Annual report readability and corporate agency costs

Jin-hui Luo *, Xue Li, Huayang Chen

School of Management, Xiamen University, PR China

ARTICLE INFO

ABSTRACT

Annual reports are the main sources of information for outside investors' investment decisions and enable shareholders to supervise the management. Difficulties with the readability of these reports may therefore have serious consequences. Using 19,221 firm-year observations of Chinese A-share listed firms from 2001 to 2015, we investigate the association between annual report readability and corporate agency costs, where readability is proxied by report file length and/or file size. We find that firms with better annual report readability experience lower agency costs, and the negative association between readability and agency costs is more pronounced in firms with higher external audit quality, internal control quality or analyst coverage. These results hold after several robustness checks. The positive effect of annual report readability is stronger in private firms than in state-owned enterprises, and becomes stronger after the implementation of new accounting standards in 2007. Readable annual reports can help in monitoring corporate insiders' opportunistic behavior and thus reduce agency costs.

1. Introduction

Readability is an important attribute of textual information and has been examined extensively in various fields. Research into the importance of readability has been conducted in areas including the military,

medicine and law (Bonsall and Miller, 2017; DuBay, 2004). The value of the information in the text can only be fully realized with a high level of readability. The issue of readability has increasingly become a focus in capital markets in recent years. The changes in requirements for corporate information disclosure have resulted in a significant increase in the quantity of information disclosed and attracted the attention of regulators and investors. In 1998, the U.S. Securities and Exchange Commission (SEC) issued the *Plain English Disclosure* and *A Plain English Handbook: How to Create Clear SEC Disclosure Documents*, which aim to help public firms improve the readability of their disclosed information and to help investors better understand the information.

Theoretically, information disclosure announcements such as annual reports are an important communication bridge between management and outside stakeholders (e.g., shareholders) in joint-stock companies due to the separation of ownership and management. Outside investors and minority shareholders can learn about a company's financial status, performance and cash flow through its annual reports and thus evaluate the prospects for corporate growth and management competence. However, the increasing deterioration in levels of readability has adversely affected the communication function of corporate annual reports in recent years. One consequence of the improvements in the information disclosure systems of capital markets is that the information disclosed in annual reports includes many professional terms and specific notes and also much non-financial information, which makes them increasingly complicated and hard to understand in listed companies, particularly those in China. One major trend is that the length of corporate annual reports is increasing (see Fig. 1), and thus the readability of these reports has become an intractable problem, particularly considering the current explosion in the volume of information and shallow network reading. The economic consequences of annual report readability have therefore attracted the attention of scholars and regulators alike.

Many studies find that annual report readability can affect the quality of resulting information. For example, poor readability may result in serious earnings management problems, poor earnings persistence, low analyst forecast quality, weak market reactions to annual reports and a high risk of stock price crashes (Ertugrul et al., 2017; Kim et al., 2017a, 2017b; Lang and Stice-Lawrence, 2015; Lawrence, 2013; Li, 2008; Lo et al., 2017; Rennekamp, 2012). Most of these studies are based on the context of the U.S. or other English-speaking countries, and few have explored the economic consequences of annual report readability in China. You and Yi (2010) and Ji et al. (2016) initially discuss the readability problem of internal control reports and corporate social responsibility reports in China. Studies also mainly focus on the direct effects of annual report readability, such as the transmission or hiding of information, and few have explored the potential impact of readability on stakeholders in achieving their economic goals. In terms of the usefulness of annual reports in decision making, their core value is to help investors make scientific economic decisions. These reports are the main information sources through which shareholders learn about management competence and firm perfor-
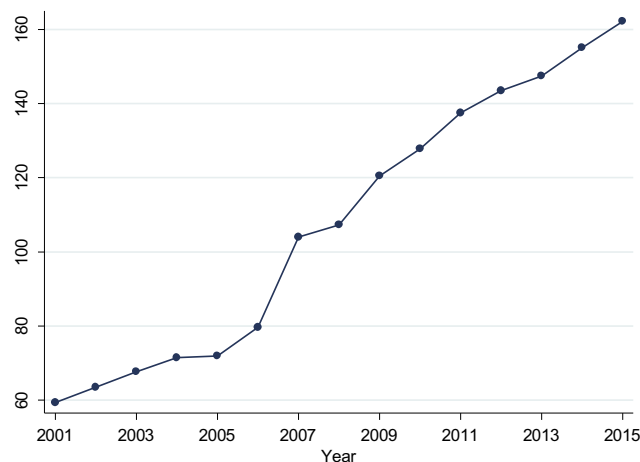

Fig. 1. The average number of pages of the annual reports of China's listed companies between 2001 and 2015.

mance, so they can better supervise and motivate the management. The question then emerges of whether and how annual report readability plays a role in governance, but few studies, if any, investigate this issue.

Using 19,221 firm-year observations of Chinese A-share listed firms from 2001 to 2015, this study investigates the association between annual report readability and corporate agency costs, where readability is proxied by hand-collected file length and/or the digital file size of annual reports. We find that (1) firms with higher levels of readability in their annual reports suffer from lower agency costs resulting from interest conflicts both between shareholders and managers and between large and minority shareholders, indicating that annual report readability contributes to the improvement of monitoring corporate insiders; and (2) the positive governance effect of annual report readability is more pronounced in firms with higher external audit quality/ internal control quality or analyst coverage, suggesting that external auditors, internal control system and professional analysts can strengthen the effect of annual report readability on reducing agency costs. These findings hold for a series of robustness checks, including the adoption of simultaneous and firm fixed effects models to tackle endogeneity problems and alternative measurements for key variables. Furthermore, we find that the governance role of annual report readability is stronger in private firms than in state-owned enterprises and becomes stronger after the implementation of the new accounting standards in 2007.

This study makes several contributions to the literature. First, to the best of our knowledge, our study is among the first to take advantage of a large sample to examine the effect of annual report readability on corporate agency costs, thereby contributing to the literature on annual report readability, the economic consequences of which have become a recent focus. Second, most research on the readability of annual reports originates from English-speaking countries such as the U.S., and our study extends this research into the Chinese context. In this study, we construct readability indexes of Chinese annual reports, which can be a valuable reference for future research into readability in China. Third, we find that the readability of annual reports is significant in corporate agency problems. Alleviating agency problems and improving corporate governance have long been concerns in China's capital markets, and thus our study contributes to and extends the field of corporate governance. Finally, plain disclosure has become an important reform direction when disclosing information in capital markets worldwide, so our findings have implications for this reform and provide insights for regulators.

2. Literature review and hypotheses development

2.1. Literature review

According to Chall (1958), readability refers to a combination of various factors involving interest, legibility and ease of understanding for readers. Yan and Sun (2002) argue that readability refers to the level of reading difficulty of an article. Readers can generate interest from readable articles and *vice versa*. In addition to the application of the text analysis method in corporate finance, recent empirical studies use large sample data to investigate the economic consequences of annual report readability (Loughran and McDonald, 2014).

Annual reports are one of the main channels through which companies communicate with external stakeholders, and readability is an important feature. Many scholars have explored potential determinants of annual report readability in recent years. Unlike previous studies based on small volumes of sample data (e.g., Baker and Kare, 1992; Barnett and Leoffler, 1979; Courtis, 1986; Smith and Smith, 1971; Subramanian et al., 1993), Li (2008) is the first to explore the association between annual report readability and corporate current and future performance by using a large volume of sample data from U.S. capital markets. He finds that the annual reports of firms with lower current earnings are less readable, and firms providing annual reports that are more readable are linked to more earnings persistence. Ajina et al. (2016) further find that companies with earnings manipulation tend to issue less readable annual reports to hide their manipulation behavior. Similarly, Lo et al. (2017) focus on the readability of the management and discussion and analysis (MD&A) of annual reports, and find that firms with strong motivations to manipulate earnings have more complex MD&A. These findings suggest that management can strategically manipulate the readability of annual reports, which contributes to the literature on strategic disclosure theory (Kim et al., 2017a; Schrand and Walther, 2000). In addition, Nelson and Pritchard (2007) investigate the effect of litigation risk to firms on

information readability and find that firms with higher risk of shareholder lawsuits have more readable disclosure documents.

Many scholars have explored the economic consequences of readability. Biddle et al. (2009) find a positive relationship between annual report readability and corporate investment efficiency, indicating that firms with better annual report readability face less serious problems of overinvestment and underinvestment. Kim et al. (2017a) and Hwang and Kim (2017) find that readability can significantly affect firm value. The trust investors place in the information disclosed by firms decreases if annual reports are less readable, thereby doing harm to firm value. Bonsall and Miller (2017) and Ertugrul et al. (2017) find that firms with less readable 10-K files have lower credit ratings, stricter loan contract terms and greater risk of stock price crashes. Similarly, Kim et al. (2017b) provide evidence that firms with poorer readability of their annual reports have a higher risk of future stock price crashes. Unlike studies that use the Fog Index or other related indexes as their measurement of readability, Loughran and McDonald (2014) use the digital file size of 10-K filings. They demonstrate that larger file sizes are linked to higher earnings volatility and lower accuracy of earnings forecasts. Asay et al. (2016) extend the literature by conducting an experimental study and find that investors search for information from outside resources when corporate disclosed documents are less readable. Lang and Stice-Lawrence (2015) use a large volume of sample data from non-American companies and find that improving information disclosure quality is beneficial to focal firms.

In addition to consequences at the firm level, the readability of annual reports also has significant market-level effects. For example, You and Zhang (2009), Miller (2010), Rennekamp (2012), Lee (2012) and Lawrence (2013) reveal that investors react more weakly to less readable disclosure announcements, indicating that readability may affect capital market efficiency. If investors need more time and cost to extract value-related information from longer and more complex documents disclosed by firms, less trading volume and slower market responses will result (Bloomfield, 2002; Grossman and Stiglitz, 1980; Hirshleifer and Teoh, 2003). Lehavy et al. (2011) and Bozanin and Thevenot (2015) investigate the effect of readability on analysts' behavior and find that analysts need more time and energy to follow less readable annual reports and provide forecast reports with lower forecast accuracy and higher forecast volatility. Likewise, Qiu et al. (2016) find that in China, even if analysts pay more attention to firms with less readable annual reports, they cannot improve the information content or quality of their forecast reports for these firms. Bonsall and Miller (2017) find that annual report readability also affects the behavior of bond rating agencies, as less readable annual reports result in higher divergence in the rating scores of bond rating agencies. De Franco et al. (2015) focus on the readability of analyst forecast reports and find that better readability can help investors reduce their costs of information collection and thus affect the stock trading volume. Tan et al. (2015) use the experimental research method to examine the effect of readability on investors' judgment. They find that better readability can improve investors' understanding of current performance, if the performance is not consistent with earnings preannouncements, and thus improves the investors' judgment of corporate future performance. In addition, Tan et al. (2014) find that investors are more subject to the tone of management when the readability of information disclosure announcements is poorer.

Taken together, studies suggest that serious economic consequences for shareholders can result if management strategically manipulates the readability of annual reports, such as by deliberately reducing readability to obfuscate and hide bad news. These types of strategic disclosure behavior are in essence managerial opportunism, and thus inevitably result in agency costs to the focal firms. However, few studies, if any, have investigated the effect of annual report readability on agency costs. In this study, we therefore aim to investigate the effect of annual report readability on corporate agency costs in the context of China.

2.2. Hypotheses development

In emerging markets such as China, concentrated ownership structures are common in listed companies. This can lead to two main types of agency problem: agency conflicts between shareholders and managers and between large and minority shareholders (Shleifer and Vishny, 1997). In this study, we argue that annual report readability can help to alleviate both types of agency problem and thus reduce the respective agency costs.

First, firms with annual reports that are more readable have higher levels of information disclosure quality, which can reduce the degree of information asymmetry faced by shareholders and help them better supervise management. The fundamental cause of agency problems is the information asymmetry between the principals and agents, which prevents the principals from properly evaluating the agents' competence and efforts. Therefore, various information disclosure mechanisms need to be set up to alleviate the problem of information asymmetry faced by both parties (Akerlof, 1970; Hart, 1995). Information announcements (e.g., annual reports) are the main communication channels and mechanisms for revealing information in listed firms. Outside (minority) shareholders usually depend on annual reports to supervise and motivate management (Healy and Palepu, 2001; Jensen and Meckling, 1976). However, if corporate annual reports are less readable, shareholders spend more time and costs on processing the information, which impedes them in extracting value-related information from the reports (Rennekamp, 2012). Shareholders may then have less exact knowledge of the competence and performance of management, which may weaken management supervision and exacerbate agency conflicts between principles and agents. Similarly, as majority shareholders can override the interests of minority shareholders, mainly through unfair related party transactions (Jiang et al., 2010), less readable annual reports may hamper minority shareholders in extracting accurate information about such transactions, thus reducing their abilities to effectively supervise the expropriation of majority shareholders.

Second, highly readable annual reports facilitate the transmission of valuable information to potential investors, which in turn may bring about strong market pressure on corporate insiders. According to the (weak) efficient market hypothesis, value-related information from disclosure announcements (e.g., annual reports) is quickly reflected in the stock price, which may lead investors' trading behavior and thus optimize the resource allocation in capital markets (Malkiel and Fama, 1970). This market mechanism can put significant market supervision pressure on corporate insiders and thus help to constrain their shirking and other opportunistic behaviors. Otherwise, poor performance will result in management turnover and even hostile takeover threats from the market (Parrino et al., 2003). Studies show that if annual reports are less readable, the quality of analyst forecasts is lower and investors will reduce their dependence on the reports and are less willing to purchase stocks of firms providing less readable annual reports (Lawrence, 2013; Lehavy et al., 2011; Rennekamp, 2012; Qiu et al., 2016). Poor readability thus prevents annual reports from effectively transmitting value-related information to the market, thereby weakening the market pressure faced by corporate insiders and exacerbating corporate agency costs.

In summary, annual reports that are more readable are beneficial in reducing the information asymmetry faced by shareholders and potential investors, and thus can improve the supervision of corporate insiders and reduce agency costs. Therefore, we put forward our first testable hypothesis as follows:

Hypothesis 1. *Ceteris paribus*, firms with higher annual report readability have lower corporate agency costs.

Li (2008) suggests that the quality of information disclosed by companies can be divided into disclosure quality and earnings quality. The readability of annual reports belongs to disclosure quality, while the value of disclosure quality depends on earnings quality. Therefore, we anticipate that the earnings quality of annual reports will reinforce the effect of annual report readability on corporate agency costs. In listed companies, an external independent audit is a main mechanism for guaranteeing the earnings quality of annual reports. External auditors provide professional auditing service and issue audit reports to reflect the earnings quality of their client firms' annual reports. Many studies have shown that higher external audit quality is associated with weaker earnings management based on accruals and/or real activities and with higher annual report earnings quality (Becker et al., 1998; Francis and Yu, 2009). We thus put forward our second hypothesis as follows:

Hypothesis 2. External audit quality strengthens the negative relationship between annual report readability and corporate agency costs.

In addition to external independent audit, internal control system is another important mechanism to ensure the earnings quality of corporate annual reports. This mechanism has increasingly become a focus of attention, particularly since the passing of the Sarbanes–Oxley Act in the U.S. in 2002. Regulators worldwide have issued a series of standards on the establishment of corporate internal control system, aiming to both improve corporate internal control and exert effective control over the production of financial reports,

thereby improving the quality of corporate disclosed information. In China, both the Shanghai and Shenzhen Stock Exchanges issued "*Internal Control Guidelines for Listed Companies*" in 2006, and the Ministry of Finance along with five other ministries and commissions jointly issued the "*Basic Rules of Enterprise Internal Control*" in 2008. Studies provide consistent evidence that internal control system can inhibit earnings management and improve earnings persistence and earnings quality (Doyle et al., 2007; Ashbaugh-Skaife et al., 2008; Ye et al., 2015). Therefore, as firms with higher internal control quality are likely to have higher earnings quality, we predict a stronger effect of annual report readability on reducing corporate agency costs in these firms. We thus put forward the third hypothesis:

Hypothesis 3. Internal control quality strengthens the negative relationship between annual report readability and corporate agency costs.

Given their earnings quality and disclosure quality characteristics, the information value of annual reports is only relevant when stakeholders use the information to make business decisions. Annual reports can thus only help reduce corporate agency costs if their users play a governance role. Security analysts are important information intermediaries in capital markets and are primary users of corporate annual reports. External minority investors and institutional investors usually acquire information regarding target companies' performance and development prospects from analyst reports (Healy and Palepu, 2001; Yezegel, 2015; Zhu et al., 2007). A higher level of analyst coverage then represents a higher probability that focal firms' annual reports will be used by analysts, and that they will be used more frequently. The readability of annual reports, which improves the efficiency of information transmission, then becomes more significant. Studies also find that annual report readability may affect analysts' forecast behavior. Analysts need to invest more time and effort to interpret less readable annual reports, but still fail to provide high quality earnings forecasts (Lehavy et al., 2011; Qiu et al., 2016). Accordingly, we put forward our fourth hypothesis:

Hypothesis 4. Analyst coverage strengthens the negative relationship between annual report readability and corporate agency costs.

3. Research design

3.1. Sample and data

Our initial sample includes 27,550 firm-year observations of all Chinese A-share listed companies during the period 2001–2015. To reduce the influence of abnormal observations, we successively exclude 2254 firm-year observations where the focal firms also issued B- or H-share stocks, 1751 where the focal firms have ST or *ST status, 228 for financial firms, 74 for firms issuing debts exceeding asset value and 4032 firm-year observations for firms with missing data. We thus have a final sample of 19,211 firm-year observations, and the sample distribution by year and industry is shown in Table 1.

We hand-collected data on the file length and file size of corporate annual reports from the CNINFO website (www.cninfo.com.cn), the officially designated disclosure website for information announcements of listed companies in China. The data on internal control quality comes from the DIB Internal Control and Risk Management database (www.ic-erm.com). All other data including agency costs, financial indicators and corporate governance comes from the China Stock Market & Accounting Research (CSMAR) database.

3.2. Measures

3.2.1. Dependent variables

Self-serving managers tend to maximize their utility through opportunistic behavior including inefficient investment and excessive perk consumption, so we follow the method of Ang et al. (2000) and use the operating expense ratio, denoted by *Agency_cost1*, to measure the agency costs between shareholders and managers. This can capture managerial perk consumption and thus reflects the efficiency in controlling agency costs between shareholders and managers. Specifically, the operating expense ratio equals the ratio of the

Table1
Sample distribution by year and industry.

Industry code															Year	Total
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	
A	12	24	22	24	29	26	23	23	24	29	41	41	46	45	48	457
B	4	11	10	12	15	17	16	18	21	22	29	34	31	30	32	302
C0	21	36	42	42	42	45	39	48	52	52	67	78	83	84	83	814
C1	25	40	39	39	49	50	46	54	52	50	71	81	79	79	100	854
C2	1	2	2	1	1	2	1	2	4	4	6	10	12	12	30	90
C3	8	18	15	17	20	18	19	24	27	32	40	43	49	49	98	477
C4	69	101	98	111	118	117	115	118	138	135	179	213	241	235	241	2229
C5	15	21	24	28	31	34	34	48	52	55	91	111	130	132	133	939
C6	50	72	78	84	92	92	82	92	98	105	134	159	167	167	170	1642
C7	74	118	123	135	149	144	140	171	185	200	306	385	435	433	434	3432
C8	30	53	53	56	69	70	60	64	69	75	100	117	127	129	129	1201
C9	7	12	10	12	13	12	13	16	17	21	26	29	28	31	30	277
D	18	35	34	38	43	52	45	44	46	44	54	55	59	50	49	666
E	9	12	12	15	15	19	23	23	31	29	36	46	48	46	45	409
F	15	21	22	24	27	37	37	37	40	41	52	53	54	54	41	555
G	33	50	51	56	55	51	50	60	70	89	131	167	201	200	198	1462
H	56	69	66	70	75	70	72	70	78	82	91	99	106	105	102	1211
J	18	31	32	32	35	34	32	34	39	40	40	42	42	39	40	530
K	13	29	30	30	30	29	32	39	37	45	51	61	64	64	64	618
L	4	7	6	6	6	5	6	7	7	8	15	21	23	23	25	169
M	46	61	57	58	58	57	51	51	54	56	61	64	66	65	72	877
Total	528	823	826	890	972	981	936	1043	1141	1214	1621	1909	2091	2072	2164	19,211

Industry codes from A to M represent Agribusiness (A), Mining (B), Manufacturing (C), Public utilities (D), Construction (E), Transportation (F), Information technology (G), Wholesale, retail and trade (H), Real estate (J), Social service (K), Communication and culture (K) and Conglomerate (M) respectively.

sum of administrative expense and sales expense over revenue. The higher the ratio, the higher the agency costs between shareholders and managers.

Following Jiang et al. (2010), we measure agency costs between majority and minority shareholders based on the ratio of other receivables over total assets, denoted by *Agency_cost2*. Large shareholders tend to expropriate the interest of minority shareholders through the occupation of listed firms' funds in China. Such expropriation is usually in the form of the receivable of related sales and other receivables of temporary borrowing in financial statements. Due to their disguised feature, other receivables are the main means for large shareholders to expropriate the interest of minority shareholders (Jiang et al., 2010). Therefore, the magnitude of other receivables can capture and reflect the extent to which large shareholders expropriate the interest of minority shareholders. In other words, the higher the proportion of other receivables, the higher the agency costs between majority and minority shareholders.

3.2.2. Independent variables

Most studies of annual report readability cover the context of English-speaking countries. Due to the features of the English language, scholars typically use the Fog Index, Flesch Reading Ease, Flesch–Kincaid Indices and other similar indexes to measure annual report readability based on the number of sentences, average sentence length and the proportion of complex words (Hwang and Kim, 2017; Li, 2008; Lo et al., 2017). However, Loughran and McDonald (2014) point out that these measures may fail to measure the readability of business documents (e.g., financial statements) because average sentence length and the proportion of complex words are not significant factors of readability in many situations. The findings based on these measurements are thus biased. Accordingly, the authors recommend that the file size of annual reports should be a better proxy of annual report readability. In China, research on annual report readability has only just emerged, and there is not a generally accepted proxy to measure the readability of Chinese annual reports (Ji et al., 2016; Yan and Sun, 2002; You and Yi, 2010). Considering the distinct differences between Chinese

and English, we cannot directly use the Fox Index or Flesh–Kincaid indices to measure the readability of annual reports in Chinese.

According to Loughran and McDonald (2014) and the measurements of Li (2008) and De Franco et al. (2015), we take the length of Chinese annual reports to measure their readability. The efficiency of information transmission largely depends on how easy it is for users to obtain the information. The magnitude of information disclosed in annual reports is directly reflected in the length of those reports. The more information is disclosed in annual reports, the more difficult it is for users to extract value-related information. However, the interest of readers is an important factor of readability (Chall, 1958). The content of annual reports as a type of commercial document is relatively boring and difficult to understand. Long annual reports are unlikely to arouse the reading interest of users, particularly with the current availability of mass information and shallow reading networks. We generate three independent variables, that is, *Pages, Words* and *Characters*, to measure the length of annual reports.[1] For better understanding and interpretation, we take the natural logarithm of the above three variables, measure their respective reciprocals and then adjust them according to range standardization. Finally, we get *Readability1, Readability2* and *Readability3*, respectively, as the proxies of annual report readability. A greater value of each variable represents a higher readability of annual reports.

3.2.3. Moderating variables

To test Hypotheses 2–4, we construct three moderating variables: the quality of external audit, the quality of internal control and analyst coverage. Following Becker et al. (1998) and Francis and Yu (2009), we generate a dummy variable, *BIG4*, to measure the quality of external audit, which equals 1 if a focal firm's external auditor belongs to the Big 4 auditors and 0 otherwise. According to the method of Ye et al. (2015), we take the natural logarithm of the sum of the Internal Control Index of listed companies from the DIB Internal Control and Risk Management database and one to measure the quality of internal control, denoted by *INTERNAL*. We follow Zhu et al. (2007) to measure analyst coverage, *ANALYST*, as the natural logarithm of the number of analysts (team) following a focal listed firm. If data are missing, we set *ANALYST* to be 0.

3.2.4. Control variables

Following previous studies (e.g., Jiang et al., 2010), we control a set of factors that may systematically relate to corporate agency costs. Specifically, we include firm size (*SIZE*, measured as the natural logarithm of total assets), financial leverage (*LEVERAGE*, measured as the ratio of total liabilities over total assets), fixed assets (*TANGIBLE*, measured as the ratio of fixed assets over total assets), sales growth (*GROWTH*, measured as the sales growth rate from year t − 1 to year t), firm age (*LISTAGE*, measured as the number of years since IPO), ownership concentration (*TOP1*, measured as the ratio of the shares held by the largest shareholder to total shares), multiple large shareholder structure (*BLOCKS*, equal to 1 if the shareholding held by the second largest shareholder is no less than 5% and 0 otherwise), board size (*BSIZE*, measured as the natural logarithm of the number of directors on the board), board independence (*INDBOARD*, measured as the ratio of independent directors on the board), executive compensation (*COMP*, measured as the natural logarithm of the compensation of top three executives), managerial ownership (*MSHARE*, measured as the ratio of the shares held by executives over total shares) and CEO duality (*DUALITY*, equal to 1 if one person serves as both the chairman and CEO and 0 otherwise). In addition, we include industry and year dummy variables to control for industry and time fixed effects, respectively. The definitions of all variables are listed in Table 2.

3.3. The models

To test our Hypotheses 1–4, we construct the following four regression models:

$$Agency\_cost_{t+1} = \alpha_0 + \alpha_1 Readability_t + \sum control_t + \sum Industry + \sum Year + \varepsilon \qquad (1)$$

---

[1] The initial document format of annual reports is the portable document format (PDF). We cannot obtain the information of total words or characters of annual reports in PDF. Therefore, we use the Solid Converter PDF software to convert PDF file format into WORD document format. The annual reports in WORD format can then calculate the number of words or characters automatically.

Table 2
Variable definitions.

Variable	Definition
Agency_cost1	The ratio of the sum of administrative expense and sales expense to revenue in year t + 1
Agency_cost2	The ratio of other receivables to total assets in year t + 1
Pages	The natural logarithm of total pages of annual report in year t
Words	The natural logarithm of total words of annual report in year t
Characters	The natural logarithm of total characters of annual report in year t
Readability1	The range standardization of the reciprocal of the natural logarithm of pages of annual reports in year t
Readability2	The range standardization of the reciprocal of the natural logarithm of words of annual reports in year t
Readability3	The range standardization of the reciprocal of the natural logarithm of characters of annual reports in year t
BIG4	An indicator variable that equals 1 if a firm is audited by Big 4 auditors in year t, and 0 otherwise
INTERNAL	The natural logarithm of the sum of internal control index and 1 in year t
ANALYST	The natural logarithm of the number of analysts following a focal firm in year t; the missing value is replaced by 0
SIZE	The natural logarithm of total assets in year t
LEVERAGE	The ratio of liabilities to assets in year t
TANGIBLE	The ratio of fixed-assets to total assets in year t.
GROWTH	The sales growth rate from year t − 1 to year t
LISTAGE	The natural logarithm of the number of years since IPO in year t
TOP1	The ratio of the shares held by the first shareholder to total shares in year t
BLOCKS	An indicator variable that equals 1 if the shares held by the second shareholder beyond 5% in year t, and 0 otherwise
BSIZE	The natural logarithm of the board membership in year t
INDBOARD	The ratio of the number of independent directors to total directors in year t
COMP	The natural logarithm of the compensation of the top three executives in year t
MSHARE	The ratio of the shareholdings held by executives to total shares
DUALITY	An indicator variable that equals 1 if the chairman and general manager are one person, and 0 otherwise
Industry	Each indicator variable corresponds to an industry and equals 1 if a firm belongs to this industry, and 0 otherwise
Year	Each indicator variable corresponds to a sample year and equals 1 if an observation comes from this year, and 0 otherwise
Readabiltiy4	The range standardization of the reciprocal of the natural logarithm of file size of annual reports in year t
Ab_readability1	The range standardization of the reciprocal of the natural logarithm of pages of annual reports in year t, where pages is divided by the natural logarithm of focal firms' total assets
Ab_readability2	The range standardization of the reciprocal of the natural logarithm of words of annual reports in year t, where pages is divided by the natural logarithm of focal firms' total assets
Ab_readability3	The range standardization of the reciprocal of the natural logarithm of characters of annual reports in year t, where pages is divided by the natural logarithm of focal firms' total assets
Ab_Agency_cost1	The difference between *Agency_cost1* and its industry-year mean in year t + 1
Ab_Agency_cost2	The difference between *Agency_cost2* and its industry-year mean in year t + 1
Cost1_1	The ratio of business entertainment expense to revenue in year t
Cost1_2	The ratio of the sum of office, travelling, business entertainment, communication, overseas training, director, car and conference expenses to revenue in year t
Cost1_3	The ratio of the sum of business entertainment and other expenses to revenue in year t
Cost2_1	The ratio of funds occupied by controlling shareholders to total assets in year t
Cost2_2	The ratio of the amount of debt guarantee provided by listed companies to their controlling shareholders to total assets
Cost2_3	The ratio of the total amount of related party transactions between listed companies and their controlling shareholders to total assets in year t
MEDIANREAD	The year-industry median of corresponding variable of annual report readability in year t
MKT	The marketization index of the province where a focal firm is registered, compiled by Fan et al. (2011)
STATE	An indicator variable that equals 1 if the ultimate controlling shareholders is the state, and 0 otherwise
NEW	An indicator variable that equals 1 if a firm-year observation comes from the years after 2007, and 0 otherwise

$$Agency\_cost_{t+1} = \beta_0 + \beta_1 Readability_t + \beta_2 Readability_t * BIG4_t + \beta_3 BIG4_t + \sum control_t$$
$$+ \sum Industry + \sum Year + \varepsilon \tag{2}$$

$$Agency\_cost_{t+1} = \gamma_0 + \gamma_1 Readability_t + \gamma_2 Readability_t * INTERNAL_t + \gamma_3 INTERNAL_t + \sum control_t$$
$$+ \sum Industry + \sum Year + \varepsilon \tag{3}$$

$$Agency\_cost_{t+1} = \lambda_0 + \lambda_1 Readability_t + \lambda_2 Readability_t * ANALYST_t + \lambda_3 ANALYST_t + \sum control_t$$
$$+ \sum Industry + \sum Year + \varepsilon \qquad (4)$$

Agency_cost represents either *Agency_cost1* or *Agency_cost2*. *Readability* represents any one of the three variables of annual report readability, that is, *Readability1*, *Readability2* or *Readability3*. *Control* represents all control variables. α_0, β_0, γ_0, λ_0 is the constant of the corresponding model and ε is the stochastic disturbance item of the models. According to the expectation of H1, the coefficient α_1 in model (1) should be significantly negative. According to the expectations of H2–H4, the coefficient β_2, γ_2, λ_2 in models (2)–(4) should also be significantly negative. To control for the potential endogeneity problem of reverse causality, we lag all independent, moderating and control variables. In addition, we winsorize the top and bottom 1% of each continuous variable to control the influence of outliers.

4. Results

4.1. Descriptive statistics

Table 3 reports the descriptive statistics of the main variables used in this study. As Table 3 shows, the mean and standard deviation of *Agency_cost1* are 0.170 and 0.147, respectively, and the mean and standard deviation of *Agency_cost2* are 0.027 and 0.046, respectively, showing a huge difference in the extent of agency costs across listed firms, thereby providing considerable variation for regression estimation. After the transformation of the logarithm, *Ln_Pages*, *Ln_Words* and *Ln_Characters* are equal to 113.296 ($=e^{4.730}$), 65,512.750 ($=e^{11.090}$) and 93,901.350 ($=e^{11.450}$), respectively. These values indicate that the annual reports of listed companies are relatively long on average, and the readability of annual reports has become a significant practical problem due to the current shallow reading behavior in China. Of the sample firms, 3.0% are audited by Big 4 auditors. The mean and standard deviation of *INTERNAL* are 6.469 and 0.387, respectively, which means the construction of the internal control system of listed companies displays convergence. The

Table 3
Descriptive statistics of main variables.

Variables	N	Mean	S.D.	Min	P25	Median	P75	Max
Agency_cost1	19,211	0.170	0.147	0.016	0.080	0.129	0.206	0.910
Agency_cost2	19,211	0.027	0.046	0.000	0.004	0.011	0.028	0.291
Pages	19,211	4.730	0.386	3.091	4.443	4.812	5.024	5.908
Words	19,211	11.090	0.383	8.428	10.850	11.180	11.370	13.620
Characters	19,211	11.450	0.356	9.319	11.230	11.520	11.710	13.920
Readability1	19,211	0.159	0.093	0.041	0.090	0.127	0.212	0.445
Readability2	19,211	0.070	0.032	0.030	0.048	0.059	0.084	0.168
Readability3	19,211	0.117	0.049	0.051	0.083	0.102	0.140	0.273
BIG4	19,211	0.030	0.171	0	0	0	0	1
INTERNAL	19,211	6.469	0.387	2.876	6.452	6.525	6.572	6.823
ANALYST	19,211	1.052	1.160	0	0	0.693	2.079	3.584
SIZE	19,211	21.650	1.076	19.550	20.870	21.520	22.270	24.880
LEVERAGE	19,211	0.454	0.201	0.052	0.301	0.460	0.610	0.871
TANGIBLE	19,211	0.255	0.173	0.003	0.121	0.226	0.364	0.746
GROWTH	19,211	0.194	0.444	−0.594	−0.020	0.125	0.301	2.823
LISTAGE	19,211	2.026	0.674	0.693	1.609	2.079	2.565	3.091
TOP1	19,211	0.371	0.157	0.092	0.245	0.350	0.488	0.750
BLOCKS	19,211	0.566	0.496	0	0	1	1	1
BSIZE	19,211	2.180	0.205	1.609	2.079	2.197	2.197	2.708
INDBOARD	19,211	0.350	0.0780	0.000	0.333	0.333	0.375	0.556
COMP	19,211	13.690	0.894	11.210	13.150	13.780	14.300	15.690
MSHARE	19,211	0.081	0.171	0	0	0	0.022	0.668
DUALITY	19,211	0.182	0.386	0	0	0	0	1

All variables are defined in Table 2.

mean and standard deviation of *ANALYST* are 1.052 and 1.160, respectively, suggesting that the average number of analysts following a focal listed firm is 2.863 ($=e^{1.052}$). The values of all other variables have a good distribution without abnormal outliers.

4.2. Pearson correlation matrix

Table 4 displays the Pearson correlation coefficients of the variables included in the regression models. As Table 4 shows, *Agency_cost1* and *Agency_cost2* are highly correlated ($r = 0.251$, $p < 0.01$). Firms with weak

Table 4
Pearson correlation matrix.

Variables	1	2	3	4	5	6	7
1 *Agency_cost1*	1						
2 *Agency_cost2*	0.251***	1					
3 *Readability1*	−0.004	0.298***	1				
4 *Readability2*	0.008	0.301***	0.915***	1			
5 *Readability3*	0.024***	0.289***	0.908***	0.985***	1		
6 *BIG4*	−0.034***	−0.025***	−0.030***	−0.013*	−0.023***	1	
7 *INTERNAL*	−0.106***	−0.075***	0.039***	0.031***	0.020***	0.027***	1
8 *ANALYST*	−0.019***	−0.239***	−0.428***	−0.439***	−0.446***	0.080***	0.142***
9 *SIZE*	−0.291***	−0.154***	−0.362***	−0.379***	−0.433***	0.184***	0.097***
10 *LEVERAGE*	−0.265***	0.172***	0.041***	0.011	−0.023***	0.028***	−0.070***
11 *TANGIBLE*	−0.146***	−0.098***	0.209***	0.210***	0.207***	0.012	−0.013*
12 *GROWTH*	−0.109***	−0.047***	−0.014*	−0.020***	−0.022***	−0.009	0.109***
13 *LISTAGE*	−0.073***	0.039***	−0.091***	−0.120***	−0.142***	0.032***	−0.052***
14 *TOP1*	−0.165***	−0.055***	0.143***	0.152***	0.144***	0.111***	0.085***
15 *BLOCKS*	0.106***	0.013*	−0.081***	−0.077***	−0.060***	−0.033***	−0.024***
16 *BSIZE*	−0.094***	0.019***	0.122***	0.096***	0.076***	0.048***	0.058***
17 *INDBOARD*	−0.008	−0.198***	−0.440***	−0.449***	−0.443***	0.020***	−0.020***
18 *COMP*	−0.026***	−0.270***	−0.614***	−0.627***	−0.639***	0.110***	0.060***
19 *MSHARE*	0.134***	−0.127***	−0.301***	−0.270***	−0.247***	−0.053***	0.004
20 *DUALITY*	0.091***	−0.087***	−0.189***	−0.178***	−0.162***	−0.029***	−0.013*

Variables	8	9	10	11	12	13	14
8 *ANALYST*	1						
9 *SIZE*	0.413***	1					
10 *LEVERAGE*	−0.100***	0.404***	1				
11 *TANGIBLE*	−0.080***	0.011	0.059***	1			
12 *GROWTH*	0.084***	0.053***	0.062***	−0.060***	1		
13 *LISTAGE*	−0.087***	0.335***	0.361***	0.028***	−0.056***	1	
14 *TOP1*	0.017**	0.172***	0.020***	0.092***	0.037***	−0.131***	1
15 *BLOCKS*	0.053***	−0.161***	−0.130***	−0.065***	0.038***	−0.256***	−0.369***
16 *BSIZE*	0.036***	0.172***	0.139***	0.151***	0.006	0.038***	0.026***
17 *INDBOARD*	0.203***	0.134***	−0.019***	−0.093***	0.005	0.097***	−0.081***
18 *COMP*	0.471***	0.451***	−0.023***	−0.221***	0.021***	0.154***	−0.095***
19 *MSHARE*	0.213***	−0.167***	−0.326***	−0.173***	0.034***	−0.486***	−0.133***
20 *DUALITY*	0.107***	−0.084***	−0.150***	−0.107***	0.002	−0.159***	−0.077***

Variables	15	16	17	18	19	20
15 *BLOCKS*	1					
16 *BSIZE*	0.003	1				
17 *INDBOARD*	0.016**	−0.299***	1			
18 *COMP*	0.061***	−0.021***	0.334***	1		
19 *MSHARE*	0.262***	−0.193***	0.170***	0.149***	1	
20 *DUALITY*	0.096***	−0.170***	0.149***	0.130***	0.272***	1

All variables are defined in Table 2.
*** Significance at the 1% level (two-sided).
** Significance at the 5% level (two-sided).
* Significance at the 10% level (two-sided).

corporate governance thus appear to suffer from agency costs both between shareholders and managers and between large and minority shareholders simultaneously. None of the three variables of readability, that is, *Readability1*, *Readability2* or *Readability3*, are consistently and significantly correlated with *Agency_cost1*; however, they are positively and significantly correlated with *Agency_cost2*, which is not in accordance with our expectations in the hypotheses. Thus, we should depend more on multiple regression analysis to test our hypotheses. The three moderators, that is, *BIG4*, *INTERNAL* and *ANALYST*, are significantly and negatively correlated with both *Agency_cost1* and *Agency_cost2*, which means that combining external audit, internal control and analyst coverage as a governance mechanism could help to alleviate corporate agency problems. In addition, most of the correlation coefficients for the control variables are less than 0.5, implying that the problem of multicollinearity is weak in the regression analysis when these variables are included in the models.

4.3. Multiple regression results

Table 5 reports the OLS regression results for testing the association between annual report readability and corporate agency costs for Hypothesis 1. When *Agency_cost1* is taken as the dependent variable, all three variables of annual report readability, including *Readability1*, *Readability2* and *Readability3*, produce negative regression coefficients at the 1% significance level (Model 1: $\beta = -0.111$, $t = -4.626$; Model 2: $\beta = -0.418$, $t = -5.730$; Model 3: $\beta = -0.240$, $t = -5.364$). Furthermore, we calculate that the standardized regression coefficient of *Readability1* in Model 1 is -0.070, suggesting that a one-standard-deviation increase in *Readabiltiy1* reduces agency costs between shareholders and managers by 0.070 standard deviation. Similarly, when *Agency_cost2* is taken as the dependent variable, all three variables of annual report readability also produce negative coefficients at the 1% significance level (Model 4: $\beta = -0.038$, $t = -4.382$; Model 5: $\beta = -0.124$, $t = -4.773$; Model 6: $\beta = -0.089$, $t = -5.566$). Likewise, we calculate that the standardized regression coefficient on *Readability1* in Model 4 is -0.076, suggesting that a one-standard-deviation increase in *Readability1* reduces agency costs between large and minority shareholders by 0.076 standard deviation. Together, these results suggest that annual report readability has a statistically and economically negative effect on corporate agency costs.

In conclusion, high levels of annual report readability can help to curb agency costs between not only shareholders and managers, but also large and minority shareholders, providing good support for Hypothesis 1. That is, annual reports with high readability can better alleviate the extent of information asymmetry faced by stakeholders including shareholders and improve corporate information transparency, thereby promoting the power and efficiency of their supervision on corporate insiders' opportunistic behavior.

Table 6 reports the OLS regression results for testing the moderating effect of external audit quality for Hypothesis 2. As shown in Table 6, all three variables of annual report readability consistently produce significant and negative coefficients in all of the models. More importantly, regardless of whether *Agency_cost1* or *Agency_cost2* is taken as the dependent variable, the coefficients on the three interactions, including *Readability1*BIG4*, *Readability2*BIG4* and *Readability3*BIG4*, are negative at the 1% significance level. These results suggest that annual report readability has a stronger effect on reducing corporate agency costs if focal firms' annual reports are audited by Big 4 auditors. Therefore, Hypothesis 2 is fully supported, which indicates that the effect of annual report readability on alleviating information asymmetry and improving information transparency may be stronger owing to the improved earnings quality of Big 4 audited annual reports.

Table 7 reports the OLS regression results for testing the moderating effect of internal control quality for Hypothesis 3. As shown in Table 7, regardless of whether *Agency_cost1* or *Agency_cost2* is taken as the dependent variable, the coefficients on the three interactions, including *Readability1*INTERNAL*, *Readability2*INTERNAL* and *Readability3*INTERNAL*, are negative at the 1% significance level, indicating that the quality of internal control intensifies the effect of annual report readability on reducing corporate agency costs, consistent with the expectation of Hypothesis 3. These results suggest that high-quality internal control can improve corporate earnings quality and enhance the effect of annual report readability on enabling annual report users to get access to earnings information and monitor corporate insiders' opportunistic behavior (Doyle et al., 2007; Ye et al., 2015).

Table 8 reports the OLS regression results for testing the moderating effect of analyst coverage for Hypothesis 4. As Table 8 shows, regardless of whether *Agency_cost1* or *Agency_cost2* is taken as the

Table 5
OLS regression results for the association between annual report readability and corporate agency costs.

	Dependent variable: *Agency_cost1*			Dependent variable: *Agency_cost2*		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Readability1	−0.111***			−0.038***		
	(−4.626)			(−4.382)		
Readability2		−0.418***			−0.124***	
		(−5.730)			(−4.773)	
Readability3			−0.240***			−0.089***
			(−5.364)			(−5.566)
BIG4	0.019***	0.020***	0.020***	−0.000	0.000	0.000
	(3.434)	(3.641)	(3.688)	(−0.267)	(0.042)	(0.109)
INTERNAL	−0.028***	−0.028***	−0.028***	−0.006***	−0.006***	−0.006***
	(−7.073)	(−7.070)	(−7.075)	(−6.048)	(−6.055)	(−6.072)
ANALYST	0.011***	0.011***	0.011***	0.000	0.000	0.000
	(10.434)	(10.383)	(10.442)	(1.493)	(1.440)	(1.481)
SIZE	−0.033***	−0.033***	−0.034***	−0.005***	−0.005***	−0.005***
	(−20.970)	(−21.079)	(−21.169)	(−10.278)	(−10.425)	(−10.991)
LEVERAGE	−0.084***	−0.085***	−0.085***	0.043***	0.043***	0.043***
	(−11.882)	(−12.079)	(−12.066)	(19.698)	(19.676)	(19.559)
TANGIBLE	−0.057***	−0.056***	−0.056***	−0.041***	−0.041***	−0.041***
	(−8.369)	(−8.167)	(−8.195)	(−19.882)	(−19.824)	(−19.731)
GROWTH	−0.030***	−0.030***	−0.030***	−0.007***	−0.007***	−0.007***
	(−9.601)	(−9.676)	(−9.611)	(−8.416)	(−8.447)	(−8.434)
LISTAGE	0.014***	0.015***	0.014***	0.006***	0.006***	0.006***
	(7.303)	(7.396)	(7.295)	(10.278)	(10.370)	(10.417)
TOP1	−0.077***	−0.077***	−0.076***	−0.020***	−0.020***	−0.019***
	(−11.112)	(−11.111)	(−10.974)	(−8.839)	(−8.869)	(−8.672)
BLOCKS	0.001	0.001	0.001	0.002***	0.002***	0.002***
	(0.633)	(0.515)	(0.640)	(2.969)	(2.877)	(2.969)
BSIZE	−0.006	−0.008	−0.007	−0.006***	−0.006***	−0.006***
	(−1.125)	(−1.369)	(−1.312)	(−3.150)	(−3.354)	(−3.376)
INDBOARD	0.017	0.016	0.017	−0.003	−0.004	−0.004
	(0.823)	(0.745)	(0.799)	(−0.491)	(−0.538)	(−0.532)
COMP	−0.002	−0.003	−0.003	−0.004***	−0.005***	−0.005***
	(−1.295)	(−1.574)	(−1.630)	(−8.457)	(−8.683)	(−8.912)
MSHARE	0.018**	0.018***	0.018**	0.004**	0.004***	0.004**
	(2.553)	(2.594)	(2.559)	(2.550)	(2.672)	(2.525)
DUALITY	0.007**	0.007**	0.007**	−0.001*	−0.001*	−0.001*
	(2.526)	(2.529)	(2.557)	(−1.934)	(−1.931)	(−1.891)
Constant	1.235***	1.268***	1.278***	0.296***	0.303***	0.314***
	(26.812)	(26.817)	(26.905)	(20.389)	(20.701)	(21.141)
Observations	19,211	19,211	19,211	19,211	19,211	19,211
F value	121.780	122.108	123.489	55.892	55.770	55.966
Adj. R^2	0.252	0.253	0.253	0.253	0.254	0.254

The industry and year indicators are included in all regression models but omitted from the table to save space. T-statistics, based on standard errors adjusted for Huber-White, are in parentheses. All variables are defined in Table 2.
*** Significance at the 1% level (two-sided).
** Significance at the 5% level (two-sided).
* Significance at the 10% level (two-sided).

dependent variable, the coefficients on the three interactions, including *Readability1*ANALYST, Readability2*ANALYST* and *Readability3*ANALYST*, are negative at the 1% significance level. These results suggest that higher analyst coverage may enhance the negative association between annual report readability and corporate agency costs, thereby supporting Hypothesis 4. That is, as high analyst coverage increases the frequency of use and effect of corporate annual reports, the value of annual report readability may be better reflected, thereby strengthening the negative association between annual report readability and corporate agency costs.

Table 6
OLS regression results for the moderating effect of external audit quality.

	Dependent variable: *Agency_cost1*			Dependent variable: *Agency_cost2*		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Readability1	−0.107***			−0.036***		
	(−4.467)			(−4.217)		
Readability2		−0.415***			−0.122***	
		(−5.673)			(−4.716)	
Readability3			−0.234***			−0.086***
			(−5.213)			(−5.390)
*Readability1*BIG4*	−0.299***			−0.103***		
	(−6.066)			(−6.489)		
*Readability2*BIG4*		−0.756***			−0.322***	
		(−4.571)			(−6.290)	
*Readability3*BIG4*			−0.495***			−0.194***
			(−5.324)			(−6.218)
BIG4	0.014***	0.018***	0.017***	−0.002	−0.001	−0.001
	(2.727)	(3.413)	(3.214)	(−1.326)	(−0.432)	(−0.707)
INTERNAL	−0.028***	−0.028***	−0.028***	−0.006***	−0.006***	−0.006***
	(−7.009)	(−7.025)	(−7.020)	(−5.950)	(−5.965)	(−5.971)
ANALYST	0.011***	0.011***	0.011***	0.000	0.000	0.000
	(10.354)	(10.286)	(10.339)	(1.373)	(1.265)	(1.306)
Control variables	Included	Included	Included	Included	Included	Included
Constant	1.234***	1.269***	1.276***	0.296***	0.303***	0.313***
	(26.797)	(26.832)	(26.875)	(20.368)	(20.738)	(21.090)
Observations	19,211	19,211	19,211	19,211	19,211	19,211
F value	119.721	119.813	121.210	55.003	54.906	55.095
Adj. R^2	0.253	0.254	0.254	0.255	0.255	0.256

The industry and year indicators and all control variables are included in all regression models but omitted from the table to save space. The interaction terms are mean-centered before they are included in the regression models. T-statistics, based on standard errors adjusted for Huber–White, are in parentheses. All variables are defined in Table 2.
*** Significance at the 1% level (two-sided).
** Significance at the 5% level (two-sided).
* Significance at the 10% level (two-sided).

Taken together, the OLS regression results of Tables 5–8 provide empirical evidence to support Hypotheses 1–4. The results suggest that annual report readability helps to alleviate corporate agency costs. Firms with more readable annual reports have higher information transparency and weaker information asymmetry, thereby improving stakeholders' supervision over corporate insiders' opportunistic behavior. External audit, internal control system and analyst coverage can also strengthen the effect of annual report readability on reducing corporate agency costs.

4.4. Robustness checks

4.4.1. Tests for endogeneity concerns
A major concern in this study is that our findings may be subject to the problem of endogeneity. The regression results demonstrate only a statistical relationship and not a causal relationship between annual report readability and corporate agency costs. Corporate agency costs may have an impact on annual report readability (Li, 2008). Our findings may thus be affected by the problem of reverse causality. Meanwhile, we may miss factors or variables that potentially affect the relationship between annual report readability and agency costs. It means that our findings may be also affected by the problem of endogeneity resulting from missing variables. Therefore, we use the simultaneous-equations and firm fixed-effects models to address the above two endogeneity problems, respectively. The analysis results are shown in Tables 9 and 10.

Table 9 displays the regression results of the 3SLS simultaneous-equations model. To improve the recognition and estimation validity of this model, we introduce two variables, that is, the year-industry median of

Table 7
OLS regression results for the moderating effect of internal control quality.

	Dependent variable: *Agency_cost1*			Dependent variable: *Agency_cost2*		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Readability1	−0.105***			−0.036***		
	(−4.436)			(−4.193)		
Readability2		−0.393***			−0.115***	
		(−5.396)			(−4.442)	
Readability3			−0.238***			−0.088***
			(−5.343)			(−5.527)
*Readability1*INTERNAL*	−0.671***			−0.222***		
	(−8.043)			(−6.096)		
*Readability2*INTERNAL*		−1.856***			−0.634***	
		(−5.211)			(−4.791)	
*Readability3*INTERNAL*			−1.056***			−0.329***
			(−5.507)			(−4.713)
INTERNAL	−0.056***	−0.053***	−0.048***	−0.015***	−0.014***	−0.012***
	(−8.814)	(−6.956)	(−7.245)	(−5.902)	(−5.082)	(−5.087)
BIG4	0.019***	0.020***	0.021***	−0.000	0.000	0.000
	(3.581)	(3.744)	(3.854)	(−0.176)	(0.114)	(0.247)
ANALYST	0.010***	0.010***	0.010***	0.000	0.000	0.000
	(9.580)	(9.466)	(9.523)	(0.252)	(0.062)	(0.231)
Control variables	Included	Included	Included	Included	Included	Included
Constant	1.347***	1.362***	1.356***	0.333***	0.335***	0.338***
	(25.657)	(24.525)	(25.329)	(18.076)	(17.764)	(18.989)
Observations	19,211	19,211	19,211	19,211	19,211	19,211
F value	120.456	121.385	122.775	57.473	57.212	56.868
Adj. R^2	0.266	0.265	0.264	0.268	0.267	0.265

The industry and year indicators and all control variables are included in all regression models but omitted from the table to save space. The interaction terms are mean-centered before they are included in the regression models. T-statistics, based on standard errors adjusted for Huber–White, are in parentheses. All variables are defined in Table 2.
*** Significance at the 1% level (two-sided).
** Significance at the 5% level (two-sided).
* Significance at the 10% level (two-sided).

annual report readability (*MEDIANREAD*) and regional market environment (*MKT*), as instrumental variables. The variable *MKT* equals the marketization index of the province where a focal firm is registered, which is compiled byFan et al. (2011). As Table 9 shows, after controlling for the effect of agency costs between shareholders and managers on annual report readability, *Readability1* still gets a negative coefficient at the 5% significance level (Model 1: $\beta = -0.453$, $t = -2.209$). Similarly, after controlling for the potential influence of agency costs between large and minority shareholders on annual report readability, *Readability1* still gets a negative coefficient at the 1% significance level (Model 3: $\beta = -0.327$, T = −4.830). The results based on the other two variables of annual report readability, that is, *Readability2* and *Readability3*, are highly similar and not reported here to save space. These results together suggest that our findings hold after using the simultaneous-equations model to address the endogeneity problem of reverse causality.

Table 10 reports the regression results of the firm fixed-effects model. As Table 10 shows, when *Agency_cost1* is taken as the dependent variable in Models 1–3, *Readability1* consistently has significant and negative coefficients in all of the models, and the coefficients on the interactions, including *Readability1*BIG4*, *Readability1*INTERNAL* and *Readability1*ANALYST*, are significantly negative. When *Agency_cost2* is used as the dependent variable in Models 4–6, although *Readability1* does not show significant coefficients, its interactions with each of the three moderating variables produce negative coefficients at the 1% significance level, consistent with the predictions in Hypotheses 2–4. The results based on the other two variables of annual report readability, that is, *Readability2* and *Readability3*, are highly similar and not reported here to save space. Overall, these results suggest that there is still a significant and negative association between annual report readability and corporate agency costs after controlling for the endogeneity problem of missing vari-

Table 8
OLS regression results for the moderating effect of analyst coverage.

	Dependent variable: Agency_cost1			Dependent variable: Agency_cost2		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Readability1	−0.139[***]			−0.057[***]		
	(−5.903)			(−7.090)		
Readability2		−0.501[***]			−0.177[***]	
		(−7.097)			(−7.242)	
Readability3			−0.303[***]			−0.121[***]
			(−6.996)			(−8.171)
Readability1*ANALYST	−0.059[***]			−0.041[***]		
	(−4.687)			(−10.819)		
Readability2*ANALYST		−0.195[***]			−0.125[***]	
		(−4.980)			(−10.120)	
Readability3*ANALYST			−0.147[***]			−0.076[***]
			(−5.957)			(−9.870)
ANALYST	0.008[***]	0.008[***]	0.007[***]	−0.001[***]	−0.001[***]	−0.001[***]
	(7.628)	(7.131)	(6.755)	(−4.015)	(−4.324)	(−4.065)
BIG4	0.018[***]	0.019[***]	0.019[***]	−0.001	−0.000	−0.000
	(3.246)	(3.536)	(3.505)	(−0.859)	(−0.264)	(−0.326)
INTERNAL	−0.028[***]	−0.028[***]	−0.028[***]	−0.006[***]	−0.006[***]	−0.006[***]
	(−7.077)	(−7.079)	(−7.088)	(−6.041)	(−6.057)	(−6.059)
Control variables	Included	Included	Included	Included	Included	Included
Constant	1.230[***]	1.262[***]	1.274[***]	0.293[***]	0.299[***]	0.312[***]
	(26.692)	(26.648)	(26.818)	(20.127)	(20.355)	(20.943)
Observations	19,211	19,211	19,211	19,211	19,211	19,211
F value	120.794	121.169	122.497	55.566	55.393	55.644
Adj. R^2	0.253	0.254	0.254	0.257	0.257	0.257

The industry and year indicators and all control variables are included in all regression models but omitted from the table to save space. The interaction terms are mean-centered before they are included in the regression models. T-statistics, based on standard errors adjusted for Huber–White, are in parentheses. All variables are defined in Table 2.

[***] Significance at the 1% level (two-sided).
[**] Significance at the 5% level (two-sided).
[*] Significance at the 10% level (two-sided).

ables using the firm fixed-effects model. This negative association is still stronger in firms with higher external audit quality, higher internal control quality or higher analyst coverage than in their counterparts.

4.4.2. Tests for estimation methods

The sample data in this study are classical unbalanced panel data. The panel data may suffer from cross-sectional correlation and time series autocorrelation, which may bias the OLS regression estimations, so we follow Petersen's (2009) method to run regressions by clustering firm and year dimensions simultaneously. The results are shown in Table 11. Similar to the results of Tables 5–8, regardless of whether Agency_cost1 or Agency_cost2 is taken as the dependent variable, Readability1 has negative coefficients in all of the models at the 5% significance level at least, and its interactions with each of the three moderating variables consistently have negative coefficients at the 1% significance level. The results based on the other two variables of annual report readability, that is, Readability2 and Readability3, are highly similar and not reported here to save space. Therefore, our findings are robust to a different estimation method.

4.4.3. Tests for the measurement of annual report readability

Annual report readability is a key variable in this study. Due to the lack of generally accepted measurements of readability in the Chinese context, we must use other measurements to enhance the robustness of our findings. Loughran and McDonald (2014) recommend a 10-k file size as the best measure of readability after comparing the various readability measures. File size is straightforward, is easily replicated and suffers less from measurement errors. More importantly, it is closely related to alternative readability measures and

Table 9
Regression results of the simultaneous-equations model.

	Model 1 *Agency_cost1*	Model 2 *Readability1*	Model 3 *Agency_cost2*	Model 4 *Readability1*
Agency_cost1		−0.287***(−9.312)		
Agency_cost2				−0.087(−1.043)
Readability1	−0.453**(−2.209)		−0.327***(−4.830)	
BIG4	0.011*(1.864)	−0.007***(−3.028)	−0.004**(−1.995)	−0.012***(−6.236)
INTERNAL	−0.029***(−11.645)	−0.008***(−5.813)	−0.006***(−6.684)	−0.000(−0.136)
ANALYST	0.009***(8.271)	0.001**(2.060)	−0.000(−0.954)	−0.002***(−5.929)
SIZE	−0.040***(−15.136)	−0.022***(−17.867)	−0.008***(−9.547)	−0.012***(−17.357)
LEVERAGE	−0.084***(−12.334)	−0.037***(−10.471)	0.039***(17.308)	−0.011**(−2.502)
TANGIBLE	−0.044***(−5.740)	0.008**(2.277)	−0.033***(−13.057)	0.022***(5.322)
GROWTH	−0.032***(−13.643)	−0.014***(−10.438)	−0.009***(−11.067)	−0.006***(−6.016)
LISTAGE	0.023***(7.058)	0.019***(24.364)	0.011***(9.933)	0.016***(22.171)
TOP1	−0.050***(−7.712)		−0.019***(−8.400)	
BLOCKS	0.004**(2.210)		0.002***(3.183)	
BSIZE	−0.009*(−1.729)	−0.009***(−4.111)	−0.008***(−4.583)	−0.009***(−4.653)
INDBOARD	0.003(0.142)	−0.029***(−3.733)	−0.014*(−2.152)	−0.036***(−5.661)
COMP	0.007***(3.758)		−0.004***(−5.686)	
MSHARE	0.033***(4.879)		0.006**(2.152)	
DUALITY	0.004**(1.995)		−0.001(−1.506)	
MEDIANREAD		0.471***(9.937)		0.534***(13.118)
MKT		0.000(0.398)		0.000(1.320)
Observations		19,211		19,211
Chi2 value	40643.90***	43651.18***	5976.81***	312204.41***

The industry and year indicators are included in all regression models but omitted from the table to save space. Z-statistics, based on standard errors adjusted for Huber–White, are in parentheses. The regression results based on *Readability2*, *Readability3* are highly similar and consistent but not reported here for saving space. All variables are defined in Table 2.

*** Significance at the 1% level (two-sided).
** Significance at the 5% level (two-sided).
* Significance at the 10% level (two-sided).

thus is a better measure of readability in a given information environment. Therefore, we follow Loughran and McDonald (2014) and take file size as another proxy of annual report readability, *Readability4*, which is calculated as the range standardization of the reciprocal of the natural logarithm of the digital file size of the annual report. The results of the regression analysis are shown in Table 12. Those based on *Readability4* are consistent with those in Tables 5–8. Therefore, our findings hold to an alternative measurement of annual report readability.

Significant differences in business complexity and operating activities are found between large and small firms, so the magnitude of information included in annual reports should vary considerably across firms of different sizes. That is, large firms' annual reports are generally longer than those of small firms. The firm size effect should thus be eliminated when calculating the length of corporate annual reports. We thus further adjust the length of annual reports, that is, *Pages*, *Words* and *Characters*, by dividing the natural logarithm of the focal firms' total assets before calculating our three variables of annual report readability. We then get three new variables of annual report readability, including *Ab_Readability1*, *Ab_Readability2* and *Ab_Readability3*, and rerun the related regression models. The regression analysis results are shown in Table 13. Similar to those in Tables 5–8, *Ab_Readability1* consistently has negative coefficients at the 1% significance level in all of the models, and its interactions with each of the three moderating variables also have negative coefficients at the 1% significance level, regardless of whether *Agency_cost1* or *Agency_cost2* is taken as the dependent variable. The results based on the other two variables of annual report readability, that is, *Ab_Readability2* and *Ab_Readability3*, are highly similar and not reported here to save space. These results suggest that our findings still hold after eliminating the effect of firm size in the measurement of annual report readability.

Table 10
Regression results of firm fixed-effects model.

	Dependent variable: Agency_cost1			Dependent variable: Agency_cost2		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Readability1	−0.088***	−0.086***	−0.107***	0.007	0.006	−0.012
	(−4.198)	(−4.113)	(−4.820)	(0.824)	(0.783)	(−1.451)
Readability1*BIG4	−0.105**			−0.097***		
	(−2.294)			(−5.574)		
Readability1*INTERNAL		−0.436***			−0.177***	
		(−14.994)			(−16.007)	
Readability1*ANALYST			−0.027**			−0.027***
			(−2.223)			(−5.909)
BIG4	−0.002	−0.002	−0.004	−0.006**	−0.007***	−0.008***
	(−0.312)	(−0.382)	(−0.620)	(−2.526)	(−2.972)	(−3.297)
INTERNAL	−0.011***	−0.029***	−0.011***	−0.004***	−0.011***	−0.004***
	(−5.402)	(−12.674)	(−5.408)	(−5.108)	(−13.024)	(−5.113)
ANALYST	0.002**	0.002*	0.001	0.002***	0.001***	0.001
	(2.148)	(1.658)	(1.240)	(3.866)	(3.360)	(1.567)
Control variables	Included	Included	Included	Included	Included	Included
Constant	0.772***	0.884***	0.785***	0.086***	0.132***	0.100***
	(17.221)	(19.592)	(17.380)	(5.070)	(7.721)	(5.822)
Observations	19,211	19,211	19,211	19,211	19,211	19,211
F value	37.097	44.662	37.085	122.498	131.387	122.649
Adj. R^2	0.1523	0.1813	0.1562	0.2724	0.2986	0.2814

The year indicators and all control variables are included in all regression models but omitted from the table to save space. The interaction terms are mean-centered before they are included in the regression models. T-statistics, based on standard errors adjusted for Huber–White, are in parentheses. The regression results based on Readability2, Readability3 are highly similar and consistent but not reported here for saving space. All variables are defined in Table 2.
*** Significance at the 1% level (two-sided).
** Significance at the 5% level (two-sided).
* Significance at the 10% level (two-sided).

4.4.4. Tests for the measurement of agency costs

In the preceding analyses, we measure agency costs between shareholders and managers based on the operating expense ratio, and those between large and minority shareholders based on the ratio of other receivables to total assets. These measurements may vary significantly across firms and industries. For example, high-tech firms have a high operating expense ratio due to high R&D investment rather serious agency conflicts between shareholders and managers. Similarly, firms in highly competitive industries or downstream firms may have a high ratio of other receivables over total assets, due to fierce market competition or low market positions rather serious agency conflicts between majority and minority shareholders. We therefore generate two new variables, including the abnormal operating expense ratio (Ab_Agency_cost1) and abnormal ratio of other receivables over total assets (Ab_Agency_cost2), to measure corporate agency costs. Specifically, Ab_Agency_cost1 equals the difference between Agency_cost1 and its industry-year mean, and Ab_Agency_cost2 equals the difference between Agency_cost2 and its industry-year mean. We then rerun the related regression models based on these two new dependent variables. The results are shown in Table 14. Similar to those in Tables 5–8, regardless of whether Ab_Agency_cost1 or Ab_Agency_cost2 is taken as the dependent variable, Readability1 has negative coefficients at the 1% significance level in all of the models, and its interactions with each of the three moderating variables also have negative coefficients at the 1% significance level. The results based on the other two variables of annual report readability, that is, Readability2 and Readability3, are highly similar and not reported here to save space. These results again give good support to our four hypotheses, suggesting that our findings hold to the firm size-adjusted measurements of agency costs.

Neither the operating expense ratio nor the ratio of other receivables to total assets is a direct measure of agency costs, and thus they may be biased in measurement due to noise. To reduce this bias risk, we use several specific items of operating expenses and other receivables to measure agency costs. Business entertainment,

Table 11
OLS regression results by clustering both firm and year dimensions.

	Dependent variable: *Agency_cost1*			Dependent variable: *Agency_cost2*		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Readability1	−0.107[***]	−0.105[***]	−0.139[***]	−0.036[**]	−0.036[**]	−0.057[***]
	(−3.020)	(−2.892)	(−3.472)	(−2.454)	(−2.356)	(−3.568)
Readability1*BIG4	−0.299[***]			−0.103[***]		
	(−3.234)			(−4.896)		
Readability1*INTERNAL		−0.671[***]			−0.222[***]	
		(−4.809)			(−2.994)	
Readability1*ANALYST			−0.059[***]			−0.041[***]
			(−2.740)			(−4.275)
BIG4	0.014	0.019	0.018	−0.002	−0.000	−0.001
	(1.151)	(1.496)	(1.366)	(−0.738)	(−0.106)	(−0.501)
INTERNAL	−0.028[***]	−0.056[***]	−0.028[***]	−0.006[**]	−0.015[***]	−0.006[**]
	(−3.440)	(−5.889)	(−3.450)	(−2.128)	(−2.916)	(−2.163)
ANALYST	0.011[***]	0.010[***]	0.008[***]	0.000	0.000	−0.001
	(5.272)	(5.376)	(3.587)	(0.379)	(0.082)	(−1.223)
Control variables	Included	Included	Included	Included	Included	Included
Constant	1.234[***]	1.347[***]	1.230[***]	0.296[***]	0.333[***]	0.293[***]
	(10.497)	(13.409)	(10.300)	(4.783)	(5.818)	(4.665)
Observations	19,211	19,211	19,211	19,211	19,211	19,211
F value	119.721	120.456	120.794	55.003	57.473	55.566
Adj. R^2	0.253	0.266	0.253	0.255	0.268	0.257

The industry and year indicators and all control variables are included in all regression models but omitted from the table to save space. The interaction terms are mean-centered before they are included in the regression models. T-statistics are in parentheses. The regression results based on *Readability2*, *Readability3* are highly similar and consistent but not reported here for saving space. All variables are defined in Table 2.

[***] Significance at the 1% level (two-sided).
[**] Significance at the 5% level (two-sided).
[*] Significance at the 10% level (two-sided).

traveling, overseas training and car expenses are identified as the main forms of corporate resource abuse that management may use to benefit themselves (Chen et al., 2005). As such, we generate three new variables, including *Cost1_1*, *Cost1_2* and *Cost1_3*, to measure the agency costs between shareholders and managers. *Cost1_1* equals the ratio of business entertainment expense to revenue. *Cost1_2* equals the ratio of the sum of office, traveling, business entertainment, communication, overseas training, car, conference and directors' expenses to revenue. *Cost1_3* equals the ratio of the sum of business entertainment and other expenses to revenue. As large shareholders expropriate minority shareholders mainly through fund occupation, debt guarantee and other related party transactions, we follow previous studies (Zheng, 2009) to generate three new variables, *Cost2_1*, *Cost2_2* and *Cost2_3*, to measure the agency costs between large and minority shareholders. *Cost2_1* equals the ratio of funds occupied by controlling shareholders to total assets. *Cost2_2* equals the ratio of the amount of debt guarantee provided by listed companies to their controlling shareholders over total assets. *Cost2_3* equals the ratio of the total amount of related party transactions between listed companies and their controlling shareholders to total assets. We then rerun the related regression models using these new specific measurements of agency costs. The results are given in Table 15. *Readability1* has negative coefficients in all of the models at the 1% significance level regardless of whether any of the new specific measurements of agency costs is taken as the dependent variable. The results based on the other two variables of annual report readability, that is, *Readability2* and *Readability3*, are highly similar and not reported here to save space. Therefore, our findings are robust after controlling for the measurement bias of agency costs.

Table 12
OLS regression results for alternative measurement of annual report readability.

	Dependent variable: Agency_cost1			Dependent variable: Agency_cost2		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Readability4	−0.093***	−0.084***	−0.105***	−0.035***	−0.032***	−0.044***
	(−3.102)	(−2.762)	(−3.640)	(−3.100)	(−2.783)	(−4.174)
Readability4*BIG4	−0.284***			−0.140***		
	(−3.561)			(−5.548)		
Readability4*INTERNAL		−0.713***			−0.246***	
		(−5.851)			(−5.213)	
Readability4*ANALYST			−0.055***			−0.041***
			(−3.332)			(−7.807)
BIG4	0.021***	0.020***	0.020***	0.001	0.000	0.000
	(3.842)	(3.672)	(3.690)	(0.405)	(0.098)	(0.187)
INTERNAL	−0.028***	−0.049***	−0.028***	−0.006***	−0.013***	−0.006***
	(−7.072)	(−7.691)	(−7.112)	(−6.023)	(−5.486)	(−6.104)
ANALYST	0.011***	0.010***	0.010***	0.000	0.000	−0.000
	(10.526)	(10.036)	(9.218)	(1.525)	(0.810)	(−1.543)
Control variables	Included	Included	Included	Included	Included	Included
Constant	1.135***	1.211***	1.133***	0.235***	0.261***	0.234***
	(25.174)	(24.087)	(25.132)	(16.940)	(15.405)	(16.852)
Observations	19,199	19,199	19,199	19,199	19,199	19,199
F value	118.933	119.767	119.729	54.634	56.430	54.859
Adj. R^2	0.252	0.259	0.252	0.254	0.262	0.255

The industry and year indicators and all control variables are included in all regression models but omitted from the table to save space. The interaction terms are mean-centered before they are included in the regression models. T-statistics, based on standard errors adjusted for Huber–White, are in parentheses. All variables are defined in Table 2.
*** Significance at the 1% level (two-sided).
** Significance at the 5% level (two-sided).
* Significance at the 10% level (two-sided).

4.5. Further analysis

4.5.1. Effect of the nature of property rights[2]

The literature shows that agency problems in state-owned enterprises (SOEs) are substantially different from those in non-SOEs (NSOEs) in China. SOEs mainly suffer from agency problems between shareholders and managers due to the absence of actual owners and the problem of insider control, while NSOEs mainly suffer agency problems between majority and minority shareholders due to the widespread tunneling behavior of controlling shareholders (Lei et al., 2013). Agency problems vary greatly depending on the nature of property rights, so the association between annual report readability and corporate agency costs may also differ according to property rights. We therefore further investigate the moderating effect of the nature of property rights. The results are shown in Table 16.

When Agency_cost1 is taken as the dependent variable, for both SOEs and NSOEs, Readability1 has significant and negative coefficients (Model 1: $\beta = -0.062$, $t = -2.501$; Model 2: $\beta = -0.183$, $t = -3.389$), while the difference in coefficients between SOEs and NSOEs is statistically significant (Chi2 = 4.15, p < 0.05). When Agency_cost2 is taken as the dependent variable, Readability1 also has negative coefficients at the 1% significance level (Model 3: $\beta = -0.034$, $t = -3.248$; Model 4: $\beta = -0.058$, $t = -3.563$), while the difference in coefficients between SOEs and NSOEs is statistically insignificant (Chi2 = 1.53, p > 0.10). The results based on the other two variables of annual report readability, that is, Readability2 and Readability3, are highly similar and not reported here to save space. These results suggest that the effect of annual report readability on reducing corporate agency costs holds in both SOEs and NSOEs. Regardless of whether the agency costs are between

---

[2] The sample period of the regression results of the moderating effect of the nature of property rights is from 2003 to 2015, because the data is judged by the ultimate controlling shareholder from ownership structure chats of listed companies. These ownership structure chats have been mandatorily disclosed since 2003 in China.

Table 13
OLS regression results based on the firm size-adjusted measurement of annual report readability.

	Dependent variable: *Agency_cost1*			Dependent variable: *Agency_cost2*		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Ab_Readability1	−0.116***	−0.107***	−0.140***	−0.045***	−0.042***	−0.061***
	(−4.711)	(−4.385)	(−5.839)	(−5.002)	(−4.677)	(−7.370)
*Ab_Readability1*BIG4*	−0.312***			−0.101***		
	(−5.889)			(−6.165)		
*Ab_Readability1*INTERNAL*		−0.733***			−0.259***	
		(−8.317)			(−6.776)	
*Ab_Readability1*ANALYST*			−0.045***			−0.036***
			(−3.347)			(−9.013)
BIG4	0.017***	0.019***	0.018***	−0.001	−0.000	−0.001
	(3.167)	(3.646)	(3.296)	(−0.803)	(−0.113)	(−0.739)
INTERNAL	−0.028***	−0.058***	−0.028***	−0.006***	−0.016***	−0.006***
	(−6.994)	(−9.129)	(−7.063)	(−5.926)	(−6.382)	(−6.024)
ANALYST	0.011***	0.010***	0.009***	0.000	0.000	−0.001***
	(10.319)	(9.719)	(8.470)	(1.310)	(0.355)	(−2.809)
Control variables	Included	Included	Included	Included	Included	Included
Constant	1.204***	1.324***	1.199***	0.285***	0.328***	0.280***
	(26.596)	(25.879)	(26.400)	(19.647)	(17.877)	(19.250)
Observations	19,211	19,211	19,211	19,211	19,211	19,211
F value	119.581	120.177	121.214	55.066	57.672	55.591
Adj. R^2	0.254	0.265	0.253	0.255	0.270	0.256

The industry and year indicators and all control variables are included in all regression models but omitted from the table to save space. The interaction terms are mean-centered before they are included in the regression models. T-statistics, based on standard errors adjusted for Huber–White, are in parentheses. The regression results based on *Ab_Readability2*, *Ab_Readability3* are highly similar and consistent but not reported here for saving space. All variables are defined in Table 2.
*** Significance at the 1% level (two-sided).
** Significance at the 5% level (two-sided).
* Significance at the 10% level (two-sided).

shareholders and managers or between large and minority shareholders, the effect is stronger in NSOEs than in SOEs. Stakeholders probably depend more on fundamental information disclosed in annual reports when investing in NSOEs than in SOEs. NSOEs survive and develop by establishing their competitiveness from their operation efficiency and innovative products or brands, while SOEs mainly rely on their monopoly status and preferential policies to develop and establish their competitive advantages in the market. In short, as the information disclosed in NSOE annual reports is more relevant than that in SOE annual reports, the effect of annual report readability is more pronounced in NSOEs than in SOEs.

4.5.2. Effects of the implementation of new accounting standards

On January 1, 2007, the Chinese Ministry of Finance implemented the new Accounting Standards, which are broadly convergent with international accounting standards. The requirements and specifications for the financial reporting of Chinese listed companies have thus been adapted and refined since 2007. As Fig. 1 shows, between 2006 and 2007 the length of annual reports increased remarkably. The association between annual report readability and corporate agency costs might have been influenced by this exogenous institutional shock. Therefore, we further investigate the moderating effect of the adoption of the new accounting standards. We generate an indicator variable, *New*, which equals 1 if a focal firm-year observation comes from the years after 2007, and 0 otherwise. We then introduce its interactions with each of the three variables of annual report readability into the regression model. The results are shown in Table 17.

As Table 17 shows, no matter whether *Agency_cost1* or *Agency_cost2* is taken as the dependent variable, the three interactions, including *Readability1*NEW*, *Readability2*NEW* and *Readability3*NEW*, have negative coefficients at the 5% significance level at least in all of the models except Model 1, indicating that the effect of annual report readability on reducing corporate agency costs became stronger after the implementa-

Table 14
OLS regression results based on industry-adjusted measurement of agency costs.

	Dependent variable: *Ab_Agency_cost1*			Dependent variable: *Ab_Agency_cost2*		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Readability1	−0.102***	−0.100***	−0.134***	−0.034***	−0.033***	−0.055***
	(−4.302)	(−4.279)	(−5.719)	(−3.951)	(−3.904)	(−6.930)
*Readability1*BIG4*	−0.316***			−0.087***		
	(−6.336)			(−5.387)		
*Readability1*INTERNAL*		−0.656***			−0.218***	
		(−7.784)			(−6.040)	
*Readability1*ANALYST*			−0.058***			−0.043***
			(−4.599)			(−11.404)
BIG4	0.015***	0.020***	0.018***	−0.002	−0.001	−0.002
	(2.794)	(3.679)	(3.354)	(−1.482)	(−0.546)	(−1.244)
INTERNAL	−0.028***	−0.055***	−0.028***	−0.006***	−0.015***	−0.006***
	(−6.978)	(−8.640)	(−7.049)	(−6.180)	(−5.935)	(−6.263)
ANALYST	0.011***	0.010***	0.009***	0.000	0.000	−0.001***
	(10.915)	(10.169)	(8.183)	(1.487)	(0.352)	(−4.248)
Control variables	Included	Included	Included	Included	Included	Included
Constant	1.003***	1.112***	1.001***	0.219***	0.255***	0.216***
	(21.746)	(21.127)	(21.689)	(14.991)	(13.884)	(14.848)
Observations	19,211	19,211	19,211	19,211	19,211	19,211
F value	46.258	46.789	46.54	26.898	27.458	31.361
Adj. R^2	0.120	0.134	0.119	0.087	0.104	0.091

The industry and year indicators and all control variables are included in all regression models but omitted from the table to save space. The interaction terms are mean-centered before they are included in the regression models. T-statistics, based on standard errors adjusted for Huber–White, are in parentheses. The regression results based on *Readability2*, *Readability3* are highly similar and consistent but not reported here for saving space. All variables are defined in Table 2.
*** Significance at the 1% level (two-sided).
** Significance at the 5% level (two-sided).
* Significance at the 10% level (two-sided).

Table 15
OLS regression results for alternative measurement of agency costs.

	Cost1_1	*Cost1_2*	*Cost1_3*	*Cost2_1*	*Cost2_2*	*Cost2_3*
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Readability1	−0.004***	−0.013***	−0.018***	−0.013***	−0.054***	−0.253***
	(−8.738)	(−5.984)	(−8.694)	(−3.324)	(−6.473)	(−5.002)
BIG4	−0.000	−0.000	0.000	0.000	−0.011***	0.015
	(−0.666)	(−0.068)	(0.578)	(0.258)	(−5.509)	(1.012)
INTERNAL	−0.000	−0.003***	−0.001***	−0.002**	−0.002	−0.031***
	(−1.579)	(−4.790)	(−2.737)	(−2.331)	(−1.572)	(−4.070)
ANALYST	0.000	0.001***	0.001***	−0.001***	−0.002***	−0.008***
	(1.528)	(4.627)	(6.801)	(−3.535)	(−3.939)	(−3.398)
Control variables	Included	Included	Included	Included	Included	Included
Constant	0.016***	0.087***	0.095***	−0.010	−0.012	0.235***
	(14.036)	(13.445)	(16.652)	(−1.532)	(−0.737)	(2.709)
Observations	19,211	19,211	19,211	19,211	19,211	19,211
F value	135.116	207.587	175.463	10.980	22.745	52.85
Adj. R^2	0.282	0.305	0.367	0.024	0.063	0.110

The industry and year indicators and all control variables are included in all regression models but omitted from the table to save space. T-statistics, based on standard errors adjusted for Huber–White, are in parentheses. The regression results based on *Readability2*, *Readability3* are highly similar and consistent but not reported here for saving space. All variables are defined in Table 2.
*** Significance at the 1% level (two-sided).
** Significance at the 5% level (two-sided).
* Significance at the 10% level (two-sided).

Table 16
OLS regression results in sample firms with different natures of property rights.

	Dependent variable: *Agency_cost1*		Dependent variable: *Agency_cost2*	
	Model 1 STATE = 1	Model 2 STATE = 0	Model 3 STATE = 1	Model 4 STATE = 0
Readability1	−0.062**(−2.501)	−0.183***(−3.389)	−0.034***(−3.248)	−0.058***(−3.563)
BIG4	0.007(1.247)	0.043***(3.740)	−0.001(−0.584)	0.000(0.093)
INTERNAL	−0.013***(−3.023)	−0.033***(−5.574)	−0.003***(−3.389)	−0.005***(−3.694)
ANALYST	0.008***(6.254)	0.009***(5.679)	−0.000(−0.960)	−0.001(−1.486)
Control variables	Included	Included	Included	Included
Constant	1.002***(18.364)	1.242***(15.669)	0.229***(11.988)	0.265***(11.053)
Observations	9118	8729	9118	8729
F value	58.611	65.993	25.519	22.800
Adj. R^2	0.239	0.269	0.214	0.247
Difference test on coefficients		*Chi2* = 4.15**		*Chi2* = 1.53

The industry and year indicators and all control variables are included in all regression models but omitted from the table to save space. T-statistics, based on standard errors adjusted for Huber–White, are in parentheses. The regression results based on *Readability2*, *Readability3* are highly similar and consistent but not reported here for saving space. All variables are defined in Table 2.
*** Significance at the 1% level (two-sided).
** Significance at the 5% level (two-sided).
* Significance at the 10% level (two-sided).

Table 17
OLS regression results for the moderating effect of the adoption of new accounting standards.

	Dependent variable: *Agency_cost1*			Dependent variable: *Agency_cost2*		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Readability1	−0.101*** (−3.120)			−0.015 (−1.167)		
Readability2		−0.295*** (−3.054)			−0.055 (−1.494)	
Readability3			−0.169*** (−2.883)			−0.045** (−1.977)
*Readability1*NEW*	−0.027 (−0.601)			−0.063*** (−4.212)		
*Readability2*NEW*		−0.343*** (−2.584)			−0.190*** (−4.244)	
*Readability3*NEW*			−0.201** (−2.471)			−0.124*** (−4.518)
NEW	−0.072*** (−4.395)	−0.054*** (−3.172)	−0.052*** (−3.064)	−0.046*** (−7.817)	−0.044*** (−7.077)	−0.043*** (−6.957)
BIG4	0.019*** (3.399)	0.020*** (3.636)	0.020*** (3.647)	−0.001 (−0.521)	0.000 (0.041)	0.000 (0.014)
INTERNAL	−0.028*** (−7.065)	−0.028*** (−7.024)	−0.028*** (−7.023)	−0.006*** (−5.990)	−0.006*** (−5.963)	−0.006*** (−5.946)
ANALYST	0.011*** (10.395)	0.010*** (10.167)	0.010*** (10.226)	0.000 (1.108)	0.000 (0.967)	0.000 (0.906)
Control variables	Included	Included	Included	Included	Included	Included
Constant	1.225*** (26.203)	1.253*** (26.039)	1.266*** (26.317)	0.287*** (19.014)	0.293*** (19.175)	0.306*** (19.814)
Observations	19,211	19,211	19,211	19,211	19,211	19,211
F value	119.87	120.205	121.76	55.731	55.97	56.274
Adj. R^2	0.252	0.253	0.253	0.254	0.254	0.255

The industry and year indicators and all control variables are included in all regression models but omitted from the table to save space. The interaction terms are mean-centered before they are included in the regression models. T-statistics, based on standard errors adjusted for Huber–White, are in parentheses. All variables are defined in Table 2.
*** Significance at the 1% level (two-sided).
** Significance at the 5% level (two-sided).
* Significance at the 10% level (two-sided).

tion of the new accounting standards. The information environment improved in capital markets after the implementation of the standards, and investors might have thus paid more attention to the basic information disclosed by listed companies. Corporate annual reports became more relevant than before, which strengthened the effect of annual report readability. However, listed companies might have passively disclosed more information to satisfy the requirements of the new accounting standards without considering the usefulness of the information, which might have significantly reduced the readability of corporate annual reports. The effect of annual report readability might have then increased. Although we cannot be entirely sure which of these views is correct, our results show strong evidence that the negative association between annual report readability and corporate agency costs became more pronounced after the implementation of the new accounting standards in 2007.

5. Summary and conclusions

The problem of annual report readability has recently attracted much attention from regulators and scholars, but little is known about its economic consequences. We investigate the governance role of annual report readability from the perspective of agency costs by using 19,221 firm-year observations of Chinese A-share listed firms from 2001 to 2015, measuring readability based on hand-collected file length and file size data. We find that firms with higher annual report readability have lower agency costs both between shareholders and managers and between majority and minority shareholders, and the negative association between annual report readability and corporate agency costs is stronger in focal firms with higher external audit quality, higher internal control quality or higher analyst coverage. These findings stand up to a series of robustness checks including the simultaneous-equations model, the firm fixed-effects model, the two-way cluster regression model and alternative measurements of readability and agency costs. These results indicate that more readable annual reports may help to improve corporate information transparency and reduce the extent of information asymmetry facing external stakeholders, which enables them to evaluate corporate performance and value more accurately and monitor corporate insiders' opportunistic behavior more efficiently. In addition, external audit and internal control are important mechanisms for improving the quality of focal firms' accounting earnings, and security analysts are the main users of corporate annual reports. These governance mechanisms help to strengthen the effect of annual report readability on reducing corporate agency costs. We further find that the effect of annual report readability is more pronounced in NSOEs and after the implementation of the new accounting standards in 2007.

This study has several theoretical and practical implications. First, although there is a broad consensus among regulators worldwide on plain disclosure, the problem of corporate annual report readability is worsening. The effects of this on corporate agency problems must therefore be emphasized. Our findings suggest that improving the readability of annual reports can assist investors in evaluating and supervising corporate insiders. Second, our results show that soon after the implementation of new accounting standards the length of annual reports increases, while the effect of annual report readability on reducing corporate agency costs becomes stronger. Regulators should thus carefully consider writing plain disclosure into relevant laws and requirements when reforming corporate information announcement systems, to create a reasonable trade-off between integrity and the readability of disclosed information. Finally, earnings quality and use value are based on the premise that annual report readability may have economic consequences. As external audit and internal control can improve the quality of earnings, and high levels of analyst coverage can guarantee the use value of corporate annual reports, regulators should continue to encourage corporations to improve their internal control system, expand the external audit services market and develop the professional security analyst industry.

Our study may suffer from limitations. The readability of annual reports is a very abstract concept, and we recognize that our measurements are relatively rough, as they are based simply on the length of annual reports. Constructing more accurate measurements of readability in the Chinese context is essential for future studies. Research into the economic consequences of readability is still in its infancy and deserves more attention in the future.

References

Ajina, A., Laouiti, M., Msolli, B., 2016. Guiding through the fog: does annual report readability reveal earnings management? Res. Int. Bus. Finance 38, 509–516.

Akerlof, G.A., 1970. The market for "lemons": quality uncertainty and the market mechanism. Quart. J. Econ. 84 (3), 488–500.

Ang, J.S., Cole, R.A., Lin, J., 2000. Agency costs and ownership structure. J. Finance 55 (1), 81–106.

Asay, H.S., Elliott, W.B., Rennekamp, K., 2016. Disclosure readability and the sensitivity of investors' valuation judgments to outside information. Account. Rev. 92 (4), 1–25.

Ashbaugh-Skaife, H., Collins, D.W., Kinney Jr., W.R., LaFond, R., 2008. The effect of SOX internal control deficiencies and their remediation on accrual quality. Account. Rev. 83 (1), 217–250.

Baker, E.H., Kare, D.D., 1992. Relationship between annual report readability and corporate financial performance. Manage. Res. News 15 (1), 1–4.

Barnett, A., Leoffler, K., 1979. Readability of accounting and auditing messages. J. Bus. Commun. 16 (3), 49–59.

Becker, C.L., Defond, M.L., Jiambalvo, J., Subramanyam, K.R., 1998. The effect of audit quality on earnings management. Contemp. Account. Res. 15 (1), 1–24.

Biddle, G.C., Hilary, G., Verdi, R.S., 2009. How does financial reporting quality relate to investment efficiency? J. Account. Econ. 48 (2), 112–131.

Bloomfield, R.J., 2002. The "Incomplete Revelation Hypothesis" and financial reporting. Account. Horizons 16 (3), 233–243.

Bonsall, S.B., Miller, B.P., 2017. The impact of narrative disclosure readability on bond ratings and the cost of debt. Rev. Acc. Stud. 22 (2), 608–643.

Chall, J.S., 1958. Readability: An Appraisal of Research and Application. Ohio State University Press, Columbus, OH.

Chen, D., Chen, X., Wan, H., 2005. Regulation and non-pecuniary compensation in Chinese SOEs. Econ. Res. J. 2, 92–101 (in Chinese).

Courtis, J.K., 1986. An investigation into annual report readability and corporate risk-return relationships. Account. Bus. Res. 16 (64), 285–294.

De Franco, G., Hope, O.-K., Vyas, D., Zhou, B., 2015. Analyst report readability. Contemp. Account. Res. 32 (1), 76–104.

Doyle, J.T., Ge, W., McVay, S., 2007. Accruals quality and internal control over financial reporting. Account. Rev. 82 (5), 1141–1170.

DuBay, W.H., 2004. The Principles of Readability. Impact Information, Costa Mesa, CA.

Ertugrul, M., Lei, J., Qiu, J., Wan, C., 2017. Annual report readability, tone ambiguity, and the cost of borrowing. J. Financ. Quant. Anal. 52 (2), 811–836.

Fan, G., Wang, X., Zhu, H., 2011. NERI Index of Marketization of China's Provinces. Economic Science Press, Beijing (in Chinese).

Francis, J.R., Yu, M.D., 2009. Big 4 office size and audit quality. Account. Rev. 84 (5), 1521–1552.

Grossman, S.J., Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. Am. Econ. Rev. 70 (3), 393–408.

Hart, O., 1995. Firms, Contracts, and Financial Structure. Oxford University Press, New York, NY.

Healy, P.M., Palepu, K.G., 2001. Information asymmetry, corporate disclosure, and the capital markets: a review of the empirical disclosure literature. J. Account. Econ. 31 (1–3), 405–440.

Hirshleifer, D., Teoh, S.H., 2003. Limited attention, information disclosure, and financial reporting. J. Account. Econ. 36 (1), 337–386.

Hwang, B.-H., Kim, H.H., 2017. It pays to write well. J. Financ. Econ. 124 (2), 373–394.

Jensen, M.C., Meckling, W.H., 1976. Theory of the firm: managerial behavior, agency costs, and ownership structure. J. Financ. Econ. 3 (4), 305–360.

Ji, L., Zhang, L., Tian, J., 2016. An analysis on the readability of the CSR information disclosed by the Chinese public firms: a perspective from managerial power and restraint mechanism. J. Account. Econ. 1, 21–33 (in Chinese).

Jiang, G., Lee, C.M.C., Yue, H., 2010. Tunneling through intercorporate loans: the China experience. J. Financ. Econ. 98 (1), 1–20.

Kim, J., Kim, Y., Zhou, J., 2017a. Languages and earnings management. J. Account. Econ. 63 (2–3), 288–306.

Kim, C., Wang, K., Zhang, L., 2017b. Readability of 10-K reports and stock price crash risk. SSRN Working Paper.

Lang, M., Stice-Lawrence, L., 2015. Textual analysis and international financial reporting: large sample evidence. J. Account. Econ. 60 (2–3), 110–135.

Lawrence, A., 2013. Individual investors and financial disclosure. J. Account. Econ. 56 (1), 130–147.

Lee, Y.J., 2012. The effect of quarterly report readability on information efficiency of stock prices. Contemp. Account. Res. 29 (4), 1137–1170.

Lehavy, R., Li, F., Merkley, K., 2011. The effect of annual report readability on analyst following and the properties of their earnings forecasts. Account. Rev. 86 (3), 1087–1115.

Lei, Q., Lin, B., Wei, M., 2013. Types of agency cost, corporate governance and liquidity. J. Account. Public Policy 32 (3), 147–172.

Li, F., 2008. Annual report readability, current earnings, and earnings persistence. J. Account. Econ. 45 (2–3), 221–247.

Lo, K., Ramos, F., Rogo, R., 2017. Earnings management and annual report readability. J. Account. Econ. 63 (3), 1–25.

Loughran, T., Mcdonald, B., 2014. Measuring readability in financial disclosures. J. Finance 69 (4), 1643–1671.

Malkiel, B.G., Fama, E.F., 1970. Efficient capital markets: a review of theory and empirical work. J. Finance 25 (2), 383–417.

Miller, B.P., 2010. The effects of reporting complexity on small and large investor trading. Account. Rev. 85 (6), 2107–2143.

Nelson, K.K., Pritchard, A.C., 2007. Litigation risk and voluntary disclosure: the use of meaningful cautionary language. Soc. Sci. Electron. Publ. 37 (1), 223–228.

Parrino, R., Sias, R.W., Starks, L.T., 2003. Voting with their feet: institutional ownership changes around forced CEO turnover. J. Financ. Econ. 68 (1), 3–46.

Petersen, M.A., 2009. Estimating standard errors in finance panel data sets: comparing approaches. Rev. Financ. Stud. 22 (1), 435–480.

Qiu, X., Zheng, X., Deng, K., 2016. Can analysts play an effective role in professional information interpretation? Evidence based on a complexity/readability index of Chinese corporate annual reports. China Econ. Q. 15 (4), 1483–1506 (in Chinese).

Rennekamp, K., 2012. Processing fluency and investors' reactions to disclosure readability. J. Account. Res. 50 (5), 1319–1354.

Schrand, C.M., Walther, B.R., 2000. Strategic benchmarks in earnings announcement: the selective disclosure of prior-period earnings components. Account. Rev. 75 (2), 151–177.

Shleifer, A., Vishny, R.W., 1997. A survey of corporate governance. J. Finance 52 (2), 737–783.

Smith, J.E., Smith, N.P., 1971. Readability: a measure of the performance of the communication function of financial reporting. Account. Rev. 46 (3), 552–561.

Subramanian, R., Insley, R.G., Blackwell, R.D., 1993. Performance and readability: a comparison of annual reports of profitable and unprofitable corporations. J. Bus. Commun. 30 (1), 49–61.

Tan, H.T., Wang, E.Y., Zhou, B.O., 2014. When the use of positive language backfires: The joint effect of tone, readability, and investor sophistication on earnings judgments. J. Account. Res. 52 (1), 273–302.

Tan, H.T., Wang, E.Y., Zhou, B., 2015. How does readability influence investors' judgments? Consistency of benchmark performance matters. Account. Rev. 90 (1), 371–393.

Yan, D., Sun, M., 2002. Study on the readability of share B's annual reports in the Shenzhen Exchange. Account. Res. 5, 10–17 (in Chinese).

Ye, K., Cao, F., Wang, H., 2015. Can internal control information disclosure reduce stock price crash risk? J. Financ. Res. 2, 192–206 (in Chinese).

Yezegel, A., 2015. Why do analysts revise their stock recommendations after earnings announcements? J. Account. Econ. 59 (2–3), 163–181.

You, H., Zhang, X.J., 2009. Financial reporting complexity and investor under reaction to 10-K information. Rev. Acc. Stud. 14 (4), 559–586.

You, T.T., Yi, S.J., 2010. Readability analysis of internal control self-assessment report. Commun. Finance Account. 3, 112–114 (in Chinese).

Zheng, G., 2009. The relation between related party transactions and earning quality from both perspectives of efficiency and tunneling. Account. Res. 10, 68–76 (in Chinese).

Zhu, H., He, X., Tao, L., 2007. Can the Chinese securities analysts improve the efficiency of stock market? Evidence from stock price synchronicity and price information content. J. Financ. Res. 2, 110–121 (in Chinese).