portugal	1
Classic French Literature	1
Leituras de 2022-Portugal	1
Best books of the 18th century	1
Books that everyone should read at least once	1

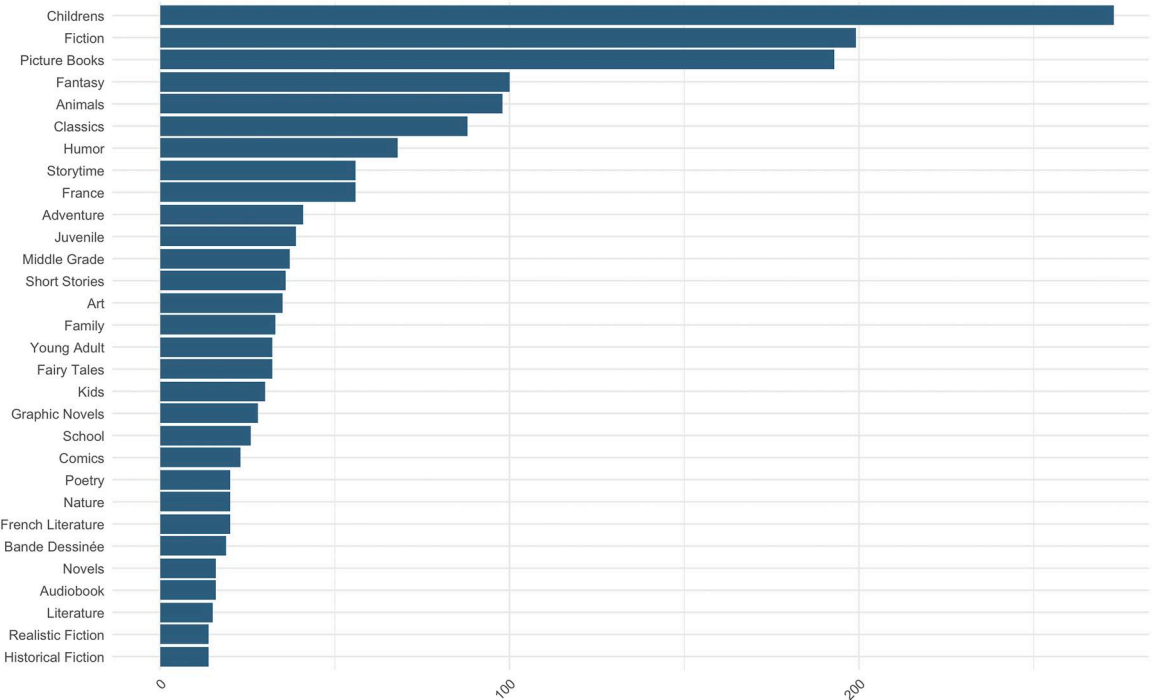


Figure 1. The thirty most frequently assigned tags to works from the French canon. Using the tidytext package, the thirty labels most commonly associated with works in the French canon are illustrated. As can be seen, the tags reflect terminology commonly used in literary criticism and theory, incorporating both geographical and generic descriptors.

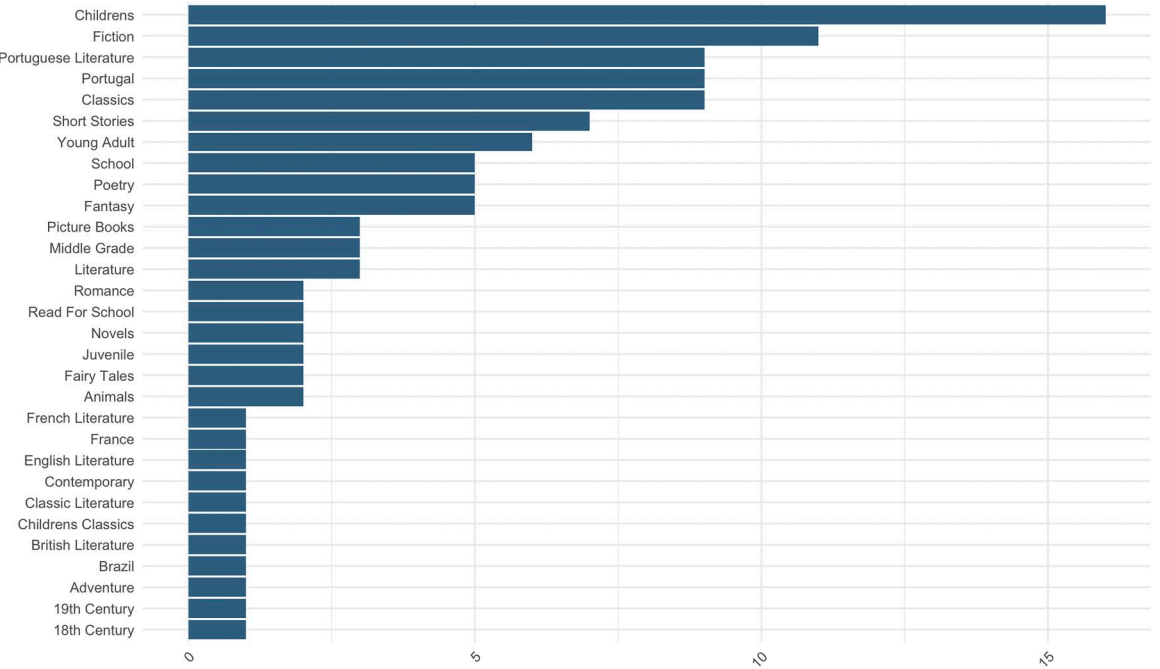


Figure 2. The thirty most frequently assigned tags to works from the Portuguese canon. Using the tidytext package, the thirty labels most commonly associated with works in the Portuguese canon are illustrated. This classification reveal notable similarities to the French corpus.

both canons, the Portuguese corpus is more closely associated with narrative genres (“Short Stories”), while the French corpus leans more towards “Picture Books.” In any event, both taxonomies exhibit certain degree of homogeneity, as well as the use of tags established by agents such as critics or academia.

3.4 Cross-tabulation of variables

After analysing whether the quantitative variables—average rating, number of ratings, number of reviews, users who marked “want to read,” “currently reading,” and “added to shelves”—varied based on factors associated with the works, the Student’s *t*-test revealed no significant differences. These factors included whether the works were national or international, the type of canon (French or Portuguese), or the fictional nature of the entries, as detailed in the data repository 1. However, the award variable (Table 13) emerged as significant. This variable was treated as nominal to assess the effects of winning an award or being nominated. Since none of the variables met the assumption of homoscedasticity, Welch’s correction was applied. The results indicated that award-winning works had higher averages in the number of ratings ($M = 37586.43$, $SD = 123332.395$; $t(107,957) = 2.25$; $P = .027$), number of reviews ($M = 1174.59$, $SD = 2687.298$; $t(123,214) = 3.017$; $P = .003$), number of users who marked “want to read” ($M = 15734.66$, $SD = 48234.971$; $t(107,319) = 2.444$; $P = .016$), the number of users who marked “currently reading” ($M = 2603.11$, $SD = 11120.058$; $t(96.2) = 2.018$; $P = .046$) and number of users who added the work to their shelves ($M = 61556.2$, $SD = 198455.799$; $t(108,183) = 2.282$; $P = .024$). However, Cohen’s *d* test indicated that the explanatory power of these relationships is small in all cases. To determine whether the variables associated with the impact of the entries on Goodreads varied based on the literary genre of the works, a one-way ANOVA was conducted. The results concluded that these differences were insignificant.

The fact that, among all the variables associated with the works only winning an award or being nominated proved significant may once again point to the influence of institutional factors in these environments.

While the cross-tabulation of variables associated with the works and those related to their impact on Goodreads yielded few significant relationships, the interaction among the variables related to impact produced notable results in most cases. The only exception was the comparison between the quantitative variables and the list variable, which had been coded as a dichotomous variable based on whether the entry had been included in any user-generated list. After categorizing the tagged variable as well (Table 14), a Student's t-test with Welch's correction was applied, as the assumption of homoscedasticity was not met. This test revealed that entries receiving any type of descriptor achieved higher averages in the number of ratings, ($M = 26563.66$, $SD = 150256.241$; $t(371.016) = 3.403$; $P = .001$), number of reviews ($M = 878.9$, $SD = 4503.188$; $t(371.284) = 3.729$; $P = .000$), number of users who marked "want to read" ($M = 10783.16$, $SD = 57663.221$; $t(371.009) = 3.602$; $P = .000$), number of users who marked "currently reading" ($M = 1249.21$, $SD = 7488.783$; $t(371.249) = 3.19$; $P = .002$) and number of users who added the work to their shelves ($M = 43782.41$, $SD = 243295.345$; $t(371.013) = 3.464$; $P = .001$). As with the award variable, the average rating did not prove significant. Additionally, the effect size of these comparisons is large in all cases, except for the "currently reading" variable.

The relationship between the quantitative variables was calculated using Pearson's correlation. As shown in Fig. 3, the average rating variable remained unaffected by interactions with other variables. In contrast, the number of reviews increased alongside the number of ratings, the number of users who indicated an intention to read the work, and the number of users who added the work to their shelves, with a substantial effect size. Additionally, although with a lower

Table 13. T-student of the award variable.

Variable	Award	M	SD	<i>t</i>	df	<i>P</i>	Cohen's effects
Average rating	Yes	4.0717	.23945	-.385	795	.7	–
	No	4.8535	19.68768				
No. of ratings	Yes	37586.43	123332.395	2.25	107.957	.027	.2985
	No	8113.95	94891.408				
No. of reviews	Yes	1174.59	2687.298	3.017	123.214	.003	.3005
	No	282.04	2989.521				
Want to read	Yes	15734.66	48234.971	2.444	107.319	.016	.3296
	No	3231.81	36277.002				
Currently reading	Yes	2603.11	11120.058	2.018	96.2	.046	.4691
	No	287.67	3425.045				
Added to shelves	Yes	61556.2	198455.799	2.282	108.183	.024	.3010
	No	13424.93	153887.119				

Table 14. T-student of the tagged variable.

Variable	Tagged	M	SD	<i>t</i>	df	<i>P</i>	Cohen's effects
Average rating	Yes	5.921	27.06572	1.544	370.365	.124	–
	No	3.7513	.64416				
No. of ratings	Yes	26563.66	150256.241	3.403	371.016	.001	.4998
	No	51.63	807.489				
No. of reviews	Yes	878.9	4503.188	3.729	371.284	.000	.6069
	No	8.1	102.242				
Want to read	Yes	10783.16	57663.221	3.602	371.009	.000	.5177
	No	14.88	234.745				
Currently reading	Yes	1249.21	7488.783	3.19	371.249	.002	.3649
	No	10.56	158.675				
Added to shelves	Yes	43782.41	243295.345	3.464	371.013	.001	.5124
	No	90.58	1186.765				

explanatory capacity, a relationship was observed between the number of reviews and the number of users who indicated they were reading the work. These findings confirm a strong correlation between the variables for the number of reviews, the number of ratings, the number of users who expressed an intention to read the work, and the number of users who added the work to their shelves, as evidenced by the correlation coefficients for these relationships. While these variables also correlate with the number of users who indicated they were reading the work, the effect size is consistently smaller in these cases.

Based on this association, a multiple linear regression model was tested (Table 15) to determine which variables predict the number of reviews. The coefficient of determination indicates an effect size of 91.1% ($R^2 = .911$, $F(3) = 2694.892$), which, according to Cohen's classification, represents a substantial magnitude. Since the Durbin-Watson test (1.596) yields a value close to two, it can be concluded that the model meets the assumption of independence. The assumption of non-collinearity is also met, given that the values obtained were >0.2 in the Tolerance test and <5 in the VIF test. Among the variables introduced, the number of ratings, the average rating, and the number of users who indicated they were reading the work significantly and positively predict the number of reviews. Thus, active user participation is determined by variables related to the act of reading or having read the title, rather than by markers associated with the creation and display of reader identity. However, as previously noted, there is a strong correlation between these variables.

4. Discussion

An increasing number of studies demonstrate how social networks and other digital platforms are redefining the concept of the literary canon (Salgaro 2022). Within the framework of systemic theories, this study

employs a quantitative approach to measure the impact of institutional repertoires on ordinary readers by analysing social networks for readers, which serve as privileged spaces for exploring society's reading preferences. For this purpose, the school canons established by the Ministries of Education of France and Portugal were selected as the institutionalized corpus, since these countries prescribe specific reading lists, unlike others. Goodreads, the largest and most globally recognized social network for readers, was chosen as the platform for analysis.

The analysis began by examining both canons, revealing distinct approaches. The French canon, which is more contemporary and includes a broader range of paraliterary alternatives (Lluch 2023), adopts a more flexible approach aligned with recent models of literary education. In contrast, the Portuguese school canon adheres to a more traditional framework, characterized by a predominance of national and narrative works. Unlike the French selection, the Portuguese canon contains no trace of non-fiction texts, pointing to the limited role of such works in educational settings (La Croix *et al.*, 2024). These findings highlight how the canonical selections of both countries are deeply rooted in their historical and national identities. This is particularly evident in the case of Portugal, where the school canon established after the Salazar dictatorship was shaped by the desire to unite national authors who contributed to building a democratic Portuguese identity. As this analysis shows, while the current school canon is more diverse in terms of forms and authors, it remains indebted to these earlier approaches (Bernardes 2005). When evaluating this issue, it is also important to consider that the Portuguese list has been developed to be used in Portuguese lessons, whereas the French list does not possess such a limitation.

Regarding the impact of both canons on Goodreads, both the French and Portuguese school canons enjoy significant acceptance among ordinary readers, which

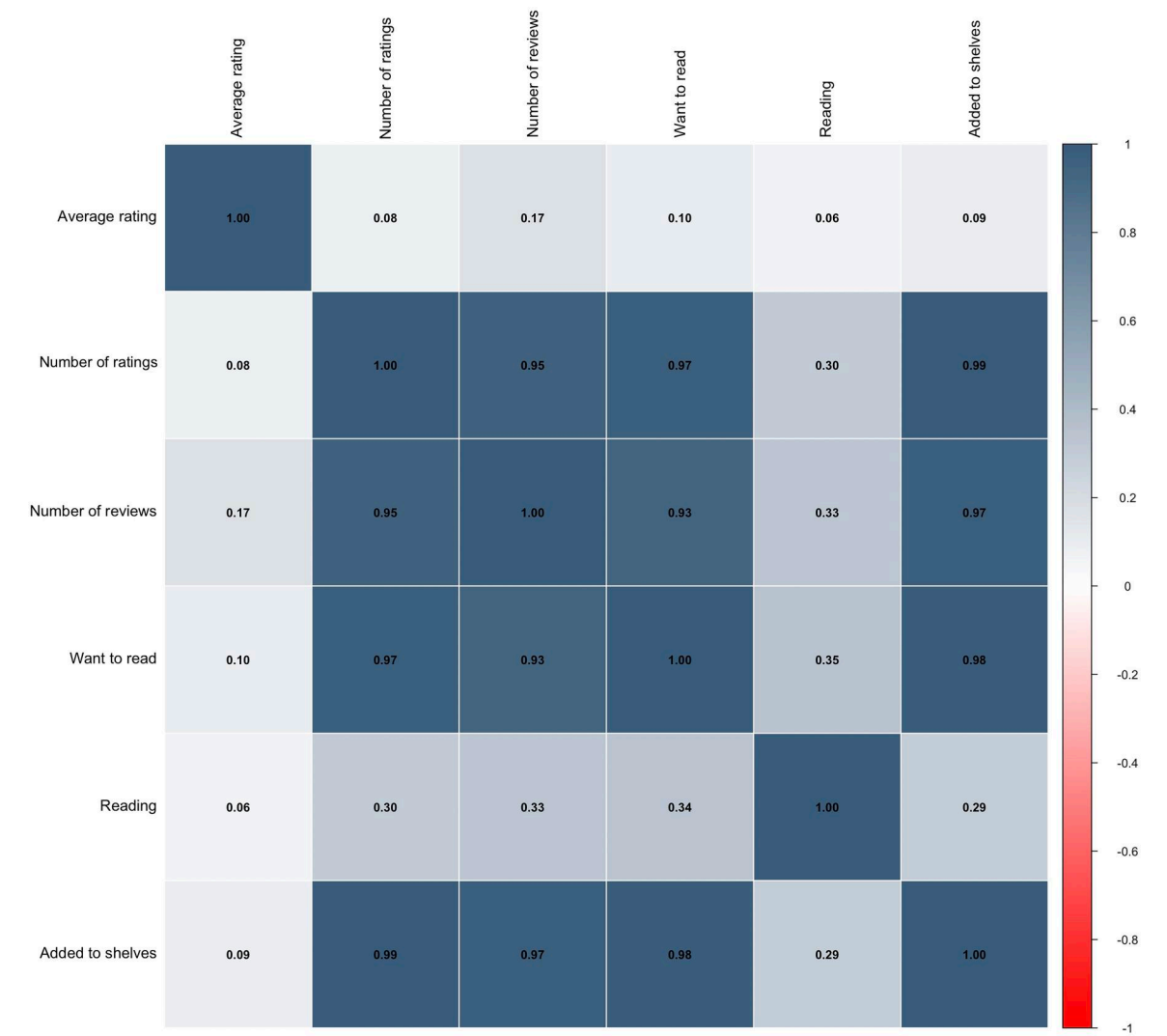


Figure 3. Correlation among the variables related to the impact of the school canon on Goodreads. Pearson's statistical method was applied to verify the correlations between the quantitative variables of Goodreads. This correlation matrix was then produced using a Code that can be consulted in the data repository. The more blue the labels appear, the stronger the positive correlation. The more red the labels appear, the greater the negative correlation, although no relationships of the latter type were identified.

Table 15. Linear regression model for the dependent variable: number of reviews.

Predictor variables	Estimate	SE	P	Tolerance	VIF	F	R ²
No. of ratings	.929	.000	.000	.905	1.104	2694.892	.911
Average rating	.092	1.790	.00	.993	1.007		
Currently reading	.44	.007	.000	.908	1.101		

may be related to the results of some studies showing how children's and young adult literature is the favourite genre of the Goodreads audience (Bourrier and Thelwall 2020; Dera 2024). Readers not only value the more classic titles (Lippolis 2023), but also

embrace contemporary alternatives, such as non-fiction works, wordless picture books and interactive materials. This resonates with findings from other studies that point to readers' interest in the materiality of the book and the physical experience of reading

- Alghamdi, A. M., and Ishaiah, S. (2021) 'The Use and Impact of Goodreads Rating and Reviews, for Readers of Arabic Books', *International Journal of Business Information Systems*, 37: 442–466. <https://doi.org/10.1504/IJBIS.2021.117004>
- Antoniak, M. et al. (2024) 'The Afterlives of Shakespeare and Company in Online Social Readership', *Journal of Cultural Analytics*, 9: 1–20. <https://doi.org/10.22148/001c.116919>
- Benton, M. (2000) 'Canons Ancient and Modern: The Text We Teach', *Educational Review*, 52: 269–77.
- Bernardes, J. A. (2005) *Como Abordar... a Literatura no Ensino Secundário*. Porto: Areal Editores.
- Bloom, H. (1995) *The Western Canon: The Books and School of the Ages*. New York: Riverhead Books.
- Bourdieu, P. (1998) *La Distinción. Criterio y Bases Sociales Del Gusto*. Madrid: Taurus.
- Bourrier, K., and Thelwall, M. (2020) 'The Social Lives of Books: Reading Victorian Literature on Goodreads', *Journal of Cultural Analytics*, 5: 1–34. <https://doi.org/10.22148/001c.12049>
- Ceballos, A. (2019) 'La Recepción Empírica de Benito Pérez Galdós en la Red Social Digital Goodreads', *Anales Galdosianos*, 54: 31–48. <https://dx.doi.org/10.1353/ang.2019.0001>.
- Cordón-García, J. A. (2018) 'Combats for the Book: A Play on Words with the Digital Mode', *El Profesional de la Información*, 27: 467–81. <https://doi.org/10.3145/epi.2018.may.02>
- Dera, J. (2024) 'Embarking on the Online Reading Challenge: Adolescents' Participation Motives, Gains and Impacts on Reading Routines', *Literacy*, 58, 301–11. <https://doi.org/10.1111/lit.12380>
- Driscoll, B., and Rehberg, D. (2019) 'Faraway, So Close: Seeing the Intimacy in Goodreads Reviews', *Qualitative Inquiry*, 25, 248–59. <https://doi.org/10.1177/107780041880137>
- Eagleton, R. (1999) *Doing English. A Guide for Literature Students*. London: Routledge.
- Even-Zohar, I. (1997) 'Factors and Dependencies in Culture: A Revised Outline Polysystem Culture Research', *Canadian Review of Comparative Literature*, 24, 15–34.
- Fish, S. (1980) *Is There a Text in This Class?*. Harvard: Harvard University Press.
- Fokkema, D. (1993) 'A European Canon of Literature?', *European Review*, 1, 21–29. <https://doi.org/10.1017/S1062798700000351>
- Francés Díez, M. Á., and Lluch, G. (2024) 'Canon y Premio Nobel: un Estado De La Cuestión', *Interlitteraria*, 29, 109–26. <https://doi.org/10.12697/IL.2024.29.1.8>
- Francés-Díez, Á. (2021) 'Dialogic Creation of the Popular Canon and Social Capital on Twitter', *Bulletion of Hispanic Studies*, 98, 981–99. <https://doi.org/10.3828/bhs.2021.57>
- García-Roca, A. (2016) 'Reading Practices in Affinity Spaces: Participatory Forms of Digital Culture', *Ocnos*, 15, 42–51. https://doi.org/10.18239/ocnos_2016.15.1.979
- Ghasemi, A., and Zahediasl, S. (2012) 'Normality Tests for Statistical Analysis: A Guide for Non-Statisticians', *International Journal of Endocrinology and Metabolism*, 10, 486–89. <https://dx.doi.org/10.5812/ijem.3505>
- Gómez-Díaz, R., and García-Rodríguez, A. (2024) 'Establishing as Children's and Young People's Literary Canon Based on Literary Awards', *Ocnos*, 23, 1–15. https://doi.org/10.18239/ocnos_2024.23.1.452
- Hegel, A. (2018) *Social Reading in the Digital Age*. Doctoral Dissertation, University of California. <https://escholarship.org/uc/item/193451g2>
- Jenkins, H., Ito, M., and Boyd, D. (2015) *Participatory Culture in a Networked Era: A Conversation on Youth, Learning, Commerce and Politics*. Cambridge: Polity Press.
- Kuijpers, M. et al. (2023) 'Absorption in Online Reviews of Books: Presenting the English-Language AbsORB Metadata Corpus and Annotation Guidelines', *Journal of Open Humanities Data*, 9, 1–7. <https://doi.org/10.5334/johd.116>
- La Croix, L. et al. (2024) 'A Snapshot of Early Childhood Teacher's Read-Aloud Selections', *Early Childhood Education Journal*, 52, 1433–45. <https://doi.org/10.1007/s10643-023-01534-3>
- Lippolis, A. S. (2023) 'Italian Nostalgia: National and Global Identities of the Italian Novel', *Journal of Cultural Analytics*, 8, 1–21. <https://doi.org/10.22148/001c.68341>
- Lluch, G. (2023) 'A Methodological Proposal to Study the Non-Academic Literary Canon', *Revista de Literatura*, 85, 349–70. <https://doi.org/10.3989/revliteratura.2023.02.014>
- Lluch, G. (2025) *Los Libros Que Leemos. ¿Quién, Dónde y Cómo se Decide? La Construcción Del Canon no Académico*. Barcelona: Tirant humanidades.
- López-Martín, E., and Ardura-Martínez, D. (2023) 'The Effect Size in Scientific Publication', *Educación XX1*, 26, 9–17. <https://doi.org/10.5944/educxx1.3627600>
- Lotman, I. (1996) *La Semiosfera. Semiótica de la Cultura y del Texto*. Madrid: Cátedra
- Mignolo, W. (1991) 'Canons A(nd)Cross-Cultural Boundaries (Or, Whose Canon Are We Talking about?)', *Poetics Today*, 12, 1–28.
- Nakamura, L. (2013) 'Words with Friends: Socially Networked Reading on Goodreads', *PMLA*, 128, 238–43.
- Nodelman, P. (2008) *The Hidden Adult: Defining Children's Literature*. Baltimore: Johns Hopkins University Press.
- Pianzola, F. (2025) *Digital Social Reading: Sharing Fiction in the Twenty-First Century*. Cambridge: The MIT Press.
- Salgaro, M. (2022) 'Literary Value in the Era of Big Data. Operationalizing Critical Distance in Professional and Non-Professional Reviews', *Journal of Cultural Analytics*, 7, 1–23. <https://doi.org/10.22148/001c.36446>
- Sánchez-García, R. (2020) 'De Poesía, Marketing Editorial y Poéticas. Canon y Contra-Canon en la Poesía Española Última', *Artifarra*, 20, 159–69. <http://dx.doi.org/10.13135/1594-378X/4423>
- Sánchez-García, P., Hernández-Ortega, J., and Rovira-Collado, J. (2021) 'Reading the Social Reader: evolution of Spanish Children's and Young Adult Literature in Goodreads', *Ocnos*, 20, 7–22. https://dx.doi.org/10.18239/ocnos_2021.20.1.2446
- Santoli, P., and Elaine, W. (2004) 'Promoting Young Adult Literature: The Other "Real" Literature', *American Secondary Education*, 33, 66–75.
- Torre-Espinosa, M. (2021) 'La Interferencia Intersistémica Como Aproximación al Canon', *Revista Signa*, 30, 713–30.
- Walsh, M., and Antoniak, M. (2021) 'The Goodreads Classics: A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism', *Journal of Cultural Analytics*, 6, 243–87. <https://doi.org/10.22148/001c.22221>
- Zheng, J., and Fan, W. (2023) 'Reader's Reception of Translated Literary Work: *Fortress Besieged* in the Digital English World', *Digital Scholarship in the Humanities*, 38, 408–19. <https://doi.org/10.1093/llc/fqac017>

© The Author(s) 2025. Published by Oxford University Press on behalf of EADH.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Digital Scholarship in the Humanities, 2025, 00, 1–15

<https://doi.org/10.1093/llc/fqaf130>

Full Paper