

Distribution Characteristics of Best-selling E-books and Online Review Analysis in B2C Platforms: A Case Study of JD.com Postprint

Authors: Lei Bing, Zhong Zhen

Date: 2023-08-27T00:00:00+00:00

Abstract

[Purpose/Significance]Through data analysis of e-book characteristics on Jingdong Mall, this study identifies the common features and differential patterns of best-selling e-books across different categories in terms of pricing, editing, publishing, sales, and user experience.[Method/Process]Employing web crawlers, we collected data on the top 60 best-selling e-books in seven categories from Jingdong Mall, encompassing information on e-book price, paper edition list price, paper edition selling price, file format and size, review count, and review text. Through methods including text mining and descriptive statistics, we investigated the pricing strategies, sales volume, text formats, and user experience of best-selling e-books.[Results/Conclusion]Best-selling e-books are generally priced at low levels, with modest variations across different categories and types; professional books command higher prices than popular titles. The target audiences for e-books and paper books differ, and e-book sales are not directly proportional to those of their paper counterparts. EPUB flowable files and PDF fixed-layout files dominate the market, with no significant difference in user acceptance between the two formats. User experience regarding best-selling e-books primarily concentrates on three dimensions: “satisfaction” , “price” , and “convenience” .

Full Text

Distribution Characteristics and Online Review Analysis of Bestselling E-books in B2C Malls: A Case Study of JD.com

Lei Bing^{1,2}, Zhong Zhen^{1,2} ¹School of Management, Henan University of Technology, Zhengzhou 450001 ²Business Intelligence & Knowledge Engineering Laboratory, Henan University of Technology, Zhengzhou 450001

Abstract

[Purpose/Significance] This study analyzes data on e-book features from JD.com to identify common characteristics and differences among different types of bestselling e-books regarding pricing, editing, publishing, sales, and user experience. **[Method/Process]** Using web crawlers, we collected information on the top 60 bestselling e-books across seven categories on JD.com, including e-book prices, paper book list prices, paper book actual prices, file formats and sizes, review counts, and review texts. Through text mining and descriptive statistical methods, we examined the pricing strategies, sales volumes, text formats, and user experiences of these bestselling e-books. **[Result/Conclusion]** Bestselling e-books are generally priced low, with slight variations across categories and features—professional books command higher prices than popular books. The audiences for e-books and paper books differ, and e-book sales are not positively correlated with their paper counterparts. EPUB flowable files and PDF fixed-layout files dominate the market, with no significant difference in user acceptance between the two formats. User experiences with bestselling e-books primarily concentrate on three dimensions: “satisfaction,” “price,” and “convenience.”

Keywords: e-books; bestselling books; pricing strategies; sales quantity; text format; user experience; review mining

1. Introduction

With the development of electronic document technology and the proliferation of portable reading devices, the e-book market has expanded continuously. According to data from the China Press and Publication Research Institute, China’s e-book market revenue reached 5.2 billion yuan in 2016. However, compared with the overall scale of the digital publishing industry, e-books account for only a small share of 0.91% [1], indicating that the book publishing industry needs to accelerate its digital integration efforts. E-book development is influenced by multiple factors, including supply-side elements such as reading terminals, file formats, pricing strategies, content production, copyright protection, and market operations [2], as well as reader behavior factors like reading habits and technology adoption [3]. For library and information science, key research concerns include the usage behaviors, reading habits, and thematic preferences of e-book readers.

JD.com is a well-known B2C marketplace in China. Its book category has rapidly expanded its online market share through advantages in brand, traffic, logistics, and promotion. In the third quarter of 2017, JD.com’s book market accounted for 36.2% of China’s online book sales market, ranking first in domestic online book sales [4]. In the e-book sector, unlike Amazon’s approach of building a relatively independent e-book ecosystem centered on the Kindle reader, JD.com emphasizes compatibility: on the one hand, it uses mainstream file formats such as EPUB and PDF, along with portable “e-ink” e-book read-

ers, to adapt to various reading environments and reduce barriers and costs for readers; on the other hand, it has developed multi-version clients for different operating systems (such as Windows, Android, and iOS), lowering the threshold and costs for publishers to produce e-books.

From the perspective of e-book reader adoption behavior, this study uses JD.com as a case to conduct empirical research on e-book sales in B2C malls through data collection, statistical analysis, and review text mining, analyzing the common characteristics and differences of current bestselling e-books across different categories in terms of pricing strategies, sales volume, text format, and user experience, thereby providing empirical evidence for theoretical research on e-book reading behavior and motivation.

2. Literature Review

2.1 Research on E-book Pricing

Pricing is a hot topic in e-book research, with existing studies primarily focusing on pricing power, pricing influencing factors, and pricing strategies. E-book sales methods mainly include three types: buyout, wholesale, and agency models [5]. L. Hao and M. Fan studied e-book pricing under wholesale and agency models, concluding that the high-price agency model reduces total e-book sales revenue [6]. Q. Lu et al. established a supply chain model comprising authors, publishers, physical bookstores, and e-bookstores [7]. B. D. L. Santos and M. R. Wildenbeest used sales data from B2C malls like Amazon to empirically test the impact of the agency model on e-book pricing in the B2C market [8]. Domestically, Hu Xingqiu and Qu Wenfeng argue that e-book pricing factors include cost, pricing power, and copyright protection [9], while Wu Guoyi suggests additional factors such as profit objectives, market demand, and market competition [10]. Furthermore, Wang Lulu and Xu Zhenyun analyzed the motivations for free e-books, including supply, demand, and branding aspects [11]. Dong Xiaomei proposed independent pricing, subscription pricing, and split pricing strategies for publishers [12]. Ye Lan et al. categorized e-book pricing into seven models, including sales by integrated database, per-copy purchase, and annual subscription [13]. Liu Yindi, through a comparison of Chinese and American e-book pricing systems, suggested adopting a fixed-price system, formulating e-book prices based on printed book prices, and further optimizing prices using big data [14].

2.2 Research on E-book Sales Volume

This research area primarily focuses on how pricing, promotion, and paper book sales affect e-book sales volume. D. Boung et al. used econometric methods to compare bestsellers in paper and electronic formats on Amazon from November 2007 to July 2010, finding no consistency between e-book and paper book sales [15]. H. Chen et al. employed a natural experiment approach to compare books with paper versions released before electronic versions versus simultane-

ous releases on Amazon in 2010, discovering that delayed e-book release reduced e-book sales without increasing paper book sales, suggesting that e-books have limited impact on paper books in the short term [16]. Regarding domestic e-book retail platforms, Xu Jie et al., based on statistical analysis of pricing information on the “Duokan Reading” e-book platform, found that book production costs have no significant impact on e-book pricing, but discount ratios affect e-book sales volume [17]. Shi Xi used marketing 4Ps theory to compare e-book marketing strategies across e-commerce platforms such as Taobao, Dangdang, JD.com, and Amazon, proposing corresponding recommendations [18].

2.3 Research on E-book Formats

E-book formats are diverse and complex, including both proprietary and open formats, making standardization difficult. Major formats include PDF, LIT, AZW, EPUB, TXT, and CAJ, with PDF and EPUB being the most widely used [19]. PDF is a fixed-layout format suitable for professional books with numerous figures, tables, and formulas but cannot adapt to different display sizes, resulting in poor mobile experience. EPUB is a flowable format that can automatically reflow according to reading terminal screen sizes but has limitations in displaying figures, tables, and formulas [20]. Currently, EPUB 3.0 and above, based on HTML5 technology, can support SVG (Scalable Vector Graphics) and MathML (Mathematical Markup Language), enhancing support for educational and technical documents [21].

2.4 Research on E-book User Experience

E-book user experience research primarily focuses on the impact of reading terminals on user experience and purchase experience. T. Nakajima et al. [22] tested six EPUB e-book portable reading terminals covering iOS, Android, Windows, and Kindle Linux operating systems, with results showing that the iBooks 3 reader on iPad tablets performed best. Y. Zeng et al. [23], based on reading experiments and questionnaire surveys, found that fixed-layout formats were more popular on traditional large displays such as PCs and laptops, while flowable formats were preferred on portable mobile terminals. N. Amblee and T. Bui, using Amazon’s low-price e-books (Shorts E-Books) as an example, noted that online word-of-mouth can convey e-book content, author reputation, and similar e-books, recommending that both sellers and readers should prioritize online word-of-mouth as the primary source of social shopping experience [24].

In summary, domestic and international scholars have conducted extensive research on e-books, delving into pricing, sales volume, format, and user experience. Nevertheless, some improvements can still be made: (1) foreign empirical research primarily uses data from mainstream online malls like Amazon with in-depth content, while domestic research mainly consists of qualitative elaboration with insufficient empirical studies; (2) there is a partial absence of correlation analysis between online reviews and user experience, lacking feature extraction analysis regarding e-book user experience in review texts.

3. Research Design

JD.com officially launched its e-book business in February 2012, synchronizing its sales platform with smartphone/PC reading client software LeBook. After years of development, it has become China's largest e-book sales platform. The rich variety of e-books and relatively broad audience make JD.com an excellent data source for studying China's e-book publishing and sales status.

3.1 Data Source and Collection Method

This study's data were obtained from JD.com's digital content channel (<http://e.jd.com>) on January 20, 2018, using the Octopus Collector version 7.2.4 (<http://www.bazhuayu.com>). Through configuration of collection rules and cloud collection, we obtained the raw data for this study. The collection process consisted of two steps: basic information collection and review text collection, with sample sizes shown in Table 1.

The first step collected basic information on bestselling e-books, including nine data items: book title, ISBN, category, e-book actual price, paper book list price and actual price, file format, file size, e-book review count, and paper book review count. According to JD.com's digital content channel classification, we collected the top 60 books by review count in each category (JD.com displays 60 products per page). JD.com does not use the Chinese Library Classification but instead divides e-books into seven categories: literature, management, social sciences, industrial technology, home life, textbook/examination, and children's books, resulting in 420 records. Among the seven categories, literature, home life, and children's books belong to popular books; textbook/examination books belong to educational books; and industrial technology, management, and social sciences belong to professional books. After data collection, manual review revealed that the 48th entry in literature, the 54th entry in social sciences, and the 10th entry in children's books were advertisements, which were removed. Therefore, the sample actually contained 417 e-books.

The second step collected user review texts for these bestsellers. Among the collected bestsellers, the highest review count was 4,100, while the lowest was 10. Considering research needs and JD.com's anti-scraping mechanism limitations, we collected all reviews for books with fewer than 100 reviews and the first 100 reviews for those with more than 100.

3.2 Data Analysis Dimensions and Methods

Based on the research objectives and collected data records and attributes, we analyzed the characteristic distribution of different types of bestselling e-books across four dimensions: pricing, sales volume, format, and user experience. The following explains the analysis methods for pricing, sales volume, and format dimensions; the user experience analysis method is described in section 3.3.

- (1) **Pricing Analysis.** Through collected data, we obtained e-book prices

and corresponding paper book list prices and actual prices. Using paper book list prices as a benchmark, we compared pricing strategies across different e-book categories through descriptive statistical analysis. In statistical analysis, we used “e-book price = paper book list price \times e-book discount” and “paper book price = paper book list price \times paper book discount” for description.

- (2) **Sales Volume Analysis.** We analyzed sales volumes of e-books across different categories and price ranges through collected data. JD.com does not provide cumulative sales volumes but offers cumulative review counts. Product reviews are a function set for consumers who have purchased products, including ratings and text comments. According to voluntary principles, consumers can comment on products, generating a review text and increasing the cumulative review count by 1; consumers may also choose not to comment, leaving the cumulative review count unchanged. This study uses cumulative review counts as an approximation for cumulative e-book sales for four reasons: (1) After consumers purchase products, JD.com uses methods such as sending reminder messages and rewarding “Jingdou” (JD.com’s virtual currency) to encourage post-purchase reviews, resulting in relatively high review rates; (2) The target consumer groups in this study are all e-book consumers, belonging to similar review groups with comparable post-purchase review rates; (3) J. Chevalier and D. Mayzlin found through regression analysis of Amazon book data that post-purchase review counts have a significant positive correlation with book sales [25], and domestic scholar Xu Jie also used review counts as a proxy for sales [17]; (4) Analysis of our data found that among the 417 bestselling e-books, the minimum review count was 10, the maximum was 4,100, with a minimum interval of 10 and a standard deviation of 410.9, indicating large dispersion in review counts, making it highly probable that books with higher review counts have higher sales.
- (3) **Format Analysis.** JD.com e-books come in two formats: EPUB (flowable) and PDF (fixed-layout). Through descriptive statistics, we analyzed the proportion, file size, and sales volume differences between the two formats across different book categories.

3.3 User Experience Analysis Method

This study obtains e-book user experience conditions through analysis of post-purchase review texts, including the characteristic distribution of user experiences across different categories, price ranges, and formats. JD.com e-book user review texts mainly include two types: those targeting book content and those targeting user experience. This study focuses on the user experience of the e-book medium, so the emphasis of review mining is on feature word analysis related to overall e-book satisfaction. The primary tool is the R language and related open-source packages, divided into three steps:

- (1) **Chinese Word Segmentation.** For the collected 29,863 review texts, we used the open-source segmentation package “Jieba Chinese Word Segmentation” (jiebaR 0.9.1), selecting its hybrid model algorithm for segmentation. To improve accuracy, we built a local lexicon, adding five proper nouns and merging synonyms such as “e-book,” “JD.com reader,” etc. Simultaneously, we used the “Sichuan University Machine Learning Intelligent Laboratory Stopword Lexicon” to remove stopwords, ultimately obtaining 16,029 different keywords.
- (2) **Feature Word Screening Based on Word Frequency and Review Coverage.** First, we conducted word frequency statistics on the 16,029 keywords, used the wordcloud package to create word clouds, and performed word frequency statistics. Then, we selected the top 380 keywords by frequency, all with review coverage above 0.2% (i.e., appearing at least once in every 500 reviews), as the candidate set for feature words. Finally, we manually screened and removed words commenting on book content and meaningless words, obtaining 33 keywords related to e-book user experience as user experience feature words.
- (3) **Feature Word Classification.** Through analysis of the 33 high-frequency feature words related to e-book user experience (Experience) in Table 2, we found they could be divided into three single categories: satisfaction (positive + negative), price perception, and convenience experience. Among them, satisfaction (Satisfaction) represents overall evaluation of purchased e-books, including both positive and negative cases, such as “The e-book is good, saves the trouble of paper version,” “E-version, I like it haha,” “Never buy e-books! Very bad, and no refund!” etc. To extract satisfaction feature words related to “user experience” rather than “book content,” this study used the “evaluation object” vocabulary in the leftmost column of Table 2 to determine whether satisfaction keywords were user experience feature words based on the “co-occurrence” of the two word groups (whether they appear in the same clause). Price (Price) and convenience (Handle) are relatively direct user experience vocabulary and no longer require co-occurrence comparison with “evaluation object” for identification. When price-related feature words appear in a text, Price=1, otherwise Price=0; the Handle variable is assigned similarly. For any review involving any of the three types of feature words, this study assigns the review’s Experience value as 1, otherwise 0 (see formula (1)).

To evaluate the effectiveness of the feature classification method, this study randomly selected 100 review texts and manually counted the precision, recall, and F-values (calculated as: $\text{precision} \times \text{recall} \times 2 / (\text{precision} + \text{recall})$) of the three types of feature words based on “true semantics,” reaching 89.83%, 91.38%, and 90.60% respectively, proving the method’s effectiveness.

4. Results and Analysis

4.1 E-book Price Distribution Characteristics

Pricing strategy is an important factor affecting e-book sales, with price and discount being two key indicators representing pricing strategy. As shown in Table 3, the average price of 417 bestselling e-books is 7.17 yuan, with an average discount of 20% relative to paper book list prices. Currently, domestic B2C malls generally offer paper book discounts of 60-80%, making such low discounts (low prices) indicate that e-books are still in a promotion period, with low-price strategy being one of the main promotion methods. Among the seven categories, professional books such as management, industrial technology, and social sciences have higher overall discounts (relatively higher prices), all above 24%; textbooks/examinations, literature, home life, and children's books have lower discounts (relatively lower prices), none exceeding 16%. This difference in discounts indicates that professional book consumers have higher price acceptance for e-books than popular and educational book readers.

As shown in Table 4, regarding specific discount choices, over 50% of e-books have discounts at or below 20%; about 20% maintain pricing in the 20-40% discount range, and less than 20% fall in the 40-60% and above 60% ranges. In other words, over 80% of JD.com's bestselling e-books are sold at less than half their "list price."

JD.com's bestselling e-books have such low discount rates because they include a large number of "ultra-low price" e-books priced between 0-1 yuan. These books total 104, accounting for 24.94% of the total, mainly consisting of "copyright-expired" books, predominantly children's books, as well as literary classics like *Romance of the Three Kingdoms* and *The Count of Monte Cristo*. Understandably, paper book production costs mainly comprise copyright fees, production fees, and material fees. If copyright is free, e-books have near-zero marginal costs and can be sold at ultra-low prices. Table 4 also shows that although limited in number, some e-books have discounts above 60%, with prices similar to paper books' "actual prices." High-discount (i.e., high-price) e-books mainly fall into two categories: one is e-versions of hot-selling paper books, such as *The Long Tail Theory* and *Tencent Biography (1998-2016: The Evolution of Chinese Internet Companies)*, where market demand drives higher pricing; the other is niche professional books whose paper versions have ceased reprinting but still have certain market demand, such as *OFDM Communication System Baseband Design Based on Xilinx FPGA* published by Zhejiang University Press in 2009 and *Newbie Color Matching Bible 4: Interior Design* published by Posts & Telecommunications Press in 2011. The distribution of "1 yuan cap" and above-60% discount books is shown in Figure 1 [Figure 1: see original paper].

4.2 E-book Sales Volume Distribution Characteristics Based on Review Counts

Using review counts, we can understand the sales characteristics of JD.com' s bestselling e-books across different categories. Comparing the median review counts for e-versions and paper versions in Table 5 reveals three characteristics of JD.com' s bestselling e-books:

- (1) The ratio of e-version to paper version review count means is as high as 2.26%, indicating that paper books remain the mainstream form of book consumption at present.
- (2) E-book sales are not significantly positively correlated with paper book sales. Review count means show that management and social science categories have the highest e-book sales, while literature, social sciences, and children' s books have the highest paper book sales.
- (3) The impact of e-book sales on total book sales varies significantly across categories. Industrial technology has a review count mean of 112, but its e-version to paper version review count ratio is as high as 27.11%; social sciences, despite having higher e-book sales, has only a 2.00% e-version to paper version ratio, making its impact on total book sales smaller than some categories with lower sales.

Regarding the second and third characteristics of e-book sales, and without considering the influence of book content themes, one explanation is that e-books are currently still in a promotion period with small market share, and readers' acceptance of e-books affects e-book sales.

Notably, in comparing bestselling e-books with their paper versions, we also identified two interesting types of books: "out-of-print" paper books (i.e., no paper version available on the platform) and e-books that exceed paper version sales (see Figure 2 [Figure 2: see original paper]): (1) E-books without paper version sales total 74, with an average price of 9.37 yuan, higher than the 7.17 yuan average for bestselling e-books, accounting for 17.75% of the total. These books are mainly characterized by early publication years without reprinting, and low market sales are insufficient to drive publishers to reprint (such as *New Greek Mythology* by Hubei People' s Publishing House, 1979). Textbook/examination categories have more such books, while literature has the fewest. (2) E-books exceeding paper version sales total 124, with an average price of 4.47 yuan, lower than the overall average for bestselling e-books, accounting for 29.74% of the total. Although paper versions of these books are still sold, their sales are very limited, so publishers and online platforms are willing to sell their e-versions at low prices, guiding readers to accept e-book consumption formats on the one hand and boosting sales of slow-selling paper books on the other. Among these low-price bestsellers, industrial technology has the most with 30 books, accounting for half of that category; social sciences has the fewest with only 10. Understandably, for "out-of-print" and "low-sales" paper books, their "elec-

tronic” resale actually expands the consumer market rather than “replacing” it as commonly perceived.

Although low-price strategies can theoretically increase sales, e-book promotion should be based on meeting consumers’ content needs and providing good experiences rather than simply relying on low-price penetration. In fact, by comparing review count means across different price ranges, we find that JD.com’ s bestselling e-book pricing and sales do not form a negatively sloped curve. As shown in Figure 3 [Figure 3: see original paper], bestselling e-books were divided into seven price intervals: (0,1], (1,2], (2,5], (5,10], (10,20], (20,40], and >40 (yuan). Books priced between 1-2 yuan have the lowest review count mean (117 reviews per book). When pricing is lowered to within 1 yuan, the review count mean increases 1.74 times; however, as pricing increases to the right, the review count mean does not decrease but instead shows linear growth. Further examination of e-books with over 1,000 reviews (24 total) reveals they are mainly composed of social sciences, management, and industrial technology books. It is conceivable that consumers of these categories have higher acceptance of e-versions and purchase books mainly for knowledge learning, making them relatively less sensitive to pricing or discount ratios. For example, the management book *Singularity Series (6-book set)* has a paper list price of 200 yuan, an e-version price of 147.1 yuan (71.6% discount), yet has 1,700 reviews, ranking fourth among the 417 e-books.

4.3 E-book File Format Distribution Characteristics

Among the 417 bestselling e-books, 239 are in EPUB format (57.31%) and 178 in PDF format (42.69%). Regarding file size, the mean for all 417 e-books is 21.97 MB, with EPUB format averaging 6.98 MB and PDF format averaging 42.08 MB—six times larger than EPUB. This mainly stems from differences in content storage methods between the two formats: EPUB text storage approximates “text,” while PDF text storage approximates “images.”

As shown in Table 6 , literature, management, and social science books predominantly use EPUB format. Their content is primarily text, and using EPUB format meets readers’ needs for adaptive reading across various screen sizes. In contrast, industrial technology, textbook/examination, and home life books predominantly use PDF format, mostly containing numerous images and mathematical formulas that do not display well in EPUB format, leading publishers to choose PDF fixed-layout structure. Additionally, some text-based books use PDF layout structure, such as some children’ s books and literary classics. These paper versions have early publication dates, and JD.com sells them at ultra-low prices (average price 0.79 yuan) by simply scanning paper books to create e-versions without complex EPUB format conversion.

It should be noted that this study compared pricing and review count means between EPUB and PDF format books and found no significant differences, indicating that file format has limited impact on book pricing and sales.

4.4 Distribution Characteristics and Influencing Factors of E-book User Experience

Based on the method described in section 3.2, we segmented 29,863 review texts and created a word cloud of JD.com bestselling e-book reviews based on high-frequency vocabulary (see Figure 4 [Figure 4: see original paper]), where font size represents keyword frequency. Word frequency statistics show the top 10 words are “good,” “very good,” “e-book,” “okay,” “great,” “worthwhile,” “like,” “very,” “JD.com,” and “convenient,” all positive terms, indicating high consumer satisfaction with bestselling books—consistent with the result that the 417 books have an average rating of 4.85 (out of 5).

Among these reviews, 5,380 are related to user experience, accounting for 18.02% of total reviews, while over 80% of reviews either have no actual meaning (such as default positive reviews) or only evaluate book content, making it difficult to measure user experience. This study calls reviews with “user experience characteristics” “experiential reviews,” and the proportion of experiential reviews among total reviews is called the “experiential review rate.” Table 7 lists the numbers of three types of experiential reviews and the experiential review rates for seven e-book categories. By category, industrial technology has the lowest proportion of experiential reviews at only 13.28%, possibly because consumers in this category have higher e-book acceptance, leading to more content-focused reviews. Among the three types of experiential reviews, satisfaction reviews total 1,616, with a review rate of 5.41%, including 1,548 positive reviews and 68 negative reviews. The high proportion of positive satisfaction indicates overall good e-book experience. Among the few “dissatisfied” reviews, the focus is mainly on reader limitations and relatively poor PDF file experience characteristics, such as “only supports JD.com reader” and “poor PDF reading experience.” However, limiting use to specific readers is common in e-book retail markets (such as Amazon), and the choice between PDF and EPUB formats depends on the original file format provided by publishers and book content. Price and convenience categories have overall review rates of 8.28% and 8.15%, respectively, higher than average satisfaction, because many reviews simply state “e-book is good” without indicating whether satisfaction comes from price or convenience.

Statistical results show that product pricing, text format, and file size are the main factors affecting experiential reviews. Regarding price, as shown in Table 8, low-priced books under 10 yuan have experiential review rates below 20%, while books above 10 yuan have review rates above 44%. One possible explanation for this significant difference is that when e-book prices are low, consumers have lower expectations for e-books, resulting in fewer related reviews, as seen in some experiential review texts like “50 cents is very cost-effective, support JD.com reading, the book is also good” and “E-book, content is okay, just for casual reading.” When e-book prices are higher, consumers pay more attention to the user experience of the e-book medium, comparing e-versions with paper versions, such as “E-book is more portable than paper version, not bad” and “Okay, has illustrations, e-version price is actually not very cheap.”

Regarding file format, EPUB and PDF format e-books have experiential review rates of 18.00% and 18.06%, respectively, with no significant difference. However, in terms of satisfaction, EPUB format has slightly better positive reviews than PDF format, 1.07% higher, which is related to EPUB format's high adaptability to different screen sizes. Regarding file size, books within 50 MB and those above 50 MB show slight differences in convenience reviews, such as one consumer's comment after purchasing a book with 185.78 MB file size: "Downloading consumes too much data," but overall there are few review texts targeting file size, which is closely related to the current popularity of WiFi and 4G in China and fast internet speeds, making file size's impact on user experience minimal. Related statistics are shown in Table 9 and Table 10 .

5. Conclusion and Discussion

Based on JD.com data, this study examined the common characteristics and differences of bestselling e-books across different categories from four aspects: pricing strategy, sales volume, text format, and user experience. The main conclusions are as follows:

- (1) Bestselling e-books are generally priced low, with slight differences across categories and features. The current e-book market is in a promotion period, and readers' acceptance is not yet high compared to paper books. Low-price strategies can promote e-book sales, such as copyright-free books sold at ultra-low "1 yuan cap" prices to expand the e-book market. Professional book reader groups, such as management, industrial technology, and social sciences, have higher e-version acceptance than other book reader groups, with pricing slightly higher than popular and educational books like literature and textbooks.
- (2) Bestselling e-book sales are not positively correlated with their paper version sales. It is generally believed that e-book promotion will cannibalize paper book markets. However, this view is not supported by this study's empirical evidence. Among the 417 books, nearly half of e-book sales do not depend on paper books—these e-books either sell more than paper versions or have no paper versions sold on JD.com. This empirical conclusion indicates that e-book and paper book audiences differ, and promoting e-books can expand book market space and increase sales revenue for publishers and B2C malls.
- (3) Among bestselling e-books, EPUB flowable files and PDF fixed-layout files dominate, with no significant difference in user acceptance and satisfaction between the two formats. Publishers' choice of e-book format mainly depends on book content: text-heavy books mostly use EPUB format, while books with numerous images and formulas primarily use PDF format. Another phenomenon exists among bestselling e-books: for books with early publication dates whose paper versions are no longer sold, publishers produce PDF paper-scanned versions, which B2C malls sell at ultra-low

prices as a promotional method to promote e-book consumption markets.

- (4) User experience characteristics of bestselling e-books mainly focus on three aspects: satisfaction, price, and convenience. Readers have high overall satisfaction with e-books, over 95%; readers pay great attention to e-book price and convenience, with “cheap” and “convenient” being readers’ most intuitive perceptions. Different book category consumer groups have slightly different attention to e-book user experience, with industrial technology book readers showing the lowest attention, possibly because this group has higher e-book acceptance. High-priced e-books have significantly higher user experience attention than low-priced e-books, possibly because when e-book prices are higher, they are more easily compared with paper versions, thereby increasing e-version acceptance.

Due to technical limitations, this study has certain limitations: (1) The selected data only include bestselling e-books, not bestselling paper versions, resulting in insufficiently rich sample types; (2) Using e-book review counts as a proxy for cumulative sales reduces statistical analysis accuracy; (3) Using online reviews as observation points for e-book user experience cannot fully reflect consumers’ true thoughts. These issues will be addressed in future research.

References

- [1] China Digital Publishing Industry Annual Report Research Group. China’ s digital publishing in the 13th Five-Year Plan’ s first year—Main report (abstract) of the 2016-2017 China digital publishing industry annual report[J]. Publishing Research, 2017(7): 5-10.
- [2] Zhang Meijuan, Pan Han, Huang Longzhou. Review of China’ s e-book industry research in recent years[J]. Publishing Science, 2013, 21(6): 48-50.
- [3] Zhao Wenjun, Chen Huanzhi, Jiang Weijin. Review of digital reading research at home and abroad in the past 10 years[J]. Library and Information Service, 2017, 61(18): 128-136.
- [4] Analysys. China B2C market quarterly monitoring report Q3 2017[EB/OL]. [2018-03-10] <https://www.analysys.cn/analysis/22/detail/1000995/>.
- [5] Zheng Lifan. Viewing e-book industry development from pricing games[J]. China Publishing, 2014(4): 11-14.
- [6] HAO L, FAN M. An analysis of pricing models in the electronic book market[J]. MIS quarterly, 2014, 38(4): 1017-1032.
- [7] LU Q, SHI V, HUANG J. Who benefit from agency model: a strategic analysis of pricing models in distribution channels of physical books and e-books[J]. European journal of operational research, 2018, 264(3): 1074-1091.
- [8] SANTOS B D L, WILDENBEEST M R. E-book pricing and vertical restraints[J]. Quantitative marketing & economics, 2017, 15(2): 85-122.

- [9] Hu Xingqiu, Qu Wenfeng. Discussion on price discrimination strategies for e-book pricing[J]. Science-Technology & Publication, 2013, 32(8): 14-18.
- [10] Wu Guoyi. Analysis of e-book pricing strategies[J]. Science-Technology & Publication, 2014(10): 95-98.
- [11] Wang Lulu, Xu Zhenyun. Analysis of motivations for free e-books[J]. Publishing Friends, 2015(7): 36-39.
- [12] Dong Xiaomei. Exploration of innovative e-book pricing strategies for publishers[J]. Science-Technology & Publication, 2015(3): 56-58.
- [13] Ye Lan, Sun Tan, Huang Guobin, et al. Research on e-book pricing models[J]. Library and Information Service, 2007, 51(12): 81-83.
- [14] Liu Yindi. Comparative study on the status quo and pricing systems of Chinese and American e-books[J]. Publishing Friends, 2015(4): 105-108.
- [15] BOUNIE D, EANG B, SIRBU M, et al. Superstars and outsiders in online markets: an empirical analysis of electronic books[J]. Electronic commerce research and applications, 2013, 12(1): 52-61.
- [16] CHEN H, HU Y J, SMITH M D. The impact of e-book distribution on print sales: analysis of a natural experiment[EB/OL]. [2017-12-14]. <http://dx.doi.org/10.2139/ssrn.1966115>.
- [17] Xu Jie, Ding Lehua, Peng Wei. Data mining analysis of e-book pricing patterns based on “Duokan Reading” [J]. Journal of Information Resources Management, 2017(1): 89-96.
- [18] Shi Xi. Analysis of e-book publishing based on e-commerce platforms—Taking Taobao Reading, Dangdang Digital, JD.com E-books, and Kindle E-bookstore as examples[J]. Press Circles, 2015(2): 56-59.
- [19] Zheng Yingyu, Cong Ting, Miao Jie. Research on e-book format standards[J]. Publishing Science, 2010, 18(4): 90-93.
- [20] Chen Min. Characteristics of professional e-books and analysis of their consumer behavior—Empirical research based on Chemical Industry Press[J]. Science-Technology & Publication, 2012(5): 64-67.
- [21] IDPF. EPUB 3.1[EB/OL]. [2018-03-10]. <http://idpf.org/epub31/>.
- [22] NAKAJIMA T, SHINOHARA S, TAMURA Y. Typical functions of e-textbook, implementation, and compatibility verification with use of epub3 materials[J]. Procedia computer science, 2013, 22: 1344-1353.
- [23] ZENG Y, BAI X, XU J, et al. The influence of e-book format and reading device on users’ reading experience: a case study of graduate students[J]. Publishing research quarterly, 2016, 32(4): 319-330.
- [24] AMBLEE N, BUI T. Harnessing the influence of social proof in online shopping: the effect of electronic word of mouth on sales of digital micro-products[J].

International journal of electronic commerce, 2011, 16(2): 91-114.

[25] CHEVALIER J A, MAYZLIN D. The effect of word of mouth on sales: online book reviews[J]. Journal of marketing research, 2006, 43(3): 345-354.

Author Contributions

Lei Bing: Responsible for determining the paper framework, data collection and organization, data analysis, and drafting the initial manuscript.

Zhong Zhen: Responsible for proposing the research topic, designing the research plan, writing the paper, and revising and finalizing the manuscript.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.