Research Article

American Journal of Science Education Research

Bayesian Prediction of Popularity in Educational and Cultural Books: Investigating the Mediating Roles of Temporal Spatial Interest Factors and Renewal Behavior in Readers' Reading Patterns

Jiexuan Liu* [ID]

Nanjing Normal University, Nanjing, Jiangsu, China

*Corresponding author: Jiexuan Liu, Email: 79821434@qq.com

Abstract

This study investigates the popularity prediction of educational and cultural books (categorized under G-type in the Chinese Library Classification) and explores the underlying mechanisms shaping readers' reading patterns, using longitudinal borrowing data from Nanjing Normal University Library spanning three periods: 32,760 G-type books with 102,566 borrowing transactions involving 10,824 readers for 2016–2018, 23,622 G-type books with 52,690 borrowing transactions involving 7,234 readers for 2019–2021, and 14,710 G-type books with 28,657 borrowing transactions involving 4,978 readers for 2022–2024. It employs a Tree-Augmented Naive (TAN) Bayesian model to predict book popularity. This model focuses on books with borrowing volumes exceeding the 75th percentile, determined via quantile analysis. It also uses a moderated regression model that integrates temporal, spatial, and interest-based factors including G0–G8 subcategories, with renewal behavior as a moderator and controls for gender and reader type. Key findings reveal that predictive variables influence popularity through multi-step pathways, with "G Subcategory" as the most critical predictive variable and "Reader Gender" as the least, highlighting that content attributes are fundamental to determining popularity, while moderated regression shows temporal-spatial-interest factors (and their evolutionary shift toward temporal-interest factors) significantly affect borrowing volume, with renewal behavior moderating these relationships by strengthening the impact of interest-based preferences and weakening temporal effects. This study contributes to theoretical advancements in reader behavior research and provides data-driven insights for library collection management, personalized book recommendations, and targeted reading promotion in educational and cultural domains.

1. Introduction

The ebb and flow of popularity in educational and cultural books have long been intertwined with the dynamic interplay of historical context, spatial arrangements, and evolving reader interests. As early as the 18th century, private libraries in rural Scotland served as vital hubs for exchanging texts on female education and child-rearing, fostering community-specific reading practices that laid the groundwork for interest-driven renewal behavior—hinting at how spatial contexts have historically mediated sustained engagement with educational content (Towsey, 2013). Similarly, medieval classification systems, such as the catalog of the Abbey of Saint-Victor, which integrated the "seven liberal arts," shaped readers' knowledge frameworks and directly influenced the dissemination scope of educational works, underscoring how content organization has long mediated books' educational impact and popularity (Bodemer, 2012). Even societal shifts, as highlighted by initiatives like Scotland's Information for Scotland conferences (1993–1998), reveal that temporal updates to knowledge systems drive "democratic renewal," paralleling how evolving public interests reshape the popularity of educational and cultural books over time (Newton & Dixon, 2000). These historical examples collectively illustrate a enduring truth: the popularity of such books is not static but emerges from complex interactions between when (temporal), where (spatial), and why (interest) readers engage with content—interactions that remain poorly understood in contemporary digital and decentralized reading landscapes.

The educational and cultural (G-category) book consumption is undergoing profound transformation now. Digital access, decentralized reading habits, and shifting institutional demands have altered how readers discover, engage with, and revisit books, challenging traditional models of popularity prediction. Against this backdrop, several critical questions emerge, guiding the present study:

Q1: How can Bayesian network models effectively predict the popularity of educational and cultural (G-category) books by integrating conditional relationships among reader attributes (e.g., gender, department affiliation), behavior-time variables (e.g., renewal behavior, month), and book content attributes (e.g., G subcategory)?

Q2: To what extent do content attributes (G subcategory) function as the foundational predictor of G-category book popularity, and how do they interact with reader-related variables (e.g., reader type, department) to influence the dissemination potential of such books across different periods?

Q3: What mediating roles do temporal-spatial-interest (TSI) factors and their evolutionary shift toward temporal-interest (TI) factors play in linking reader behavior, book content characteristics, and the popularity of educational and cultural books?

Q4: How does renewal behavior moderate the relationship between TSI/TI factors and borrowing volume (LogTotal) across the periods 2016–2018, 2019–2021, and 2022–2024, and

what underlying mechanisms explain the observed temporal variations in the strength of this moderating effect?

Q5: How do the most popular theme phrases and title topics of G-category books—encompassing educational theory, teaching practices, psychology, and cultural studies—interact with TSI/TI factors to shape reader demand and book popularity?

The study employs longitudinal data across three periods (2016–2024) with fine-grained classification of G-type books into secondary categories (G0–G8) to capture long-term trends and subfield differences; identifies popular books using the 75th percentile of lending volume for each period and categorizes them by cross-period overlap to distinguish stable and short-term popularity; and adopts advanced methods like the TAN Bayesian model (accommodating tree-like feature dependencies to better fit real data) and a moderated regression framework integrating time, space, and interest factors with renewal behavior as a moderator, thereby uncovering complex interactive effects on borrowing behavior.

The significance of this study lies in its focus on the popularity of educational and cultural books, using Bayesian models to predict their popularity trends while exploring the mediating roles of temporal-spatial factors, interest preferences, and renewal behavior in readers' reading patterns. Theoretically, it uncovers the dynamic formation mechanism of the popularity of educational and cultural books, clarifies the relational pathways between temporal-spatial characteristics, readers' interests, and borrowing behaviors (e.g., renewal), thereby providing new empirical support for theories of reader behavior in library and information science. Practically, it offers data-driven references for libraries and related institutions in collection development and optimization of book recommendation strategies, facilitating precise alignment with readers' needs and improving resource utilization efficiency. Additionally, it holds guiding value for practices in publishing, distribution, and reading promotion within the educational and cultural fields.

2. Previous Research

The popularity of educational and cultural books is a dynamic phenomenon shaped by multifaceted factors, including temporal shifts, spatial contexts, reader interests, and technological advancements. Existing research has explored these dimensions individually, revealing critical insights into how each factor influences book circulation, engagement, and sustained relevance. However, gaps remain in understanding their interconnectedness and the role of renewal behavior as a mediating mechanism—limitations that the present study aims to address.

2.1. Temporal Factors: Context, Technology, and Demand Fluctuations

Temporal dynamics play a pivotal role in determining book popularity, with historical, technological, and cyclical contexts driving shifts in reader engagement. Munteanu (2020) highlights how 17th–18th century European educational books on "Christian civility" gained traction amid religious reforms, as elites promoted standardized behavioral models to align with societal changes. Similarly, van Beek (2009) notes that 1870–1930 German experimental kits and manuals rose in popularity due to educational reforms emphasizing practical experimentation, reflecting how institutional temporal demands shape content relevance.

Technological adoption further amplifies temporal effects: Wynter et al. (2019) observe that online question banks have increasingly supplemented traditional medical textbooks, signaling evolving acceptance of digital resources over time. For cultural content, Bissenbayeva et al. (2024) attribute the popularity of Chinese online novels among adolescents to temporal convenience—on-demand access enabled by internet technology—underscoring how digital platforms compress time barriers.

Cyclical and stress-driven temporal patterns also matter. Sysło (2024) identifies seasonal shifts and curriculum cycles as key mediators of long-term educational book usage, while McLean (2022) finds that societal precariousness in Mexico drives repeated engagement with self-help books, sustaining their popularity. Collectively, these studies confirm that temporal factors—whether historical, technological, or cyclical—are integral to understanding when and why books resonate with readers.

2.2. Spatial Factors: Accessibility, Hubs, and Organizational Dynamics

Spatial contexts, including geographic reach, physical accessibility, and organizational design, significantly influence book popularity. Cross-regional dissemination, as seen in British educational books circulating in Transylvanian private libraries (Munteanu, 2020), demonstrates how spatial market reach expands a book's influence. Closer to home, specific spatial hubs—such as classroom libraries in African schools (Duncan & Kayoro, 2022; Fleisch & Schöer, 2023), and 18th-century rural Scottish private libraries (Towsey, 2023)—concentrate access and foster community-specific reading practices, directly boosting borrowing rates.

Organizational spatial factors also shape engagement. Zhai and Wang (2016) and Mahwasane (2017) show that library classification systems affect accessibility—accurate, standardized arrangements promote renewal behavior, while misclassification reduces borrowing. Technological tools like GIS-based library space models (LSIM) further optimize spatial resource mapping, enhancing circulation (Shen, 2018). Additionally, Moore (2018) highlights how spatial diversity—such as peripatetic teachers using libraries, homes, or gardens—creates varied utilization patterns, emphasizing that "where" readers engage matters as much as "what" they read.

2.3. Interest Factors: Preferences, Identity, and Emotional Resonance

Reader interests, shaped by preferences, identity, and psychological needs, are powerful mediators of book popularity. Ngafeeson and Sun (2015) find that technological innovativeness and system exposure influence e-textbook acceptance, while Mennenga (2016) notes that nursing students' preference for print textbooks reflects how ingrained habits dampen digital adoption. Content alignment with developmental or cultural interests is equally critical: Duncan and Kayoro (2022) link the success of the Inspiring Readers program to age-appropriate books that resonate with local cultural values.

Beyond content, psychological and identity-based factors drive engagement. Brewster (2017) attributes the sustained popularity of crime fiction to its ability to meet needs for narrative closure and predictability, while Menadue and Jacups (2018) observe that genre affinity—such as science fiction fans engaging across books and media—fosters long-term interest. McLean and

Vermeylen (2019) further show that identity performance, such as male self-help readers framing engagement as "advancement," mediates renewal behavior. Even linguistic features matter: Westbury et al. (2015) find that emotional valence in text (e.g., positive words like "celebration") correlates with heightened engagement and renewal.

2.4. Technological and Algorithmic Influences

Digital technologies and algorithms increasingly mediate book popularity by shaping access, visibility, and engagement. Shin and Valente (2020) note that recommendation algorithms (e.g., Amazon's) cluster ideologically similar books, creating echo chambers that amplify certain cultural texts. Conversely, augmented reality (AR) and indoor positioning technologies enhance spatial access, guiding readers to books efficiently and potentially boosting borrowing (Huang et al., 2016).

Algorithms also optimize resource distribution. Su and Zhao (2023) use differential evolution algorithms to balance regional access to educational books, reducing allocation time and improving utilization. For sustained engagement, Žitnik and Gordon Smith (2024) demonstrate that NLP-driven analysis of online book club discussions keeps focus on educational materials, supporting group renewal behavior. These studies highlight technology's dual role as both a facilitator and a filter of popularity.

## 2.5. Limitations of Existing Research and Significance of the Present Study

While prior work illuminates the roles of temporal, spatial, and interest factors—alongside technology—in shaping book popularity, critical gaps persist. First, most studies examine these factors in isolation: few investigate how temporal patterns (e.g., curriculum cycles) interact with spatial contexts (e.g., library accessibility) or reader interests (e.g., genre affinity) to drive popularity. Second, renewal behavior—a key indicator of

sustained engagement—is often treated as a secondary outcome rather than a mediator that connects these factors. For example, while Alao (2002) and Tu et al. (2020) link renewals to engagement, they do not explore how renewals are shaped by the interplay of time, space, and interest. Third, predictive models of popularity, particularly in educational and cultural domains, rarely integrate Bayesian approaches to account for uncertainty and dynamic relationships between variables. This limits understanding of how factors co-evolve to influence book circulation.

This study addresses these gaps. By (1) modeling the joint effects of temporal, spatial, and interest (TSI) factors; (2) positioning renewal behavior as a critical mediator between these factors and popularity; and (3) using Bayesian methods to capture uncertainty and dynamic relationships, the study offers a more holistic framework for predicting popularity. This approach not only advances theoretical understanding of interconnected drivers of book engagement but also provides actionable insights for libraries, educators, and policymakers seeking to optimize resource allocation and promote sustained reader engagement.

3. Data and Method

The sample encompasses data across three distinct time periods in the library of NNU. For the period of 2016–2018, there were 32,760 G Books in the dataset. These books were accessed by 10,824 readers, who collectively generated 102,566 book borrowing transactions. For the period of 2019–2021, the sample included 23,622 G Books. A total of 7,234 readers interacted with these books, resulting in 52,690 book borrowings. For the period of 2022–2024, the sample consisted of 14,710 G Books. With 4,978 readers involved, the number of book borrowings amounted to 28,657.
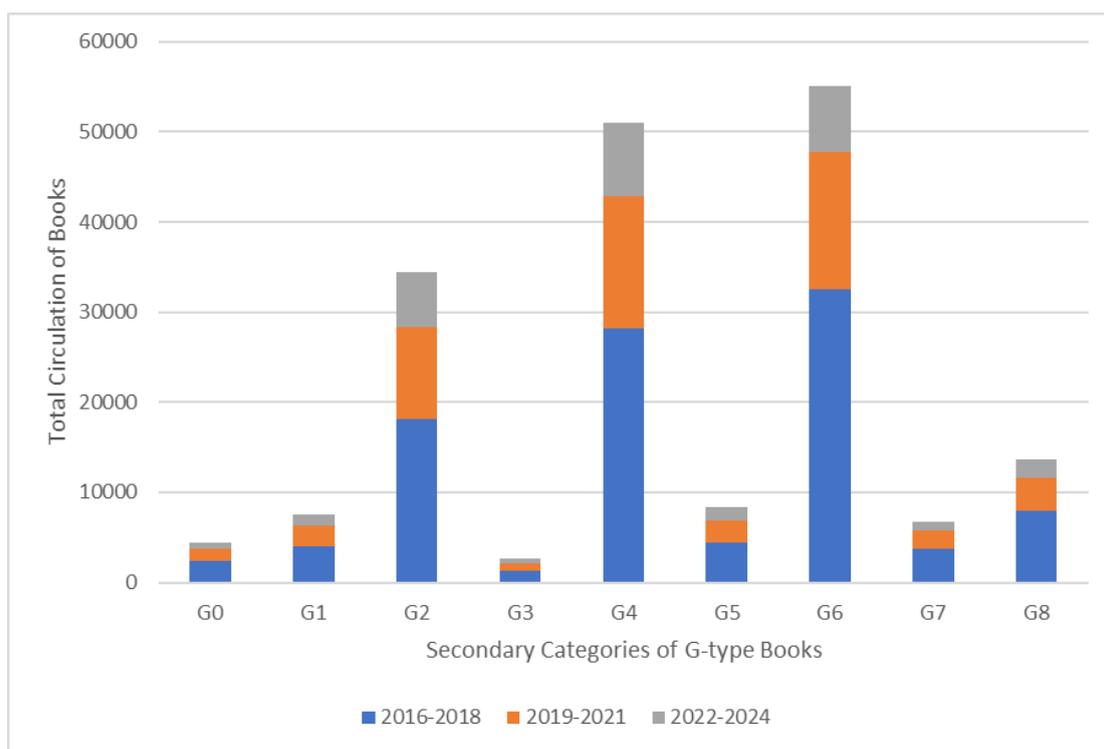
Figure 1. The borrowing volume of books in the secondary categories of G.

Figure 1 presents the total circulation of G - type books across different secondary categories (G0–G8) over three time periods: 2016–2018 (blue), 2019–2021 (orange), and 2022–2024 (gray). G2 and G4, especially G6, show significant circulation. For G6, the circulation grows steadily across the three periods, with the 2016–2018 period already having a high base, and further increases in the subsequent periods. This indicates that these categories, particularly G6, are highly popular among readers over time. G2 pertains to information and knowledge communication, incorporating theories, journalism, broadcasting, publishing, mass culture, librarianship, museology, and archival science, spanning multiple communication fields and institutions. G4 encompasses education, including educational theories, ideological education, teaching theories, educational technology, educational psychology, teacher-student issues, administration, management, and global educational undertakings across all levels (preschool to higher) and types (vocational, adult, etc.), comprehensively covering education. G6 refers to education at all levels: preschool, primary, secondary, higher, and teacher education, involving theories, ideological education, teaching methods, materials, management, and global overviews, presenting the educational system from early childhood to higher education.

G8 has a moderate circulation. Compared to the least popular categories (like G0, G1, G3, G5, G7), its total circulation is relatively higher. Over the three periods, there is a small but noticeable increase. The 2016–2018 period forms a base, with 2019–2021 and 2022–2024 seeing incremental growth, suggesting a gradually rising interest in G8 - type books. G8 covers sports, including theories, global sports undertakings, venues/equipment, technical principles, and diverse sports (track and field, gymnastics, martial arts, etc.) and recreational activities (chess, fishing, etc.), encompassing sports theories and practices.

G0, G1, G3, G5, and G7 have relatively low total circulations. Their bars are short across all three periods, suggesting less reader interest or lower availability compared to the dominant categories. G0 denotes cultural theory, encompassing cultural philosophy, cultural nationality, comparative culturology, cultural interdisciplinary relations, cultural geography, and cultural history, focusing on fundamental theories and interdisciplinary studies in culture. G1 covers global and national cultures and cultural undertakings, including thematic research, cultural industries, exchanges, and history, reflecting cultural development and affairs across scopes. G3 involves science and scientific research, covering theories (e.g., science studies, futurology, methodologies), research practices (management, methods), global scientific undertakings, and information science, focusing on research theories, practices, and global overviews. G5 focuses on global educational undertakings, including policies, reforms, systems, international

organizations, teacher-student status, statistics, history, and overviews of China and global regions (Asia, Africa, Europe, Oceania, Americas), reflecting overall educational development. G7 includes diverse education types: vocational-technical, adult, overseas Chinese, ethnic minority, special, social, family, and self-study education, addressing needs beyond conventional levels for varied groups.

3.1. Book Popularity Prediction
3.1.1 Books Above the 75th Percentile
After opening the data in SPSS, navigate to the menu bar, select "Transform" followed by "Visual Binning" to bring up the initial window. From the left variable list, choose the quantile-categorized variable (Reader), move it to the "Variables to Bin" box on the right using the arrow button, and click "Continue". In the main "Visual Binning" settings window, locate the "Cut Points" area on the left, input the "Number of quantiles" (e.g., "4" for quartiles).

Books with a lending volume exceeding the 75th percentile are selected for thematic analysis: For 2016–2018, books with a lending volume of 6 or higher are included, amounting to 5,089 titles. For 2019–2021, the threshold is set at a lending volume of 4 or above, encompassing 3,964 book titles. For 2022–2024, books with a lending volume of 4 or higher are selected, with a total of 1,686 titles meeting this criterion.

3.1.2 TAN (Tree-Augmented Naive) Bayesian Model
TAN Bayesian networks retain the core advantages of Bayesian networks while overcoming the strict feature conditional independence assumption of naive Bayes. TAN allows for a tree-like dependency structure among features (i.e., each feature, in addition to depending on the class label, depends on at most one other feature). This structure better aligns with the inherent feature correlations in actual data, thereby improving the model's ability to fit complex datasets. TAN's tree structure limits the number of dependencies. This makes parameter learning and inference more efficient.

The TAN Bayesian model is designed to mine patterns from book borrowing data across different periods, quantify book popularity using probabilistic methods, and visually present the differentiation of popularity levels (especially focusing on the 75th percentile as a key threshold), providing a data - driven approach for understanding book borrowing trends and popularity characteristics.

The TAN Bayesian model design for book - related analysis, as depicted in Figure 2, comprises multiple interrelated components. First, it takes the fundamental data source related to book borrowing (encompassed in the G Books Borrowed part, which categorizes borrowing records into three time periods: 2016–2018, 2019–2021, and 2022–2024) as input.
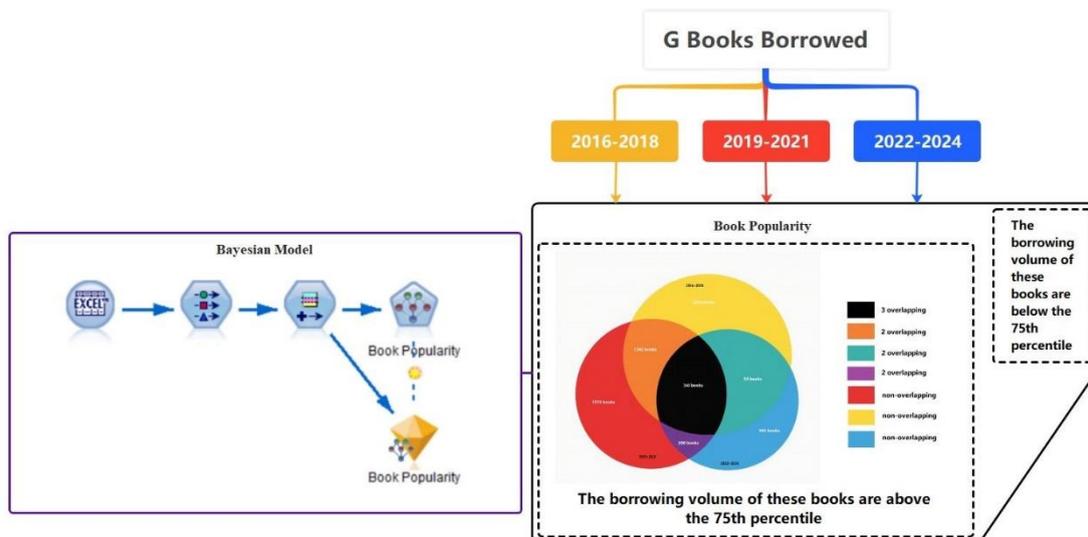
Figure 2. TAN Bayesian model design

The model then processes these data through a series of operations. The data is split into 70% as the training set and 30% as the test set for training and validating the Bayesian model. Through probabilistic reasoning and learning, it aims to analyze "Book Popularity". The analysis of book popularity is further visualized and differentiated via the Venn - diagram - based "Book Popularity" part. This visualization distinguishes books into two key categories: those with borrowing volumes above the 75th percentile (the overlapping and colored regions in the Venn diagram, representing relatively popular books) and those below (the area described as "the borrowing volume of these books are below the 75th percentile").

Books involved in three - period overlaps (360 books) are assigned a value of 1; those in two - period overlaps (1,723 books) are assigned a value of 2; and single - period books (6,213 books) are assigned a value of 3. Books with borrowing volumes below the 75th percentile are assigned a value of 4.

Regarding the college/department classification: the College of Education and Science is coded as 1, the College of Teacher Education as 2, the College of Journalism and Communication as 3, the College of Sports Science as 4, and other colleges or departments as 5.

For user types: undergraduate students are coded as 1, postgraduate students as 2, faculty and staff as 3, and other user types as 4.
For gender: male is coded as 1 and female as 2.

For the G - category classification: G0 is coded as 1, G1 as 2, G2 as 3, G3 as 4, G4 as 5, G5 as 6, G6 as 7, G7 as 8, and G8 as 9.

3.2. The Impact of Time, Space, Interest Factor, Renewal Intention, and Their Interaction on Readers' Reading Behavior
The moderated regression design systematically integrates temporal, spatial, and interest - based dimensions of borrowing data to explore how renewal behavior moderates their influence on total borrowing volume, with rigorous control for confounding user characteristics. The moderated regression model, as visualized in Figure 3, is structured to analyze book borrowing data spanning 2016–2024, segmented into three periods: 2016–2018, 2019–2021, and 2022–2024. The core

dataset integrates multi - dimensional attributes of borrowing behavior:
Time: Captures monthly borrowing distributions across periods.
Space: Includes physical location metrics.
Interest: Reflects borrowing preferences via categories (G0–G8) to represent thematic or genre - based engagement.
User & Transaction Information: Contains Reader ID, Gender, Renewal status, Reader Type, and Total Borrowing Volume.
The model tests whether and how the impact of time - space - interest factors on borrowing volume (LogTotal) are modified by renewal behavior (moderator). By including covariates, it accounts for demographic/user type influences, ensuring the regression isolates the effect of the independent variables (via PCA) on borrowing, as adjusted by renewal and controlled for gender/reader type.

The model incorporates several key components, each serving a distinct role in analyzing the book - borrowing data:
Independent variables are constructed through principal component analysis (PCA) - derived combinations, namely "Time + Space + Interest", "Space + Interest", and "Time + Interest". The maximum variance method is adopted. Research on G category book borrowing behavior involves large-sample data. Compared with the oblique rotation method, the maximum variance method does not require estimating the correlation coefficients between factors, resulting in a more concise calculation process. By enforcing factor independence, the maximum variance method enables the "characteristic variables" of each potential dimension to be more concentrated. By applying PCA, these composites distill complex, multi - dimensional data (encompassing temporal borrowing patterns, spatial borrowing locations, and thematic borrowing interests) into parsimonious, interpretable factors. The factor with the largest eigenvalue (F1) act as the primary predictors in the regression framework, enabling the quantification of their collective influence on borrowing - related outcomes.

The "Renewal" variable functions as the moderator. Coded dichotomously (0 indicating no renewal, 1 indicating renewal), it moderates the associations between the PCA - based independent variables and the dependent variable. Specifically, it scrutinizes the extent to which renewal behavior modulates the effects of temporal, spatial, and interest - based factors on borrowing patterns. In doing so, it illuminates the conditional

nature of these relationships, i.e., how the influence of the independent variables on borrowing outcomes varies depending on whether a renewal occurs.

"Gender" and "Reader Type" (e.g., undergraduate students, faculty and staff) are included as covariates. Their inclusion serves to statistically control for potential confounding effects. By isolating these variables, the model ensures that the estimated relationships between the independent variables (PCA composites) and the dependent variable are not spuriously influenced by demographic or institutional - role - related

differences among borrowers. This enhances the internal validity of the regression analysis.

The dependent variable is operationalized as "LogTotal", which represents the log - transformed Total Borrowing Volume. This transformation addresses potential skewness in the borrowing volume data and linearizes relationships for regression modeling. As the outcome variable, it captures the magnitude of borrowing behavior, allowing the model to assess how effectively the independent variables (and their moderation by renewal) explain variations in the total volume of book borrowings.
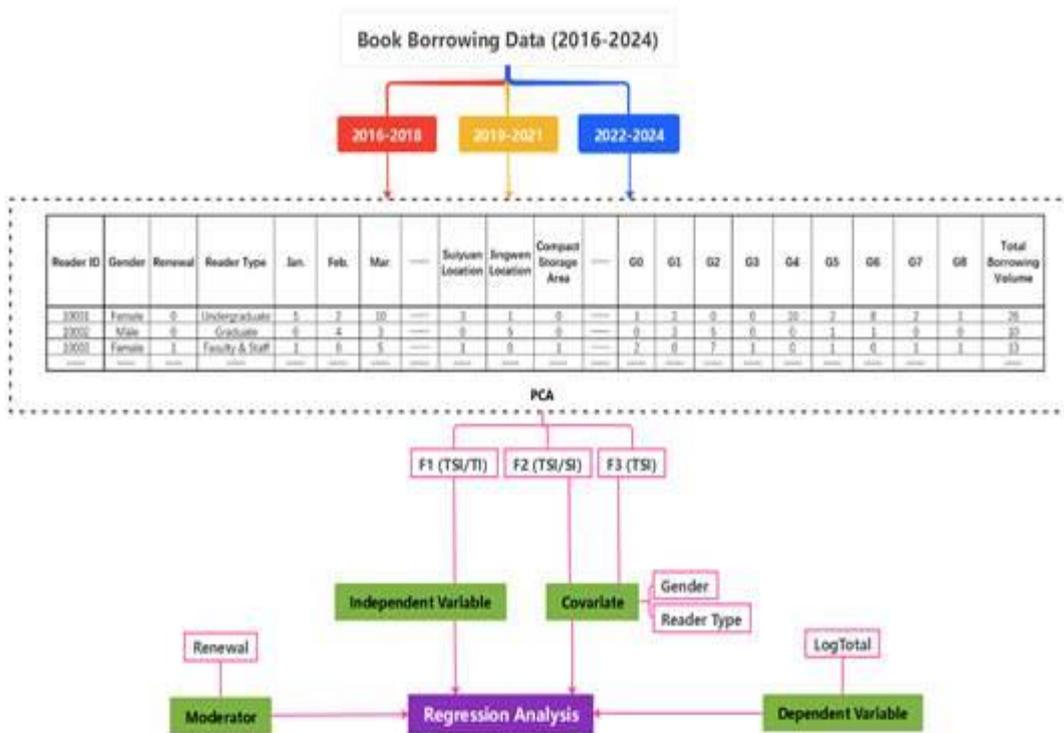

Figure 3. Moderated regression design.

4. Result
4.1 Description of the Characteristics of Popular Educational Books
4.1.1 The Most Popular Theme Phrase of G Books
Figure 4 Venn diagram illustrates the overlap in books above 75th percentile reader borrowing rates across three periods. The

data reveals limited cross-period continuity (small three-period overlap) and varying correlation strengths between different period pairs. A total of 360 books overlap across all three periods, indicating sustained reader attention.
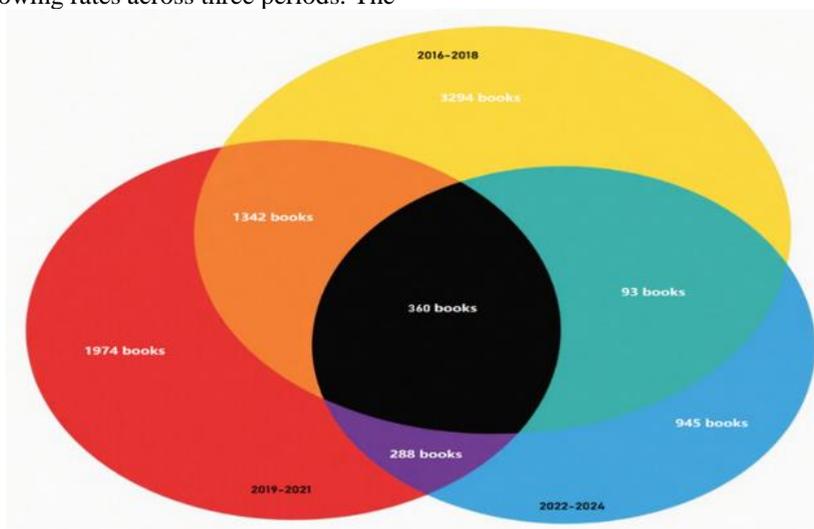
Figure 4. Repeated borrowing of books above the 75th percentile.

The most popular subject phrases, extracted from the Standardized and Controlled Subject Terms within the 6XX fields of CNMARC and UNIMARC for the 360 books in the overlapping zone across the three periods, are as follows:

Educational thought and theory studies: Education-theory-China-ancient times, education-thought-France-modern times, education-theory-Britain-modern times, education-science-research-methods, education-case study-research-methods, early education-theory, affective education-theory, teaching-theory, education-thought-research, education-policy, comparative education-institutions of higher education.

Teaching and curriculum studies: School-education, literary appreciation-middle school-China, Chinese class-classroom-teaching-research-primary and secondary schools, middle school-geography class, teaching-research-primary and secondary schools-teacher training, teaching design, educational history-institutions of higher education-textbooks-foreign countries, aesthetic education-research, ideological and political education, moral education-research, badminton-sports.

Psychology and learning studies: Philosophers-pragmatism-education-thought-United States, children-education-research-united states, learning-theory-psychology, teaching-psychology, analysis-statistics, statistical procedures-application, citation analysis, survey methods.

Cultural and ethnic studies: Culture-theory-research, culture-postmodernism-research, culture-ethnicity-psychology-research-Japan, culture-research-world, research-western-countries, literary and artistic criticism-research-world.

Ancient books and documents studies: Ancient books-collation-bibliography-research-China, research-ancient books-editions, ancient philology, communication and media studies, mass-communication-childhood-sociology-media-research, women-relation-communication-media-research, communication-media-culture-research, China-communication studies-research, communication studies-intellectual history-world, communication studies-history.

Journalism and publishing undertakings: Newspapers-journalism-undertaking-history-research, journalism-undertaking-history-institutions of higher education-world, journalism-intellectual history-China-modern times, journalism-work-cultural history-China, journalism-work-research, internet-public opinion-research, publishing-work.

Other historical studies: World War I-1914–1918, history-research-China, late Qing dynasty.

4.1.2 The Most Popular Title Topic of G Books

The most popular title topics of G Books are extracted using the LAS (Latent Semantic Analysis) model from the books at or above the 75th percentile of circulation. In comparison with the analysis of Popular Theme Phrases presented above, the range of terms selected from titles is broader and more flexible, without being controlled.

This LSA model encompasses 10 topics. The total explained variance of the model is 8.36%, and the average similarity between topics is 10.47%. The most similar pair of topics is Topic 4 and 5, with a similarity of 25.00%. Figure 5 presents the connections and clustering situations among different topics, helping to intuitively understand the similarity and independence relationships between topics.

Blue Cluster (Topics 1, 4, 5): Topics 4 and 5 have a relatively high similarity of 25.00%, the highest among all topic pairs. This indicates a strong semantic connection between them. Topic 1 is also part of this cluster, linked to both 4 and 5, implying that Topic 1 shares some semantic overlap with these two, though likely to a lesser extent than the 4 and 5 pair.

Red Cluster (Topics 0, 3, 6, 7): These topics form a connected cluster. While the exact pairwise similarities aren't all specified, their connection in the network shows that there is semantic relatedness among them. They likely share common semantic features that make them group together.

Orange Cluster (Topics 8, 9): Topics 8 and 9 are connected, indicating a semantic link. Their similarity level, though not given as a specific number like the 4 and 5 pair, is above the minimum similarity threshold (15%) to be visualized as connected.

Isolated Topic (Topic 2): Topic 2 is green and isolated, meaning it has no detected similarity above the 15% threshold with other topics. This shows a high level of semantic independence, having distinct content compared to the rest.
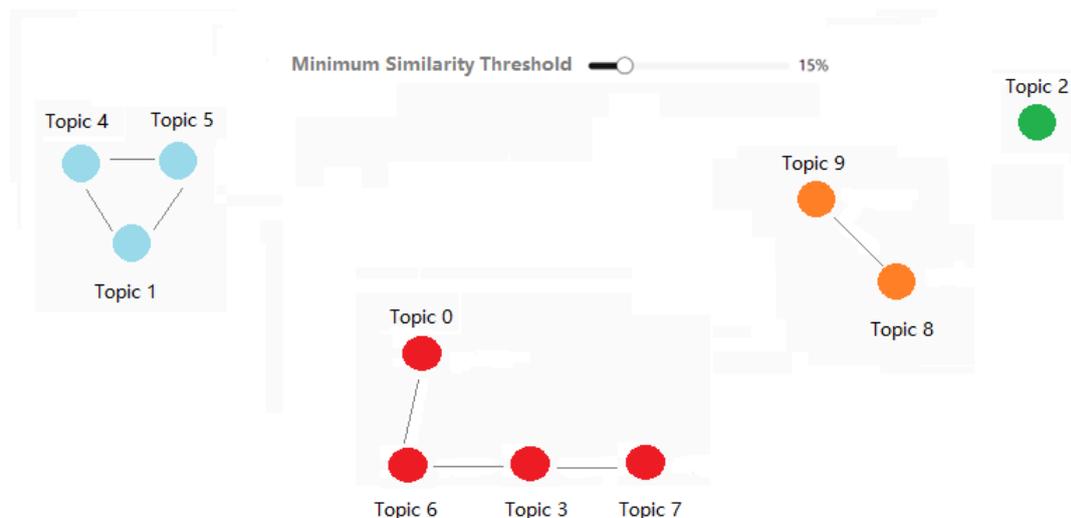
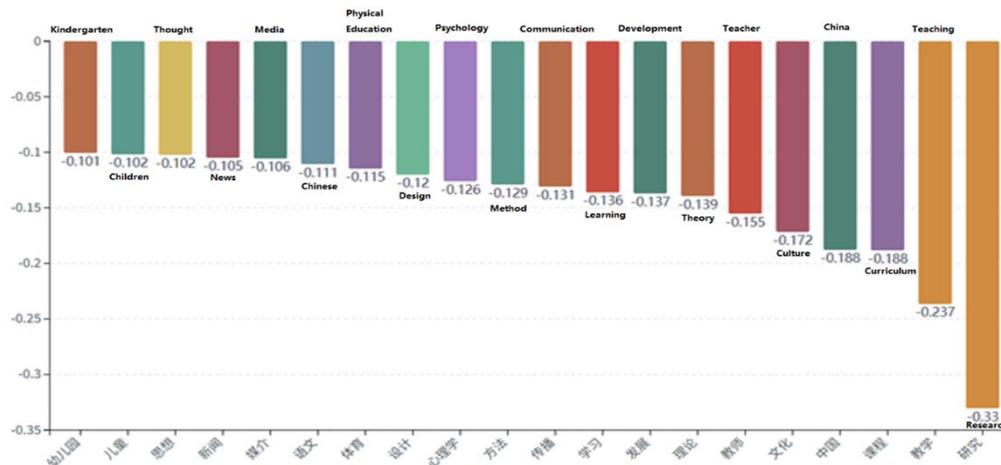**Figure 5.** Topic relationship network.

The LSA identified 10 distinct topics (in Table 1), with a total of 106 unique keywords distributed across these topics. The top 20 keywords for each topic are shown in Figure 6. Details of the topics are as follows:

Table 1. The title topic of G books

| Topic	Interpretation
Topic 0	This topic primarily focuses on content related to research, teaching, and curriculum.
Topic 1	This topic mainly centers on content associated with curriculum, kindergarten, and principles.
Topic 2	This topic is predominantly concerned with content related to design, teaching, and cases.
Topic 3	This topic mainly focuses on content associated with teachers, professionalism, and development.
Topic 4	This topic primarily centers on content related to China, kindergarten, and activities.
Topic 5	This topic is mainly concerned with content associated with methods, theories, and research.
Topic 6	This topic primarily focuses on content related to knowledge, psychology, and China.
Topic 7	This topic mainly centers on content associated with teaching materials, research, and special purposes.
Topic 8	This topic is predominantly concerned with content related to science, higher education, and knowledge.
Topic 9	This topic mainly focuses on content associated with knowledge, subject specialties, and learning.

Topic 0
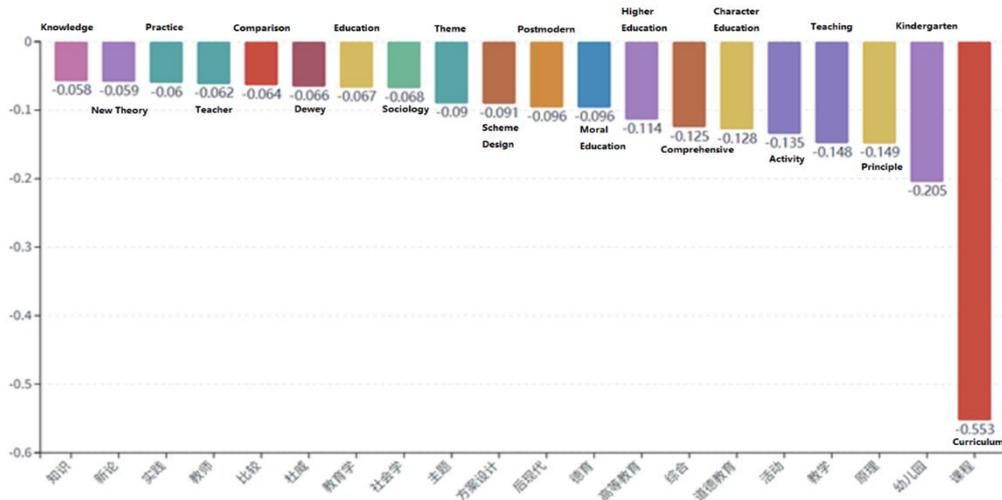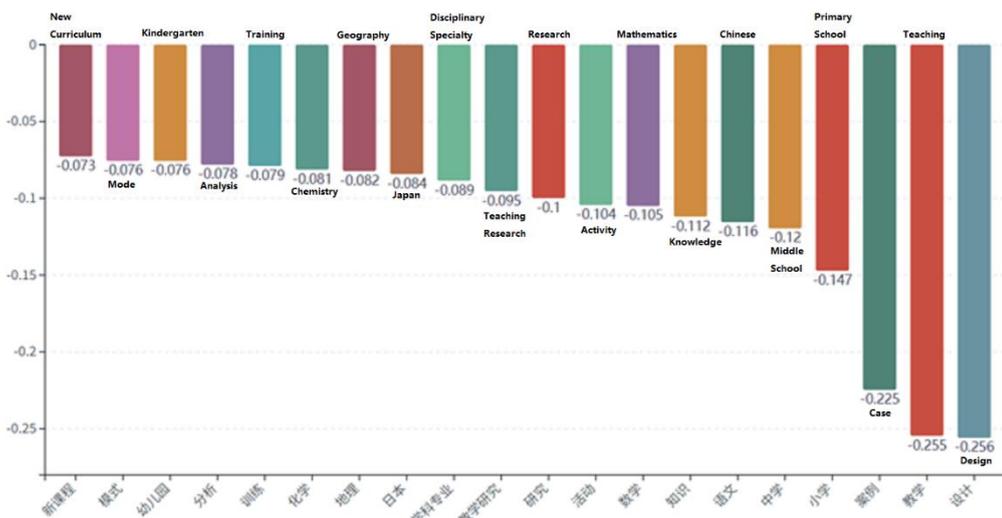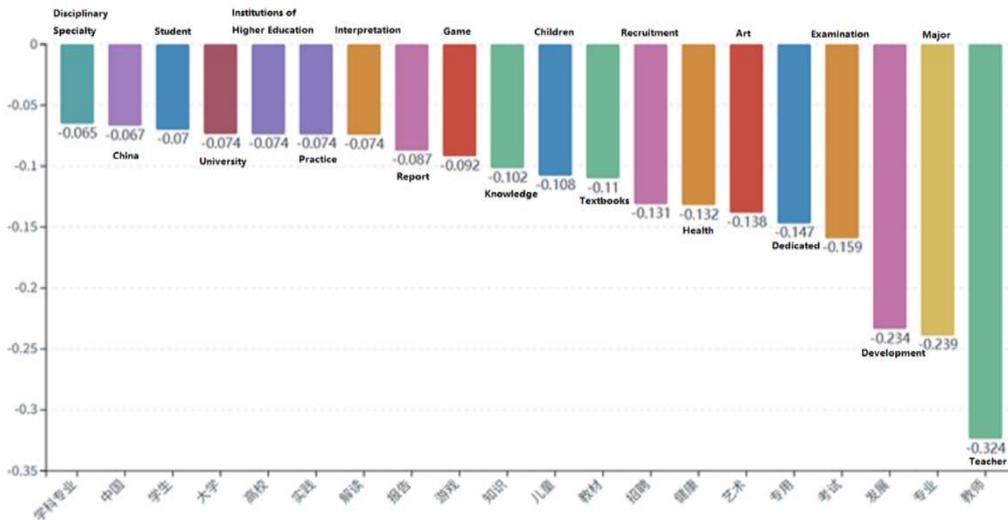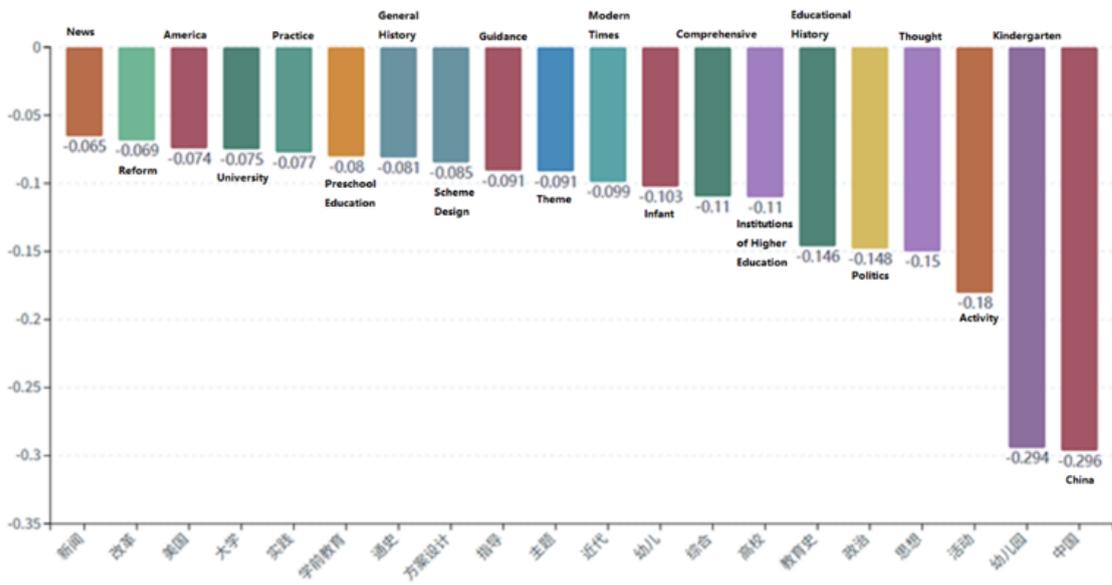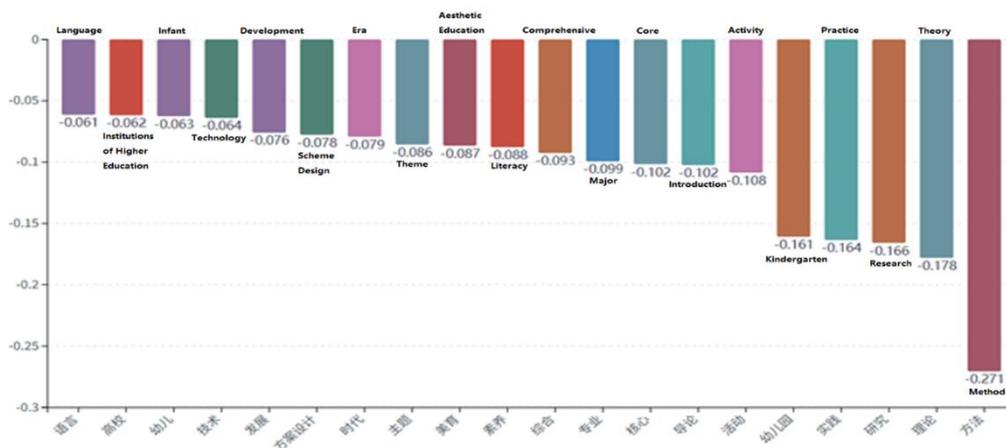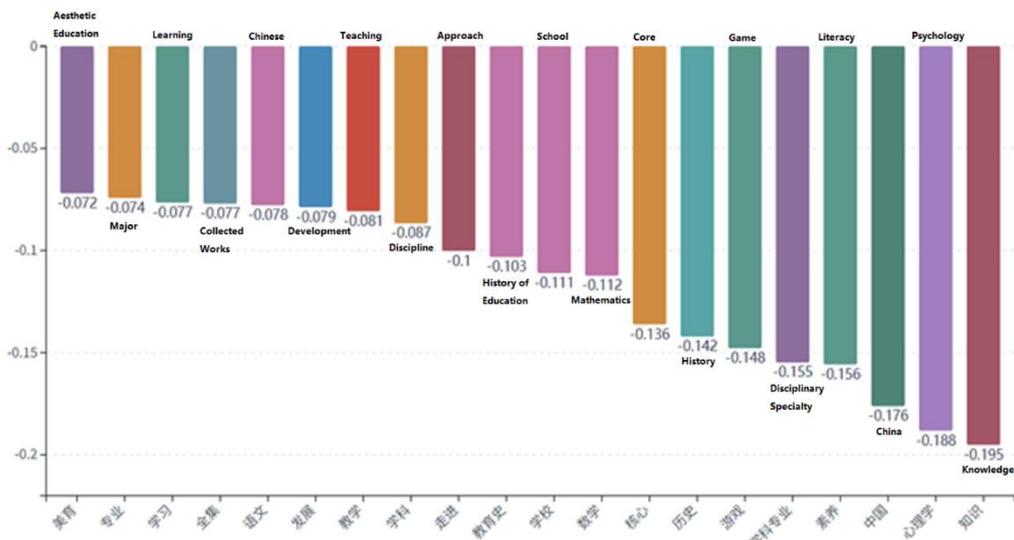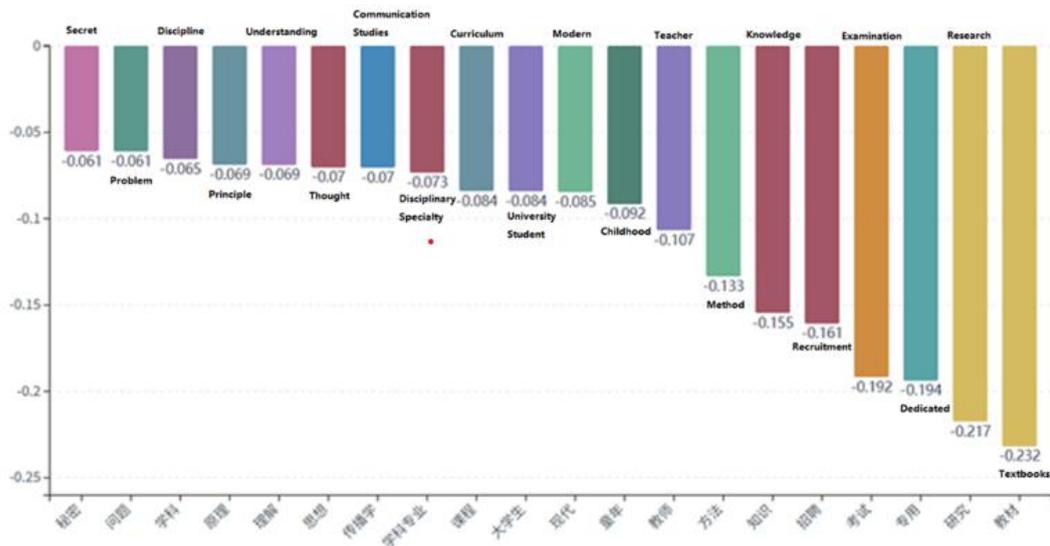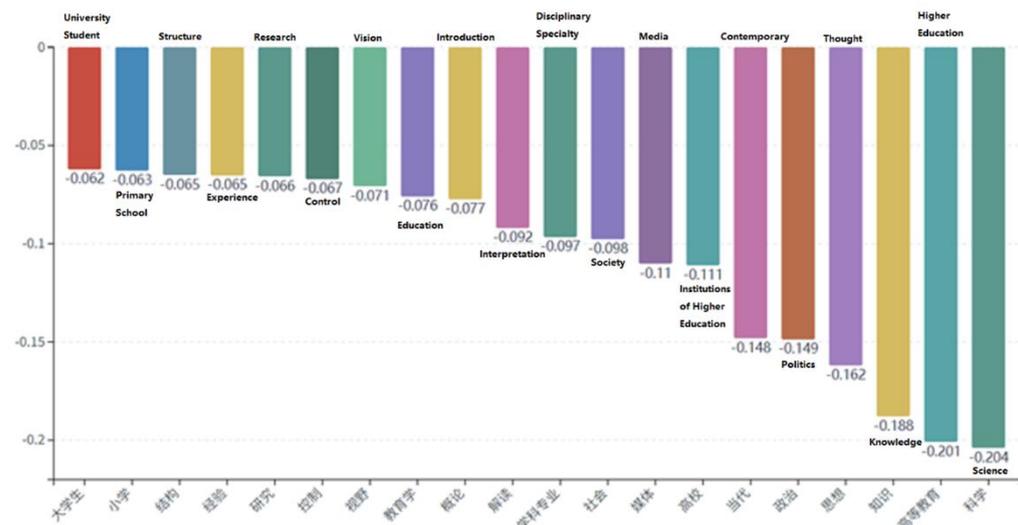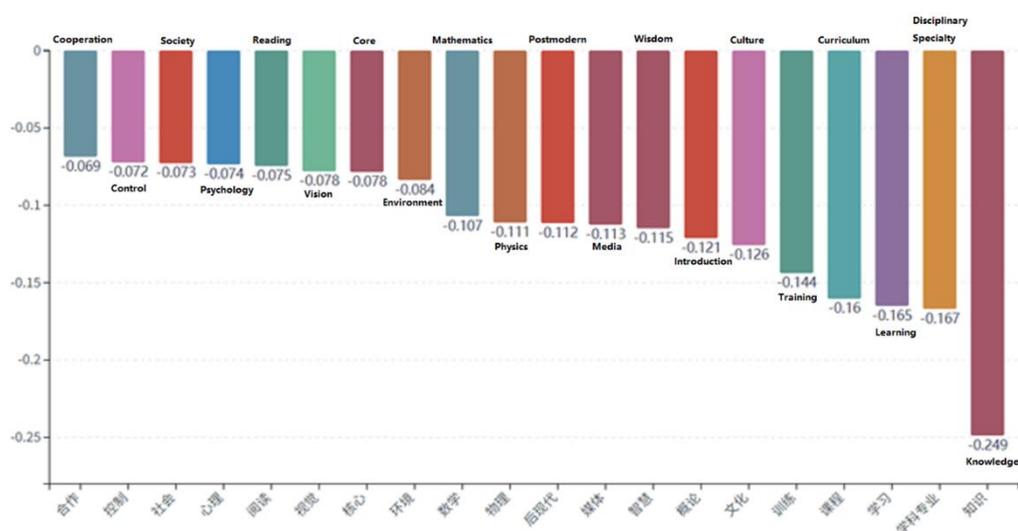
Figure 6. The top 20 words of the topic

4.2 The Conditional Relationships in Bayesian Networks

The Bayesian network in Figure 7 centers on predicting Book Popularity. Six blue nodes (e.g., Reader Gender, Renewal Behavior) act as predictive variables. They cover three dimensions:

Reader attributes: Reader Gender, Reader Type, Department (describing "who reads").

Behavior - time: Renewal Behavior, Month (reflecting "how and when to read").

Book content: G Subcategory (representing "what content is read").

Book Popularity points to all blue nodes, implying that high - popularity books can reshape variables like reader behavior and attributes (e.g., popular books may attract specific - gender readers or increase renewal rates).

Predictors interact to influence Book Popularity through multi - step paths: Gender - specific reading habits affect renewal behavior, which then impacts popularity; Reader identities, like student/teacher, link to departmental reading preferences, further shaping popularity; Seasonal cycles, such as semester/vacation, adjust demand for book subcategories, influencing popularity.

Renewal Behavior, Reader Type (darker colors) have stronger direct explanatory power for Book Popularity (e.g., renewal behavior directly signals reader demand intensity). Reader Gender, Month (lighter colors) have weaker direct impacts but still matter via synergistic effects.

The content attributes inherent to book subcategories serve as foundational variables that underpin the determination of popularity. Notably, high-popularity books exert a reverse reinforcing effect on their content labels, thereby shaping the collective perception of popularity within their respective subcategories—a phenomenon that embodies the role of popularity in reshaping content classification systems.

Furthermore, within the Bayesian predictive framework, readers' affiliated departments (Department) function as a primary determinant of their demand for specific subcategories, while reader types (e.g., students, faculty) act as moderating factors that amplify such demand. This hierarchical interplay of variables ultimately feeds into the prediction of book popularity, as captured by the probabilistic dependencies modeled in the Bayesian network.
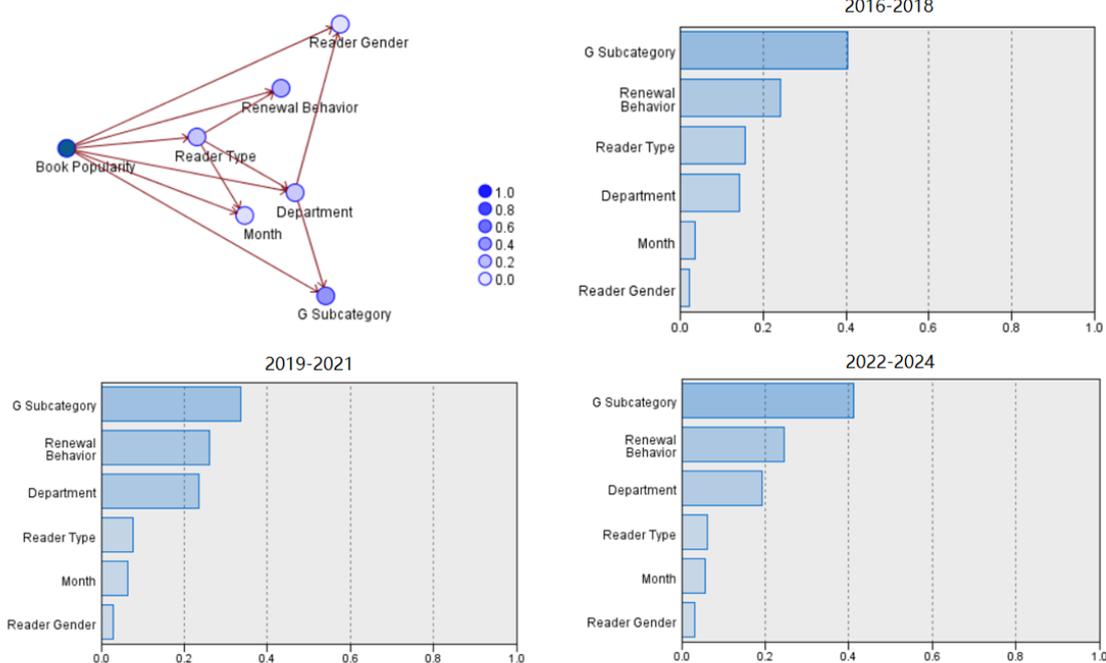
Figure 7. TAN Bayesian network.

Among the three core variables and importance ranking picture in Figure 7, "G Subcategory" (book sub - category) has always been the most important predictive variable (with a proportion of nearly 0.4), and "Reader Gender" (reader gender) has always been the least important (with a proportion close to 0). This indicates that:

Content attributes of books are fundamental: The sub - category of a book itself has the most stable and crucial impact on "popularity". Different types of book content inherently determine the breadth of the audience and the potential for dissemination (for example, popular social science books are more likely to gain popularity, while niche academic books have difficulty in achieving popularity).

The impact of reader gender is extremely weak: The direct explanatory power of gender on book popularity is extremely low. This may be because the gender differences in reading needs are covered by variables such as "sub - category" and "reader type" (for example, the preference of female readers for romance novels has been reflected by "G Subcategory").

By comparing the three time periods of 2016–2018, 2019–2021, and 2022–2024, there are slight fluctuations in the importance of core variables. "Renewal Behavior" (renewal behavior): The proportion in 2019–2021 (Approximately 0.25) is higher than that in other periods (Approximately 0.2). This may be because the correlation between readers' renewal behavior and popularity was closer during this period (for example, changes

in borrowing habits during the pandemic, and books with more renewals were more likely to be disseminated).

"Department" (affiliated department / college): The proportion in 2022–2024 (Approximately 0.2) is slightly higher than that in the previous periods (Approximately 0.15), suggesting that the trend of "specialized segmented reading" has strengthened in recent years. The differences in the demand for professional books among different colleges and departments have a more significant impact on popularity.

"Month" (month): The importance has always been low (except for gender), meaning monthly variations fail to significantly explain or predict the book popularity. Education and culture (G-category) books lack a strong tie to monthly cycles.

4.3 The Quantitative Relationship Between Various Factors and Borrowing Volume
4.3.1 The Extraction of Common Factors
Table 1 presents extracted factors (F1, F2, F3) across three time periods (2016–2018, 2019–2021, 2022–2024), with loadings indicating the strength of association between variables (e.g., lending rooms, book categories, months) and each factor. Notably, the number of factors evolved from singularity to diversification. A key development within this trend is the emergence of the TI (Time-Interest) factor, which supplanted the TSI (Time-Space-Interest) factor as the primary F1 factor with the highest eigenvalue. During the 2022–2024 period, the Time-Interest dimension underlying the lending of educational books assumed a particularly prominent role.

Table 1. The extracted factors.

Factor	2016-2018	2019-2021	2022-2024
F1	**TSI Factor:** Suiyuan Chinese Book Lending Room (0.842), G4 (0.767), September (0.632), October (0.606), December (0.602), May (0.600), G5 (0.598), April (0.594), June (0.587), March (0.574), Suiyuan Chinese Sample Book Lending Room (0.560), November (0.545), January (0.508)	**TSI Factor:** Suiyuan Chinese Book Lending Room (0.729), November (0.694), December (0.690), September (0.663), G4 (0.646)	**TI Factor:** G5 (0.758), G4 (0.674), February (0.603), April (0.519)
F2	-	**SI Factor:** Social Science Library (0.793), G1 (0.707), G0 (0.706)	**TSI Factor:** Teacher Education Professional Lending Room (0.947), G6 (0.886), June (0.500)
F3	-	-	**TSI Factor:** G2 (0.840), Suiyuan Chinese Book Lending Room (0.726), November (0.514)

In 2016–2018 period, education-related books (G4 and G5) are strongly associated with Factor 1, with high loadings (G4: 0.767; G5: 0.598). This factor is dominated by the Suiyuan Chinese Book Lending Room (0.842), indicating that education books were primarily borrowed from this core lending space. The factor also includes multiple months (September, October, December, May, April, June, March), suggesting that education book lending showed obvious seasonal patterns during this period, possibly linked to academic schedules (e.g., start/end of semesters, exam periods).

In 2019–2021 period, education-related books (G4) remain in Factor 1, with a slightly reduced but still high loading (0.646). The Suiyuan Chinese Book Lending Room (0.729) remains the primary borrowing location, consistent with the previous period. However, the factor includes fewer months (only November, December, September), indicating weakened seasonal fluctuations in education book lending. This may reflect changes in academic rhythms or reading habits (e.g., more dispersed study schedules). Moreover, Social science-related books (G0 and G1) are the main focus of Factor 2, with high loadings (G1:

0.707; G0: 0.706). This factor is dominated by the Social Science Library (0.793), indicating that social science books were primarily borrowed from this specialized space. The factor includes few months with low loadings (November: 0.144, September: 0.091, December: 0.019), indicating minimal seasonal fluctuations in social science book lending. This may reflect steady demand for social science resources independent of academic schedules.

In 2022–2024 period, education-related books show a more fragmented distribution across three factors, reflecting diversified borrowing behaviors. Factor 1 still includes G4 (0.674) and G5 (0.758), with loadings higher than in 2019–2021, indicating sustained demand for general education books. Months like February and April are associated with this factor, suggesting a partial return to seasonal patterns. Factor 2 introduces G6 (0.886) and links it to the Teacher Education Professional Lending Room (0.947), highlighting a surge in demand for specialized teacher education materials, which are now concentrated in a dedicated professional collection space. Factor 3 features G2 (0.840), a category likely related to

educational technology or applied education, with the Suiyuan Chinese Book Lending Room (0.726) still playing a role in its circulation, alongside November (0.514), implying seasonal demand for applied education resources.

4.3.2 Factors Affecting Readers' Borrowing Volume

Across the three time periods, the core variables including the dependent variable (Y: LogTotal) and moderator (W: Renewal) maintain a consistent framework. However, notable shifts are observed in the independent variable, composition of covariates, and sample size. Regarding the independent variable (X), there is a structural change in F1, evolving from TSI (Time-Space-Interest) in the 2016–2018 and 2019–2021 models to TI (Time-Interest) in the 2022–2024 model. In terms of covariates, the number increases incrementally over time, rising from 3 (Reader, Depa, Gender) in the 2016–2018 model to 5 in the 2022–2024 model, with the inclusion of additional factors (F2, F3) in the latter periods. Concurrently, the sample size demonstrates a declining trend, decreasing from 10,824 in 2016–2018 to 7,234 in 2019–2021 and further to 4,978 in 2022–2024.

Table 2. Regression coefficient (2016-2018).

	Coeff	se	t	p	LLCI	ULCI
Constant	0.6239	0.0201	31.0952	0.0000	0.5846	0.6632
F1 (TSI)	0.5935	0.0081	73.6374	0.0000	0.5777	0.6093
Renewal	0.3674	0.0083	44.5260	0.0000	0.3512	0.3836
Int_1	-0.4193	0.0086	-48.5842	0.0000	-0.4362	-0.4024
Reader Type	0.0387	0.0057	6.7988	0.0000	0.0275	0.0498
Department	-0.0110	0.0021	-5.1491	0.0000	-0.0152	-0.0068
Gender	-0.0137	0.0080	-1.7235	0.0848	-0.0293	0.0019

Table 3. Regression coefficient (2019-2021).

	Coeff	se	t	p	LLCI	ULCI
Constant	0.6508	0.0278	23.4273	0.0000	0.5963	0.7053
F1 (TSI)	0.2835	0.0076	37.4254	0.0000	0.2687	0.2984
Renewal	0.3970	0.0119	33.3357	0.0000	0.3737	0.4204
Int_1	-0.1480	0.0093	-15.9456	0.0000	-0.1662	-0.1298
F2 (SI)	0.0380	0.0042	8.9885	0.0000	0.0297	0.0463
Reader Type	0.0436	0.0079	5.5388	0.0000	0.0282	0.0590
Department	-0.0561	0.0027	-21.1100	0.0000	-0.0613	-0.0509
Gender	-0.0140	0.0106	-1.3293	0.1838	-0.0347	0.0067

Table 4. Regression coefficient (2022-2024).

	Coeff	se	t	p	LLCI	ULCI
Constant	0.4359	0.0281	15.5299	0.0000	0.3809	0.4910
F1 (TI)	0.3702	0.0112	32.9148	0.0000	0.3482	0.3923
Renewal	0.3121	0.0107	29.0585	0.0000	0.2911	0.3332
Int_1	-0.2750	0.0121	-22.7740	0.0000	-0.2987	-0.2514
F2 (TSI)	0.1251	0.0043	28.9800	0.0000	0.1167	0.1336
F3 (TSI)	0.1294	0.0044	29.6430	0.0000	0.1208	0.1380
Reader Type	0.0670	0.0082	8.1977	0.0000	0.0510	0.0830
Depart	-0.0195	0.0026	-7.3871	0.0000	-0.0247	-0.0143
Gender	-0.0301	0.0105	-2.8626	0.0042	-0.0507	-0.0095

Tables 2–4 present the regression coefficients of the variables. These tables offer detailed insights into the magnitude and significance of the relationships between the variables. The 2016–2018 model exhibits the strongest explanatory power for the dependent variable (R^2=59.48%), which declines significantly in 2019–2021 (R^2=45.10%) and partially recovers in 2022–2024 (R^2=55.13%), possibly due to the inclusion of additional covariates. All models are statistically significant ($p<0.0001$) as indicated by their F-values, confirming the overall validity of the model specifications.

The interaction term (F1×Renewal) demonstrates temporal variations in its contribution to the model's explanatory power. It contributes most significantly in the 2016–2018 period (R^2 Change=8.84%), followed by the 2022–2024 period (R^2 Change=4.68%), and least in the 2019–2021 period (R^2 Change=1.93%). This pattern highlights that the moderating effect of Renewal on the relationship between F1 and LogTotal fluctuates over time, with the strongest influence observed in the earliest period and a moderate resurgence in the latest period.

F1 exerts a significant positive impact on LogTotal across all periods ($p<0.0001$), with fluctuating strength: peaking in 2016–2018 (0.5935), declining by nearly half in 2019–2021 (0.2835), recovering to 0.3702 in 2022–2024, reflecting phase-specific changes in the factor's influence.

Renewal consistently shows a significant positive effect ($p<0.0001$), with the strongest impact in 2019–2021 (0.3970) and the weakest in 2022–2024 (0.3121), suggesting a gradual weakening of its independent role over time. The negative and significant interaction coefficients indicate that **Renewal consistently exerts a negative moderating effect on the relationship between F1 and LogTotal** (i.e., higher Renewal weakens the positive effect of F1). The strength of moderation

follows the order: 2016–2018 (-0.4193) > 2022–2024 (-0.2750) > 2019–2021 (-0.1480). When Renewal=0 (low moderator level), the positive effect of F1 on LogTotal peaks in all periods, with the 2016–2018 effect (0.5935) significantly higher than in other periods. When Renewal=1 (high moderator level), the effect of F1 weakens significantly, with the smallest effect observed in 2022–2024 (0.0952), indicating the strongest "weakening effect" of Renewal during this period. All conditional effects are statistically significant ($p<0.0001$), confirming the stability of the moderating relationship.

The effects of covariates exhibit distinct patterns across the three periods. For the Reader variable, its positive effect strengthens over time, with the largest impact observed in the 2022–2024 period, which indicates an increasingly promotional role in explaining LogTotal. In contrast, the Gender variable only shows a significant negative effect in 2022–2024 ($p=0.0042$), while its effects in other periods are non-significant, a trend that may reflect changes in gender characteristics within the sample. Department variable consistently exhibit negative and significant effects across all periods, with the strongest impact noted in 2019–2021 (-0.0561), suggesting a more pronounced inhibitory role during this period. Additionally, the added factors (F2, F3) all demonstrate significant positive effects after their inclusion in the models. Specifically, F2 and F3 in the 2022–2024 period show strong effects (0.1251 and 0.1294, respectively), highlighting their non-negligible explanatory power for LogTotal.

5. Discussion

5.1 Bayesian networks offer a robust framework for predicting G-category book popularity by capturing complex conditional relationships

Bayesian rationality frameworks, which align with probabilistic modeling of book selection and integrate temporal/spatial factors (Oaksford & Chater, 2009), lay a solid foundation for this framework. It enables the integration of various factors that affect the popularity of G-category books in a probabilistic manner, considering both the changes in reader preferences over time and the differences in reading habits across regions.

Content attributes (G Subcategory) as the most stable predictor (with a contribution of nearly 0.4) underscores that the intrinsic thematic and disciplinary nature of educational books fundamentally shapes their popularity. Bayesian models effectively quantify how reader attributes (e.g., department affiliation), behavior-time variables (e.g., renewal behavior), and content characteristics interact hierarchically—for instance, how departmental specialization amplifies demand for specific subcategories, or how renewal behavior signals demand intensity.

This framework also captures temporal cascades in user behavior (e.g., early ratings shaping subsequent popularity) (Salganik & Watts, 2009), a factor critical to long-term predictions of G-category book popularity. Early ratings of G-category books can sway subsequent readers' choices, and Bayesian networks excel at modeling this time-dependent chain reaction. Furthermore, big data analysis (e.g., circulation, online access) and standardized metadata enhance the accuracy of G-category book popularity predictions, with necessary attention to privacy considerations (Travis & Ramirez, 2020; Burke et al., 2022). Analyzing large volumes of circulation data and online access behavior related to G-category books yields more comprehensive insights, while standardized metadata ensures the reliability of data used in the model.

This framework not only identifies direct drivers (e.g., renewal behavior's strong direct effect) but also reveals indirect pathways (e.g., gender's weak direct impact, mediated by subcategory preferences), making it superior for modeling the multi-layered dynamics of popularity. Bayesian networks can handle these complex conditional relationships, organically combining various factors to form a comprehensive and accurate prediction model, thus providing reliable methodological support for predicting the popularity of G-category books.

5.2 Content attributes (G Subcategory) act as a foundational anchor for popularity, with subcategory-specific differences in audience breadth and dissemination potential

The consistent dominance of G Subcategory in predictive importance highlights that educational books' popularity is not arbitrary but rooted in their disciplinary and thematic scope. Social science-related subcategories (e.g., G0, G1) exhibit broader appeal due to their relevance to general educational discourse, while niche academic subcategories may attract smaller, more specialized audiences. This anchor effect explains why even across shifting periods, certain subcategories (e.g., G4, G5) maintain sustained relevance—their content addresses enduring educational needs (e.g., teaching theory, curriculum design), whereas others (e.g., G6 in 2022–2024) gain prominence as specialized demands (e.g., teacher education) emerge.

A survey of Chinese academic libraries demonstrates that user satisfaction with books is closely linked to library classification systems (Zhai & Wang, 2016). For G-category books, inaccurate classification, such as issues with descriptive cataloging and subject indexing, reduces readers' efficiency in finding and accessing them, thereby affecting their borrowing rates and popularity. This indicates that classification systems, as a form of spatial organizational factor, mediate the popularity of G-category books. Research on collection management in foundation phase school libraries also finds that disorganized arrangement of educational books hinders access for teachers and students, leading to low utilization rates (Mahwasane, 2017). Conversely, standardized classification and arrangement of G-category books improve their accessibility and promote renewal behaviors such as repeated borrowing, which in turn enhances their popularity. Additionally, automated classification technologies (e.g., generative language models) furnish structured data for analyzing G-category book popularity trends (Luo et al., 2025; Qingkui et al., 2018). These technologies aid in classifying and annotating G-category book content, transforming unstructured text into structured data to streamline the analysis and modeling process.

For G-category books, which cover a wide range of educational topics, well-structured and standardized metadata makes it easier for readers to find relevant materials. This directly affects how readers encounter and engage with educational books, as enhanced discoverability increases the likelihood of these books being accessed and utilized. Thus, the quality of metadata mediates the popularity of G-category books, as higher-quality metadata can lead to greater visibility and engagement. CNMARC (Chinese Machine-Readable Cataloging) and UNIMARC (Universal Machine-Readable Cataloging) are two important metadata formats that play significant roles in the management and dissemination of bibliographic information. CNMARC ensures the accurate and consistent description of Chinese-language materials. It facilitates efficient organization and retrieval of these resources within Chinese libraries and information systems, enabling readers to quickly locate relevant educational materials that meet their specific needs. This standardized format enhances the accessibility of G-category books, thereby contributing to their potential popularity by reducing the barriers to discovery. UNIMARC as a universal standard, promotes international exchange and sharing of bibliographic data. For G-category books with cross-border relevance or those intended for an international readership, UNIMARC ensures that their metadata is understandable and usable across different countries and information systems. This global interoperability expands the reach of G-category books beyond local boundaries, increasing their exposure to a wider audience and potentially boosting their popularity on an international scale. Both CNMARC and UNIMARC, through their standardized structures, contribute to the quality and consistency of metadata, which in turn supports the effective prediction and understanding of G-category book popularity by providing reliable and comparable data for analysis.

5.3 The evolution from TSI to TI factors reflects a shifting focus in reader behavior: from spatial constraints to temporal-interest alignment

The structural shift in F1 from TSI (Time-Space-Interest) in 2016–2018 to TI (Time-Interest) in 2022–2024 indicates a diminishing role of physical lending spaces (e.g., Suiyuan Chinese Book Lending Room) and a growing emphasis on how

temporal patterns (e.g., academic cycles) align with reader interests. This transition likely stems from increased digital access and decentralized reading habits, reducing reliance on specific libraries. By 2022–2024, the TI factor's dominance underscores that when and what readers want to engage with (e.g., seasonal demand for teacher education materials in June) has become more critical than where they borrow, reflecting a more fluid and interest-driven consumption model.

For instance, the transnational circulation of educational texts—such as Benjamin Franklin's life story adapted from English to Hindi—has been shown to be jointly shaped by spatial dissemination and temporal colonial-era educational demands, with these factors interacting to influence both popularity and reinterpretation (Gandotra, 2023). Similarly, Chinese online novels' cross-regional spread highlights how spatial borderlessness and temporal convenience interact with content attributes like value influence to determine their reach (Bissenbayeva et al., 2024). In F1's context, the declining emphasis on "Space" mirrors this shift toward prioritizing accessibility across geographic boundaries, as digital access reduces reliance on fixed physical locations.

Temporal dynamics have long played a key role in content engagement, as seen in 19th-century British university novels, which merged traditional education with emerging visual technologies to reflect temporal shifts in pedagogy—appealing to readers' interest in reconciling tradition and modernity through the interaction of temporal context and content relevance (Bunzel, 2024). This resonates with F1's focus on temporal patterns like academic cycles, where seasonal demand (e.g., for teacher education materials in June) underscores how "when" aligns with "what" readers seek. In F1's 2022–2024 phase, the dominance of TI (Time-Interest) similarly reflects an interest-driven model, where reader engagement is less tied to specific spaces and more to the alignment of content with their needs and timelines.

Overall, F1's transition to TI dominance likely stems from increased digital access and decentralized reading habits, reducing reliance on specific libraries. This shift underscores that when and what readers want to engage with has become more critical than where they borrow, reflecting a more fluid, interest-driven consumption model—one that echoes the dynamic interplay of time, space, and interest observed in diverse content circulation contexts globally.

5.4 Renewal behavior exerts a dynamic moderating effect, with its influence peaking during periods of disrupted reading habits

The fluctuating strength of renewal behavior's moderation—strongest in 2016–2018 (R² Change=8.84%), weakest in 2019–2021 (R² Change=1.93%), and partially recovered in 2022–2024 (R² Change=4.68%)—suggests it is sensitive to external conditions. This pattern aligns with broader research on renewal dynamics, which highlights how such behavior is shaped by a mix of user needs, institutional policies, and spatial constraints. For instance, studies in medical libraries have shown that overdue books—often preceding renewals—stem from unmet needs or forgetfulness, with "providing for renewals" being the most favored measure, indicating that renewal reflects sustained engagement with resources (Alao, 2002). This resonates with the observed fluctuations in F1's context: periods of stronger renewal moderation (e.g., 2016–2018) may correspond to heightened unmet demand, where repeated use signals genuine

interest, while weaker moderation (2019–2021) could reflect disruptions in such engagement, possibly exacerbated by external factors like the pandemic.

Institutional practices also play a role: renewal and cancellation of e-resources, based on usage evaluation, mediate resource availability and popularity (Patra, 2017). This mirrors the nuanced relationship between renewal and borrowing volume in F1, where renewal behavior is not merely a passive indicator but an active institutional response that shapes access patterns. Similarly, spatial constraints—such as portability needs in veterinary education—affect willingness to reuse resources, thereby mediating renewal behavior (Swarts & VanNorman, 2008). In F1's framework, the diminishing role of physical spaces (as noted in prior analysis) may partially explain the recovery of renewal's moderation in 2022–2024, as reduced reliance on fixed locations could alter how users decide to renew or reuse materials.

Additionally, spatial factors like faculty development centers have been found to indirectly influence repeated borrowing of related educational books (Hitchcock et al., 1992), suggesting that even as physical lending spaces decline in importance, other spatial hubs may still shape renewal tendencies. Meanwhile, bibliomining of self-service library data reveals borrowing patterns (e.g., stable preferences for educational books) that inform renewal behavior predictions (Tu et al., 2020)—a finding that supports the idea that renewal in F1 is not random but tied to consistent user preferences, even amid external disruptions.

The 2019–2021 peak in renewal's predictive importance (Approximately 0.25) aligns with pandemic-era shifts, where limited access to physical books may have increased reliance on renewals as a proxy for demand. Notably, renewal consistently weakens the positive effect of F1 on borrowing volume, indicating that high renewal rates may signal constrained access (e.g., limited copies) rather than unqualified popularity—adding nuance to its interpretive role, as also observed in contexts where renewal reflects both sustained engagement and systemic limitations.

5.5 Popular theme phrases and title topics form interconnected clusters that mirror the multi-dimensionality of reader demand for educational content

The overlap between standardized subject phrases (e.g., education-theory-China) and LSA-derived title topics (e.g., clusters linking curriculum, teaching, and psychology) reveals that reader interest in G-category books is not fragmented but forms a cohesive ecosystem. Strong semantic connections between topics (e.g., the blue cluster in LSA) reflect interdisciplinary demand—for example, readers interested in teaching methods (Topic 4) also engage with educational theories (Topic 5). Meanwhile, isolated topics (e.g., Topic 2 on design and cases) highlight niche but impactful interests, demonstrating that popularity in educational books arises from both broad interdisciplinary appeal and targeted relevance to specific pedagogical or research needs.

The formation of this interest ecosystem is closely intertwined with the diverse characteristics of reader groups, where multidimensional demographic differences mediate the demand for and use of educational books. For instance, demographic differences (such as students' year of study, gender, and region) affect library use, with first-year students showing a stronger correlation between borrowing frequency and academic

performance—indicating that demographics mediate temporal engagement with educational materials (Brazier & Conroy, 1996). This suggests that analyzing demand for G-category books requires attention to temporal engagement variations across readers of different academic years and backgrounds: novice readers, for example, may rely more on frequent borrowing to build foundational knowledge, while advanced readers might focus on in-depth exploration of specific topics.

Furthermore, readers' skill levels modulate demand through quantile-based patterns: high-fluency readers (upper quantiles) exhibit distinct engagement with educational books, and bilingual reader's vocabulary knowledge correlates differentially with reading achievement across quantiles—underscoring that skill levels mediate demand for language-specific books (Van Norman et al., 2024; Marc Goodrich et al., 2023). This aligns with the interdisciplinary ecosystem of G-category books, where readers of varying skill levels may form differentiated demand within the same topic cluster. For example, junior readers might prioritize the practicality of teaching cases, while advanced readers may focus on critical analysis of educational theories.

Spatial and sociocultural factors also shape demand patterns through group characteristics. Socioeconomic status clusters high-achieving students spatially, driving demand for advanced educational materials in high-SES areas—where spatial inequality interacts with reader ability to influence book popularity (Barnes et al., 2024). Similarly, gender norms in Southern/Eastern Africa reduce reading score gaps, with girls performing better in gender-equal countries—indicating that spatial-cultural factors influence demand for gender-responsive educational/cultural books (Usui et al., 2024). These findings enrich our understanding of the thematic ecosystem of G-category books: the coherence of interests stems not only from the intrinsic connections between knowledge domains but also from the mediation of reader differences in demographics, skill levels, SES, and cultural backgrounds. These factors, in turn, interact with temporal engagement and spatial distribution to collectively shape the demand landscape for educational books.

6. Conclusion

This study employs Bayesian networks, factor extraction, and moderated regression models to explore the popularity dynamics of educational and cultural (G-category) books. Bayesian networks effectively model the hierarchical drivers of popularity, with content attributes (G Subcategory) as the most stable predictor. This framework captures both direct effects (e.g., renewal behavior signaling demand intensity) and indirect pathways (e.g., gender's influence mediated by subcategory preferences), highlighting the multi-layered nature of popularity. The evolution of core factors from TSI (Time-Space-Interest) to TI (Time-Interest) reflects a shift in reader behavior: physical lending spaces (e.g., Suiyuan Library) have diminished in importance, while temporal patterns aligned with interests (e.g., seasonal demand for teacher education materials) have become dominant, mirroring more fluid, digital-era reading habits. Renewal behavior exerts a dynamic moderating effect, weakening the impact of TSI/TI factors on borrowing volume. Its influence peaked in 2019–2021 (likely due to pandemic-related disruptions) and underscores that high renewals may signal constrained access rather than unqualified popularity. Popular themes and title topics form an interconnected ecosystem, with interdisciplinary clusters (e.g., teaching

methods and educational psychology) coexisting with niche interests (e.g., design and cases), reflecting diverse reader demands.

The study's generalizability is limited by its single-institution (NNU) sample, declining transaction volumes (102,566 in 2016–2018 to 28,657 in 2022–2024), and reliance on physical borrowing data. Simplified temporal (monthly) and spatial (lending rooms) metrics, along with static G subcategory classifications, further constrain depth. These limitations highlight the need for future research to expand sampling across institutions, integrate digital resource data, refine temporal/spatial metrics, and incorporate external contextual variables to enhance the robustness and generalizability of findings.

Future research should advance by expanding the sampling frame to encompass diverse institutional contexts, such as public libraries and specialized colleges, to validate the generalizability of observed patterns. It is critical to integrate data on digital resource usage to mitigate selection bias. Temporal analyses should be refined to account for granular patterns (e.g., exam cycles), while spatial metrics need to be extended to include digital access channels (e.g., library websites, mobile applications). Additionally, adopting flexible analytical models (e.g., deep learning-enhanced Bayesian networks) will better capture multi-layered interactions among variables. External contextual factors, such as curriculum reforms and pandemic-related disruptions, should be incorporated to explain temporal variations in popularity dynamics. Finally, implementing adaptive subcategory classifications will enable tracking of evolving content within categories (e.g., educational technology within G4). Addressing these gaps will enhance the understanding of educational book popularity, providing evidence-based insights to inform library collection development and reader-centric service design.

List of abbreviations
CLC=Chinese Library Classification
CNMARC=China Machine-Readable Cataloging Format
LAS=Latent Semantic Analysis
NNU=Nanjing Normal University
PCA=Principal Component Analysis
TAN Bayesian= Tree Augmented Naive Bayes
TSI=Time-Space-Interest
UNIMARC=Universal Machine-Readable Cataloging Format

Availability of data and materials
All the data and supporting information are provided within the article.

Conflict of interest
The authors declare no conflict of interest, financial or otherwise.

References

1. Alao, I. A. (2002). Students and overdue books in a medical library. Journal of the Medical Library Association, 90(3), 294–297.

2. Barnes, Z. T., Edwards, A. A., Strachota, S., Feng, Y., & Logan, J. (2024). Understanding the relation between socioeconomic status and elementary science achievement: A quantile regression approach. Infant and Child Development, 33(4), e2502. https://doi.org/10.1002/icd.2502

3. Bissenbayeva, A. P., Cheng, C. J., Abdurakyn, N., Saparbayeva, N. B., & Serikbayeva, G. J. (2024). The influence of Chinese online novels on the values of adolescents. Integration of Education, 28(1), 125–139. https://doi.org/10.15507/1991-9468.114.028.202401.125-139

4. Bodemer, B. (2012). Rabelais and the Abbey of Saint-Victor revisited. Information and Culture, 47(1), 4–17. https://doi.org/10.1353/lac.2012.0004

5. Brazier, H., & Conroy, R. M. (1996). Library use and academic achievement among medical students. Medical Education, 30(2), 142–147. https://doi.org/10.1111/j.1365-2923.1996.tb00732.x

6. Brewster, L. (2017). Murder by the book: Using crime fiction as a bibliotherapeutic resource. Medical Humanities, 43(1), 62–67. https://doi.org/10.1136/medhum-2016-011069

7. Bunzel, J. L. (2024). Trick of the eye: Prospect gazing, illusion, and the university novel. Journal of Victorian Culture, 29(1), 140–155. https://doi.org/10.1093/jvcult/vcad028

8. Burke, M., Tarver, H., Phillips, M. E., & Zavalina, O. (2022). Using existing metadata standards and tools for a digital language archive: a balancing act. Electronic Library, 40(5), 579–593. https://doi.org/10.1108/EL-02-2022-0028

9. Duncan, R., & Kayoro, C. (2022). The Role of African School Libraries in Supporting Quality Education. Publishing Research Quarterly, 38(1), 71–80. https://doi.org/10.1007/s12109-022-09866-5

10. Fleisch, B., & Schöer, V. (2023). Storybooks in hand: A randomised control trial of a classroom library model. International Journal of Educational Development, 96, 102705. https://doi.org/10.1016/j.ijedudev.2022.102705

11. Gandotra, S. (2023). Reading Benjamin Franklin's life story in Bareilly. South Asian History and Culture, 14(4), 450–466. https://doi.org/10.1080/19472498.2022.2141260

12. Hitchcock, M. A., Stritter, F. T., & Bland, C. J. (1992). Faculty development in the health professions: Conclusions and recommendations. Medical Teacher, 14(4), 295–309. https://doi.org/10.3109/01421599209018848

13. Huang, T. C., Shu, Y., Yeh, T. C., & Zeng, P. Y. (2016). Get lost in the library? An innovative application of augmented reality and indoor positioning technologies. Electronic Library, 34(1), 99–115. https://doi.org/10.1108/EL-08-2014-0148

14. Luo, P., Hong, L., & Nie, L. (2025). Automatic classification of research data sets into the Chinese Library Classification with generative large language model. Electronic Library. https://doi.org/10.1108/EL-02-2025-0042

15. Mahwasane, N. P. (2017). Managing the school library collection for ease of access in the foundation phase school library. Journal of Social Sciences, 53(2), 81–87. https://doi.org/10.1080/09718923.2017.1368220

16. Marc Goodrich, J., Fitton, L., & Thayer, L. (2023). Relations between oral language skills and English reading achievement among Spanish–English bilingual children: a quantile regression analysis. Annals of Dyslexia, 73(1), 6–28. https://doi.org/10.1007/s11881-022-00257-1

17. McLean, S. (2022). The cultural logic of precariousness and the marginalization of the sociological imagination: Signs from Mexican self-help books. Canadian Review of Sociology, 59(1), 115–134. https://doi.org/10.1111/cars.12371

18. McLean, S., & Vermeylen, L. (2019). From getting ahead to getting back on one's feet: Performing masculinity as a self-help reader. Men and Masculinities, 22(4), 716–737. https://doi.org/10.1177/1097184X17724188

19. Menadue, C. B., & Jacups, S. (2018). Who reads science fiction and fantasy, and how do they feel about science? Preliminary findings from an online survey. SAGE Open, 8(2). https://doi.org/10.1177/2158244018780946

20. Mennenga, H. A. (2016). Nursing student perceptions of digital textbooks: A pilot study. Nursing Education Perspectives, 37(2), 107–109. https://doi.org/10.5480/14-1377

21. Moore, M. (2018). Work discussion as a method for supporting peripatetic teachers of vulnerable children. Infant Observation, 21(1), 88–97. https://doi.org/10.1080/13698036.2018.1542832

22. Munteanu, A. I. (2020). "Christ, the way, the truth, and the life": Samuel von Brukenthal's British books on Christian civility. Brukenthal. Acta Musei, 15(5), 1037–1044.

23. Newton, R., & Dixon, D. (2000). Information for Scotland 1993-1998: a review. Library Review, 49(1), 11–21. https://doi.org/10.1108/00242530010307616

24. Ngafeeson, M. N., & Sun, J. (2015). The effects of technology innovativeness and system exposure on student acceptance of e-textbooks. Journal of Information Technology Education, 14(1), 55–71.

25. Oaksford, M., & Chater, N. (2009). Précis of Bayesian Rationality: The probabilistic approach to human reasoning. Behavioral and Brain Sciences, 32(1), 69–84. https://doi.org/10.1017/S0140525X09000284

26. Patra, N. K. (2017). Electronic resource management (ERM) in libraries of management institutes in India. Electronic Library, 35(5), 1013–1034. https://doi.org/10.1108/EL-01-2016-0020

27. Qingkui, X., Wenliang, H., Xinquan, L., Zhang, Q., & Feng, N. (2018). Resource co-construction in the CADAL and an evaluation model. Science and Technology Libraries, 37(4), 420–436. https://doi.org/10.1080/0194262X.2018.1506725

28. Salganik, M. J., & Watts, D. J. (2009). Web-based experiments for the study of collective social dynamics in cultural markets. Topics in Cognitive Science, 1(3), 439–468. https://doi.org/10.1111/j.1756-8765.2009.01030.x

29. Shen, Y. (2018). Library space information model based on GIS — a case study of Shanghai jiao tong University. Information Technology and Libraries, 37(3), 99–110. https://doi.org/10.6017/ital.v37i3.10308

30. Shin, J., & Valente, T. (2020). Algorithms and health misinformation: A case study of vaccine books on Amazon. Journal of Health Communication, 25(5), 394–401. https://doi.org/10.1080/10810730.2020.1776423

31. Su, Y., & Zhao, H. (2023). Infiltration approach of green environmental protection education in the view of sustainable development. Sustainability (Switzerland), 15(6), 5287. https://doi.org/10.3390/su15065287

32. Swarts, J., & VanNorman, M. (2008). Effective information design for PDAs in veterinary medical education. Journal of Veterinary Medical Education, 35(1), 118–128. https://doi.org/10.3138/jvme.35.1.118

33. Sysło, M. M. (2024). Pascal as the driving force of informatics education in schools in Poland. Informatics in Education, 23(4), 869–882. https://doi.org/10.15388/infedu.2024.31

34. Towsey, M. (2013). "I can't resist sending you the book": Private libraries, elite women, and shared reading practices in Georgian Britain. Library and Information History, 29(3), 210–222. https://doi.org/10.1179/1758348913Z.00000000040

35. Travis, T. A., & Ramirez, C. (2020). Big data and academic libraries: The quest for informed decision-making. Portal, 20(1), 33–47. https://doi.org/10.1353/pla.2020.0003

36. Tu, Y. F., Chang, S. C., & Hwang, G. J. (2020). Analysing reader behaviours in self-service library stations using a bibliomining approach. Electronic Library, 39(1), 1–16. https://doi.org/10.1108/EL-01-2020-0004

37. Usui, S., Blevins, B. K., & Kawata, K. (2024). Social gender inequality and the gender gap in the classroom: evidence from Southern and Eastern Africa. Journal of Contemporary African Studies, 42(4), 489–510. https://doi.org/10.1080/02589001.2024.2375658

38. van Beek, V. (2009). "Let these things speak for themselves". Experiment kits, instruction manuals, and narratives between 1870 and 1930. NTM International Journal of History and Ethics of Natural Sciences, Technology and Medicine, 17(4), 387–414. https://doi.org/10.1007/s00048-009-0356-z

39. Van Norman, E. R., Truman, K. J., & Clemens, N. H. (2024). Further exploration of the predictive validity of growth on early reading curriculum-based measures among students at risk for reading difficulties. Journal of School Psychology, 107, 101360. https://doi.org/10.1016/j.jsp.2024.101360

40. Westbury, C., Keith, J., Briesemeister, B. B., Hofmann, M. J., & Jacobs, A. M. (2015). Avoid violence, rioting, and outrage; approach celebration, delight, and strength: Using large text corpora to compute valence, arousal, and the basic emotions. Quarterly Journal of Experimental Psychology, 68(8), 1599–1622. https://doi.org/10.1080/17470218.2014.970204

41. Wynter, L., Burgess, A., Kalman, E., Heron, J. E., & Bleasel, J. (2019). Medical students: What educational resources are they using? BMC Medical Education, 19(1), 36. https://doi.org/10.1186/s12909-019-1462-9

42. Zhai, X., & Wang, J. (2016). Improving relations between users and libraries: A survey of Chinese academic libraries. Electronic Library, 34(4), 597–616. https://doi.org/10.1108/EL-03-2015-0041

43. Žitnik, S., & Gordon Smith, G. (2024). Automated analysis of postings in fourth grade online discussions to help teachers keep students on-track. Interactive Learning Environments, 32(8), 4587–4612. https://doi.org/10.1080/10494820.2023.2204327