



Does the annual report readability improve corporate R&D investment? Evidence from China

Ya-Guang Du¹ · Shu Li^{2,3} · Nan-Ting Kuo⁴ · Dan Li⁵

Received: 23 February 2023 / Accepted: 15 March 2024
© The Author(s), under exclusive licence to Springer Nature Limited 2024

Abstract

Our study investigates the association between annual report readability and corporate R&D investment decisions. We argue that a more readable annual report enhances investors' disclosure processing fluency and lowers information asymmetry, which mitigates insufficient funds and thus improves R&D investment levels. By exploring a panel sample of 26,359 firm-year observations for the period of 2007–2019 based on China's A-share listed firms, we find that increased readability of annual reports is associated with higher R&D investment. These results hold after a series of robustness checks. Mediation analysis reveals that annual report readability prompts R&D investment mainly through reducing financial constraints and supervising tunneling behaviors. Cross-sectional tests suggest that the positive association between readability and R&D investment is more pronounced in lower managerial manipulating incentives and lower disclosure processing ability of investors and exists peer spillover effects. Our results provide a novel insight that the economic impact of qualitative discourses can extend to management's financial decisions. Our study contributes to the literature on annual report readability and R&D under-investment governance.

Keywords Annual report readability · R&D investment · Disclosure processing costs · Financial constraints · Tunneling

Introduction

How financial reporting affects managers' investment decisions is a fundamental question in accounting. Financial reporting contributes to mitigating information asymmetry between firms and investors, as well as among investors. Thus, it will alleviate external financial constraints and enhance external supervision. Consequently, this will reduce agency costs from adverse selection and moral hazards and encourage managers to make better investment decisions (Biddle and Hilary 2006; Biddle et al. 2009). Although the actual effect of financial reporting is well recognized, most studies in this field focus on quantitative disclosure in the annual report, such as earnings management (McNichols and Stubben 2008), accounting conservatism (Balakrishnan et al. 2016) or disclosure transparency (Zhong 2018). Whether this real effect can extend to qualitative disclosure remains unknown. Motivated by the claims that few papers examine the "real effects" of disclosure processing costs on corporate actions, our study fills this gap by exploring the association between the textual readability of the annual report and research and development (R&D) investment.

✉ Ya-Guang Du
410520816@qq.com

Shu Li
lishu999@126.com

Nan-Ting Kuo
kuonantin@hotmail.com

Dan Li
nkolidan@163.com

- ¹ School of Economics and Management, Beijing University of Posts and Telecommunications, Xitucheng Road No.10, Haidian District, Beijing, People's Republic of China
- ² Business School, Nankai University, Tianjin, People's Republic of China
- ³ China Academy of Corporate Governance, Nankai University, Tianjin, People's Republic of China
- ⁴ Business School, Shandong University of Technology, Zibo, People's Republic of China
- ⁵ School of Management, Tianjin University of Technology, Tianjin, People's Republic of China



Specifically, a growing body of the literature has contributed to the information content of annual report readability in recent years. Many studies find that poor readability of textual information decreases the disclosure processing fluency of investors and increases their disclosure processing costs (You and Zhang 2009; Lee 2012; Rennekamp 2012). Consequently, it induces adverse market reactions, for example, fewer and less profitable trades among small investors (Miller 2010), less accurate analyst forecasts (Loughran and McDonald 2014), higher stock return volatility, stock price crashes (Bonsall et al. 2017; Kim et al. 2019), and higher cost of equity capital (Rjiba et al. 2021). Additionally, some focus on the effect of disclosure readability on executive actions. For example, Lo et al. (2017) find that managers have strong incentives to manage earnings when the disclosures are more complex. Luo et al. (2018) point out that more readable annual reports experience lower agency costs through reducing information asymmetry, which reflects an indirect effect of disclosure readability.

R&D investment is a critical managerial decision for enhancing a firm's competitiveness. However, the under-investment phenomenon of R&D expenditures is popular among listed companies in China. Because R&D activities have high input expenses, long duration, and high uncertainty of innovation outputs, abundant cash reserves are necessary to support R&D investment (Hall 2002). The agency theory suggests that external financial constraints due to information asymmetry induce managerial myopia and cut R&D investment (Roychowdhury 2006). Moreover, cash is the most liquid asset of firms and is easily misused or tunneled by managers or large shareholders, which reduces resource allocation efficiency and harms R&D investment (Chen et al. 2017). In the Chinese capital market, considering higher equity capital costs and strict requirements for seasoned equity offerings, listed companies are difficult to obtain external financial resources (Ju et al. 2013). Particularly, the weak corporate governance and legal protection for small investors' rights induce tunneling behaviors of large shareholders rampantly, so cash holdings are more likely to be misused (Kuo et al. 2022). Accordingly, most Chinese listed companies face insufficient cash or resources to support R&D investment.

Most research attributes R&D under-investment to managerial myopia and finds that high-quality disclosure can alleviate myopic R&D cuts (Shroff 2017; Bae et al. 2017; Heitzman and Huang 2019). We develop our argument from the perspective that increased annual report readability mitigates the information asymmetry of investors and it will improve R&D investment levels. Specifically, a more readable annual report helps reduce disclosure processing costs of investors, which mitigates their information asymmetry. Lower information symmetry implies fewer financial constraints and tunneling behaviors from large shareholders (La

Porta et al. 1999; O'Hara 1999; Asciglu et al. 2008). Furthermore, since fewer financial constraints and tunnelings expand cash or resources invested in R&D activities, it will encourage managers to take on more risky R&D investment projects. We argue that the higher readability of the annual report enhances the level of R&D investment.

We choose China's data to verify our argument for the following reasons. First, most studies rely on the analysis of the USA or other English-speaking countries, and few have explored the real economic effects of annual report readability in the Chinese context. Chinese expression has many differences from English in linguistics and logic structures, which implies that we cannot use prior results to test Chinese disclosure readability. Second, because small or non-professional investors are the dominant component in China's capital market, they usually face higher disclosure processing costs than developed capital markets. In addition, China's capital market is still weak on investor legal protection (Ke and Zhang 2021), which induces listed firms with lower information disclosure quality and increased agency costs. In particular, listed firms are challenged to realize financing at lower capital costs, and tunnelings of large shareholders are common. Accordingly, investors may be more sensitive to annual report readability, which will significantly impact management R&D investment decisions.

Taking 26,359 firm-year observations for Chinese A-share listed firms from 2007 to 2019, we explore the association between annual report readability and corporate R&D investment. According to the prior literature, we use the linguistic complexity of hand-collected yearly reports to measure readability. We find that higher readability of annual reports is associated with increased R&D investment, consistent with our argument. Moreover, our finding exists after several robustness tests, such as the instrumental variable method, Heckman's two-step, sample matching, and other robustness checks.

Through mediation analysis, we find that a more readable annual report improves R&D investment by reducing financial constraints and restraining the tunnelings of large shareholders. Cross-sectional tests suggest that the lower manipulating incentives of management and lower information processing ability of investors enhance the real effect of annual report readability. These results help mitigate the endogeneity concern from omitted variables. Furthermore, we find that there is a positive spillover effect of peer annual report readability for the R&D investment of target firms.

Overall, our results demonstrate that increased annual report readability is conducive to improving the level of R&D investment. Our study makes the following contributions to the literature. First, we contribute to the literature on how firm disclosure quality affects real investment decisions, such as earnings quality (Biddle et al. 2009), mandatory disclosure regulations (Chen et al. 2013; Albuquerque and



Zhu 2019), and weakness of internal control over financial reporting (Cheng et al. 2013; Feng et al. 2015). We explore this research from the novel insight of qualitative disclosure quality. Our results confirm that the real effect of disclosure quality not only exists in financial reporting but also extends to qualitative disclosure as manifested in readability.

Second, our study contributes to the literature on the economic consequences of annual report readability. On the one hand, we construct a readability index for the annual report based on the Chinese context, which is significantly different from the English context, and enlarge the research on China's annual report readability. On the other hand, we extend the results of annual report readability from investors' valuation to management's financial activities, which responds to the claims that few papers examine the "real effect" of disclosure processing costs on corporate actions (Blankespoor et al. 2020).

Finally, our study enlarges the literature on the channels of annual report readability impacting managerial R&D investment decisions. Our study confirms an association between annual report readability, agency costs, and R&D investment. Additionally, we prove that there exist considerable differences in information processing ability across professional and non-professional investors in China's capital market and negative information content of manipulating readability. These results clarify how annual report readability shapes managerial investment decisions and varies with information environments.

The remainder of our paper is organized as follows. "Literature review and hypothesis development" section reviews the literature and develops our hypothesis. "Research design" section introduces our research design and sample. "Empirical results" section discusses the empirical results and robustness tests. "Additional analysis" section presents additional analyses, and "Conclusion" section draws our conclusion.

Literature review and hypothesis development

The effect of financial reporting on real investment

It is well documented that financial reporting affects corporate investment decisions. Roychowdhury et al. (2019) attribute their mechanisms to information friction, such as information asymmetry and information uncertainty. According to the principal-agent theory, the agent conflict caused by incentive incompatibility can easily induce managers' moral hazard behaviors (Jensen and Meckling 1976; Jensen 1986). High-quality financial reporting helps external shareholders to understand and supervise the performance of the manager's fiduciary responsibility timely, which in turn

alleviates the information asymmetry and reduces the manager's overinvestment or opportunistic behavior in building a business empire (Biddle and Hilary 2006; Hope and Thomas 2008; Bushman et al. 2011; Delgado-Domonkos and Zeng 2023). Moreover, high-quality financial reporting can better describe the asset value of existing investment opportunities and reduce the information asymmetry between external investors and internal managers; thus, it helps reduce external financial constraints and expand capital expenditures (Biddle et al. 2009; Chen et al. 2011; Balakrishnan et al. 2016; Shroff 2020; Goldstein et al. 2023), but crowds out private firm's capital investment (Liu et al. 2023).

Managers may also face the challenge of information uncertainty when making investment decisions because they do not fully understand all current and future investment opportunities (Roychowdhury et al. 2019). However, the management is motivated to learn incremental information through private investors, analyst reports, product market adjustments, and peer firms in the same industry due to information acquisition and processing costs. As a result, they can alleviate information uncertainty and optimize investment decisions (Durnev and Mangen 2009; Shroff et al. 2014; Goldstein and Yang 2019; Heitzman and Huang 2019).

There is ample evidence that financial reporting disclosures influence corporate R&D investment. Since the accounting standard requires that firms disclose R&D expenditures as expenses during the operating period, management can increase the current income by reducing R&D activities to meet earnings performance targets, resulting in insufficient R&D investment (Bushee 1998; Stein 2003; Roychowdhury 2006). However, some argue that high-quality financial reporting helps to improve R&D investment. For example, Park (2018) reports a negative association between earnings management and innovation efforts. Zhong (2018) finds that higher financial reporting transparency mainly promotes management's innovation activities by reducing career concerns, alleviating the uncertainty of corporate long-term R&D investment, and improving the R&D investment level. Laux and Ray (2020) find more conservative accounting policies prompting corporate innovation activities. Chircop et al. (2020) show that the accounting comparability among peer firms is conducive to enhancing the ability to predict future cash flows generated by R&D investment and improving the efficiency of R&D investment.

Other evidence suggests that qualitative disclosure in the annual report impacts corporate investment. Dyreng et al. (2016) find that requiring companies to disclose the geographical location of subsidiaries helps reduce their motivation to invest and set up subsidiaries in tax havens. Christensen et al. (2017) find that positively declaring coal mine safety performance by coal mine owners in their annual reports is beneficial to increasing investment in coal mine



safety. Cheng et al. (2012), based on Chinese listed companies, suggest that non-financial disclosures can alleviate external financial constraints and reduce under-investment but increase over-investment of managements. Huang et al. (2014) report that the tone management level of the annual report is positively associated with merger and acquisition activities. Durnev and Mangen (2020) find that the positive MD&A tone of peers is conducive to promoting the target firm's real investment. De Simone and Olbert (2022) reveal that the mandatory disclosure of private country-by-country reporting improves corporate capital investment.

Linguistic properties and disclosure processing costs

Under the practical and imperfect capital market, investors need to spend time and effort to obtain and understand the information content of listed companies, that is, disclosure processing costs. Blankespoor et al. (2020) divide disclosure processing costs into three types such as awareness costs, acquisition costs, and integration costs. They further point out that the rapid development of the new information era has made investors access information conveniently, significantly reducing awareness and acquisition costs. How investors integrate information more efficiently has become an important research topic for disclosure processing costs. Some evidence finds that the linguistic properties of qualitative disclosure in the firm's annual report significantly impact the cost of disclosure processing. For example, the annual report tone can reflect management's sentiment and convey value-related information to external information users, which alleviates integration costs of disclosure processing. However, disclosure tone lacks adequate external supervision, and management has the selfish motivation to manipulate disclosure tone, thus expanding disclosure integration costs (Li 2010; Huang et al. 2014; Brochet et al. 2019).

Linguistic readability or complexity is another important disclosure property, which reflects the ease of comprehending textual information. Psychological research shows that individual information processing fluency depends on the specific context and subconsciously feels the degree of subjective comfort (Schwarz 2004). Individuals tend to feel higher credibility, more positive sentiment, and more substantial confidence in disclosure with higher fluency (Alter and Oppenheimer 2009). So, the readability of written texts is closely related to the fluency of processing, which determines disclosure processing costs. Existing research finds that the worse readability of the annual report induces higher post-earnings announcement drift (You and Zhang 2009), lower trading activities and investment return of retail investors (Miller 2010; Lawrence 2013), more efforts in writing analysts (Loughran and McDonald 2014), greater

equity mispricing (Chen et al. 2023), and lower stock return synchronicity (Gangadharan and Padmakumari 2023). Rennekamp (2012) explains the mechanism between readable disclosure and investors' decision-making through experimental research. They find that a readable annual report increases the fluency of investors' disclosure processing, which enhances the perception of management's reliability and thus produces a positive market reaction. Tan et al. (2019) find that the impact of linguistic processing fluency on investor decision-making significantly depends on the level of individual investors' industry knowledge. Moreover, the less readable annual report can predict the probability of a firm's bankruptcy (Le Maux and Smaili 2021). However, Dalwai et al. (2021) find no significant relationship between annual report readability and firm performance.

To further explore the motivation of textual readability, researchers struggle to isolate manipulating readability from the non-manipulating readability of the annual report. Firms that have ethical and legal responsibilities tend to disclose more readable and understandable annual reports (Bajaj et al. 2023; Xu et al. 2022). On the one hand, the complexity of corporate business operations, disclosure rule changes, and stricter supervision have led to a textual readability decrease normally. On the other hand, qualitative disclosure primarily consists of descriptive texts that are fuzzy and difficult to identify (Huang et al. 2014). Therefore, compared with quantitative disclosure, management motivated by self-interest prefers to conceal negative information by manipulating complex qualitative disclosure (Nadeem 2022; Sun et al. 2022a). Bonsall et al. (2017) find that if controlling non-discretionary complexity factors, some findings of disclosure readability are diminished. Bushee et al. (2018) report that discretionary complexity reduces stock liquidity. DeHaan et al. (2021) suggest that discretionary complexity will exacerbate the disclosure processing costs of investors and result in higher fund fees. However, there has been no consensus on clearly distinguishing between textual manipulating readability and non-manipulating readability.

Overall, higher-quality disclosure is beneficial to promoting real corporate investment by information mechanism. The readability of qualitative disclosure, as a critical information property, represents either high disclosure quality or discretionary behavior of management. However, few studies have focused on the effect of annual report readability on corporate investment decisions. Consequently, we, therefore, aim to investigate the effect of annual report readability on R&D investment in the context of China.

Hypothesis development

According to the background of the R&D under-investment problem resulting from a lack of financial support, we argue that increased readability of annual reports



facilitates reducing financial constraints and tunnelings, and then providing more available funds to improve corporate R&D investment. Based on the information processing fluency theory, the information will become difficult to process if one text contains vague and obscure words or complex sentences (Alter and Oppenheimer 2009). Investors regard the relatively fluent information as a subconscious and heuristic decision-making clue, which can improve their perception of the reliability of management, regardless of whether the disclosure is positive or negative (Hafner and Stapel 2010, Reneekamp 2012). So, readable disclosure facilitates investors to constrain information integration costs and obtain valuable information that reduces information asymmetry between corporate and external investors. Specifically, the readable disclosure can help many retail investors who are noise and non-professional traders in China's capital market minus the information gap to professional investors.

Moreover, with the decrease in information asymmetry, investors will make positive evaluations of management and reduce the cost of equity capital that induces external financial constraints (O'Hara 1999; Ascioğlu et al. 2008). From the perspective of resource dependency theory, corporate R&D investment is a costly, long-lasting, and hazardous activity that needs abundant cash reserves and financial support (Hall 2002). Therefore, a more readable annual report mitigates the under-investment of R&D activities by reducing financial constraints.

In addition, a readable annual report also plays a governance role in corporate innovation activities. The tunneling behavior by large shareholders, mainly through unfair related party transactions or internal group loans (Cheung et al. 2006; Jiang et al. 2010), is common in China. Because cash is the most liquid and transferable asset, severe tunnelings will harm or misuse cash required for innovation activities, which frustrates management from investing more in R&D projects (Chen et al. 2017). However, higher readability disclosure can provide value-related information and helps reduce information asymmetry between large and small shareholders (Luo et al. 2018), supervising large shareholders' tunnelings on cash misuses and then improving accessible funds to support corporate R&D investment.

Some of China's institutional features highlight the effects of annual report readability on R&D investment. First, the structure of investors in China is dispersed, contrary to the US capital market. Due to the limited ability of information acquisition and integration, retail or individual investors will respond more actively to higher readability disclosure than large investors (Miller 2010; Vieru et al. 2006; Lawrence 2013). Thus, annual report readability might impact R&D investment decisions significantly through information and governance effects. Second, China's economy is in a transition period and needs large-scale innovation to drive

high-quality economic development. Under the encouragement of favorable policies for innovation, firms prefer to invest in R&D activities rather than other real investments to obtain government subsidies and realize durable development. This particular institutional feature enhances the sensitivity of corporate R&D investment on available resources, highlighting the value of a readable annual report in solving investment problems.

The above discussions imply that increased readability of annual reports reduces the information asymmetry of investors, relieving insufficient funds due to financial constraints and tunneling, and then encourages corporate R&D investment activities. Overall, we expect increased readability of annual reports to promote corporate R&D investment levels. This leads to the following hypothesis.

H1 *Ceteris paribus*, higher annual report readability is associated with increased corporate R&D investment.

Moreover, two counterarguments add tension to H1. First, to the extent that information users perceive qualitative disclosure as "cheap talk," managements lack external pressure to conform disclosure readability to real corporate situations. In this case, manipulating the readability of annual reports cannot provide valuable information to investors, even misleading their decisions, which results in higher information asymmetry and will fail to alleviate financial constraints and tunnelings that hinder corporate R&D investment. So, managers are more likely to cut R&D expenditures when higher manipulating readability increases information asymmetry. Second, disclosure processing ability depends on the information user's educational background, industry knowledge, information environment, etc. There is a vast difference between different investors. For example, professional investors, such as institutional investors, sell-side analysts, or short sellers, use disclosure information more sophisticatedly than non-professional investors (Campbell et al. 2009; Battalio et al. 2012; Lee and Zhu 2022). In this view, the positive effect of a more readable annual report on corporate R&D investment might change when considering investors' properties.

Research design

Sample selection

Our initial sample consists of all A-share listed firms from 2007 to 2019. Our sample period begins in 2007 because the Ministry of Finance of the People's Republic of China implemented the new accounting standards for business enterprises in 2007, which has converged to IFRS and impacted corporate information disclosure significantly.



We collect all firm-level financial data from the CSMAR databases except for the textual information of the annual report from the Shanghai Stock Exchange and Shenzhen Stock Exchange, the officially designated disclosure platform. In addition, we exclude observations (1) within the financial or banking industry; (2) labeled as ST or ST* status because they face severe operational risk, so their decisions may differ from ordinary firms; and (3) with missing data for estimating equations. After imposing these filtering criteria, our final sample consists of 26,359 firm-year observations, where the number of observations varies across different tests considering the data availability required by the tests. To mitigate the disturbance of outliers, we winsorize all continuous variables at the top and bottom 1%.

Measure of annual report readability

Some studies use specific linguistic characteristics, including Fog and Bog score, Flesch Reading Ease, and jargon, to describe annual report readability (Bonsall et al. 2017; Lo et al. 2017; Tan et al. 2019). Other proxies focus on the overall processing costs of textual disclosure, such as the number of words, the average length of sentences, and the file size of annual reports (Miller 2010; Loughran and McDonald 2014). However, research on annual report readability has only just emerged in China, and there is no consensus about proxies to measure the readability of Chinese firms' annual reports (Luo et al. 2018).

On the one hand, there is a significant difference between Chinese- and English-speaking countries in logic, linguistics, the structure of words and sentences, and so on. Therefore, the Fog Index, Flesch Reading Ease, complex words, and other similar indexes to measure annual report readability are not comfortable in the Chinese context. On the other hand, Bonsall et al. (2017) assert that a large file size may include unrelated information to the underlying text in the disclosure, so the file size measurement is a rather noisy proxy for readability. Given the above reasons, we cannot directly use the prior methods to measure Chinese annual report readability.

According to Wang et al. (2018), we construct annual report readability proxies from the perspective of understandable ease of disclosure.¹ Pretorius (2006) argues that individuals struggle to understand written texts with complex logical relationships. Considering adversative sentences are hard to understand, we measure readability based on the proportion of disjunction numbers over every 100 total

words in the annual report. Referring to Wang et al. (2018), as “rare words” in the text will reduce the fluency of readers' reading and increase reading difficulties, annual report readability will decrease if it consists of many rare words. We take the intensity of “rare words” numbers over every 100 total words as the second proxy for readability. Additionally, jargon refers to management's use of technical or obscure language in disclosures, which impedes the processing fluency of readers and the readability of management disclosures (Bonsall et al. 2017; Tan et al. 2019). Most small investors are non-professional. They are unfamiliar with professional accounting terms disclosed in the annual report and usually look at them as jargon.² Thus, we take the number of accounting terms to account for every 100 words of the annual report to proxy readability. Finally, we adjust three variables through one minus their range standardization and get *Read_reverse*, *Read_rare*, and *Read_jargon*. On this base, we calculate a comprehensive variable *Read_comp* which equals the sum of *Read_reverse*, *Read_rare*, and *Read_jargon*. A greater value of each variable represents more readable annual reports. We use textual analysis techniques in Python programming language to capture and process qualitative information.³

Regression specification

To test H1 and evaluate the relative importance of annual report readability on R&D investment, we construct Eq. (1) as follows:

² The adversative conjunction dictionary comes from “Connective Elements in Modern Chinese Texts,” proposed by Liao (1986). The “rare words” dictionary comes from the 1000 secondary commonly used words specified in the “List of Commonly Used Words in Modern Chinese” formulated by the National Education Commission and the National Language and Writing Working Committee in 1988. The accounting terms dictionary comes from the Sogou cell thesaurus recommending the “A Complete of Accounting Vocabulary,” totaling 7631 entries.

³ The textual analysis process is as follows: First, we retrieve the documents of annual reports from 2007 to 2019 in batches from the Shanghai and Shenzhen Stock Exchange, the officially designated disclosure platform by Python programming language, and get 33,485 initial annual reports. After excluding incomplete disclosure, repeated disclosure (the annual report before the final update), English annual reports, and samples with incorrect format, we obtained 33,208 annual reports. Second, for pure output texts, we use Python to process textual data, such as excluding special characters, punctuation marks, picture information, numbers, English characters, spaces, and non-Chinese characters in the annual report. Finally, with the help of Python, using the Jieba library to accurately segment the Chinese contexts of the yearly report and count word frequencies of readability based on their dictionaries.

¹ There are also studies following Wang et al. (2018)'s method to measure annual report readability in the Chinese context, such as Wang et al. (2020), Sun et al. (2022b), Yu et al. (2022), and so on. Accordingly, we follow Wang et al. (2018)'s method to construct indexes that can proxy for annual report readability effectively.



Table 1 Descriptive statistics

Variables	N	Mean	Sd	Min	Q1	Median	Q3	Max
R&D	26,359	0.0153	0.0181	0.0000	0.0000	0.0103	0.0239	0.0900
Read_reverse	26,359	0.5861	0.1960	0.0000	0.4643	0.6002	0.7226	1.0000
Read_rare	26,359	0.8483	0.1469	0.0000	0.8144	0.8866	0.9339	1.0000
Read_jargon	26,359	0.5429	0.1931	0.0000	0.4248	0.5550	0.6755	1.0000
Read_comp	26,359	1.9773	0.3278	0.0000	1.7813	1.9911	2.1905	3.0000
Size	26,359	22.1122	1.3038	19.2889	21.1816	21.9351	22.8639	26.0082
Lev	26,359	0.4441	0.2102	0.0500	0.2786	0.4399	0.6013	0.9865
Cash	26,359	0.1835	0.1313	0.0121	0.0918	0.1478	0.2381	0.7100
Netfin	26,359	-0.1138	0.7565	-4.0363	-0.1158	0.0146	0.2433	0.7509
Soe	26,359	0.4203	0.4936	0.0000	0.0000	0.0000	0.0000	1.0000
Age	26,359	2.1670	0.7397	0.0000	1.6094	2.3026	2.7726	3.2581
AH	26,359	0.0311	0.1737	0.0000	0.0000	0.0000	0.0000	1.0000
Rev	26,359	21.4372	1.4745	17.4446	20.4549	21.3079	22.3023	25.3859
Employ	26,359	7.6566	1.3030	3.9703	6.8167	7.6104	8.4621	11.0724
KL	26,359	0.5711	0.9909	0.0103	0.1430	0.2738	0.5450	6.9133
BM	26,359	0.4440	0.3081	0.0119	0.2248	0.3675	0.5759	1.5851
Roa	26,359	0.0422	0.0607	-0.2222	0.0150	0.0385	0.0702	0.2370
DRoa	26,359	-0.0051	0.0571	-0.2676	-0.0202	-0.0023	0.0103	0.2432
Ret	26,359	0.2090	0.7237	-0.7076	-0.2670	0.0108	0.4457	3.2707
Tobinq	26,359	2.2532	1.4723	0.9230	1.3324	1.7716	2.6050	9.3085
HHI	26,359	0.1046	0.1035	0.0152	0.0471	0.0727	0.1331	0.6657
GDP	26,359	10.9775	0.5458	8.8414	10.6254	11.0568	11.3791	12.0090

This table reports the descriptive statistics results of the main variables. *N* is the number of usable observations for the variable. *Sd* is the standard deviation. *Q1* is the first quartile. *Q3* is the third quartile

$$\begin{aligned}
R\&D_{t+1} = & \beta_0 + \beta_1 \text{Readability}_t + \beta_2 \text{Size}_t \\
& + \beta_3 \text{Lev}_t + \beta_4 \text{Cash}_t + \beta_5 \text{Netfin}_t \\
& + \beta_6 \text{Soe}_t + \beta_7 \text{Age}_t + \beta_8 \text{AH}_t \\
& + \beta_9 \text{Rev}_t + \beta_{10} \text{Employ}_t + \beta_{11} \text{KL}_t \\
& + \beta_{12} \text{BM}_t + \beta_{13} \text{Roa}_t + \beta_{14} \text{DRoa}_t \\
& + \beta_{15} \text{Ret}_t + \beta_{16} \text{Tobinq}_t + \beta_{17} \text{HHI}_t \\
& + \beta_{18} \text{GDP}_t + \text{Fixed Effects} + \varepsilon
\end{aligned} \quad (1)$$

In Eq. (1), *R&D* represents the next period's research and development investment level, calculated as R&D expenditures divided by total assets (Zhong 2018). The explanatory *Readability* means annual report readability, including previous variables *Read_reverse*, *Read_rare*, *Read_jargon*, and *Read_comp*. In addition, we incorporate the following control variables that influence corporate R&D investment decisions. *Size* is the natural logarithm of total assets at the end of the period. *Lev* is financial leverage, defined as liabilities divided by total assets at the end of the period. *Cash* is cash and cash equivalents over total assets at the end of the year. *Netfin* is the sum of net finance, calculated as the difference between net cash received from equity and bonds issued and cash paid for purchasing financial assets scaled by total assets. *Soe* is a dummy variable that equals one

for state-owned enterprises and zeroes otherwise. *Age* is the natural logarithm of one plus the listed years in the A-share market. *AH* is an indicator variable that equals one for firms cross-listing of A and H share and zeroes otherwise. *Rev* is the natural logarithm of major revenue. *Employ* is the natural logarithm of one plus the number of staff. *KL* is net fixed assets divided by total employees at the end of the year. *BM* is the book value of equity divided by the market value of equity. *Roa* is the ratio of net income over average total assets. *DRoa* is a change in net income scaled by average total assets. *Ret* is the annual stock return. *Tobinq* is the market value of equity scaled by total assets. *HHI* is the industry Herfindahl index, calculated as the sum of the square of the operating income divided by the total operating income in the same industry and taking its square root. *GDP* is the natural logarithm of the province's per capita gross domestic product during the fiscal year. *Fixed Effects* represent the year and industry fixed effects. The standard errors of regression coefficients are two-way clustered by firm and year.

We include *Size* to control for any size effects. We include *Lev*, *Cash*, and *Netfin* because the financial resource is a critical element for innovation. We include *Soe*, *Age*, *AH*, *Employ*, and *KL* to control the effects of firms' nature. We include *Rev*, *Roa*, and *DRoa* because better financial



performance affects firm innovation activities. We include *BM*, *Ret*, and *Tobinq* as a proxy for firm value because the market evaluations induced by investors encourage managers to innovate. We also include *HHI* and *GDP* to capture the effects of external product market competence and macro-economic development on corporate R&D investment decisions. According to the expectation of H1, the coefficient β_1 in Eq. (1) should be significantly positive.

Table 1 exhibits the descriptive statistics. The mean R&D of 0.0153 implies that the proportion of R&D investment in total assets is 1.53%, close to 1.50% of the US listed firms (Zhong 2018). The mean of *Read_reverse*, *Read_rare*, *Read_jargon*, and *Read_comp* are equal to 0.5861, 0.8483, 0.5429, and 1.9773 respectively. These values indicate that most of China's listed firms' annual reports are readable but vary widely across listed firms, providing statistically powerful tests to examine our arguments.

Table 2 reports the matrix of Pearson correlation coefficients. *Read_rare* and *Read_jargon* are positively and significantly correlated with R&D. However, *Read_Reverse* and *Read_comp* are negatively correlated with R&D, inconsistent with our expectations of the previous hypothesis. Thus, we should rely on multiple regression analysis to test our hypothesis accurately. Most of the correlation coefficients between R&D and control variables are statistically significant, so our model is reasonable. Furthermore, the average of variance inflation factors (VIF) is 2.13, and the highest VIF is 6.50, suggesting that multicollinearity is not a concern.

Empirical results

Main empirical results

Table 3 reports the OLS regression results for examining the relationship between annual report readability and corporate R&D investment for Hypothesis 1. According to Column (1), the coefficient on *Read_reverse* is positive and significant at the 1% level (coeff. = 0.0024, $t=3.09$), indicating the lower density of adversative words in the annual report promotes R&D investment levels in the future. Column (2) reports that the coefficient on *Read_rare* is positive and significant at the 5% level (coeff. = 0.0025, $t=2.40$), suggesting that management's use of less obscure or rare words in the annual report facilitates investing in research and development activities. In Column (3), the coefficient on *Read_jargon* is also positive and significant at the 1% level (coeff. = 0.0043, $t=4.25$), showing the lower frequency of accounting terms in the annual report helps improve R&D investment levels in the next period. Moreover, the results presented in Column (4) show that the coefficient on *Read_comp* is significant and positive at the 1% level (coeff. = 0.0028, $t\text{-stat}=5.21$). For economic significance, this result implies that the annual

report readability improves the level of future R&D investment by 27.18% of its sample median (0.0028/0.0103). In conclusion, these results support H1 that higher readability of the annual report enhances the level of R&D investment in the future.

Turning to our control variables, consistent with prior studies, we find that firms with lower financial leverage, lower listed age, lower book-to-market value of equity, and lower industry competence facilitate expanding R&D expenditures, which is consistent with Zhong (2018). In addition, We also find a positive and statistically significant coefficient on *Rev*, *Roa*, *Employ*, *KL*, and *GDP*, confirming that firms with higher profitability, abundant human capital, and highly developed macro-economy help inspire firms' inputs of innovation.

Robustness tests

Endogeneity concern

Factors that impact corporate innovation decisions are numerous and complex, and our study may face the problem of omitting latent variables. Additionally, annual report disclosure and R&D investment are the results of the managerial decision, which implies the manager makes dynamic adjustments for the readability of the annual report according to current R&D expenditures. Thus, our findings may be challenged by the problem of reverse causality. Therefore, we take several tests to mitigate the above endogeneity concerns, where the results are exhibited in Table 4. Control variables results are not presented for brevity since they are similar to those in Table 3. Additionally, *Read_comp* is the sum of *Read_reverse*, *Read_rare*, and *Read_jargon*, which is more representative than the three variables on readability. Given this, we use *Read_comp* as the primary explanatory variable.

We use the instrumental variable (IV) method by 2SLS regression to relieve the variable omitted problem. We take the educational background of one firm's Chief Financial Officer (CFO) as an instrumental variable of annual report readability. First, innovation activities, as an essential strategy, are primarily decided by the chairman of the board and CEO. So the CFO's educational background is unlikely to affect R&D investment directly, which meets the exogenous assumption of instrumental variables. Second, the CFO has major responsibilities for the content and quality of the annual report. A better CFO's educational background represents richer professional theoretical knowledge, careful thinking, and a more vital ability to understand and refine the critical information from complex situations so that they can express statements to the outside more concisely. Thus, it meets the correlation assumption of instrumental variables. We use the final educational degree of the CFO



Table 2 Pearson correlation matrix

	R&D	Read_reverse	Read_rare	Read_jargon	Read_comp	Size	Lev	Cash	Netfin	Soe	Age	AH		
R&D	1.0000													
Read_reverse	-0.1319***	1.0000												
Read_rare	0.0199***	-0.0138**	1.0000											
Read_jargon	0.1105***	0.1033***	0.0547***	1.0000										
Read_comp	-0.0049	0.6528***	0.4721***	0.6755***	1.0000									
Size	-0.1591***	0.1185***	-0.1318***	-0.1455***	-0.0739***	1.0000								
Lev	-0.2872***	0.1551***	-0.0540***	-0.1495***	-0.0195***	0.4291***	1.0000							
Cash	0.1928***	-0.0181***	0.0603***	0.0516***	0.0466***	-0.2103***	-0.3825***	1.0000						
Netfin	-0.0596***	0.1334***	-0.0227***	0.0100	0.0756***	0.0149**	0.1219***	-0.0476***	1.0000					
Soe	-0.2732***	0.1643***	0.0522***	-0.1418***	0.0381***	0.2881***	0.2801***	-0.1196***	0.0581***	1.0000				
Age	-0.2920***	0.1000***	0.0024	-0.2387***	-0.0798***	0.3425***	0.3508***	-0.2311***	-0.1318***	0.4005***	1.0000			
AH	-0.0526***	0.0882***	-0.0522***	0.0966***	0.0865***	0.3317***	0.1020***	-0.0643***	0.0343***	0.1442***	0.0264***	1.0000		
Rev	-0.0622***	0.1119***	-0.1176***	-0.1595***	-0.0797***	0.8799***	0.4180***	-0.1661***	0.0139**	0.2810***	0.2835***	0.2961***	1.0000	
Employ	0.0369***	0.1080***	-0.0946***	-0.1122***	-0.0439***	0.6925***	0.2801***	-0.1461***	0.0475***	0.2298***	0.1731***	0.2796***	0.1099***	1.0000
KL	-0.2167***	0.0632***	-0.0558***	-0.0265***	-0.0028	0.2907***	0.1412***	-0.2108***	0.0385***	0.1915***	0.1537***	0.1099***	0.2320***	0.1099***
BM	-0.1805***	0.0871***	-0.0738***	-0.0553***	-0.0136**	0.5593***	0.1547***	-0.1575***	0.0114*	0.2133***	0.2308***	0.2320***	0.2308***	0.2320***
Roa	0.1550***	-0.0048	0.0375***	0.0714***	0.0560***	0.0124**	-0.3634***	0.2594***	-0.0804***	-0.0983***	-0.1797***	-0.0021	-0.1797***	-0.0021
DRoa	-0.0348***	0.0162***	0.0155**	-0.0255***	0.0016	0.0249***	-0.0062	0.0063	-0.0157**	0.0487***	0.0878***	0.0031	0.0878***	0.0031
Ret	-0.0418***	0.0432***	0.0092	-0.0288***	0.0130**	-0.0870***	0.0423***	0.0125**	0.0398***	0.0294***	0.0023	-0.0043	0.0294***	-0.0043
Tobinq	0.1629***	-0.0835***	0.0513***	0.0248***	-0.0123**	-0.4155***	-0.2476***	0.1286***	-0.0818***	-0.1483***	-0.0066	-0.0780***	-0.1483***	-0.0066
HHI	-0.2219***	0.0579***	0.0280***	0.0047	0.0499***	0.0471***	0.0579***	-0.0358***	0.0469***	0.1155***	0.0338***	0.0764***	0.1155***	0.0338***
GDP	0.2803***	-0.2285***	-0.0375***	-0.0423***	-0.1784***	0.1457***	-0.1456***	0.0801***	-0.1722***	-0.2258***	-0.0477***	0.0289***	-0.2258***	-0.0477***
	Rev	Employ	KL	BM	Roa	DRoa	Ret	Tobinq	HHI	GDP				
Rev	1.0000													
Employ	0.7780***	1.0000												
KL	0.1561***	-0.0796***	1.0000											
BM	0.4612***	0.3425***	0.2598***	1.0000										
Roa	0.0893***	0.0586***	-0.0682***	-0.1396***	1.0000									
DRoa	0.0489***	0.0138**	0.0200***	-0.0333***	0.4777***	1.0000								
Ret	-0.0644***	-0.0321***	-0.0204***	-0.3503***	0.1078***	0.1727***	1.0000							
Tobinq	-0.3967***	-0.2924***	-0.1520***	-0.5685***	0.1469***	0.0515***	0.3176***	1.0000						
HHI	0.0389***	0.0701***	0.0478***	0.0396***	-0.0403***	0.0021	0.0307***	-0.0441***	1.0000					
GDP	0.1077***	-0.0272***	-0.0096	0.0965***	0.0347***	-0.0195***	-0.2025***	-0.0091	-0.0931***	1.0000				

This table presents the Pearson correlation matrix of the main variables in Eq. (1). ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively



Table 3 Annual report readability and corporate R&D investment

	(1) R&D _{t+1}	(2) R&D _{t+1}	(3) R&D _{t+1}	(4) R&D _{t+1}
Read_reverse	0.0024*** (3.09)			
Read_rare		0.0025** (2.40)		
Read_jargon			0.0043*** (4.25)	
Read_comp				0.0028*** (5.21)
Size	- 0.0044*** (- 8.42)	- 0.0043*** (- 8.31)	- 0.0043*** (- 8.44)	- 0.0043*** (- 8.39)
Lev	- 0.0048*** (- 4.19)	- 0.0047*** (- 4.05)	- 0.0046*** (- 3.95)	- 0.0048*** (- 4.10)
Cash	0.0028 (1.57)	0.0030 (1.64)	0.0030 (1.63)	0.0027 (1.49)
Netfin	- 0.0000 (- 0.17)	- 0.0000 (- 0.01)	- 0.0001 (- 0.38)	- 0.0001 (- 0.32)
Soe	- 0.0000 (- 0.01)	- 0.0001 (- 0.17)	0.0001 (0.13)	0.0000 (0.01)
Age	- 0.0024*** (- 4.86)	- 0.0024*** (- 4.72)	- 0.0022*** (- 4.49)	- 0.0023*** (- 4.86)
AH	0.0000 (0.03)	0.0001 (0.09)	- 0.0007 (- 0.54)	- 0.0004 (- 0.33)
Rev	0.0028*** (6.42)	0.0028*** (6.46)	0.0029*** (6.67)	0.0029*** (6.57)
Employ	0.0023*** (6.52)	0.0023*** (6.57)	0.0023*** (6.54)	0.0023*** (6.55)
KL	0.0005*** (2.89)	0.0005*** (2.97)	0.0005*** (2.85)	0.0005*** (2.94)
BM	- 0.0043*** (- 4.78)	- 0.0043*** (- 4.70)	- 0.0042*** (- 4.76)	- 0.0043*** (- 4.83)
Roa	0.0278*** (6.67)	0.0276*** (6.60)	0.0270*** (6.48)	0.0270*** (6.49)
DRoa	- 0.0148*** (- 5.61)	- 0.0147*** (- 5.58)	- 0.0143*** (- 5.44)	- 0.0145*** (- 5.53)
Ret	- 0.0005 (- 0.96)	- 0.0005 (- 0.96)	- 0.0005 (- 0.92)	- 0.0005 (- 0.96)
Tobinq	0.0006*** (3.22)	0.0006*** (3.27)	0.0006*** (3.27)	0.0006*** (3.28)
HHI	0.0077*** (2.78)	0.0080*** (2.88)	0.0084*** (2.97)	0.0079*** (2.83)
GDP	0.0022*** (4.49)	0.0022*** (4.46)	0.0023*** (4.80)	0.0023*** (4.70)
Constant	0.0136* (1.88)	0.0112 (1.50)	0.0078 (1.04)	0.0064 (0.86)
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
Observations	22,328	22,328	22,328	22,328
Adj. R ²	0.4984	0.4982	0.4994	0.4999

This table presents the results of examining our hypothesis with Eq. (1), where Columns (1) measures annual report readability based on adversative conjunction density (*Read_reverse*), Columns (2) measures annual report readability based on rare words density (*Read_rare*), Columns (3) measures annual report readability based on accounting terms density (*Read_jargon*), and Columns (4) use the sum of *Read_reverse*, *Read_rare*, and *Read_jargon* proxy for annual report readability. Definitions of variables are shown in the table in Appendix. Industry and year fixed effects are included, and the results are not reported. Robust t-statistics clustered by firm and year are reported in the parentheses beneath the coefficients. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively



obtaining and set variable *CFO_degree* representing for CFO's educational background. If a CFO experiences technical secondary school or below, the value of *CFO_degree* is 1, technical school is 2, a bachelor's degree is 3, a master's degree is 4, and a doctoral degree is 5. As shown in Column (1) of Table 4, the coefficient on *CFO_degree* is positive and significant, indicating that the selected IV better explains the annual report readability without the problem of weak instrumental variables. The second regression of 2SLS in Column (2) of Table 4 shows that our inferences are unchanged.

The readability of the annual report may be the consequence of management characteristics. We use Heckman's two-step to mitigate the self-selection problem. We select the management's male ratio (*Gender*), the average age (*Mage*), average educational background (*Degree*), average receiving salary percentage (*Paid*), management shareholdings (*Share*), average overseas back (*Oversea*), and average financial back (*Finback*) as instrumental variables for annual report readability.⁴

As shown in Column (3) of Table 4, the coefficient on management characteristics is significant, confirming that individual characteristics impact the annual report readability. Column (4) in Table 4 reports that after controlling the reverse mills ratio (*IMR*), the coefficient on *Read_comp* is still positive and significant, consistently with H1.

Moreover, we further use the propensity score matching (PSM) and entropy balance methods (Hainmueller 2012) to relieve issues of omitted variables and sample selection bias. First, set up the probability of readable annual reports as the dependent variable *D_Readability*, which equals one if firms are top 20% of the annual report readability and zero if firms are bottom 20%. Then, taking firm-level and other factors similar to control variables in Eq. (1) as the independent variable to make Probit regressions. Second, calculating the propensity score and getting the control group by one-to-one matching or calculating entropy weights for every covariant variable. After matching, the differences in covariant variables between the treatment group and control group become insignificant. Finally, we use matched samples or weighted variables repeating Eq. (1) regression. As shown in Column

(5) and Column (6) of Table 4, these results do not change our conclusions.

We also conduct the following tests to mitigate endogeneity concerns. First, the China Securities Regulatory Commission issued announcement No. 24 in 2015, which modified the disclosure standard of "Management Discussion and Analysis" and impacted corporate qualitative disclosure significantly. At the same time, there was no innovation policy shock, but the average R&D investment of listed firms increased dramatically after 2015 ($t=13.87$). Therefore, we take announcement No. 24 as an external policy shock to test the causality of the annual report readability and R&D investment. As shown in Column (7) and Column (8) of Table 4, the coefficient on *Read_comp* after policy shock is more significant and positive than the coefficient before policy shock (P -value for difference = 0.00). Second, the standard disclosure requires mandatory disclosing of R&D expenditures and voluntary disclosing relative information about innovation in the annual report, so it is necessary to exclude R&D information from the annual report to mitigate reverse causality. We use the residual of the annual report readability regressing on R&D investment as pure readability and repeat Eq. (1) regression. Third, we conduct a firm fixed effects model to control the influence of firm characteristics that are not changed over time or are unobservable. As shown in Column (9) and Column (10) of Table 4, the coefficient on *Read_comp* is positive and significant. The above tests do not change our conclusions.

Other robustness tests

We perform the following robustness tests, where the results are reported in Table 5. Koh and Reeb (2015) argue that missing R&D information in the annual report does not mean that there are no R&D activities. If we use zero to replace the missing value, it will produce inaccurate results. So, according to Zhong (2018), we use the median value of R&D filling in missed data or retaining non-missing samples to reexamine Eq. (1). As shown in Column (1) to Column (2) of Table 5, the coefficients on *Read_comp* are significant and positive at the 1% level, supporting our inferences. Then, we substitute measurements of *Read_comp*. First, the longer the sentences in the annual report, the more difficult it is to read and understand for information users. We set up *Read_sent1* and *Read_sent2* variables to represent the average length of a sentence and the ratio of long difficult sentences over total sentences, respectively, and one minus their standardization.⁵ Thus, the annual report readability is measured by a new variable

⁴ *Gender* is male management divided by total management. *Mage* is the sum of management ages divided by total management. *Degree* is the number of management obtaining master's degree and above scaled by total management. *Paid* is the amount of management receiving salary from corporate scaled by total management. *Share* is an indicator variable that equals one if managements hold firm stocks and zero otherwise. *Oversea* is the amount of experience in overseas back management scaled by total management. Finally, *Finback* is the amount of having financial back management scaled by total management.

⁵ The average length of a sentence is total words divided by total sentences. A sentence is long and difficult to understand if the sentence includes more than 15 words (Jiang and Wang 2019).



Table 4 Regression results of endogeneity concerns

	2SLS regression		Heckman's two-step		PSM and Entropy balance		Policy shock effects		Residual method		Firm fixed effects	
	(1) Read_comp _t	(2) R&D _{t+1}	(3) D_Readability _t	(4) R&D _{t+1}	(5) R&D _{t+1}	(6) R&D _{t+1}	(7) R&D _{t+1}	(8) R&D _{t+1}	(9) R&D _{t+1}	(10) R&D _{t+1}	(10) R&D _{t+1}	(10) R&D _{t+1}
Read_comp		0.2970*** (5.73)	0.0024*** (4.78)	0.0026*** (4.82)	0.0027*** (4.86)	0.0016** (2.00)	0.0031*** (3.09)	0.0029*** (5.42)	0.0013*** (3.38)			
CFO_degree	0.0054* (1.76)											
Gender			-0.0958*** (-5.06)									
Male			0.0560* (1.78)									
Degree			0.0359*** (10.03)									
Paid			0.1023*** (7.59)									
Share			0.0115** (2.46)									
Oversea			0.1025*** (4.77)									
Finback			-0.0676*** (-2.75)									
IMR												
Constant	2.6682*** (26.80)	-0.7768*** (-5.55)	1.6453*** (13.49)	-0.0006 (-0.06)	0.0106 (1.11)	0.0098 (0.92)	0.0096 (0.87)	0.0117 (1.63)	-0.0108 (-0.65)			
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Firm fixed effects	No	No	No	No	No	No	No	No	No	No	Yes	Yes
Observations	17,574	14,658	25,611	8301	8926	8004	7680	22,328	22,328			
Adj. R ²	0.2066	0.4923	0.1675	0.5031	0.5019	0.4922	0.5130	0.5001	0.7989			

This table reports the results of endogeneity concern for the association between annual readability and R&D investment. Columns (1) and (2) are the first and second stage regression of the 2SLS method, respectively. Columns (3) and (4) are the first-step regression and second-step regression of Heckman's two-step method. Column (5) conducts Propensity Scoring Matching (PSM). Column (6) uses the Entropy Balance method to mitigate biased samples. Columns (7) and (8) are the effects of policy shocks before and after, respectively. Column (9) takes the residuals of *Read_comp* regressing on *R&D* as a new measurement of annual report readability, excluding the impacts of innovation information. Column (10) adds firm-level fixed effects. *IMR* represents the reverse mills ratio calculated by the probability density function scaled by the cumulative distribution function of fitting values of the dependent variable in the first stage of Probit regression. Definitions of other variables are shown in the table in Appendix. Fixed effects and control variables are included, and the results are not reported. Robust t-statistics clustered by firm and year are reported in the parentheses beneath the coefficients. ***, **, * and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5 Robustness tests

	Change measurement of R&D investment		Change measurement of the annual report readability		Control R&D trend (5) R&D _{t+1}	Subsample tests (6) R&D _{t+1}	Tobit model (7) R&D _{t+1}
	(1) R&D1 _{t+1}	(2) R&D2 _{t+1}	(3) R&D _{t+1}	(4) R&D _{t+1}			
Read_comp	0.0022*** (3.41)	0.0030*** (4.38)			0.0009*** (3.15)	0.0024*** (3.05)	0.0035*** (5.63)
Read_comp1			0.0017*** (4.70)				
Read_comp2				0.0031*** (3.39)			
R&D					0.6596*** (13.40)		
Constant	0.0583*** (4.36)	0.0175* (1.74)	0.0130* (1.82)	0.0068 (0.87)	0.0045 (1.61)	0.0148 (1.54)	- 0.0265*** (- 2.67)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,166	16,757	22,328	22,251	22,328	11,931	22,328
Adj. R ²	0.3844	0.4249	0.4991	0.4991	0.7944	0.5458	

This table presents the results of other robustness checks for the association between annual readability and R&D investment. Columns (1) and (2) measure R&D investment with *R&D* median filling in missing data and using non-missing *R&D* samples, respectively. Columns (3) and (4) measure annual report readability by adding sentence readability and using MD&A readability, respectively. Column (5) adds a current period of *R&D* into Eq. (1) to control the impact of R&D investment trends. Column (6) uses only samples with non-high-tech firms from 2012 to 2019. Column (7) uses the Tobit model to estimate Eq. (1). Definitions of variables do not change. Control variables and fixed effects are included without being reported. Robust t-statistics clustered by firm and year are reported in the parentheses. *** and * denote statistical significance at the 1% and 10% levels, respectively

Read_comp1 which is the sum of *Read_comp*, *Read_sent1*, and *Read_sent2*, for which a more excellent value stands for higher readability. Second, the MD&A sector of the annual report consists of much qualitative information that reflects management's sentiment and judgment. Accordingly, we use MD&A readability to measure annual report readability, similar to *Read_comp* on calculation, and get a new *Read_comp2* variable.⁶ As shown in Column (3) to Column (4) of Table 5, our conclusions do not change.

Furthermore, we add current *R&D* investment levels into Eq. (1) to control the trend of innovation activities. We also take non-high-tech firms after 2012 as our subsamples to avoid the shock of abnormal R&D investment and the disclosure policy of innovation information in 2012.⁷ Finally,

⁶ We use the MD&A textual disclosures to calculate the readability of MD&A. The calculation method is consistent with the annual report readability *Read_comp* in the previous explanation.

⁷ The definition of high-tech firms from the "Statistical Yearbook of China's High-tech Industry" includes the information chemicals manufacturing industry, pharmaceutical manufacturing industry, aerospace, and other transportation equipment manufacturing industry, electronic and communication equipment manufacturing industry, computer and office equipment manufacturing industry, and medical equipment and instrument manufacturing industry. In addition, in 2012, the China Securities Regulatory Commission started to require

considering the left truncation problem of R&D expenditures, we perform the Tobit model to regress Eq. (1). The results in Columns(5) to (7) of Table 5 suggest that using different robustness tests does not affect our conclusions.

Additional analysis

The channel through which readability impacts R&D investment

In this section, we explore the causal chains that the more readable annual report enhances the level of R&D investment by improving the cash resource of innovation. Our test is motivated by the theory of resource dependency that financial constraints and tunneling from large shareholders limit managers' ability to invest in R&D activities, which constructs the causal chain underlying our hypothesis.

According to the above analysis, higher readability of the annual report can enhance the information processing

Footnote 7 (continued)

listed firms to disclose R&D expenditure information mandatory in the annual report, so the data on R&D investment will change dynamically after 2012.



Table 6 Mediation analysis

	(1) R&D _{t+1}	(2) FC_SA _{t+1}	(3) R&D _{t+1}	(4) Tunneling _{t+1}	(5) R&D _{t+1}
Read_comp	0.0028*** (5.21)	0.0263*** (2.59)	0.0027*** (5.12)	- 0.0030*** (- 3.12)	0.0027*** (5.14)
FC_SA			0.0022*** (2.67)		
Tunneling					- 0.0133*** (- 2.59)
Constant	0.0064 (0.86)	- 4.5836*** (- 18.61)	0.0217*** (2.65)	0.0305* (1.92)	0.0068 (0.92)
Control variables	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	22,328	22,328	22,328	22,328	22,328
Adj. R ²	0.4999	0.4333	0.4993	0.4999	0.5002
Sobel Z		1.86*		1.99**	

This table performs a mediation analysis to examine whether annual report readability affects R&D investment through the channel of improving resources for innovation, as shown by mitigating financial constraints (*FC_SA*) and restraining tunneling behaviors (*Tunneling*). Column (1) is the first step of mediation analysis shown in Table 3. Columns (2) and (3) are mediation analysis's second and third steps under the financial constraints channel. Columns (4) and (5) are the second and third steps of mediation analysis under the tunneling channel. Sobel's z-statistics of testing mediating effects are reported in the bottom table. Definitions of variables are shown in the table in Appendix. Industry and year fixed effects and control variables are included, and the results are not reported. Robust t-statistics clustered by firm and year are reported in the parentheses beneath the coefficients. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively

fluency of investors and reduce disclosure processing costs (Reneekamp 2012; Bonsall et al. 2017). Thus, it alleviates investors' information asymmetry (Rjiba et al. 2021). On the one hand, with information asymmetry between the firm and investors decreasing, it will improve investors' evaluation of firms and further reduce financial constraints. Moreover, due to the higher information symmetry, the tunneling behaviors of large shareholders will be constrained when the external supervision from small investors is stronger. Moreover, low financial constraints and tunneling behaviors provide a better resource base that encourages managers to spend more time and cash on risky activities (Li 2011; Chen et al. 2017), enhancing R&D investment.

We use the SA index (*FC_SA*) to measure financial constraints because it avoids endogeneity influences from cash flows, leverage, and other financial factors (Hadlock and Pierce 2010).⁸ Financial constraints decrease if the value of *FC_SA* is high. Referring to Jiang et al. (2010), we proxy for the tunneling degree with the ratio of other receivables on total assets (*Tunneling*) that a greater value represents higher levels of tunneling. We implement a mediation analysis to verify our argument's channels. We herein perform a two-stage analysis: the first stage explores whether increased annual report readability reduces financial constraints and tunneling behaviors, and the second stage tests whether

lower financial constraints and tunneling promote R&D investment accompanied by the effect of readability.

In Columns (2) and (4) of Table 6, where the dependent variable is *FC_SA* and *Tunneling*, the coefficients on *Read_comp* are significantly positive and negative at the 1% level, respectively, confirming that more readable annual reports mitigate financial constraints and tunneling behaviors by reducing information asymmetry of investors. In Columns (3) and (5), we find that *Read_comp* and *FC_SA* are positively associated with *R&D*, and *Tunneling* is negatively associated with *R&D*, which suggests that lower financial constraints and tunneling motivation improve levels of R&D investment. Sobel's z-statistics are -1.86 and -1.99 for the cases of *FC_SA* and *Tunneling*, representing that mediation effects exist and are significant.

Collectively, the results in Table 6 confirm that annual report readability enhances the level of R&D investment through channels of relieving financial constraints and tunneling behaviors. Moreover, these results show the direct governance effect of disclosure readability on investors and managerial investment decisions.

Cross-sectional tests on the association between annual report readability and R&D investment

In this section, we conduct three cross-sectional tests to examine whether the change in information processing costs

⁸ SA index = -0.737* *Size* + 0.043* *Size*² - 0.04* *Age*.



Table 7 Cross-sectional test: managerial manipulating incentives and the effect of annual report readability

	(1)	(2)	(3)	(4)	(5)	(6)
	Managerial selfish		Internal control quality		Business complexity	
	High	Low	High	Low	High	Low
	R&D _{t+1}	R&D _{t+1}	R&D _{t+1}	R&D _{t+1}	R&D _{t+1}	R&D _{t+1}
Read_comp	0.0013* (1.92)	0.0016*** (3.31)	0.0016*** (3.11)	0.0009 (1.63)	0.0018*** (3.02)	0.0006 (1.22)
Constant	- 0.0099 (- 0.48)	- 0.0085 (- 0.48)	- 0.0094 (- 0.41)	- 0.0164 (- 0.77)	0.0084 (0.33)	- 0.0151 (- 0.77)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,385	10,943	11,494	10,833	13,377	8594
Adj. R ²	0.8039	0.7886	0.8192	0.7869	0.8165	0.8054
P-value for difference	0.150		0.080		0.020	

This table reports the results of the cross-sectional tests exploring whether the effect of annual report readability varies with the degree of managerial manipulating incentives. We measure managerial manipulating incentives using levels of free cash flows (*FCF*) in Columns (1) and (2), internal control quality (*ICQ*) in Columns (3) and (4), and business complexity (*BUC*) in Columns (5) and (6), where high and low groups are identified at their median. The *P*-value of testing the difference in the coefficients on *Read_comp* between two groups is reported in the bottom table. Definitions of variables are shown in the table in Appendix. Firm and year fixed effects and control variables are included, and the results are not reported. Robust t-statistics clustered by firm and year are reported in the parentheses beneath the coefficients. *** and * denote statistical significance at the 1% and 10% levels, respectively

for investors affects annual report readability on the effect of R&D investment. The results help assess whether the effect of annual report readability exhibits cross-sectional variations and whether our findings are reasonable, mitigating the concern from omitted factors.

Managerial incentives to manipulate readability

We examine whether the effect of annual report readability on R&D investment is conditional on managerial incentives to manipulate disclosure readability. Specifically, the qualitative disclosure as “soft information” with low supervision costs, managements have incentives to manipulate textual readability to conceal negative information about firms, which increases the processing costs of investors under the biased disclosure (Leuz and Wysocki 2016; Bushee et al. 2018; DeHaan et al. 2021). However, firms with little manipulating incentives will display normal complexity of the annual report based on their business transactions and regulatory policies, which helps investors better understand firms. Therefore, we expect the association between the annual report readability and corporate R&D investment to be enhanced when management’s manipulating incentives are low.

First, we measure manipulating incentives with selfish managerial levels represented by the ratio of the free cash flow to total assets (*FCF*). Jensen (1986) argued that more free cash flow will induce higher agency costs from

managers. A firm is more likely to manipulate disclosure if the free cash flow is high. Second, the high internal control quality can detect opportunistic behaviors of management which constrains the probability of manipulating disclosure. We proxy for low manipulating incentives with a high internal control index (*ICQ*).⁹ Third, a firm with diversified businesses may disclose more complex information typically, whereas manipulating incentives will decrease. Thus, we use the natural logarithm of one adding the number of major products (*BUC*) to measure business complexity. We repeat Eq. (1) by separating the sample at the median of *FCF*, *ICQ*, and *BUC*, respectively.

As shown in Table 7, the effect of annual report readability on R&D investment is more substantial for subsamples with low management manipulating incentives. There is a significant difference in this effect between the high and low subsamples for the case of internal control quality and business complexity. These results suggest that the influence of annual report readability on R&D investment depends on managerial manipulating incentives.

⁹ The internal control index data comes from the DIB Internal Control and Risk Management database. The internal control index describes a firm’s internal control quality from a strategic level, operational level, report credibility, legal and compliance, and asset safety. Therefore, a greater value represents higher internal control quality.



Table 8 Cross-sectional test: information processing ability of investor and the effect of annual report readability

	(1)	(2)	(3)	(4)	(5)	(6)
	Institutional ownership		Analyst followings		Media coverage	
	High	Low	High	Low	High	Low
	R&D _{t+1}	R&D _{t+1}	R&D _{t+1}	R&D _{t+1}	R&D _{t+1}	R&D _{t+1}
Read_comp	0.0010** (1.97)	0.0014*** (2.76)	0.0011* (1.88)	0.0017** (2.27)	0.0006 (1.03)	0.0018*** (2.67)
Constant	- 0.0282 (- 1.21)	0.0057 (0.23)	- 0.0482 (- 1.54)	0.0403 (1.37)	- 0.0061 (- 0.31)	- 0.0139 (- 0.61)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,586	10,742	10,394	9470	11,502	10,726
Adj. R ²	0.8187	0.8169	0.8494	0.7835	0.8246	0.7856
P-value for difference	0.100		0.080		0.000	

This table presents the results of the cross-sectional tests exploring whether the effect of annual report readability varies with investors' information processing ability. We measure the information processing ability of investors using institutional ownership (*Inshold*) in Columns (1) and (2), analyst followings (*Analyst*) in Columns (3) and (4), and media coverage (*Media*) in Columns (5) and (6), where high and low groups are identified at their median. The *P*-value of testing the difference in the coefficient on *Read_comp* between two groups is reported in the bottom table. Definitions of variables are shown in the table in Appendix. Firm and year fixed effects and control variables are included, and the results are not reported. Robust t-statistics clustered by firm and year are reported in the parentheses beneath the coefficients. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively

The effect of readability on the information processing ability of investors

We then explore whether the relationship between annual report readability and R&D investment varies with different investors. Theoretically, sophisticated investors, such as institutional investors and financial analysts, have a stronger ability for information process, analysis, and integration than non-sophisticated investors (Campbell et al. 2009). Consequently, small investors are more sensitive to the readability of the annual report because they face higher disclosure processing costs than large investors (Blankespoor et al. 2020). We thus expect that increased readability of the annual report will have a more significant effect on R&D investment for small investors since there is a different ability on disclosure processing.

We measure the professional investor based on institutional investor ownership (*Inshold*), analyst followings (*Analyst*), and media coverage (*Media*). *Inshold* equals the ratio of institutional investor shareholdings. *Analyst* is the natural logarithm of one adding the number of analyst reports for the firm over the fiscal year. *Media* is the natural logarithm of one adding the amount of media coverage for the firm over the fiscal year. Table 8 shows that the effect of annual report readability is more significant for the groups with low institutional ownership, low analyst following, and decreased media coverage (i.e., lower than the median). Consistent with our expectation, the lower processing ability of investors strengthens the governance effect of the annual report readability, which in turn optimizes the information environment of investors.

The spillover effects of peer annual report readability for R&D investment

We further explore the spillover effects of peer annual report readability for R&D investment. There is a significant “peer effect” phenomenon in firm behaviors. Roychowdhury et al. (2019) point out that management is motivated to learn incremental information related to investment decisions from peer firms in the same industry to reduce information uncertainty when internal information is limited. Some studies focus on how qualitative disclosures of peers affect the target firm's real investment. Durnev and Mangen (2020) reveal that MD&A tone provides incremental information related to an investment decision for the target firm, and management can improve investment efficiency through an information learning mechanism.

R&D investment is riskier than other real investments, requiring management to collect valuable information to avoid innovation failure (Hall 2002). More readable contexts help mitigate the target firm's processing costs, which are convenient for management reading and learning relative information about innovation from peers' annual reports. As a result, the lower information uncertainty encourages managers to seize the investment opportunity and improve R&D investment efficiency. In conclusion, we expect that the higher readability of peers' annual reports is positively associated with the R&D investment levels of the target firm.

We extend Eq. (1) by introducing peer readability (*Peer_comp*) as the explanatory variable and new control variables



Table 9 Spillover effects of annual report readability for R&D investment

	(1) R&D _{t+1}	(2) R&D _{t+1}	(3) R&D _{t+1}	(4) R&D _{t+1}
Peer_comp	0.0113*** (2.93)	0.0103*** (3.59)	0.0073*** (5.49)	0.0061*** (3.69)
Read_comp	0.0033*** (4.68)	0.0028*** (5.42)	0.0027*** (5.33)	0.0009** (2.12)
Constant	- 0.0140* (- 1.68)	- 0.0143 (- 1.56)	0.0075 (0.42)	0.0192 (0.77)
Control variables of target firms	No	Yes	Yes	Yes
Control variables of peer firms	No	No	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	No
Firm fixed effects	No	No	No	Yes
Observations	22,281	22,281	22,281	22,281
Adj. R ²	0.4374	0.5006	0.5057	0.8067

This table presents the results of the spillover effects of annual report readability for R&D investment, where Column (1) performs regression without control variables, Column (2) adds control variables of target firms similar to Table 3, and Column (3) further adds control variables of peer firms in the same industry, Column (4) adds firm-level fixed effects. Definitions of variables are shown in the table in Appendix. Fixed effects and control variables are included, and the results are not reported. Robust t-statistics clustered by firm and year are reported in the parentheses beneath the coefficients. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively

reflecting the characteristics of peer firms. Following Durnev and Mangen (2020), *Peer_comp* is the average of *Read_comp* in the same industry, excluding the target firm's readability. In Columns (1) and (2) of Table 9, the coefficients on *Peer_comp* are positive and significant. After controlling characteristics of peer firms and firm fixed effects, as shown in Columns (3) and (4), respectively, the results do not change. These suggest that annual report readability has positive peer spillover effects on corporate R&D investment.

Conclusion

Our study explores the association between annual report readability and corporate R&D investment. We assert that higher readability of the annual report facilitates investors to process disclosures fluently and reduce information asymmetry, which helps to relieve insufficient funds and encourages management to invest in R&D activities. By exploiting Chinese listed firms, we find that increased readability of the annual report enhances R&D investment levels. The mechanism test reveals that the effect of readability works through the channels of mitigating financial constraints and tunneling behaviors, confirming our argument's causality. The impact of readability on R&D investment varies with the degree of managerial manipulating incentives and investor ability to process disclosures. Moreover, the spillover effects of annual report readability for R&D investment are significant.

Overall, our research suggests that the annual report readability not only affects investors' decisions directly but also

influences managers' decisions on real investment indirectly, which response to the call of Roychowdhury et al. (2019), Blankespoor et al. (2020) how firm information disclosures impact management behaviors. Moreover, our study provides novel insights into the growing literature about the real economic effect of qualitative disclosure properties. Hence, simplifying qualitative information in the annual report benefits a firm in improving innovation efforts and quality.

We notice that our study has limitations. First, the construction of indicators for annual report readability in the Chinese setting is relatively subjective and cannot include all the features reflecting disclosure readability. Future research can identify textual complexity using convolutional neural network algorithms in deep learning and WordVec2 neural network language feature engineering. Second, in common with most studies (Biddle et al. 2009; Blankespoor et al. 2020), although we make great efforts to constrain the reverse causality, omitted variable, and selected bias, we cannot wholly resolve all the potential endogeneity concerns. Future research can focus on scientific approaches to mitigate endogeneity issues. Third, we only consider the single mechanism of annual report readability impacting R&D investment from the resource dependency perspective. As suggested by Roychowdhury et al. (2019), researchers can explore mechanisms based on information uncertainty in the future.

Appendix

See Table 10.



Table 10 Definitions of variables

R&D	The research and development expenditures scaled by total assets at the end of the year
Read_reverse	1-standardization of the number of adversative conjunctions in Chinese divided by total words in the annual report and multiplied by 100
Read_rare	1-standardization of the number of rare words in Chinese divided by the total words in the annual report and multiplied by 100
Read_jargon	1-standardization of the number of accounting terms in Chinese divided by total words in the annual report and multiplied by 100
Read_comp	The sum of Read_reverse, Read_rare, and Read_jargon
Size	Natural logarithm of total assets at the end of the year
Lev	The ratio of total liabilities over total assets at the end of the year
Cash	The sum of cash and cash equivalents over total assets at the end of the year
Netfin	The sum of the difference between net cash received from equity and bonds issued and cash paid for the purchase of financial assets scaled by total assets
Soe	Indicator variable that equals one for state-owned enterprises and zero otherwise
Age	Log(1 + number of years since the firm has been listed in the A-share market)
AH	Indicator variable that equals one for firms listed in A and H share and zero otherwise
Rev	The natural logarithm of major revenue over the fiscal year
Employ	Log(1 + number of employees in the firm at the end of the year)
KL	The ratio of net fixed assets scaled by total employees at the end of the year
BM	Book value of equity divided by market value of equity at the end of the year
Roa	The ratio of net income over average total assets during the fiscal year
DRoa	The change of net income over average total assets during the fiscal year
Ret	Buy-and-hold stock return over the fiscal year
Tobinq	The market value of equity divided by the book value of total assets at the end of the year
HHI	The sum of the square of the operating income divided by the total operating income in the same industry and takes its square root
GDP	The natural logarithm of per capita GDP of the province over the fiscal year
CFO_degree	Chief Financial Officers' educational background. Taking the value from 1 to 5, if the CFO experiences technical secondary school or below, technical school, bachelor's degree, master's degree, and doctoral degree, respectively
Gender	The ratio of male management over total management at the end of the year
Mage	Average of management ages at the end of the year
Degree	The number of management obtaining master's degrees and above scaled by total management
Paid	The number of management receiving salaries from listed firms scaled by total management
Share	Indicator variable that equals one if managements hold the firm's stocks and zero otherwise
Oversea	The number of management experiencing overseas divided by total management
Finback	The number of management having financial backs divided by total management
FC_SA	SA index equals to $(-0.737 \times Size + 0.043 \times Size^2 - 0.04 \times Age)$
Tunneling	The ratio of other receivables over total assets at the end of the year
FCF	The ratio of corporate free cash flows divided by total assets at the end of the year
ICQ	The internal control index coming from the DIB database
BUC	Log(1 + number of major products at the end of the year)
Inshold	The ratio of institutional investors' shareholdings over total shares at the end of the year
Analyst	Log(1 + number of analyst reports over the fiscal year)
Media	Log(1 + number of media coverage over the fiscal year)
Peer_comp	Annual and industry average for <i>Read_comp</i> excluding the target firm
Peer_size	Annual and industry average for <i>Size</i> , excluding the target firm
Peer_lev	Annual and industry average for <i>Lev</i> , excluding the target firm
Peer_cash	Annual and industry average for <i>Cash</i> , excluding the target firm
Peer_netfin	Annual and industry average for <i>Netfin</i> , excluding the target firm
Peer_soe	Annual and industry average for <i>Soe</i> , excluding the target firm
Peer_age	Annual and industry average for <i>Age</i> , excluding the target firm
Peer_AH	Annual and industry average for <i>AH</i> , excluding the target firm
Peer_rev	Annual and industry average for <i>Rev</i> , excluding the target firm



Table 10 (continued)

R&D	The research and development expenditures scaled by total assets at the end of the year
Peer_employ	Annual and industry average for <i>Employ</i> , excluding the target firm
Peer_KL	Annual and industry average for <i>KL</i> , excluding the target firm
Peer_BM	Annual and industry average for <i>BM</i> , excluding the target firm
Peer_roa	Annual and industry average for <i>Roa</i> , excluding the target firm
Peer_Droa	Annual and industry average for <i>DRoa</i> , excluding the target firm
Peer_ret	Annual and industry average for <i>Ret</i> , excluding the target firm
Peer_tobinq	Annual and industry average for <i>Tobinq</i> , excluding the target firm

Author contributions Y-GD makes efforts in conceptualization, investigation, data analysis, methodology, writing, reviewing, and editing. SL takes charge of conceptualization, reviewing, and supervision. N-TK makes efforts in methodology, reviewing and editing, and validation. DL contributes to data curation, formal analysis, investigation, reviewing, and revision.

Funding This work was supported by the National Social Science Foundation of China under Grant 23BJY250 and the China Postdoctoral Science Foundation under Grant 2022M720515.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Albuquerque, A., and J.L. Zhu. 2019. Has Section 404 of the Sarbanes-Oxley Act discouraged corporate investment? New evidence from a natural experiment. *Management Science* 65 (7): 3423–3446.
- Alter, A.L., and D.M. Oppenheimer. 2009. Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review* 13 (3): 219–235.
- Ascioglu, A., S.P. Hegde, and J.B. McDermott. 2008. Information asymmetry and investment–cash flow sensitivity. *Journal of Banking & Finance* 32 (6): 1036–1048.
- Bae, G.S., S.U. Choi, D.S. Dhaliwal, and P.T. Lamoreaux. 2017. Auditors and client investment efficiency. *The Accounting Review* 92 (2): 19–40.
- Bajaj, A., L.N. Leonard, L. Sun, and Z. Xing. 2023. Corporate social responsibility and annual report reading difficulty. *Review of Quantitative Finance and Accounting* 60: 1393–1428.
- Balakrishnan, K., R. Watts, and L. Zuo. 2016. The effect of accounting conservatism on corporate investment during the global financial crisis. *Journal of Business Finance & Accounting* 43 (5–6): 513–542.
- Battalio, R.H., A. Lerman, J. Livnat, and R.R. Mendenhall. 2012. Who, if anyone, reacts to accrual information? *Journal of Accounting and Economics* 53 (1–2): 205–224.
- Biddle, G.C., and G. Hilary. 2006. Accounting quality and firm-level capital investment. *The Accounting Review* 81 (5): 963–982.
- Biddle, G.C., G. Hilary, and R.S. Verdi. 2009. How does financial reporting quality relate to investment efficiency? *Journal of Accounting and Economics* 48 (2–3): 112–131.
- Blankespoor, E., E. deHaan, and I. Marinovic. 2020. Disclosure processing costs, investors' information choice, and equity market outcomes: A review. *Journal of Accounting and Economics* 70 (2–3): 101344.
- Bonsall, S.B., IV., A.J. Leone, B.P. Miller, and K. Rennekamp. 2017. A plain English measure of financial reporting readability. *Journal of Accounting and Economics* 63 (2–3): 329–357.
- Brochet, F., G.S. Miller, P. Naranjo, and G. Yu. 2019. Managers' cultural background and disclosure attributes. *The Accounting Review* 94 (3): 57–86.
- Bushee, B.J. 1998. The influence of institutional investors on myopic R&D investment behavior. *The Accounting Review* 73 (3): 305–333.
- Bushee, B.J., I.D. Gow, and D.J. Taylor. 2018. Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research* 56 (1): 85–121.
- Bushman, R.M., J.D. Piotroski, and A.J. Smith. 2011. Capital allocation and timely accounting recognition of economic losses. *Journal of Business Finance & Accounting* 38 (1–2): 1–33.
- Campbell, J.Y., T. Ramadorai, and A. Schwartz. 2009. Caught on tape: Institutional trading, stock returns, and earnings announcements. *Journal of Financial Economics* 92 (1): 66–91.
- Chen, C., D. Hanlon, M. Khedmati, and J. Wake. 2023. Annual report readability and equity mispricing. *Journal of Contemporary Accounting & Economics* 19 (3): 100368.
- Chen, C., D. Young, and Z. Zhuang. 2013. Externalities of mandatory IFRS adoption: Evidence from cross-border spillover effects of financial information on investment efficiency. *The Accounting Review* 88 (3): 881–914.
- Chen, F., O.K. Hope, Q. Li, and X. Wang. 2011. Financial reporting quality and investment efficiency of private firms in emerging markets. *The Accounting Review* 86 (4): 1255–1288.
- Chen, K.J., Y.L. Kang, Q.Q. Wan, and Q. Liu. 2017. Can outside blockholders promote corporate innovation? An empirical analysis from the perspective of exit threats. *Nankai Business Review* 24 (3): 202–212 (in Chinese).
- Cheng, M., D. Dhaliwal, and Y. Zhang. 2013. Does investment efficiency improve after the disclosure of material weaknesses in internal control over financial reporting? *Journal of Accounting and Economics* 56 (1): 1–18.
- Cheng, X.S., Y.C. Tan, and J.M. Liu. 2012. The non-financial information, the external financing, and the investment efficiency: A study based on the constraint of external systems. *Journal of Management World* 7: 137–150 (in Chinese).
- Cheung, Y.L., P.R. Rau, and A. Stouraitis. 2006. Tunneling, propping, and expropriation: Evidence from connected party transactions in Hong Kong. *Journal of Financial Economics* 82 (2): 343–386.
- Chircop, J., D.W. Collins, L.H. Hass, and N.Q. Nguyen. 2020. Accounting comparability and corporate innovative efficiency. *The Accounting Review* 95 (4): 127–151.
- Christensen, H.B., E. Floyd, L.Y. Liu, and M. Maffett. 2017. The real effects of mandated information on social responsibility in



- financial reports: Evidence from mine-safety records. *Journal of Accounting and Economics* 64 (2–3): 284–304.
- Dalwai, T., G. Chinnasamy, and S.S. Mohammadi. 2021. Annual report readability, agency costs, firm performance: An investigation of Oman's financial sector. *Journal of Accounting in Emerging Economies* 11 (2): 247–277.
- De Simone, L., and M. Olbert. 2022. Real effects of private country-by-country disclosure. *The Accounting Review* 97 (6): 201–232.
- DeHaan, E., Y. Song, C. Xie, and C. Zhu. 2021. Obfuscation in mutual funds. *Journal of Accounting and Economics* 72 (2–3): 101429.
- Delgado-Domonkos, L., and A.M. Zeng. 2023. Reporting quality, stock market integration, and investment efficiency in Latin America. *Borsa Istanbul Review* 23 (4): 779–803.
- Durnev, A., and C. Mangen. 2009. Corporate investments: Learning from restatements. *Journal of Accounting Research* 47 (3): 679–720.
- Durnev, A., and C. Mangen. 2020. The spillover effects of MD&A disclosures for real investment: The role of industry competition. *Journal of Accounting and Economics* 70 (1): 101299.
- Dyregang, S.D., J.L. Hoopes, and J.H. Wilde. 2016. Public pressure and corporate tax behavior. *Journal of Accounting Research* 54 (1): 147–186.
- Feng, M., C. Li, S.E. MacVay, and H. Skaife. 2015. Does ineffective internal control over financial reporting affect a firm's operations? Evidence from firms' inventory management. *The Accounting Review* 90 (2): 529–557.
- Gangadharan, V., and L. Padmakumari. 2023. Annual report readability and stock return synchronicity: Evidence from India. *Cogent Economics & Finance* 11 (1): 2186034.
- Goldstein, I., and L. Yang. 2019. Good disclosure, bad disclosure. *Journal of Financial Economics* 131 (1): 118–138.
- Goldstein, I., S. Yang, and L. Zuo. 2023. The real effects of modern information technologies: Evidence from the EDGAR implementation. *Journal of Accounting Research* 61 (5): 1699–1733.
- Hadlock, C.J., and J.R. Pierce. 2010. New evidence on measuring financial constraints: Moving beyond the K.Z. index. *The Review of Financial Studies* 23 (5): 1909–1940.
- Hafner, M., and D.A. Stapel. 2010. Information to go: Fluency enhances the usability of primed information. *Journal of Experimental Social Psychology* 46 (1): 73–84.
- Hainmueller, J. 2012. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* 20 (1): 25–46.
- Hall, B.H. 2002. The financing of research and development. *Oxford Review of Economic Policy* 18 (1): 35–51.
- Heitzman, S., and M. Huang. 2019. Internal information quality and the sensitivity of investment to market prices and accounting profits. *Contemporary Accounting Research* 36 (3): 1699–1723.
- Hope, O., and W.B. Thomas. 2008. Managerial empire building and firm disclosure. *Journal of Accounting Research* 46 (3): 591–626.
- Huang, X., S.H. Teoh, and Y. Zhang. 2014. Tone management. *The Accounting Review* 89 (3): 1083–1113.
- Jensen, M.C. 1986. Agency costs of free cash flow, corporate finance, and takeovers. *The American Economic Review* 76 (2): 323–329.
- Jensen, M.C., and W.H. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3 (4): 305–360.
- Jiang, G., C.M.C. Lee, and H. Yue. 2010. Tunneling through intercorporate loans: The China experience. *Journal of Financial Economics* 98 (1): 1–20.
- Jiang, Y., and Z. Wang. 2019. Readability of board report, institutional environment and analyst forecast—Based on empirical from Chinese listed companies. *The Theory and Practice of Finance and Economics* 40 (3): 88–93 (in Chinese).
- Ju, X.S., D. Lo, and Y.H. Yu. 2013. Financing constraints, working capital management and the persistence of firm innovation. *Economic Research Journal* 48 (1): 4–16 (in Chinese).
- Ke, B., and X. Zhang. 2021. Does public enforcement work in weak investor protection countries? Evidence from China. *Contemporary Accounting Research* 38 (2): 1231–1273.
- Kim, C., K. Wang, and L. Zhang. 2019. Readability of 10-K reports and stock price crash risk. *Contemporary Accounting Research* 36 (2): 1184–1216.
- Koh, P.S., and D.M. Reeb. 2015. Missing R&D. *Journal of Accounting and Economics* 60 (1): 73–94.
- Kuo, N.T., S. Li, Y.G. Du, and C.F. Lee. 2022. Does individual auditor quality contribute to firm value? Evidence from the market valuation on corporate cash holdings. *International Review of Economics & Finance* 79: 135–153.
- La Porta, R., F. Lopez-de-Silanes, and A. Shleifer. 1999. Corporate ownership around the world. *The Journal of Finance* 54 (2): 471–517.
- Laux, V., and K. Ray. 2020. Effects of accounting conservatism on investment efficiency and innovation. *Journal of Accounting and Economics* 70 (1): 101319.
- Lawrence, A. 2013. Individual investors and financial disclosure. *Journal of Accounting and Economics* 56 (1): 130–147.
- Le Maux, J., and N. Smali. 2021. Annual report readability and corporate bankruptcy. *Journal of Applied Business Research* 37 (3): 73–80.
- Lee, C., and C. Zhu. 2022. Active funds and bundled news. *The Accounting Review* 97 (1): 315–339.
- Lee, Y.J. 2012. The effect of quarterly report readability on information efficiency of stock prices. *Contemporary Accounting Research* 29 (4): 1137–1170.
- Leuz, C., and P.D. Wysocki. 2016. The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of Accounting Research* 54 (2): 525–622.
- Li, D. 2011. Financial constraints, R&D investment, and stock returns. *The Review of Financial Studies* 24 (9): 2974–3007.
- Li, F. 2010. The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach. *Journal of Accounting Research* 48 (5): 1049–1102.
- Liao, Z. Q. 1986. “Connective elements in modern Chinese texts.” *Studies of the Chinese Language* 6: (in Chinese).
- Liu, J., W. Shi, C. Zeng, and G. Zhang. 2023. Does public firms' mandatory Ifrs reporting crowd out private firms' capital investment? *Journal of Accounting Research* 61 (4): 1263–1312.
- Lo, K., F. Ramos, and R. Rogo. 2017. Earnings management and annual report readability. *Journal of Accounting and Economics* 63 (1): 1–25.
- Loughran, T., and B. McDonald. 2014. Measuring readability in financial disclosures. *The Journal of Finance* 69 (4): 1643–1671.
- Luo, J.H., X. Li, and H.Y. Chen. 2018. Annual report readability and corporate agency costs. *China Journal of Accounting Research* 11 (3): 187–212.
- McNichols, M.F., and S.R. Stubben. 2008. Does earnings management affect firms' investment decisions? *The Accounting Review* 83 (6): 1571–1603.
- Miller, B.P. 2010. The effects of reporting complexity on small and large investor trading. *The Accounting Review* 85 (6): 2107–2143.
- Nadeem, M. 2022. Board gender diversity and managerial obfuscation: Evidence from the readability of narrative disclosure in 10-K reports. *Journal of Business Ethics* 179 (1): 153–177.
- O'Hara, M. 1999. Making market microstructure matter. *Financial Management* 28 (2): 83–90.
- Park, K. 2018. Financial reporting quality and corporate innovation. *Journal of Business Finance & Accounting* 45 (7–8): 871–894.



- Pretorius, E.J. 2006. The comprehension of logical relations in expository texts by students who study through the medium of ESL. *System* 34 (3): 432–450.
- Rennekamp, K. 2012. Processing fluency and investors' reaction to disclosure readability. *Journal of Accounting Research* 50 (5): 1319–1354.
- Rjiba, H., S. Saadi, S. Boubaker, and X.Y. Ding. 2021. Annual report readability and the cost of equity capital. *Journal of Corporate Finance* 67: 101902.
- Roychowdhury, S., N. Shroff, and R.S. Verdi. 2019. The effects of financial reporting and disclosure on corporate investment: A review. *Journal of Accounting and Economics* 68 (2–3): 101246.
- Roychowdhury, S. 2006. Earnings management through real activities manipulation. *Journal of Accounting and Economics* 42 (3): 335–370.
- Schwarz, N. 2004. Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology* 14 (4): 332–348.
- Shroff, N. 2017. Corporate investment and changes in GAAP. *Review of Accounting Studies* 22 (1): 1–63.
- Shroff, N. 2020. Real effects of PCAOB international inspections. *The Accounting Review* 95 (5): 399–433.
- Shroff, N., R.S. Verdi, and G. Yu. 2014. Information environment and the investment decisions of multinational corporations. *The Accounting Review* 89 (2): 759–790.
- Stein, J.C. 2003. Agency, information and corporate investment. *Handbook of the Economics of Finance* 1: 111–165.
- Sun, L., G. Johnson, and W. Bradley. 2022a. CEO power and annual report reading difficulty. *Journal of Contemporary Accounting & Economics* 18 (2): 100315.
- Sun, W. Z., J. W. Zhu, X. H. Wang. 2022b. "Do board secretaries influence annual report readability?" *Pacific Accounting Review* (ahead-of-print).
- Tan, H.T., E.Y. Wang, and G.S. Yoo. 2019. Who likes jargon? The joint effect of jargon type and industry knowledge on investors' judgments. *Journal of Accounting and Economics* 67 (2–3): 416–437.
- Vieru, M., J. Perttunen, and H. Schadewitz. 2006. How investors trade around interim earnings announcements. *Journal of Business Finance & Accounting* 33 (1–2): 145–178.
- Wang, K.M., H.J. Wang, D.D. Li, and X.Y. Dai. 2018. Complexity of annual report and management self-interest: Empirical evidence from Chinese listed firms. *Journal of Management World* 34 (12): 120–132 (in Chinese).
- Wang, Y.C., K. He, L.M. Wan, and X. Xie. 2020. Readability of annual report and stock liquidity: A perspective from text mining. *Securities Market Herald* 7: 61–71 (in Chinese).
- Xu, H., M. Dao, J. Wu, and H. Sun. 2022. Political corruption and annual report readability: Evidence from the United States. *Accounting and Business Research* 52 (2): 166–200.
- You, H., and X.J. Zhang. 2009. Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies* 14 (4): 559–586.
- Yu, M.Y., K.F. Lyu, and Y.P. Ruan. 2022. Information blurring or truth report: Corporate tax avoidance and the complexity of the annual report". *Economic Science* 3: 112–126 (in Chinese).
- Zhong, R.I. 2018. Transparency and firm innovation. *Journal of Accounting and Economics* 66 (1): 67–93.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

