Quarterly Review of Economics and Finance 91 (2023) 68–83

Contents lists available at ScienceDirect

Quarterly Review of Economics and Finance

journal homepage: www.elsevier.com/locate/qref

Language barriers, corporate site visit, and analyst forecast accuracy

Lewis H.K. Tam [a], Shaohua Tian [b,*]

[a] Faculty of Business and Administration, University of Macau, Macau
[b] School of Business, Macau University of Science and Technology, Macau

ARTICLE INFO

Keywords:
Financial analysts
Forecast accuracy
Linguistic distance
Corporate site visit

ABSTRACT

This study examines the impact of language barriers on financial analysts' decisions to perform corporate site visits and the extent to which corporate site visits help analysts overcome language barriers to improve earnings forecast accuracy. Using a sample of analysts' visits to listed firms on the Shenzhen Stock Exchange, we find that analysts are more likely to visit firms headquartered in areas where local dialects are largely different from Standard Mandarin. Moreover, corporate site visits increase analysts' forecast accuracy more for those firms. Altogether, the findings suggest that language barriers create difficulties for analysts in obtaining information via verbal communication, and corporate site visits help analysts reduce the communication noise.

1. Introduction

Financial analysts serve as the information bridge between listed firms and investors. One important responsibility of analysts is to provide investors with accurate forecasts of the listed firms' earnings. For this purpose, analysts obtain information from various channels. They rely on both financial information and nonfinancial disclosure by firms (e.g., Hope, 2003; Bhat et al., 2006; Horton et al., 2013; Chen et al., 2015) for their analyses. They also fill in the information mosaic via private communications with top executives (Soltes, 2014; Brown et al., 2015; Choy & Hope, 2022), using relevant government records (Klein et al., 2020), or visits to corporate headquarters and other facilities (Cheng et al., 2016; Han et al., 2018; Jiang & Yuan, 2018).

As opposed to many studies that have assessed available financial information from a content perspective (i.e., amount and quality of information), this study aims to assess the communication perspective. Although financial analysts widely use numerical and textual information on corporate annual reports and various disclosure forms, such information is standardized and always subject to further interpretation. Indeed, analysts frequently communicate with corporate management (mainly via phone) shortly after firm-initiated news (Soltes, 2014; Choy & Hope, 2022). Analysts believe that private communication with corporate management is more important than the analysis of public sources of information for earnings forecasts (Brown et al., 2015). Frequent interactions between financial analysts and corporate management suggest the importance of private communication to enhance analysts' understanding of firm fundamentals.

While a phone conversation with the subject firm's management (perhaps other stakeholders of the subject firm) allows analysts to solicit additional information, the communication channel could be noisy. Phone conversations generally result in information loss for two reasons. First, in a phone conversation, the listener must translate verbal information into perceived meaning, and the process is subject to the listener's attention, memory, and interpretation of information. The lack of visual cues and other visual illustrations such as graphs and diagrams in a pure verbal conversation reduces the listener's ability to extract non-verbal signals such as facial expressions and body gestures (Preston, 2005). These non-verbal signals help the listener to decode the speaker's thoughts and feelings (Reiman, 2007). They also contribute to maintaining mutual attention and regulating turn-taking in conversations (Jiang et al., 2012). Therefore, face-to-face communication involves not only linguistic but also psychological, affective, and social aspects of interaction (Dohen et al., 2010). A recent empirical study by Battiston et al. (2021) shows that face-to-face communication among co-workers enhances work productivity.[1] Second, the listener's language proficiency affects the quality of input for interpretation. If the speaker offers

* Corresponding author.
E-mail address: shtian@must.edu.mo (S. Tian).

[1] We are aware of the long-stated counterargument that visual cues and other visual information in a face-to-face meeting could distract the listener's attention from an interview (Fowler & Wackerbarth, 1980; Novick, 2008; Block & Erskine, 2012); Farooq and de Villiers (2017). Indeed, the superiority of face-to-face interviews over phone interviews is not evident in interviews/communications in many different contexts (e.g., Fenig et al., 1993; Sturges & Hanrahan, 2004; Vogl, 2013; Ward et al., 2015; Paul & Jefferson, 2019). However, those studies do not address the issue of language proficiency in phone conversations.

https://doi.org/10.1016/j.qref.2023.07.010
Received 17 September 2022; Received in revised form 28 April 2023; Accepted 26 July 2023
Available online 28 July 2023

information in the listener's second language, the listener may be able to receive part of the message only. Studies on survey interviews have widely documented that interviewees' language proficiency largely influences the quality of data collected (Kleiner et al., 2015; Wenz et al., 2021) and that the language of interview affects respondents' opinions (Viruell-Fuentes et al., 2011; Lee & Perez, 2014). During a site visit, financial analysts can talk face-to-face with corporate managers and employees. The process can help analysts to gain "soft information" that may not be clarified well in pure verbal conversation. Indeed, a survey conducted by Chen et al. (2022) shows that collecting soft information is an important objective in site visits and face-to-face interactions with managers and employees help analysts to collect such information. Thus, we expect that marginal benefits from a site visit are greater when pure verbal communication is less effective in the first place.

This study uses a sample of listed firms on the Shenzhen Stock Exchange (SZSE) in China to test the above hypotheses. Specifically, it examines whether a financial analyst is more likely to pay a corporate site visit if they expect communication difficulties with the subject firm's managers over verbal conversations and the extent to which the site visit could enhance the analyst's understanding of the firm's fundamentals and therefore forecast accuracy. China is an excellent place for studying the impacts of language barriers on analysts' site visit decisions and earnings forecast accuracy for several reasons. First, market institutions, policies, and regulations are centralized decisions in China. As a result, language barriers in China are unlikely to be related to differences in institutions that may also affect analysts' activities. Second, while Standard Mandarin is the official spoken language in China, dialects such as Cantonese are still widely used in certain local areas. As China is geographically dispersed, differences between respective dialects and Mandarins and among different dialects are sufficiently large for investigating the impacts of language barriers on information gathering. Third, although people in China speak in different dialects, they write with standardized simplified Chinese characters. As a result, dialectal variation in China is largely verbal rather than textual, which provides researchers an opportunity to identify the pure verbal effect of language barriers. Finally, comprehensive data on analysts' site visits available for the SZSE in China provide an opportunity to investigate the relation between linguistic distance and analysts' site visit decisions.

We measure the potential language barrier faced by an analyst in equity research by mapping the subject firm's headquarters with the major dialect spoken by people in the headquarters' district, as given by *The Language Atlas of China*. The atlas provides dialectal information at the administrative district (county) level and linguistic distances among dialects. The linguistic distance measure takes a value between 0 and 5, with 5 representing the largest possible difference between two dialects. In China, schools are required to teach in Standard Mandarin, and teachers must possess a certain standard of Mandarin set by the Ministry of Education.[2] As most financial analysts are young and well-educated,[3] they are expected to be proficient in Standard Mandarin. Meanwhile, corporate managers are much older, and most of them were born before China started its economic reforms in 1978.[4] As they had a small chance of receiving formal education, especially higher education, they were likely to receive little training in listening and speaking Standard Mandarin in their childhood. They tend to use their mother tongues frequently and are likely to face difficulties in Mandarin communication. Other firm stakeholders, inside and outside, are also expected to have low exposure to Mandarin, as talking in a local dialect is common in the workplace. Therefore, we expect the linguistic distance between the firm location's dominant dialect and Standard Mandarin to be positively correlated with the language barrier faced by the financial analyst in research.

For each listed firm, we collect information about analysts' site visits from the China Stock Market and Accounting Research (CSMAR) database, which provides detailed site visit data including institution names, visitor names, dates, and memos starting in 2012. Then we merge the data with the analyst forecast data by brokerage firm. We follow other studies (e.g., Duru & Reeb, 2002; Walther & Willis, 2013; Han et al., 2018) to define an analyst's forecast accuracy for a firm's earnings per share (EPS).

We hypothesize that the need for a face-to-face meeting is stronger when distant communication, such as a phone conversation, is likely to result in information loss. The main result shows that financial analysts indeed consider language barriers when they decide whether to conduct a corporate site visit. Specifically, analysts are more likely to visit firms in areas dominated by dialects that are more different from Standard Mandarin. Then, we classify firm locations into two groups according to their linguistic distances. The result shows that a site visit enhances earnings forecast accuracy only when the visited firm is in an area with a long linguistic distance, where the analyst faces a significant language barrier with the subject firm's managers and other stakeholders.

We perform a battery of robustness checks for the major finding that site visits enhance earnings forecast accuracy in areas with long linguistic distances only. First, to address the potential endogeneity problem in the regression for analysts' earnings forecast accuracy, we run a two-stage least-squares (2SLS) regression with high-speed railway (HSR) accessibility to firm headquarters location as the instrument for an analyst's decision to conduct a site visit. We also use nearest-neighbor matching to assign a non-visiting analyst to each visiting analyst (studying the same firm) and compare the two groups' earnings forecast accuracy. Our main result survives in both the 2SLS regression and the firm-matching method. Second, people growing up together could have different proficiencies in Standard Mandarin because of differences in exposure to Standard Mandarin in the workplace. Our subsample analysis shows that our main result survives in small cities but not in metropolises namely Beijing, Shanghai, Guangzhou, and Shenzhen (BSGS). Therefore, a site visit enhances an analyst's forecast accuracy only when the local firm's managers and stakeholders are nonproficient in Standard Mandarin. Finally, we replace linguistic distance with two alternative measures for language barriers and document robust results.

This study contributes to the literature in three ways. First, it provides an alternative angle to understand the value of corporate site visits. Recent studies suggest that a reduction in transportation difficulties allows analysts to perform site visits more frequently and therefore produce more accurate forecasts and investment recommendations (Chen et al., 2022; Kong et al., 2020). This study additionally suggests that available information is subject to analysts' comprehension and interpretation and that the language barrier is an important factor that affects analysts' ability to effectively communicate with corporate managers and stakeholders. Paying a visit to corporate headquarters could allow analysts to collect more accurate information by compensating for the deficiency in verbal communication. The result also sheds light on Cheng et al. (2016) by showing that corporate site visits address the information gap caused by not only physical distance but also language distance.

Second, it contributes to the literature on the economic and financial impacts of language barriers. Previous cross-country studies generally document the negative impacts of language barriers, proxied by linguistic distance, on efficiency in various contexts such as economic development (Alesina et al., 2003; Alesina & Ferrara, 2005; Nakagawa & Sugasawa, 2020), firm mergers and acquisitions (Berger et al., 2001;

[2] http://www.moe.gov.cn/srcsite/A02/s5911/moe_621/200009/t20000923_180473.html (in Chinese).

[3] According to New Fortune's "2019 Report on the Value of China Securities Research Industry," the average age of analysts is 32.5 (24–29: 28.4%, 30–34: 49.1%, 35–40: 17.7%), and 80% of the analysts hold a master's degree or above. Website: https://finance.sina.com.cn/roll/2019–12–25/doc-iihnzahi9924645.shtml (in Chinese).

[4] According to CSMAR, the average age of the top management team in Chinese listed companies is 50, with only 15.8% aged 40 or below and 34.9% holding a master's degree or above.

Baik et al., 2015; Cuypers et al., 2015), corporate governance (Kang & Kim, 2010), investor decisions (Grinblatt & Keloharju, 2001; Lundholm et al., 2018), and analysts' forecast accuracy (Cho et al., 2020). Those studies, however, are limited by difficulties in alienating the language effect from other differences in market institutions, laws, culture, polities, and policies. This study uses China, a country with a unified written language, as a platform to examine the impact of language barriers on analysts' forecast accuracy. The result suggests that even though financial analysts in China have access to all publicly available documents, their ability to collect additional information is limited by information barriers due to dialectal variations. Our study is different from Li et al. (2018) and a series of recent studies on dialect sharing (Bian et al., 2019; Du, 2019; Fu et al., 2021; Hu, Xiao et al., 2021). In the settings of these studies, direct contact and negotiation is the default and communication barriers tend to be small; while in our setting, analysts aim for information collection and conducting a site visit is a choice for them to overcome the barriers in the process.

Third, it shows that site visits can reduce communication barriers between analysts and listed firms. The result is particularly relevant to the recent surge in demand for online communication channels such as Zoom amid the COVID-19 pandemic. While it is contended that those online communication channels could reduce information barriers among people far apart, this study indicates that face-to-face meetings are still necessary and irreplaceable for the effective communication of ideas. This is especially true when verbal information requires further interpretation while people in a conversation are not equally proficient in the language of discussions.

The remainder of this chapter proceeds as follows. Section 2 describes the institutional background, literature, and hypothesis development. Section 3 outlines the data and methodology. Section 4 presents the empirical results. Section 5 discusses limitations of the study. Section 6 concludes the study.

2. Background, literature, and hypothesis development

2.1. Background

2.1.1. Chinese dialects

Although written forms of Chinese language were unified by the first emperor of the Qin Dynasty (BC 221–BC 207), spoken languages still vary largely across different areas in China. According to *The Language Atlas of China* (1987) – a joint publication by the Australian Academy of the Humanities and the Chinese Academy of Social Sciences, there are six language families in China. The Sinitic stock, under the Sino-Tibetan family, is most widely shared by people in China. The Sinitic stock consists of Mandarins and other dialects, with eight forms of Mandarins and nine major dialects still widely used. The eight forms of Mandarins include Beijing (standard) Mandarin, Northeastern Mandarin, Jilu Mandarin, Jiaoliao Mandarin, Zhongyuan Mandarin, Jianghuai Mandarin, Lanyin Mandarin, and Southwestern Mandarin, while the nine dialects include Jin dialect, Hui dialect, Gan dialect, Xiang dialect, Cantonese, Hakka, Min, Goetian, and other.[5]

To promote Standard Mandarin as a common language among citizens, i.e. *Putonghua*, the State Council of China issued the "Instruction Concerning Spreading Putonghua" in 1956. The instruction defines Putonghua as "with Beijing pronunciation as standard pronunciation, northern speech as basic dialect, and model modern vernacular prose writings as the grammatical standard." The "Law on the Standard Spoken and Written Chinese Language of the People's Republic of China" in 2000 further addresses the importance of promoting normalized and standardized Mandarin. It stipulates that Standard Mandarin should be used in government administration and education

Table 1
Dialects Distribution in China. This table presents the distribution and linguistic distances of dialects in China. There are 17 dialects containing eight Mandarins and nine non-Mandarin dialects. Column (1) reports the number of administrative districts speaking the dialect. Column (2) reports the total population speaking the dialect in millions. Column (3) reports the linguistic distance between the dialect and Standard Mandarin. Column (4) reports the main provinces or cities where the dialects are spoken. The administrative districts and population information are from *The Language Atlas of China.*

Dialects	No. of administrative districts	Population (million)	Linguistic distance	Main provinces
	(1)	(2)	(3)	(4)
Standard Mandarin	52	26.76	0	Beijing
Northeastern Mandarin	198	98.02	1	Heilongjiang, Jilin, Liaoning
Jilu Mandarin	162	88.43	1	Hebei, Tianjin, Shandong
Jiaoliao Mandarin	44	34.95	2	Liaoning, Shandong
Zhongyuan Mandarin	397	186.48	2	Anhui, Gansu, Henan, Jiangsu, Qinghai, Shandong, Shanxi, Shannxi, Xinjiang
Jianghuai Mandarin	108	86.05	2	Anhui, Hubei, Jiangsu
Lanyin Mandarin	70	16.90	3	Gansu, Ningxia, Xinjiang
Southwestern Mandarin	546	260.00	3	Guizhou, Hubei, Hunan, Sichuan, Yunnan, Chongqing
Jin dialect	194	63.05	3	Hebei, Henan, Neimeng, Shanxi, Shannxi
Hui dialect	19	3.30	4	Anhui
Gan dialect	102	48.00	4	Anhui, Hubei, Hunan, Jiangxi
Xiang dialect	64	36.37	4	Hunan
Cantonese	141	58.82	5	Guangdong
Hakka	110	42.20	5	Fujian, Guangdong, Jiangxi
Min	154	75.00	5	Fujian, Guangdong, Hainan
Goetian	160	73.79	5	Jiangsu, Shanghai, Zhejiang
Other	60	7.78	5	-

(Gao & Ren, 2019).[6] Although the popularization rate of Standard Mandarin has been increasing for economic reasons and other motivations such as further education, as of 2013, 30% of the population, mostly older adults and people in rural areas, still cannot communicate in Standard Mandarin. The standard of speaking Mandarin for the remaining 70% is generally unsatisfactory.[7]

To quantify linguistic differences between various dialects and Mandarins, *The Language Atlas of China* provides a matrix indicating the linguistic distance between any dialect pair from 0 to 5, where 0 represents no difference and 5 suggests the longest linguistic distance. A longer linguistic distance between two dialects generally means that it is more difficult for the user of one dialect to learn and understand another

[5] Beyond the eight major dialects, about 2000 minor dialects and subdialects are spoken in different regions (Li, 2006). The atlas classifies them as "other."

[6] In minority areas, schools are allowed to teach in both Standard Mandarin and ethnic minority languages.

[7] http://www.gov.cn/jrzg/2013–09/05/content_2482016.htm (in Chinese).

Table 2

Language Barriers and Popularization Rate. This table presents the popularization rates of Standard Mandarin and local dialects during childhood, in daily conversations, or at work. The data are collected from the *Survey of Language Situation in China* in 2006. The national mean is reported in column (1), and the mean values for each linguistic distance region are reported in columns (2)–(7). The definitions of all variables are available in Appendix 1.

	Nation	*Regions with Different Linguistic Distance*					
		0	*1*	*2*	*3*	*4*	*5*
Popularization (%)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Standard Mandarin-Childhood	13.5	72.7	27.8	10.6	5.5	5.5	6.9
Standard Mandarin-Talk	84.2	88.9	55.8	46.0	44.7	56.1	66.9
Standard Mandarin-Work	53.1	81.2	53.8	37.5	30.4	38.0	41.5
Dialect-Childhood	86.4	28.9	73.9	90.0	90.4	95.0	95.5
Dialect-Talk	42.0	22.3	72.6	89.8	93.2	97.7	97.4
Dialect-Work	70.1	10.4	52.0	71.3	80.2	82.0	80.4

dialect. For example, the linguistic distance between Standard Mandarin and Northeastern Mandarin is 1 whereas the linguistic distance between Standard Mandarin and Hakka is 5. The linguistic distance between "other" dialects and Standard Mandarin is also 5.

Although the concept of linguistic family and related linguistic distance measures are widely adopted by previous studies, they lack theoretical support and are therefore intuitive if not arbitrary (Tang & van Heuven, 2009). As two languages could be different in many dimensions with some differences being more important than others, more rigorous and scientific procedures are needed to measure the "distance" between two languages. One approach is to measure mutual intelligibility among related languages. Mutual intelligibility refers to how well listener A understands speaker B and vice versa if both speak using their respective (different) native languages. Intuitively, two languages should have a short distance if two people speak in their own (different) languages but manage to understand the other's meaning well.

If linguistic distances from *The Language Atlas of China* accurately reflect differences between dialects in China, then there should be little information loss or misunderstanding for a native Standard Mandarin speaker talking with a native Northeastern Mandarin speaker. Meanwhile, information loss or misunderstanding would be more pronounced if a native Standard Mandarin speaker talks with a native Hakka speaker because it is difficult for the native Standard Mandarin speaker to be proficient in Hakka or the native Hakka speaker to be proficient in Standard Mandarin. Tang and van Heuven (2009) perform intelligibility testing for 15 Chinese dialects and provide results generally consistent with the above prediction. The study shows that dialects in Chinese are extremely distinct from each other. In many cases, the users of one dialect can only recognize less than 20% of words spoken in another dialect. Another notable result is that even though Standard Mandarin has been popularized for many years, there is still significant information loss when local dialect speakers listen to it.

A fundamental assumption behind the linguistic distance in *The Language Atlas of China* is that a dialect speaker does not have exposure to other dialects. In real life, however, people have different opportunities to learn other languages or dialects.[8] For example, a local dialect speaker who receives an education in Beijing is likely to be proficient in Standard Mandarin. Meanwhile, a native Standard Mandarin speaker who works in Guangdong should have a better understanding of Cantonese than those without such exposure to the Cantonese environment. Therefore, the linguistic distance probably represents the worst scenario of miscommunication that may not reflect the actual situation in the business world. We provide a battery of tests to address how language exposure influences the information effect of linguistic distance.

Table 1 presents the distribution of the dominant dialect by region. There are 52 administrative districts with 26.76 million people speaking Standard Mandarin. Southwestern Mandarin is spoken most widely in 546 administrative districts with 260 million people. Hui dialect has the least number of speakers. In general, other Mandarins have relatively short linguistic distances from Standard Mandarin, with linguistic distance scores ranging from 1 to 3 whereas other dialects have longer linguistic distances from Standard Mandarin, with distance scores ranging from 3 to 5. Cantonese, Hakka, Min, and Goetian have the longest linguistic distances from Standard Mandarin.

It is worth noting that people speaking more distinct dialects are not located in less developed areas in China. For example, among municipalities and major business hubs, namely Beijing, Tianjin, Chongqing, Shanghai, Guangzhou, and Shenzhen, Beijing has a linguistic distance value of 0 (by default), Tianjin has a value of 1, Chongqing has a value of 3, and Shanghai, Guangzhou, and Shenzhen all have a value of 5.[9]

Table 2 presents the popularization rates for Standard Mandarin and respective dialects by linguistic distance. The data are collected from the *Survey of Language Situation in China* (2006, in Chinese) led by China's Ministry of Education. The nationwide survey was conducted in 1998–2001 and it collected information on people's usage of Standard Mandarin and local dialects in different circumstances. We extract the following province-level statistics for the purpose of our study. *Standard Mandarin-Childhood* (*Dialect-Childhood*) refers to the percentage of people speaking Standard Mandarin (own dialect) during childhood. *Standard Mandarin-Talk* (*Dialect-Talk*) refers to the percentage of people speaking Standard Mandarin (own dialect) for daily conversations. *Standard Mandarin-Work* (*Dialect-Work*) refers to the percentage of people speaking *Mandarin-Talk* (*Dialect-Talk*) at the workplace. The statistics indicate that many people still speak in their own dialects in daily life and at the workplace, especially those whose dialects are extremely different from Standard Mandarin. Meanwhile, although Standard Mandarin is the official spoken language, it is not as popular as dialects even at the workplace.

2.1.2. *Corporate site visit*

A corporate site visit refers to an analyst's/investor's trip to a firm's headquarters and operating facilities. In general, a corporate site visit includes a briefing, a presentation from firm managers, a question-and-answer (Q&A) session, and a field tour of the operating facilities (Cheng et al., 2019). Unlike a phone conversion or a media conference that involves mostly top executives, most site visits involve middle-level managers and other employees.[10] Thus, during the visit, the analyst/investor has a chance to talk to various firm stakeholders and observe the operating environment. From the discussions and observations, they could confirm the validity of public information and further understand the firm's latest development. An analyst's site visit is mostly paid for by the analyst's employer, i.e. the brokerage firm.

Opportunities for site visits should not be restricted to favored

[8] Therefore, Tang & van Heuven (2009) intentionally select rural dialect speakers who are likely to have little exposure to other dialects for their study.

[9] Guangzhou and Shenzhen are both in Guangdong Province.

[10] Cheng et al. (2019) document that middle-level managers and other employees are involved in more than two-thirds of site visits. Meanwhile, board chairpersons, CEOs, and CFOs are involved in 15% of all site visits.

market participants according to the Guidelines of Investor Relations Management in China. In Article 41 of the guidelines, the SZSE states that "[L]isted companies should try to accommodate requests from investors, analysts, and fund managers to visit company headquarters and project sites to the greatest extent." The SZSE emphasizes that "listed companies should arrange the site visits properly, so that visitors could better understand the companies' businesses and operational situations."

Although corporate site visits are important, information on corporate site visits was not mandatorily available to the public before 2008. In 2009, the SZSE issued a new disclosure rule that requires listed firms to disclose detailed information on activities for investor relationships, including data on corporate site visits. The disclosure rule allows the public to access the site visit information, including questions raised by the participants and the management's responses to the questions. In July 2012, the SZSE further required listed firms to disclose each visit separately, with details including questions and answers, within two days after the visit (Yang et al., 2020). Meanwhile, the Shanghai Stock Exchange (SHSE) only requires listed firms to submit a summary report to the exchange and the China Securities Regulatory Commission (CSRC) without public disclosure. Therefore, existing studies on corporate site visits in China focused on visits to SZSE-listed firms.

2.2. Literature review

This study is related to two streams of literature: the economic impacts of language barriers and the role of information in analyst forecasts.

2.2.1. Language barriers and their economic influences

Whorf (1956) presents the linguistic relativity principle (or the Sapir–Whorf hypothesis), which states that languages conveying cultural and cognition categories influence how humans think and act. Scholars study vocabulary, grammar, and syntax to explain differences between languages. For instance, the Russian language mandatorily differentiates dark blue from light blue, so Russian speakers are better at distinguishing the two colors than English speakers (Winawer et al., 2007). The mother tongue of non-native English speakers also influences their spoken English, sometimes confusing information that could result in serious consequences and outcomes.[11]

Language barriers hinder economic activities. For example, Easterly and Levine (1997) find that the ethnolinguistic diversity is associated with low schooling, an underdeveloped financial system, distorted foreign exchange markets, and insufficient infrastructure, all of which further result in low income in Africa. Alesina et al. (2003) find that linguistic diversity negatively affects economic development. Alesina and Ferrara (2005) point out that because of linguistic heterogeneity, it is difficult for ethnic groups to reach an agreement, especially in underdeveloped nations such as some African countries. Recently, Nakagawa and Sugasawa (2020) construct measures for domestic linguistic distance and international linguistic distance for each country. They show that while domestic linguistic distance is negatively associated with GDP per capita, international linguistic distance has an insignificant impact.

Linguistic differences among dialects within a language also affect economic and social activities. Falck et al. (2012) find that cross-regional migration flows in Germany in the 2000 s can be explained by historical dialect similarity dating back to the 19th century. Lameli et al. (2015) show that controlling for other factors, more trades take place between regions sharing similar dialects than between those sharing different dialects in Germany. Chen et al. (2014) show that migrant workers in Shanghai, China, command a higher salary and better opportunities if they can speak Shanghainese. A recent empirical study by Battiston et al. (2021) shows that face-to-face communication among co-workers enhances work productivity. All the above studies suggest dialect similarity and language familiarity can build trust among people.

Firm-level evidence also shows that linguistic distance matters in the management of multinational corporations, cross-border investment decisions, and others. Based on a qualitative analysis of 15 multinational teams from three German corporations, Tenzer et al. (2014) find that a long linguistic distance reduces trust between team members. Focusing on the entry mode of cross-border mergers and acquisitions, Cuypers et al. (2015) find that an acquirer takes a lower equity stake when the linguistic distance between the acquirer and the foreign target is long but a higher equity stake when the combined lingua franca proficiency of the parties is high. Linguistic distance also influences investors' attitude. Lundholm et al. (2018) document that U.S. institutional investors show a significant bias against firms in Quebec relative to other Canadian firms since French is the predominant language in Quebec.

Linguistic distance also affects financial outcomes in China. For instance, Li et al. (2018) find that the linguistic distance between two parties in mergers negatively affects the acquirer's abnormal return, and the effect is more pronounced if the acquirer is less likely to be proficient in Standard Mandarin. Other studies find that the CEO-board chair dialect similarity enhances bank profitability and efficiency (Bian et al., 2019) and stock price informativeness (Fu et al., 2021), reduces agency and precautionary cash holding (S. Hu, Xiao et al., 2021; G. Hu, Xiao et al., 2021), and enhances analysts' investment recommendations (Zhang, 2022). Du (2019), on the other hand, documents the dark side of dialect similarity by showing that when the CEO of a pre-IPO firm shares the same dialect as the auditor, the audit quality suffers. These studies suggest dialect similarity creates trust among members that reinforce corporate governance in some situations while having an opposite effect in some others.

2.2.2. Information and analyst forecast accuracy

Financial analysts are the main players in financial markets to gather and process information for their forecasts. Their performance depends significantly on the amount and quality of information available to analysts. Early studies show that more comprehensive financial report information reduces the information gap between firms and analysts, thereby improving analyst forecast accuracy (Hope, 2003). Besides financial information, disclosure of nonfinancial information such as corporate governance and corporate social responsibilities also enhances analyst forecast accuracy (Bhat et al., 2006; Dhaliwal et al., 2012). More recent studies show that quality disclosure of information, financial or nonfinancial, also enhances analyst forecast accuracy (Chen et al., 2015; Muslu et al., 2019; Hu, Liu et al., 2021). All these results suggest that financial analysts collect and synthesize information from different sources to develop a full picture of a firm's prospects.

While a company's information disclosure determines information availability, some barriers exist to deter analysts from collecting further information to enhance their understanding of the company. Geographic distance is a natural barrier to information access. Therefore, local analysts have information advantage over nonlocal ones and therefore can provide more accurate forecasts than nonlocal ones (Bae et al., 2008). The local advantage is more prominent in less developed markets. Recent studies show that other distances such as language barriers (Cho et al., 2020) and cultural gaps (Frijns & Garel, 2021) also hinder analysts' forecasts. Therefore, not only information unavailability but also

[11] To illustrate, the grammatical difference between Dutch and English is considered one of the reasons for the aviation disaster on the runway of Tenerife Airport in 1977. Specifically, the Dutch pilot radioed to the controller, "We are now at takeoff," where in Dutch "at takeoff" refers to an action but in English it refers to a position. In other words, by "at takeoff," the Dutch pilot meant "in the process of taking off." The controller heard that the Dutch pilot was ready at the position and would await further instructions and then replied, "OK." The Dutch pilot then rolled down the runway and collided with a Pan Am plane sitting on the same runway (Wierzbicka, 2006: 304).

language and cultural obstacles affect analysts' ability to make accurate forecasts.

Financial analysts adopt different strategies to fill in the information mosaic. In the United States, analysts could attend conference calls to collect more information and enhance their forecasts (Bowen et al., 2002; Kimbourgh, 2005). They could also privately communicate (e.g. via phone calls, meetings, etc.) with top executives (Soltes, 2014; Brown et al., 2015; Choy & Hope, 2022). However, private communication became much riskier after the Regulation Fair Disclosure (i.e., Reg FD) was passed in October 2000 to require non-discriminating disclosure of information to all market participants or impose a severe penalty otherwise. Because of data unavailability, most studies on analyst/investor communications focus on conference presentations (Bushee et al., 2011, 2017; Green et al., 2014a; b) or analyst/investor days (Kirk & Markov, 2016) and few studies have examined private meetings (Soltes, 2014; Kirk & Markov, 2016).

In China, besides conducting a video conference, financial analysts pay a firm visit for additional information (Cheng et al., 2016; Han et al., 2018). During the visit, analysts may access unique information via interactions with corporate management (Han et al., 2018; Yin et al., 2018). Empirically, corporate site visits are found to be useful for analysts to improve their forecast accuracy (Cheng et al., 2016) and to provide more differential information in research reports (Yin et al., 2018) and for institutional investors to improve their trading performance (Cheng et al., 2019).

2.3. Hypotheses development

Corporate public disclosure is the major source of information for equity research. If corporations and analysts communicate only through textual information, all analysts should have almost the same understanding of information conveyed by corporations. Nevertheless, surveys show that corporate top executives spend one-third of their working time talking with financial analysts (Soltes, 2014), and analysts consider information from private conversations more critical than those in financial reports (Brown et al., 2015). Therefore, private communication is important for analysts to improve their understanding of their subject firms.

However, verbal communication is effective only when parties in conversations are equally proficient in the communication language. Studies have analyzed communication difficulties due to language barriers from different perspectives. Cross-country linguistic studies suggest that differences exist among languages in vocabulary, grammar, and syntax, among others (e.g., Chen, 2013; Huang & Kim, 2020). According to social identity theory, people divide themselves into groups by social identity proxies. Natural language reflects individual identity and ideology, especially in personality, cognition, and psychological modes (Brewer, 1979; Hogg & Turner, 1987; Fishman, 1991; Pennebaker & Graybeal, 2001; Pennebaker et al., 2003; Holtzman et al., 2015). People speaking different languages also have habitual expressions, such as choice of word/phrase or sentence structure (Akstinaite et al., 2020). The same exists among dialects in China. Even if a native speaker of a dialect (e.g., Cantonese) can also speak Mandarin, their Mandarin accent and choices of expressions are likely to be different from those of the native Standard Mandarin speaker. Unless they have received professional training in Standard Mandarin, some information may be lost when they translate ideas from their mother tongue into Standard Mandarin or vice versa. Recent empirical studies in China also shows that sharing a common dialect promotes trust among people and enhances communications between them (Bian et al., 2019; Du, 2019). Therefore, while analysts and corporate executives can communicate in Standard Mandarin, differences in accent and choice of word/phrase could create a psychological barrier that limits the depth of conversation.

Therefore, we expect that if an analyst does not speak the dialect of the subject firm's local executives, the problem of inaccurate or ambiguous information may arise in distant communication. To overcome the information obstacle, the analyst could visit the subject firm's headquarters directly. Corporate site visits are a "seeing is believing" method for information gathering and verification. During the visit, the analyst can communicate with the firm's executives face-to-face, which could alleviate miscommunication. In addition, the analyst can observe the firm's daily operations and collect circumstantial evidence to support or correct their forecast. In sum, the information gap arising from the language barrier could be narrowed down by a site visit. Thus, we hypothesize the following:

Hypothesis 1. *A higher language barrier results in a higher probability of a corporate site visit.*

Studies show that corporate site visits enhance the forecast accuracy of analysts (Cheng et al., 2016; Han et al., 2018), indicating that corporate site visits are an effective method for analysts to obtain more precise information. If such a benefit exists, it should be larger when the information asymmetry is larger, that is, when the barrier for communication is higher. Thus, we hypothesize as follows:

Hypothesis 2. *The enhancement effect of corporate site visits on analyst forecast accuracy is more pronounced when the subject firm location's dominant dialect is more different from Standard Mandarin.*

3. Data and Methodology

3.1. Sample

Our initial sample consists of all firms listed on the SZSE in 2012–2017. The sample starts from 2012 when the CSMAR started reporting detailed corporate site visit information. Corporate site visit data includes the visiting date, meeting location, brokerage firm names, analyst names, and the memos of the Q&A section. For our research, we exclude off-site communications such as video conferences and press conferences that mainly involve verbal communications only. Analyst forecast data and corporate financial data are also obtained from the CSMAR. We hand-merge the corporate site visit data and the forecast data by brokerage firm name. We exclude the firm-year-brokerage observations where there is no analyst forecast information from the brokerage firm.

To test our hypotheses, we collect county-level dialectal information for each sample firm according to the location of the firm's headquarters. The geographic distribution of dialects and linguist distance data come from *The Language Atlas of China* and the relevant literature (Li et al., 2018). The atlas provides a matrix that indicates the linguistic distance between each dialect pair from 0 to 5, where 0 represents no difference between the two dialects and 5 represents the longest distance. Popularization rates of Standard Mandarin and local dialects in each province are collected from the *Survey of Language Situation in China*. We remove firms in the financial service industry and observations with missing variables. The final sample for main regressions consists of 48,120 firm-year-brokerage observations (7102 firm-year observations).

3.2. Measurement of analyst forecast accuracy

The dependent variable is analyst forecast accuracy. Following previous studies (e.g. Duru & Reeb, 2002; Walther & Willis, 2013; Han et al., 2018), we measure forecast accuracy as:

$$Accuracy_{ijt} = -100 \times \frac{|Forecast \quad EPS_{ijt} - Actual \quad EPS_{jt}|}{Price_{jt}} \tag{1}$$

where *Forecast* EPS_{ijt} is the last earnings per share forecast issued by brokerage firm i for firm j before the date of firm j's earnings announcement in year t. *Actual* EPS_{jt} is the actual earnings per share reported by firm j for year t. $Price_{jt}$ is the stock price of firm j at the beginning of year t. The standardized absolute difference is multiplied by -100 so that a larger value of $Accuracy_{ijt}$ represents a more accurate forecast.

3.3. Baseline model

To test Hypothesis 1, we construct baseline model 1 as follows:

$$Site\quad Visit = \alpha + \beta Linguistic + \gamma M + \varepsilon \tag{2}$$

where *Site Visit* is either (a) the natural logarithm of 1 plus the number of corporate site visits made by the brokerage firm to the subject firm (*Ln* (*NVisit*)) or (b) a dummy variable (*DVisit*) that equals 1 if the brokerage firm visits the firm this year, and 0 otherwise. *Linguistic* is either (a) the linguistic distance between the subject firm location's dominant dialect and Standard Mandarin (*LDist*) or (b) an indicator for long linguistic distance (*LongLDist*) that equals 1 for the regions with *LDist* from 3 to 5, and 0 otherwise. *M* is a set of control variables. Logit model and negative binomial model are adopted to estimate baseline model 1. Brokerage firm-, industry-, and year-fixed effects are included.[12]

To test Hypothesis 2, we construct baseline model 2 as follows:

$$Accuracy = \alpha + \beta Site\quad Visit + \gamma X + \varepsilon \tag{3}$$

where *Accuracy* is the analyst's forecast accuracy as defined in Section 3.2. *X* is a set of control variables that is similar to *M*.

Following Cheng et al. (2016) and Han et al. (2018), we include certain control variables for models (2) and (3). We use three proxies for the brokerage firm's resources and research experience. A brokerage firm is more likely to pay a site visit if it is more established and resourceful. *BrokerSize* is the natural logarithm of the number of analysts employed by the brokerage firm. *BrokerGexp* is the natural logarithm of 1 plus the number of years since the brokerage firm began providing earnings forecasts for any listed firms. *BrokerCover* is the natural logarithm of 1 plus the number of companies covered by the brokerage firm. On the other hand, paying a site visit to a listed firm is less necessary if the brokerage firm/analyst is more familiar with the firm. To capture a brokerage firm's knowledge of the firm, we define *BrokerFexp* as the natural logarithm of 1 plus the number of years since the brokerage firm started issuing earnings forecasts for the firm. Finally, paying a site visit to a firm is more necessary if an analyst always updates their forecast for the firm's earning. To capture this effect, we define *BrokerFreq* as the natural logarithm of 1 plus the number of earnings forecasts issued by the brokerage firm for the firm.

Other control variables are defined as follows. *Horizon* is the natural logarithm of 1 plus the days elapsed from the brokerage firm's latest earnings forecast date to the earnings announcement date. *Follow* is the natural logarithm of 1 plus the number of brokerage firms that cover the subject firm. *Size* is the firm's market value. *Return* is the firm's annual stock return. *Loss* is an indicator that equals 1 if the firm's net profit is negative and 0 otherwise. *Leverage* is the firm's liabilities-to-assets ratio. *Local* is an indicator that equals 1 if the brokerage firm is headquartered in the same city as the firm and 0 otherwise. *Inst* is the firm's institutional ownership. *SDNI* is the firm's standard deviation of net income over the past five years. *HSR* is an indicator that equals 1 if there is at least one HSR station in the city of the firm's headquarters and 0 otherwise. Firm-, brokerage firm-, and year-fixed effects are included to control for unobservable heterogeneities among firms and among brokers as well as unobservable time-varying macro factors that affect analysts' forecast accuracy. All continuous variables are winsorized at the 1st and 99th percentiles. Appendix 1 provides the definitions of all the variables.

[12] We do not include firm-fixed effects because of the associated incidental parameters problem pointed out by Greene (2008), among others.

Table 3
Summary Statistics. This table presents the summary statistics of key variables. All continuous variables are winsorized at the 1st and 99th percentiles to alleviate the effects of outliers. The definitions of all variables are available in Appendix 1.

Variables	Obs	Mean	SD	25%	50%	75%
DVisit	48,120	0.386	0.487	0	0	1
NVisit	48,120	0.601	0.959	0	0	1
Ln (Visit)	48,120	0.343	0.469	0	0	0.693
Accuracy	48,120	-0.669	1.049	-0.773	-0.274	-0.084
LDist	48,120	3.293	1.876	2	4	5
LongLDist	48,120	0.512	0.500	0	1	1
BrokerSize	48,120	3.676	0.635	3.332	3.829	4.127
BrokerGexp	48,120	2.371	0.325	2.197	2.398	2.565
BrokerFexp	48,120	0.928	0.810	0	0.693	1.609
BrokerCover	48,120	6.899	0.781	6.612	7.047	7.445
BrokerFreq	48,120	1.566	0.682	1.099	1.609	2.079
Horizon	48,120	4.811	0.927	4.369	5.094	5.468
Follow	48,120	2.643	0.714	2.197	2.773	3.178
Size	48,120	22.970	0.900	22.340	22.890	23.560
Return	48,120	0.260	0.591	-0.152	0.109	0.504
Loss	48,120	0.029	0.168	0	0	0
Leverage	48,120	0.384	0.193	0.227	0.364	0.528
Local	48,120	0.074	0.262	0	0	0
Inst	48,120	5.995	4.462	2.550	5.170	8.613
SDNI	48,120	0.020	0.018	0.009	0.015	0.025
HSR	48,120	0.777	0.416	1	1	1

4. Results

4.1. Descriptive statistics

Table 3 reports the statistics summary for the main variables. The mean value of *DVisit* is 0.386, suggesting that in the sample, 38.6% of financial analysts pay a site visit to the listed firms they cover. The mean value of *Accuracy* is -0.669 with a standard deviation of 1.049. The average value of *LDist* is 3.293 and the median is 4, indicating that the majority of forecasts are made for firms headquartered in counties with a long linguistic distance from Standard Mandarin. *Local* has a mean value of 0.074, suggesting that 7.4% of the forecasts are made by local brokerage firms. In other words, most forecasts in the sample are made by non-local brokerage firms. The mean value of *HSR* is 0.777, suggesting that most of the forecasts are made for firms in counties with HSR access. Because of the rapid development of HSR in the recent decade and the fact that most listed firms are located in large cities, the high mean value of *HSR* is not surprising.

4.2. Main empirical results

Table 4 reports the regression results for model 1. Columns (1) and (2) report the logit model results for an analyst's decision to visit the subject firm (*DVisit*). For each column, the coefficient of the key independent variable, that is, the dummy for long linguistic distance (*LongLDist*) or the value of linguistic distance (*LDist*), is positive and significant at the 1% level. The economic significance is such that the probability of a site visit is 4.81% higher if linguistic distance increases by one. Columns (3) and (4) report the negative binomial model results for the number of visits made by the analyst to the subject firm (*Ln* (*NVisit*)), and the coefficients of *LongLDist* and *LDist* are still both positive and significant at the 1% level. The result indicates that a longer linguistic distance increases the probability of analysts conducting a site visit, consistent with Hypothesis 1.

The coefficients of the control variables are basically consistent with our expectations. Brokerage firms with more analysts (i.e., higher *BrokerSize*) and providing more forecasts for the subject firm (i.e., higher *BrokerFreq*) are more likely to conduct a visit. Brokerage firms with a longer history of covering the subject firm (i.e., higher *BrokerFexp*) are less likely to make a site visit. A brokerage firm's experience with other

Table 4

Language Barriers and Corporate Site Visit. This table presents regression results for the effect of linguistic distance on analysts' corporate site visits. The dependent variable of columns (1) and (2) is a dummy that equals 1 if the brokerage visits the firm and 0 otherwise (*DVisit*). The dependent variable of columns (3) and (4) is the total number of visits made by the brokerage to the firm (*NVisit*). The key independent variable in columns (1) and (3) is a dummy that equals 1 if the linguistic distance is 3 or above and 0 otherwise (*LongLDist*), and the key independent variable in columns (2) and (4) is the linguistic distance between the dialect and Standard Mandarin ranging from 0 to 5 (*LDist*). Logit regression model is adopted in columns (1) and (2), and the negative binomial regression model is adopted in columns (3) and (4). Brokerage firm-, industry-, and year-fixed effects are included. All continuous variables are winsorized at the 1st and 99th percentiles. Heteroscedasticity robust standard errors, clustered by firm, are in parentheses. $p < 0.1$, $p < 0.05$, and $p < 0.01$ levels of significance are represented as * , * *, and * ** , respectively. The definitions of all variables are available in Appendix 1.

	DVisit		*NVisit*	
	(1)	(2)	(3)	(4)
LongLDist	0.216 * **		0.179 * **	
	(0.062)		(0.048)	
LDist		0.047 * **		0.037 * **
		(0.016)		(0.012)
BrokerSize	0.123 * *	0.123 * *	0.096 * *	0.097 * *
	(0.060)	(0.060)	(0.042)	(0.042)
BrokerGexp	0.380	0.373	0.339 *	0.333 *
	(0.238)	(0.238)	(0.193)	(0.193)
BrokerFexp	-0.188 * **	-0.188 * **	-0.154 * **	-0.154 * **
	(0.026)	(0.026)	(0.019)	(0.019)
BrokerCover	0.001	0.002	-0.007	-0.007
	(0.054)	(0.054)	(0.038)	(0.038)
BrokerFreq	0.349 * **	0.349 * **	0.330 * **	0.330 * **
	(0.023)	(0.023)	(0.016)	(0.016)
Follow	0.230 * **	0.232 * **	0.190 * **	0.191 * **
	(0.037)	(0.037)	(0.028)	(0.028)
Size	-0.064	-0.065	0.003	0.001
	(0.048)	(0.048)	(0.038)	(0.039)
Return	0.206 * **	0.205 * **	0.144 * **	0.144 * **
	(0.042)	(0.042)	(0.029)	(0.029)
Loss	-0.184	-0.185	-0.170 * *	-0.171 * *
	(0.114)	(0.114)	(0.085)	(0.086)
Leverage	-0.443 * **	-0.441 * **	-0.342 * **	-0.340 * **
	(0.161)	(0.161)	(0.122)	(0.122)
Local	0.393 * **	0.395 * **	0.359 * **	0.360 * **
	(0.054)	(0.054)	(0.036)	(0.036)
Inst	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
SDNI	-1.111	-1.142	-0.897	-0.932
	(1.200)	(1.208)	(0.956)	(0.963)
HSR	0.366 * **	0.358 * **	0.263 * **	0.256 * **
	(0.068)	(0.069)	(0.053)	(0.053)
Constant	-1.873 *	-1.889 *	-21.141	-22.017
	(1.101)	(1.109)	(131.145)	(.)
Observations	46,864	46,864	48,120	48,120
Pseudo R-squared	0.092	0.092	0.088	0.088
Brokerage FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

firms, however, has no impact on the probability of visiting the subject firm as indicated by the insignificant coefficients of *BrokerGexp* and *BrokerCover*.

A financial analyst is more likely to conduct a site visit when there are more of other brokerage firms following a firm (i.e., higher *Follow*). While studies have widely documented a positive relation between firm size and analyst coverage, larger firms (i.e., higher *Size*) do not attract more site visits. Meanwhile, firms with higher recent stock returns (i.e., higher *Return*) and less leveraged firms (i.e., lower *Leverage*) are more likely to be visited. Local firms (*Local* = 1) are more likely to attract local brokerage firms. In addition, firms in counties with HSR access (*HSR* = 1) are more likely to receive site visits from brokerage firms. The result is consistent with findings in recent studies (Chen et al., 2022; Kong et al., 2020). The significant coefficients of *Local* and *HSR* indicate that traveling time and cost are important factors for site visit decisions.

Table 5

Language Barriers, Corporate Site Visit, and Forecast Accuracy. This table presents ordinary-least-squares (OLS) results for the effect of corporate site visit on analyst forecast accuracy. The dependent variable is forecast accuracy (*Accuracy*). The key independent variable is a dummy that equals 1 if the brokerage visits the firm and 0 otherwise (*DVisit*) in columns (1) and (3) and is the natural logarithm of 1 plus the number of visits made by the brokerage to the firm (*Ln* (*NVisit*)) in columns (2) and (4). Columns (1) and (2) show the results for firms headquartered at locations with a short linguistic distance (*LongLDist* = 0), and columns (3) and (4) show the results for firms headquartered at locations with a long linguistic distance (*LongLDist* = 1). Firm-, brokerage firm-, and year-fixed effects are included. All continuous variables are winsorized at the 1st and 99th percentiles. Heteroscedasticity robust standard errors, clustered by brokerage firm, are reported in parentheses. $p < 0.1$, $p < 0.05$, and $p < 0.01$ levels of significance are represented as * , * *, and * ** , respectively. The definitions of all variables are available in Appendix 1.

	Short Linguistic Distance (*LongLDist* = *0*)		Long Linguistic Distance (*LongLDist* = *1*)	
	(1)	(2)	(3)	(4)
DVisit	0.016		0.048 * **	
	(0.016)		(0.010)	
Ln(*NVisit*)		0.018		0.058 * **
		(0.016)		(0.014)
BrokerSize	0.004	0.004	-0.002	-0.002
	(0.056)	(0.056)	(0.041)	(0.041)
BrokerGexp	-0.028	-0.027	-0.017	-0.017
	(0.145)	(0.146)	(0.115)	(0.115)
BrokerFexp	0.012	0.012	-0.004	-0.004
	(0.013)	(0.013)	(0.008)	(0.007)
BrokerCover	-0.048	-0.048	0.009	0.009
	(0.044)	(0.044)	(0.028)	(0.028)
BrokerFreq	-0.012	-0.012	0.001	0.000
	(0.014)	(0.014)	(0.010)	(0.010)
Horizon	-0.327 * **	-0.327 * **	-0.338 * **	-0.338 * **
	(0.010)	(0.010)	(0.008)	(0.008)
Follow	0.077 * **	0.077 * **	0.065 * **	0.066 * **
	(0.025)	(0.025)	(0.015)	(0.015)
Size	0.126 * **	0.126 * **	0.182 * **	0.182 * **
	(0.030)	(0.030)	(0.025)	(0.025)
Return	-0.143 * **	-0.143 * **	-0.120 * **	-0.119 * **
	(0.020)	(0.020)	(0.012)	(0.012)
Loss	-1.512 * **	-1.512 * **	-1.627 * **	-1.626 * **
	(0.098)	(0.098)	(0.064)	(0.064)
Leverage	-0.046	-0.047	-0.150 * *	-0.152 * *
	(0.105)	(0.105)	(0.069)	(0.069)
Local	-0.039	-0.039	0.028	0.028
	(0.031)	(0.031)	(0.021)	(0.021)
Inst	0.004 * **	0.004 * **	0.004 * **	0.004 * **
	(0.002)	(0.002)	(0.001)	(0.001)
SDNI	-3.503 * **	-3.506 * **	-6.613 * **	-6.606 * **
	(1.083)	(1.083)	(0.757)	(0.755)
Constant	-1.817 * *	-1.817 * *	-3.321 * **	-3.302 * **
	(0.757)	(0.757)	(0.609)	(0.610)
Fisher's Permutation test			-0.032 * **	-0.026 * *
Observations	15,446	15,446	32,520	32,520
R-squared	0.460	0.460	0.415	0.415
Firm FE	Yes	Yes	Yes	Yes
Brokerage firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

To test whether analysts gain from visiting firms in regions dominated by dialects largely different from Standard Mandarin, that is, *H2*, we classify firms into two groups according to *LDist* of the firm location. Table 5 reports the regression results. Columns (1) and (2) show regression results for firms with *LongLDist* equal to 0, while columns (3) and (4) show the results for firms with *LongLDist* equal to 1. In general, analysts gain from visiting firms with a long linguistic distance as indicated by the significant coefficients of *DVisit* and *Ln*(*NVisit*) in columns (3) and (4). In terms of economic significance, compared with analysts who do not conduct a visit, analysts who do can improve their forecast accuracy by 0.048 (column (3)), which is appropriately 17.5%

Table 6

Forecast Accuracy: Endogeneity. This table presents results from additional tests addressing endogeneity. Panel A reports the 2SLS results with the instrumental variable, and panel B reports the results of treatment-effect estimation. In panel A, the dependent variable for columns (1) and (3) is the natural logarithm of 1 plus the number of visits made by the brokerage to the firm (*Ln*(*NVisit*)). The dependent variable for columns (2) and (4) is forecast accuracy (*Accuracy*). The instrumental variable *HSR* in columns (1) and (3) is a dummy that equals 1 if there is at least one high-speed-railway station where the firm is headquartered. Column (2) shows the result for firms headquartered at locations with a short linguistic distance (*LongLDist* = 0), and column (4) shows the results for firms headquartered at locations with a long linguistic distance (*LongLDist* = 1). In panel B, each treatment firm is matched with a control firm by nearest-neighbor matching, after controlling for covariates same as those in column 1 of Table 4. The dependent variable is forecast accuracy (*Accuracy*). The coefficients in columns (1) and (2) show the estimate average treatment effect on the treated (*DVisit* = *1*). All continuous variables are winsorized at the 1st and 99th percentiles. Heteroscedasticity robust standard errors, clustered by firm, are in parentheses. $p < 0.1$, $p < 0.05$, and $p < 0.01$ levels of significance are represented as * , * *, and * ** , respectively. The definitions of all variables are available in Appendix 1.

Panel A 2SLS	*Short Linguistic Distance*		*Long Linguistic Distance*	
	1st: Ln (NVisit)	*2nd: Accuracy*	*1st: Ln (NVisit)*	*2nd: Accuracy*
	(1)	(2)	(3)	(4)
Pred Ln (NVisit)		0.560		1.131 * *
		(0.507)		(0.577)
HSR	0.104 * **		0.056 * **	
	(0.022)		(0.016)	
BrokerSize	0.052 * **	-0.016	0.031 * *	-0.042
	(0.018)	(0.047)	(0.013)	(0.035)
BrokerGexp	-0.032	-0.108	-0.100 * *	0.093
	(0.061)	(0.142)	(0.041)	(0.119)
BrokerFexp	-0.035 * **	-0.047 *	-0.046 * **	0.008
	(0.009)	(0.025)	(0.007)	(0.031)
BrokerCover	-0.007	-0.053	0.018 *	-0.020
	(0.013)	(0.035)	(0.011)	(0.027)
BrokerFreq	0.081 * **	-0.021	0.101 * **	-0.102 *
	(0.009)	(0.041)	(0.007)	(0.061)
Horizon	-0.008	-0.316 * **	-0.015 * **	-0.321 * **
	(0.005)	(0.019)	(0.003)	(0.015)
Follow	0.042 * **	0.129 * **	0.055 * **	0.054
	(0.014)	(0.039)	(0.010)	(0.038)
Size	0.002	0.055 *	-0.007	0.055 * *
	(0.015)	(0.031)	(0.014)	(0.027)
Return	0.057 * **	-0.132 * **	0.040 * **	-0.076 * *
	(0.016)	(0.050)	(0.012)	(0.038)
Loss	-0.032	-1.825 * **	-0.030	-1.591 * **
	(0.033)	(0.183)	(0.025)	(0.154)
Leverage	-0.082	-0.857 * **	-0.096 * *	-0.833 * **
	(0.051)	(0.156)	(0.045)	(0.105)
Local	0.044 *	-0.038	0.118 * **	-0.078
	(0.025)	(0.055)	(0.017)	(0.077)
Inst	-0.001	0.000	-0.000	0.001
	(0.000)	(0.001)	(0.000)	(0.001)
SDNI	-0.415	-2.912 * *	-0.109	-5.425 * **
	(0.388)	(1.151)	(0.314)	(1.012)
Constant	-0.198	0.305	0.052	0.941
	(0.358)	(0.738)	(0.311)	(0.631)
Observations	15,483	15,483	32,637	32,637
R-squared	0.158	0.257	0.144	0.061
F-statistic	19.27		31.33	
Brokerage FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Panel B Treatment-effects estimation	*Short Linguistic Distance*	*Long Linguistic Distance*
ATET	(1)	(2)
DVisit (1 vs 0)	0.027	0.038 * **
	(0.018)	(0.014)
Observations	13,860	29,240
Estimator	Nearest-neighbor matching 0.144	

Table 6 (*continued*)

Panel B Treatment-effects estimation	*Short Linguistic Distance*	*Long Linguistic Distance*
ATET	(1)	(2)
Control	Yes	Yes
Industry FE	Yes	Yes
Year FE	Yes	Yes

of the sample median value of *Accuracy*. Meanwhile, analysts do not gain from visiting firms in locations with a short linguistic distance.[13] The Fisher's permutation test shows that differences in coefficients for *DVisit* between columns (1) and (3) and differences in coefficients for *Ln* (*NVisit*) between columns (2) and (4) are significant at the 1% and 5% levels, respectively. Overall, the results are consistent with *H2,* that is, corporate site visit is effective in reducing language barriers between an analyst and the subject firm's stakeholders, especially when the stakeholders' mother tongue is largely different from Standard Mandarin.

The coefficients of the control variables generally meet our expectations. Specifically, analysts' forecast errors are smaller when the subject firm is larger and covered by more brokerage firms and institutional ownership is higher. High recent stock return, operation loss, and large profit volatility negatively affect analysts' forecast accuracy. Local analysts, however, do not gain an advantage over nonlocal analysts in earnings forecasts, inconsistent with Cheng et al. (2016).

4.3. Endogeneity concerns

An analyst's decision to conduct a site visit and their forecast accuracy may be endogenous; that is, there could unobservable factors that drive both site visit decision and forecast accuracy. In the above regressions, we add firm- and brokerage firm-fixed effects to capture unobservable heterogeneities among firms and among brokers. In the following, we use two other approaches to further address the endogeneity concern and report the results in Table 6.

First, we run a 2SLS model for analysts' forecast accuracy. Following Chen et al. (2022), we use the availability of an HSR station at the subject firm location as the instrument for *Ln*(*NVisit*). The availability of an HSR station significantly reduces the traveling cost and time of site visit. Kong et al. (2020) and Chen et al. (2022) show that after the development of the HSR, analysts were more likely to visit firms located in HSR-connected cities than in non-HSR-connected cities. On the other hand, the HSR development and construction plan was made by China's central government, not listed firms, and there is no reason to expect that HSR would directly affect analysts' forecast accuracy.

The results in panel A of Table 6 are consistent with those in Table 5; that is, an analyst gains more from a site visit if the subject firm location's dominant dialect is largely different from Standard Mandarin. In particular, the coefficients of *HSR* in the first-stage regressions (columns 1 and 3) are both positive and significant at the 1% significance level. The *F*-statistics in both columns are larger than 10, which are sufficiently large to reject the weak instrument hypothesis (Stock & Watson, 2003).

[13] Our sub-sample result indicates that one third of corporate site visits do not improve analysts' earnings forecast accuracy. Following the literature, our study mainly focuses on the impact of a site visit on the visiting analyst's short-term earnings forecast accuracy. We recognize this measurement of benefit is incomplete because it does not consider the possibility that site visits may enhance analysts' earnings forecast accuracy beyond the current year. A site visit could also provide insights for long-term forecasts. Recent studies show that corporate site visits reduce earnings management (Gao et al., 2022), increase pay-performance sensitivity (Wu et al., 2022), and improve ESG performance (Jiang et al., 2022; Zhou & Gan, 2022). A deeper understanding of those changes sometimes enhances analysts' long-term rather than short-term earnings forecast accuracy because some changes tend to have long-term effects on firm performance.

Table 7

Forecast Accuracy: BSGS, Provincial Capitals, and Other Locations. This table presents regression results for the effect of analysts' corporate site visit on analysts' forecast accuracy by firm headquarter location. The dependent variable is forecast accuracy (*Accuracy*). The key independent variable is a dummy that equals 1 if the brokerage visits the firm and 0 otherwise (*DVisit*) or the natural logarithm of 1 plus the number of visits made by the brokerage to the firm (*Ln*(*NVisit*)). In panel A, columns (1)–(4) show the results for firms located in Beijing, Shanghai, Guangzhou, or Shenzhen (BSGS), four main business hubs in China, and columns (5)-(8) show the results for firms located in BSGS. In panel B, columns (1)–(4) show the results for firms outside provincial capitals and BSGS, and columns (5)–(8) show the results for firms located in provincial capitals (excluding BSGS). Firm-, brokerage firm-, and year-fixed effects are included. All continuous variables are winsorized at the 1st and 99th percentiles to alleviate the effects of outliers. Heteroscedasticity robust standard errors, clustered by brokerage firm, are in parentheses. $p < 0.1$, $p < 0.05$, and $p < 0.01$ levels of significance are represented as * , * *, and * ** , respectively. The definitions of all variables are available in Appendix 1.

Panel A *Bejing, Shanghai, Guangzhou, and Shenzhen (BSGS)* vs. *others*

	Non-BSGS				*BSGS*			
	Short Linguistic Distance		*Long Linguistic Distance*		*Short Linguistic Distance*		*Long Linguistic Distance*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DVisit	0.039		0.061 * **		-0.007		0.017	
	(0.025)		(0.012)		(0.020)		(0.017)	
Ln(*NVisit*)		0.040		0.051 * **		-0.004		0.008
		(0.027)		(0.013)		(0.019)		(0.016)
Observations	8732	8732	23,090	23,090	6710	6710	9424	9424
R-squared	0.467	0.467	0.434	0.434	0.465	0.465	0.382	0.382
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Brokerage FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Panel B Provincial capitals (excluding BSGS) vs. others

	Outside Provincial Capitals				*Provincial Capitals*			
	Short Linguistic Distance		*Long Linguistic Distance*		*Short Linguistic Distance*		*Long Linguistic Distance*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DVisit	0.041		0.071 * **		0.044		0.046 * *	
	(0.031)		(0.013)		(0.039)		(0.022)	
Ln(*NVisit*)		0.033		0.062 * **		0.063		0.036
		(0.034)		(0.014)		(0.046)		(0.022)
Observations	6518	6518	14,378	14,378	2204	2204	8710	8710
R-squared	0.454	0.454	0.426	0.426	0.478	0.478	0.416	0.416
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Brokerage FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

The result suggests that *HSR* is an effective instrumental variable and is consistent with Chen et al. (2022) that HSR increases accessibility. Furthermore, results from the second-stage regressions (columns 2 and 4) show that the coefficient of *Pred Ln*(*NVisit*) is positive and significant only when the subject firm location's dominant dialect is largely different from Standard Mandarin.

Second, we adopt propensity score matching (PSM) to address the endogeneity concern. For each analyst who conducts a visit to a firm and provides an earnings forecast afterward, we match them with another analyst who provides an earnings forecast for the same firm without conducting a visit before and has the nearest characteristics to the visiting analyst. We use nearest-neighbor matching to identify the closest match with regression variables for model 1 as the inputs. We perform matching for firms with $LongLDist = 0$ and firms with $LongLDist = 1$. After matching, we calculate the average treatment effect on the treated (i.e., $DVisit = 1$) by taking the average of the difference between the observed and potential values of *Accuracy* for each observation. Panel B of Table 6 reports the result, which indicates that a site visit improves an analyst's forecast accuracy only when the subject firm location's dominant dialect is largely different from Standard Mandarin.

4.4. Additional analysis

The linguistic distance measure does not account for people's exposure to other languages or dialects. Therefore, speakers of a specific dialect could have a different proficiency in another language. In China, while people in some major business hubs such as Guangzhou and Shanghai speak dialects with the longest linguistic distance from Standard Mandarin, they commonly use Standard Mandarin in daily business communication and in the workplace to attract talents nationwide. As a result, inherent language barriers should have weaker effects on analysts' forecast accuracy in those business hubs. To test this expectation, within each group classified by *LongLDist*, we separate firms in BSGS from those in other areas. We then rerun model 2 for each group of firms and report the results in panel A of Table 7. Columns (1)–(4) report the results for the non-BSGS sample, and columns (5)–(8) report the results for the BSGS sample. The results indicate that corporate site visits enhance analysts' forecast accuracy only when visited firms are located in nonhub areas where people speak in dialects largely distinct from Standard Mandarin. Therefore, both inherent dialect and acquired exposure to Standard Mandarin affect a dialect speaker's barriers to communication in Standard Mandarin.

Besides BSGS, other large cities also adopt Standard Mandarin more widely than small cities and rural areas. For example, provincial capitals are local administrative centers, and therefore, Standard Mandarin is likely to be the working language in government bureaus and large corporations. To further test if the impact of corporate site visits depends on dialect speakers' exposure to Standard Mandarin, within each group classified by *LongLDist*, we separate firms in provincial capitals (excluding BSGS) from firms in other locations. We then rerun model 2 for each group of firms and report the results in panel B of Table 7. Columns (1)–(4) report the results for firms outside provincial capitals and BSGS, and columns (5)–(8) report the results for firms in provincial capitals (excluding BSGS). The results indicate that outside provincial capitals and BSGS, corporate site visits enhance analysts' forecast accuracy when dominant dialects in visited firms' locations are distinct

Table 8

Forecast Accuracy: Popularization Rate. This table presents regression results for the effect of analysts' corporate site visit on analysts' forecast accuracy for the regions with different popularization rate of Standard Mandarin (panel A) and dialects (panel B). The dependent variable is forecast accuracy (*Accuracy*). The key independent variable is a dummy that equals 1 if the brokerage visits the firm and 0 otherwise (*DVisit*) or the natural logarithm of 1 plus the number of visits made by the brokerage to the firm (*Ln*(*NVisit*)). In panel A, columns (1)–(4) show the results for subsamples according to the popularization rate that people speak Standard Mandarin during childhood, and columns (5)–(8) show the results for subsamples according to the popularization rate that people speak Standard Mandarin at work. In panel B, columns (1)–(4) show the results for subsamples according to the popularization rate of dialects during childhood, and columns (5)–(8) show the results for subsamples according to the popularization rate of dialects at work. Firm-, brokerage firm-, and year-fixed effects are included. All continuous variables are winsorized at the 1st and 99th percentiles. Heteroscedasticity robust standard errors, clustered by brokerage firm, are in parentheses. $p < 0.1$, $p < 0.05$, and $p < 0.01$ levels of significance are represented as * , * *, and * ** , respectively. The definitions of all variables are available in Appendix 1.

Panel A *Standard Mandarin*

	Popularization Rate of Standard Mandarin: In Childhood				*Popularization Rate of Standard Mandarin: At Work*			
	Provinces in Top 3 Deciles		*Provinces in Bottom 3 Deciles*		*Provinces in Top 3 Deciles*		*Provinces in Bottom 3 Deciles*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DVisit	-0.000		0.074 * **		0.010		0.058 * **	
	(0.019)		(0.018)		(0.020)		(0.015)	
Ln(*NVisit*)		-0.005		0.072 * **		0.002		0.052 * **
		(0.019)		(0.018)		(0.020)		(0.016)
Observations	10,889	10,889	7927	7927	10,214	10,214	13,728	13,728
R-squared	0.458	0.458	0.448	0.448	0.451	0.451	0.451	0.451
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Brokerage FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Panel B Dialects

	Popularization Rate of Dialects: In Childhood				*Popularization Rate of Dialects: At Work*			
	Provinces in Top 3 Deciles		*Provinces in Bottom 3 Deciles*		*Provinces in Top 3 Deciles*		*Provinces in Bottom 3 Deciles*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DVisit	0.047 * **		-0.007		0.044 * **		-0.006	
	(0.011)		(0.021)		(0.012)		(0.020)	
Ln(*NVisit*)		0.038 * **		-0.013		0.035 * **		-0.009
		(0.011)		(0.019)		(0.012)		(0.019)
Observations	30,182	30,182	9236	9236	25,810	25,810	9725	9725
R-squared	0.418	0.418	0.488	0.488	0.422	0.422	0.473	0.473
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Brokerage FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

from Standard Mandarin. The coefficients of *DVisit* and *Ln*(*NVisit*) are both positive and significant at the 1% level. In provincial capitals (excluding BSGS), the result is weaker with the coefficient of *DVisit* significant at the 5% level and that of *Ln*(*NVisit*) insignificant.

In sum, the results in Table 7 are consistent with our expectation that exposure to Standard Mandarin could alleviate a dialect speaker's barriers to communication in Standard Mandarin in the workplace.

To further confirm that exposure to Standard Mandarin enhances analysts' communication with local managers, we replace the linguistic distance measure with two alternative measures of language barrier. The first one is based on the *Survey of Language Situation in China*. The survey provides province-level survey data for popularization rates of Standard Mandarin in childhood, in daily conversations, and at the workplace. For our analysis, we sort provinces from highest to lowest according to the popularization rate of Standard Mandarin in childhood and select the top three deciles and bottom three deciles for comparing the effect of site visits. We use the popularization rate in childhood because the best time to learn a second language is childhood. We rerun model 2 for the two groups separately based on this alternative measurement and report the results in panel A of Table 8. Columns (1)–(4) show that the coefficients of *DVisit* and *Ln*(*NVisit*) are statistically insignificant for regions where the popularization rate of Standard Mandarin in childhood is high but positive and significant at the 1% level for regions where the popularization rate of Standard Mandarin in childhood is low. Columns (5)–(8) confirm this finding by replacing the popularization rate in childhood by the popularization rate at work.

Furthermore, we partition the sample by popularization rates of dialect in childhood (columns (1)–(4) of panel B) and at work (columns (5)–(8)) for regressions. The results show that the positive impact of corporate site visits on analysts' forecast accuracy is more pronounced for firms in locations where people speak in dialects in childhood and at work more frequently.

The second alternative measure of language barrier comes from Tang and van Heuven (2009), who test the mutual intelligibility of Chinese dialects. In their study, for each of the 15 dialects, Tang and van Heuven (2009) selected 15 participants from a representative city (e.g. Guangzhou for Cantonese). The participants listened to 150 words spoken in different dialects and then classified each word into one of the 10 designated categories. Then they listened to 60 sentences spoken in different dialects and wrote down the target word in each sentence in their own dialect. The intelligibility of dialect A to dialect B in words (sentences) was defined as the percentage of correct answers the listener of dialect A correctly makes when the words (sentences) are presented in dialect B. The intelligibility of two dialects is not necessarily symmetric. For the purpose of our study, we define the mutual intelligibility of two dialects as the average value of intelligibility in two opposite directions.[14] A low value of mutual intelligibility indicates a longer linguistic distance between two dialects.

For this test, we only include the 15 cities included in Tang and van Heuven (2009): Suzhou, Wenzhou, Guangzhou, Xiamen, Fuzhou, Chaozhou, Meixian, Nanchang, Changsha, Taiyuan, Beijing, Jinan,

[14] The values are available in Tables 2 and 3 of Tang & van Heuven (2009).

Table 9
Forecast Accuracy: Mutual Intelligibility. This table presents regression results for the effect of analysts corporate site visit on analysts' forecast accuracy for the regions with city-level mutual intelligibility scores of words (panel A) and sentences (panel B). The dependent variable is forecast accuracy (*Accuracy*). The key independent variable is a dummy that equals 1 if the brokerage visits the firm and 0 otherwise (*DVisit*) or the natural logarithm of 1 plus the number of visits made by the brokerage to the firm (*Ln*(*NVisit*)). The regressions only include cities that are included in the mutual intelligibility test done by Tang and van Heuven (2009): Suzhou, Wenzhou, Guangzhou, Xiamen, Fuzhou, Chaozhou, Meixian, Nanchang, Changsha, Taiyuan, Beijing, Jinan, Hankou, Chengdu, and Xi'an. The cities are split into two groups (above and below median) according to their values of mutual intelligibility (in words (panel A) or in sentences (panel B)) documented by Tang and van Heuven (2009). Firm-, brokerage firm-, and year-fixed effects are included. All continuous variables are winsorized at the 1st and 99th percentiles. Heteroscedasticity robust standard errors, clustered by brokerage firm, are in parentheses. $p < 0.1$, $p < 0.05$, and $p < 0.01$ levels of significance are represented as * , * *, and * ** , respectively. The definitions of all variables are available in Appendix 1.

Panel A Mutual Intelligibility in Words

	Dependent Variable: Forecast Accuracy			
	Mutual Intelligibility > Median		*Mutual Intelligibility ≤ Median*	
	(1)	(2)	(3)	(4)
DVisit	-0.003		0.087 * **	
	(0.019)		(0.028)	
Ln (NVisit)		-0.000		0.073 * **
		(0.019)		(0.025)
Fisher's Permutation test			-0.090 * **	-0.122 * **
Observations	8480	8480	4491	4491
R-squared	0.468	0.468	0.496	0.496
Control	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Brokerage firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Panel B Mutual Intelligibility in Sentences

	Dependent Variable: Forecast Accuracy			
	Mutual Intelligibility > Median		*Mutual Intelligibility ≤ Median*	
	(1)	(2)	(3)	(4)
DVisit	0.009		0.068 * *	
	(0.019)		(0.027)	
Ln (NVisit)		0.010		0.062 * *
		(0.019)		(0.025)
Fisher's Permutation test			-0.059 * **	-0.052 * **
Observations	8681	8681	4290	4290
R-squared	0.457	0.457	0.518	0.518
Control	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Brokerage firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Hankou, Chengdu, and Xi'an. We then split the cities into two groups according to the median value of mutual intelligibility in words between the city's dominant dialect and Standard Mandarin. We rerun model 2 for the two groups and report the results in panel A of Table 9. The results show that the coefficients of *DVisit* and *Ln*(*NVisit*) are statistically insignificant for firms in cities with dialects intelligible (in words) to Standard Mandarin but positive and significant at the 1% level for firms in cities with dialects unintelligible to Standard Mandarin. The results are robust when we use mutual intelligibility in sentences for the sub-sample test (panel B). Therefore, the value of a site visit is larger when the firm location's dominant dialect is more different from Standard Mandarin.

Finally, we examine whether information benefits from corporate site visits also depend on the trustworthiness of firm managers who provide information. Recent studies provide evidence for the impacts of social capital/trust on corporate decisions, governance, and performance. Among them, Jha (2019) shows that U.S. firms headquartered in counties with high social capital are less likely to misrepresent information. In China, Dong et al. (2018) also show that higher social trust is associated with lower incidence of corporate misconduct. Following those studies, the benefits of site visits for financial analysts should be larger when firm managers are less trustworthy. By paying a site visit, financial analysts could make up for the information deficit because of the incomplete information provided by firm managers.

We expect that if the trustworthiness problem is highly correlated with the language barrier problem, then the benefits of corporate site visits in long-linguistic-distance regions should be significantly weakened after controlling for the trustworthiness of firm managers. On the other hand, if corporate site visits mainly resolve the language barrier problem or if the language barrier problem is largely orthogonal to the trustworthiness problem, then the benefits of corporate site visits in long-linguistic-distance regions should remain strong after controlling for the trustworthiness of firm managers.

Trustworthiness at the individual level is difficult to observe and measure, however. Studies in the United States mostly follow Rupasingha et al. (2006) to construct county-level social capital as a proxy for social trust. In general, social capital refers to a collection of social concepts including the values, norms, and the networks that facilitate cooperation among members (Fukuyama, 1997; Woolcock, 2001), mutual trust and altruistic tendency in a community (Guiso et al., 2004), and a network of benefits (Payne et al., 2011). Local values and norms diffuse from the community to individuals because people are connected via their social networks. Being influenced by such social values and norms over time and over generations, people feel a sense of obligation to shape their behaviors to be in line with the norms (Portes, 1998).

Following Dong et al. (2018), we use the province-level social trust index from various versions of the Business Environment Index for China's Provinces Report as a proxy for trustworthiness of firm managers.[15] The index is based on a biannual survey that aims to compare and evaluate the general business environment at the provincial level. Survey participants are asked for their opinions, in a five-point scale, the trust level in the local province, with a question similar to that in the World Value Surveys.

We split provinces into quartiles according to the social trust index and define the top quartile provinces as high-trust regions and the bottom quartile provinces as low-trust regions. Both regional subsamples are then further divided into two groups according to the linguistic distance measure. We rerun model 2 for the four subsamples and report the results in Table 10. The result indicates that regardless of the trust index level, site visits significantly improve analysts' forecast accuracy for firms in areas with long linguistic distances but not for firms in areas with short linguistic distances. Meanwhile, the result does not support the expectation that corporate site visits resolve the trustworthiness problem. Therefore, linguistic distance represents a unique dimension of information asymmetry that can be alleviated by corporate site visits.

5. Limitations

5.1. Linguistic distance as a proxy for the language barrier between analysts and corporate managers

Our results indicate that corporate site visits alleviate communication barriers between financial analysts and corporate managers by showing that financial analysts can provide more accurate earnings forecasts after visiting firms in locations with a dominant dialect that is very different from Standard Mandarin. Although our results survive with additional checks, we admit two major limitations of our analysis.

[15] Because of data availability, we use the 2012, 2013, 2015 and 2017 versions for our test. For years without available data, we carry forward the values from the most recent available year.

Table 10
Forecast Accuracy: Social Trust. This table presents regression results for the effect of analysts' corporate site visit on analysts' forecast accuracy for the regions with different social trust levels. The dependent variable is forecast accuracy (*Accuracy*). The key independent variable is a dummy that equals 1 if the brokerage visits the firm and 0 otherwise (*DVisit*) or the natural logarithm of 1 plus the number of visits made by the brokerage to the firm (*Ln(NVisit)*). The full sample is separated into four groups: short linguistic distance and low social trust, short linguistic distance and high social trust, long linguistic distance and low social trust, and long linguistic distance and high social trust. The low social trust is the regions where the yearly social trust score is below the first quartile, and the high social trust is the regions where the yearly social trust score is above the third quartile. The OLS model is adopted in each column. Firm-, brokerage firm-, and year-fixed effects are included. All continuous variables are winsorized at the 1st and 99th percentiles. Heteroscedasticity robust standard errors, clustered by brokerage firm, are in parentheses. $p < 0.1$, $p < 0.05$, and $p < 0.01$ levels of significance are represented as * , * *, and * ** , respectively. The definitions of all variables are available in Appendix 1.

	Short Linguistic Distance				*Long Linguistic Distance*			
	Low Social Trust		*High Social Trust*		*Low Social Trust*		*High Social Trust*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DVisit	-0.006		-0.002		0.066 * **		0.057 * *	
	(0.044)		(0.019)		(0.020)		(0.026)	
Ln(NVisit)		-0.013		0.003		0.063 * **		0.058 * *
		(0.049)		(0.018)		(0.023)		(0.025)
Observations	4137	4137	7223	7223	7491	7491	2464	2464
R-squared	0.505	0.505	0.482	0.482	0.443	0.443	0.411	0.411
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Brokerage FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 11
Forecast Accuracy: Top Management Age. This table presents ordinary-least-squares (OLS) results for the effect of corporate site visit on analyst forecast accuracy, controlling for the average age of top management team within firms in short linguistic distance areas (*LongLDist* = 0) and long linguistic distance areas (*LongLDist* = 1). The dependent variable is forecast accuracy (*Accuracy*). The key independent variable is a dummy that equals 1 if the brokerage visits the firm and 0 otherwise (*DVisit*) or the natural logarithm of 1 plus the number of visits made by the brokerage to the firm (*Ln(NVisit)*). *AgeQuartile* is 1, 2, 3 or 4 if the average age of the top management team is at the first (youngest), second, third or fourth quartile (oldest) in the subsample. Firm-, brokerage firm-, and year-fixed effects are included. All continuous variables are winsorized at the 1st and 99th percentiles. Heteroscedasticity robust standard errors, clustered by brokerage firm, are reported in parentheses. $p < 0.1$, $p < 0.05$, and $p < 0.01$ levels of significance are represented as * , * *, and * ** , respectively. The definitions of all variables are available in Appendix 1.

	Short Linguistic Distance (*LongLDist* = 0)		Long Linguistic Distance (*LongLDist* = 1)	
	(1)	(2)	(3)	(4)
DVisit	-0.007		-0.013	
	(0.029)		(0.026)	
Ln(NVisit)		0.001		-0.021
		(0.028)		(0.026)
AgeQuartile	-0.080 * **	-0.078 * **	-0.023 * *	-0.022 *
	(0.022)	(0.022)	(0.012)	(0.011)
AgeQuartile × DVisit	0.009		0.025 * **	
	(0.012)		(0.008)	
AgeQuartile × Ln(NVisit)		0.005		0.024 * **
		(0.011)		(0.008)
Observations	15,446	15,446	32,520	32,520
R-squared	0.461	0.461	0.415	0.415
Firm FE	Yes	Yes	Yes	Yes
Brokerage firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

First, we cannot offer direct measures of dialects used by corporate managers and financial analysts and we also do not have information about their Mandarin proficiency. Second, even if a firm's managers are not familiar with Standard Mandarin, the brokerage firm can choose an analyst that grew up in the firm's region to conduct a site visit.[16]

To address the second concern, we argue that if some brokerage firms really choose analysts that grew up in the firm's region to conduct a site visit, then there will be little need to conduct a site visit and our finding should be weakened as a result. Therefore, this possibility should bias against us in finding a significant result. Indeed, Zhang (2022) infers analysts' native places based on the high schools they studied. The summary statistics show that only 6% of analysts share the same dialect as CEOs. This suggests few brokerage firms really choose analysts to follow firms in the analysts' hometowns.[17]

For the first concern, Table 2 shows that most Chinese spoke in their own dialects in childhood. This leads to our conjecture that communication barriers exist between financial analysts and corporate managers for two reasons. First, half of corporate managers are 50 years old or older. As they were born in the 1960 s, many of them were unable to receive formal education until the economic reforms in 1978. On the other hand, financial analysts are much younger and they were mostly born in the 1970–80 s and they received much better education than the older generation. Second, although the Chinese government has been promoting Standard Mandarin in government administration and education since the 1950 s, the use of local dialects is still very common in the workplace. Therefore, if a person did not receive proper education in Mandarin in childhood, it is difficult for him/her to improve their Mandarin standards in the workplace.

It could also be argued that listed firms can attract talents from other areas to be their top executives. However, several studies show that listed firms in China mostly hire talents from their local pools (Giannetti et al., 2015; Tan et al., 2021; Huang et al., 2022). In particular, Giannetti et al. (2015) find that over 80% of directors of listed firms reside in the same provinces as the firm headquarters. To provide further evidence of listed firms' reliance on local talents, we identify top managers' native places from the CSMAR. However, only few managers' native places are reported in CSMAR. Nevertheless, within the small sample of managers with available data, we show that 57.6% of them are local. Therefore, top managers of listed firms are likely to speak and be proficient in local dialects.

To confirm our expectation that older managers are less proficient in Standard Mandarin. We classify firms into quartiles according to the average age of top managers. We then interact the quartile variable with the site visit variable, rerun the baseline model, and report the result in Table 11. The result shows that corporate site visits provide greater

[16] We thank the referee for providing this insightful comment for our study.

[17] We have tried to identify analysts' background information from the data source mentioned in Zhang (2022). Unfortunately, the data is no longer available for unknown reasons.

Table 12
Forecast Accuracy: Physical Distance. This table presents regression results of Table 4 and Table 5 controlling for the physical distance between the brokerage firm and the listed firm. Panel A shows the results with the model in Table 4 and panel B reports the results with the model in Table 5. *Physical Distance* is the great-circle distance (100 kilometers) between the brokerage firm and the listed firm. Heteroscedasticity robust standard errors, clustered by brokerage firm, are reported in parentheses. $p < 0.1$, $p < 0.05$, and $p < 0.01$ levels of significance are represented as * , * *, and * ** , respectively. The definitions of all variables are available in Appendix 1.

Panel A Physical Distance and Site Visits

	DVisit		*NVisit*	
	(1)	(2)	(3)	(4)
LongLDist	0.196 * **		0.159 * **	
	(0.062)		(0.048)	
LDist		0.041 * **		0.031 * *
		(0.016)		(0.012)
Local	0.274 * **	0.278 * **	0.271 * **	0.270 * **
	(0.070)	(0.070)	(0.049)	(0.049)
Physical Distance	-0.011 * **	-0.010 * **	-0.008 * **	-0.008 * **
	(0.003)	(0.003)	(0.002)	(0.002)
Observations	46,864	46,864	48,120	48,120
R-squared	0.093	0.092	0.088	0.088
Control	Yes	Yes	Yes	Yes
Brokerage FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Panel B Physical Distance and Forecast Accuracy

	Short Linguistic Distance (*LongLDist* = 0)		Long Linguistic Distance (*LongLDist* = 1)	
	(1)	(2)	(3)	(4)
DVisit	0.022		0.041 * *	
	(0.027)		(0.019)	
Ln(*NVisit*)		0.009		0.037 * *
		(0.026)		(0.018)
Physical Distance	0.003	0.003	0.000	0.000
	(0.004)	(0.004)	(0.001)	(0.001)
Physical Distance × *DVisit*	-0.001		0.001	
	(0.002)		(0.002)	
Physical Distance × *Ln*(*NVisit*)		0.000		0.002
		(0.002)		(0.002)
Observations	15,446	15,446	32,520	32,520
R-squared	0.460	0.460	0.415	0.415
Firm FE	Yes	Yes	Yes	Yes
Brokerage firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

benefits for analysts when visited firms have older managers and are headquartered in areas with a long linguistic distance. On the other hand, the interaction term is insignificant for firms headquartered in areas with a short linguistic distance. The result is consistent with our argument that older managers are relatively less proficient in Standard Mandarin and a face-to-face meeting with them can help analysts to obtain a clearer picture of the firm's conditions.

5.2. *Linguistic distance as a noisy proxy for physical distance*

It is also possible that the linguistic distance is a noisy proxy for the physical distance between the brokerage firm and the listed firm. If two geographically close areas are more likely to share similar dialects than two distant areas, the results we obtain above could simply manifest information asymmetry due to transportation time and difficulties. To address this concern, we include the physical distance between the brokerage firm and the listed firm (*Physical Distance*) in the model for the site visit decision and the model for analysts' forecast accuracy. To measure *Physical Distance*, we collect the coordinates of the brokerage firm and the listed firm from the Baidu Map and calculate the great-circle distance (in 100 kilometers) between the two entities.

We report the results in Table 12. Panel A shows a negative coefficient of *Physical Distance* is negative and statistically significant at the 1% level. Therefore, analysts are less likely to visit listed firms that are far from the brokerage firm. The result is also consistent with the positive coefficient of *Local* in Table 4. Therefore, a longer physical distance is more likely to be associated with higher traffic costs (Cheng et al., 2016) than with information asymmetry that motivates a financial analyst to collect more information.

In Panel B, *Physical Distance* and its interaction with *DVisit* (or *Ln* (*NVisit*)) are added to the model for analysts' forecast accuracy. If the linguistic distance is a rough proxy for the physical distance, then the difference in coefficients of *DVisit* (or *Ln*(*NVisit*)) between two subsamples (i.e. *LongLDist* = 0 and *LongLDist* = 1) should disappear in the new model. Instead, the coefficient of the interaction term should be positive and statistically significant. The result, however, shows that coefficients of *Physical Distance* and its interaction term are statistically insignificant in all regressions. Meanwhile, the coefficient of *DVisit* (or *Ln*(*NVisit*)) is still positive and statistically significant when the linguistic distance is long but statistically insignificant when the linguistic distance is short. Therefore, our main findings are unlikely to be driven by the physical distance between the brokerage firm and the listed firm.

6. Conclusions

In this paper, we empirically investigate whether language barriers between analysts and corporate stakeholders create the need for conducting site visits and whether site visits enhance analysts' forecast accuracy by overcoming language barriers. We use the linguistic distance between Standard Mandarin, China's official spoken language, and the subject firm location's dominant dialect to proxy for language barriers. Empirical results show that linguistic distance is an important determinant for an analyst's site visit decision. In addition, a corporate site visit enhances an analyst's forecast accuracy mainly when the subject firm location's dominant dialect is largely different from Standard Mandarin. The latter result is robust to endogeneity tests and further analyses with alternative definitions of language barriers.

Overall, our results show that language barriers exist among dialects in the same language and their economic impacts are significant. Language barriers significantly affect financial markets by creating an additional layer of information asymmetry between information providers, i.e. listed firms, and information users, i.e. financial analysts, investors, and others. Nevertheless, the enactment of Standard Mandarin as the official spoken language has significantly reduced language barriers among people in China.

Our study provides another angle to understand the costs and benefits of alternative channels of communication between listed firms and market participants. While online meetings or remote communication with firm executives or stakeholders may help analysts verify existing information and collecting information, distant communication could be noisy when people in conversations are not equally proficient in the communication language. Conducting site visits is a solution for such situations. Our study is helpful for brokerage firms to make optimal site visit decisions and use travel budgets effectively.

Our study also provides implications for fair disclosure regulations. Fair disclosure regulations aim to promote fairness among market participants by restricting access to material private information. However, information is subject to interpretation, and barriers may exist to deter market participants from understanding information for further interpretation. The fact that U.S. corporate top executives spend much time talking with sell-side analysts (Soltes, 2014) suggests clarification of information is important to provide additional insights for investors. The SZSE requires immediate disclosure of site visit information. Meanwhile, the U.S. Securities and Exchange Commission imposes the legal burden of complying with fair disclosure on firm management without requiring a disclosure of meeting information. Our results suggest that site visits could provide insights even if only public information is involved and that the disclosure of meeting information is beneficial to financial market participants.

Declaration of Competing Interest

We declare that to our best knowledge, there are no competing financial interests or personal relationships that could have appeared to influence the objectivity and reliability of work reported in this paper.

Acknowledgements

Tian acknowledges research grant support from the Macau University of Science and Technology (FRG-22-047-MSB).

Appendix 1. Definitions of Variables

Variables	Definitions
Accuracy	Following past studies (e.g.,Duru & Reeb, 2002;Walther & Willis, 2013;Han et al., 2018), forecast accuracy is measured as: $Accuracy_{ijt} = -100 \times \frac{\lvert Forecast\ EPS_{ijt} - Actual\ EPS_{ijt}\rvert}{Price_{jt}}$ where *Forecast* EPS_{ijt} is the last earnings per share forecast issued by brokerage firm i for firm j before the date of firm j's earnings announcement in year t. *Actual* EPS_{ijt} is the actual earnings per share reported by firm j. $Price_{jt}$ is the stock price of company j at the beginning of year t.
LDist	The linguistic distance between the dialect for the county-level administrative district where the firm is located and Standard Mandarin.
LongLDist	An indicator that equals 1 if the linguistic distance is 3 or above.
Dvisit	An indicator that equals 1 if the brokerage firm visits the firm this year and 0 otherwise.
Ln(NVisit)	The natural logarithm of 1 plus the number of visits made by the brokerage firm to the firm.
Standard Mandarin-Childhood	The percentage of the population in the provincial investigation sample speaking Standard Mandarin during childhood, scaled by sample size.
Standard Mandarin-Talk	The percentage of the population in the provincial investigation sample speaking Standard Mandarin in daily conversations, scaled by sample size.
Standard Mandarin-Work	The percentage of the population in the provincial investigation sample speaking Standard Mandarin at work, scaled by sample size.
Dialect-Childhood	The percentage of the population in the provincial investigation sample speaking dialects during childhood, scaled by sample size.
Dialect-Talk	The percentage of the population in the provincial investigation sample speaking dialects in daily conversations, scaled by sample size.
Dialect-Work	The percentage of the population in the provincial investigation sample speaking dialects at work, scaled by sample size.
BrokerSize	The natural logarithm of the number of analysts in the brokerage firm.
BrokerGexp	The natural logarithm of 1 plus the current year minus the initial year that the brokerage firm began to provide earnings forecasts for any listed firms.
BrokerFexp	The natural logarithm of 1 plus the current year minus the initial year that the brokerage firm began to provide earnings forecasts for the firm.
BrokerCover	The natural logarithm of 1 plus the number of firms covered by the brokerage firm in each year.
BrokerFreq	The natural logarithm of 1 plus the number of forecasts that the brokerage firm provides to the firm in each year.
Horizon	The natural logarithm of 1 plus the days elapsed from the brokerage firm's latest earnings forecast date to the earnings announcement date.
Follow	The natural logarithm of one plus the numbers of brokerage firms that provide forecasts for the firm.
Inst	Total institutional shares scaled by total shares for the firm.
SDNI	The standard deviation of net income for the last five years.
Size	The natural logarithm of firm market value.
Return	Yearly stock return for the firm.
Loss	An indicator that equals 1 if the firm has negative profit for the year and 0 otherwise.
Leverage	Total liabilities scaled by total assets of the firm.
HSR	An indicator that equals 1 if there is at least one high-speed-railway (HSR) station in the city where the firm headquarters in this year and 0 otherwise.
Physical Distance	The great-circle distance (100 kilometers) between the brokerage firm and the listed firm.
AgeQuartile	1, 2, 3 or 4 if the average age of the top management team is at the first (youngest), second, third, or fourth (oldest) quartile in the subsample.

References

Akstinaite, V., Robinson, G., & Sadler-Smith, E. (2020). Linguistic markers of CEO hubris. *Journal of Business Ethics, 167*(4), 687–705.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth, 8*(2), 155–194.

Alesina, A., & Ferrara, E. L. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature, 43*(3), 762–800.

Bae, K. H., Stulz, R. M., & Tan, H. (2008). Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts. *Journal of Financial Economics, 88*(3), 581–606.

Baik, B., Cho, K., Choi, W., & Kang, J. K. (2015). The role of institutional environments in cross-border mergers: A perspective from bidders' earnings management behavior. *Management International Review, 55*(5), 615–646.

Battiston, D., Blanes i Vidal, J., & Kirchmaier, T. (2021). Face-to-face communication in organizations. *The Review of Economic Studies, 88*(2), 574–609.

Berger, A. N., De Young, R., & Udell, G. F. (2001). Efficiency barriers to the consolidation of the European financial services industry. *European Financial Management, 7*(1), 117–130.

Bhat, G., Hope, O. K., & Kang, T. (2006). Does corporate governance transparency affect the accuracy of analyst forecasts? *Accounting & Finance, 46*(5), 715–732.

Bian, W., Ji, Y., & Zhang, H. (2019). Does dialect similarity add value to banks? Evidence from China. *Journal of Banking & Finance, 101*, 226–241.

Block, E. S., & Erskine, L. (2012). Interviewing by telephone: Specific considerations, opportunities, and challenges. *International Journal of Qualitative Methods, 11*, 428–445.

Bowen, R. M., Davis, A. K., & Matsumoto, D. A. (2002). Do conference calls affect analysts' forecasts? *The Accounting Review, 77*, 285–316.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin, 86*(2), 307.

Brown, L. D., Call, A. C., Clement, M. B., & Sharp, N. Y. (2015). Inside the "black box" of sell-side financial analysts. *Journal of Accounting Research, 53*(1), 1–47.

Bushee, B. J., Jung, M. J., & Miller, G. S. (2011). Conference presentations and the disclosure milieu. *Journal of Accounting Research, 49*(5), 1163–1192.

Bushee, B. J., Jung, M. J., & Miller, G. S. (2017). Do investors benefit from selective access to management? *Journal of Financial Reporting, 2*(1), 31–61.

Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review, 103*(2), 690–731.

Chen, Z., Lu, M., & Xu, L. (2014). Returns to dialect: Identity exposure through language in the Chinese labor market. *China Economic Review, 30*, 27–43.

Chen, D., Ma, Y., Martin, X., & Michaely, R. (2022). On the fast track: Information acquisition costs and information production. *Journal of Financial Economics, 143*(2), 794–823.

Chen, S., Miao, B., & Shevlin, T. (2015). A new measure of disclosure quality: The level of disaggregation of accounting data in annual reports. *Journal of Accounting Research, 53*(5), 1017–1054.

Cheng, Q., Du, F., Wang, X., & Wang, Y. (2016). Seeing is believing: Analysts' corporate site visits. *Review of Accounting Studies, 21*(4), 1245–1286.

Cheng, Q., Du, F., Wang, B. Y., & Wang, X. (2019). Do corporate site visits impact stock prices? *Contemporary Accounting Review, 36*(1), 359–388.

Cho, H., Muslu, V., & Koo, M. (2020). Lost in translation? Analysts' forecasts of cross-listed firms. *Journal of International Accounting Research, 19*(2), 41–64.

Choy, S., & Hope, O.K. (2022). Private communication between managers and financial analysts: Evidence from taxi ride patterns in New York City. Working paper.

Cuypers, I. R., Ertug, G., & Hennart, J. F. (2015). The effects of linguistic distance and lingua franca proficiency on the stake taken by acquirers in cross-border acquisitions. *Journal of International Business Studies, 46*(4), 429–442.

Dhaliwal, D. S., Radhakrishnan, S., Tsang, A., & Yang, Y. G. (2012). Nonfinancial disclosure and analyst forecast accuracy: International evidence on corporate social responsibility disclosure. *The Accounting Review, 87*(3), 723–759.

Dohen, M., Schwartz, J. L., & Bailly, G. (2010). Speech and face-to-face communication–An introduction. *Speech Communication, 52*(6), 477–480.
Dong, W., Han, H., Ke, Y., & Chan, K. C. (2018). Social trust and corporate misconduct: Evidence from China. *Journal of Business Ethics, 151*(2), 539–562.
Du, X. (2019). Does CEO-auditor dialect sharing impair pre-IPO audit quality? Evidence from China. *Journal of Business Ethics, 156*(3), 699–735.
Duru, A., & Reeb, D. M. (2002). International diversification and analysts' forecast accuracy and bias. *The Accounting Review, 77*(2), 415–433.
Easterly, W., & Levine, R. (1997). Africa's growth tragedy: policies and ethnic divisions. *The Quarterly Journal of Economics, 112*(4), 1203–1250.
Falck, O., Heblich, S., Lameli, A., & Südekum, J. (2012). Dialects, cultural identity, and economic exchange. *Journal of Urban Economics, 72*(2–3), 225–239.
Farooq, M. B., & de Villiers, C. (2017). Telephonic qualitative research interviews: when to consider them and how to do them. *Meditari Accountancy Research, 25*, 291–316.
Fenig, S., Levav, I., Kohn, R., & Yelin, N. (1993). Telephone vs face-to-face interviewing in a community psychiatric survey. *American Journal of Public Health, 83*, 896–898.
Fishman, J. A. (1991). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages* (Vol. 76). Multilingual Matters.
Fowler, G. D., & Wackerbarth, M. E. (1980). Audio teleconferencing versus face-to-face conferencing: A synthesis of the literature. *The Western Journal of Speech Communication, 44*, 236–252.
Frijns, B., & Garel, A. (2021). The effect of cultural distance between an analyst and a CEO on analysts' earnings forecast performance. *Economics Letters, 205*, Article 109957.
Fu, Y., Liu, C., & Qin, Z. (2021). Does CEO-chairman dialect similarity affect stock price informativeness for Chinese listed firms? *The North American Journal of Economics and Finance, 55*, Article 101313.
Fukuyama, F. (1997). Social capital and the modern capitalist economy: Creating a high trust workplace. *Stern Business Magazine, 4*(1), 1–16.
Gao, Z., Quan, X., & Xu, X. (2022). Under watchful eyes: Analyst site visits and firm earnings management. *International Review of Financial Analysis, 83*, Article 102269.
Gao, X., & Ren, W. (2019). Controversies of bilingual education in China. *International Journal of Bilingual Education and Bilingualism, 22*(3), 267–273.
Giannetti, M., Liao, G., & Yu, X. (2015). The brain gain of corporate boards: Evidence from China. *Journal of Finance, 70*(4), 1629–1682.
Green, T. C., Jame, R., Markov, S., & Subasi, M. (2014a). Access to management and the informativeness of analyst research. *Journal of Financial Economics, 114*(2), 239–255.
Green, T. C., Jame, R., Markov, S., & Subasi, M. (2014b). Broker-hosted investor conferences. *Journal of Accounting and Economics, 58*(1), 142–166.
Greene, W. H. (2008). *Econometric analysis*. Prentice Hall.
Grinblatt, M., & Keloharju, M. (2001). How distance, language, and culture influence stockholdings and trades. *The Journal of Finance, 56*(3), 1053–1073.
Guiso, L., Sapienza, P., & Zingales, L. (2004). The role of social capital in financial development. *American Economic Review, 94*(3), 526–556.
Han, B., Kong, D., & Liu, S. (2018). Do analysts gain an informational advantage by visiting listed companies? *Contemporary Accounting Research, 35*(4), 1843–1867.
Hogg, M. A., & Turner, J. C. (1987). Intergroup behaviour, self-stereotyping and the salience of social categories. *British Journal of Social Psychology, 26*(4), 325–340.
Holtzman, N. S., Kwong, S., & Baird, K. L. (2015). Exploring political ideologies of senators with semantic analysis tools: Further validation of CASS. *Journal of Language and Social Psychology, 34*(2), 200–212.
Hope, O. K. (2003). Disclosure practices, enforcement of accounting standards, and analysts' forecast accuracy: An international study. *Journal of Accounting Research, 41*(2), 235–272.
Horton, J., Serafeim, G., & Serafeim, I. (2013). Does mandatory IFRS adoption improve the information environment? *Contemporary Accounting Research, 30*(1), 388–423.
Hu, G., Xiao, J. Z., & Zhen, Y. (2021). The effect of dialect sharing on corporate cash holdings in China. *ABACUS*.
Hu, S., Liu, M., Sohn, B. C., & Yuen, D. C. (2021). The impact of disclosure quality on analyst forecasts in China. *Accounting Forum, 45*(4), 411–434.
Huang, W., & Kim, J. (2020). Linguistically induced time perception and asymmetric cost behavior. *Management International Review, 60*(5), 755–785.
Huang, B., Yang, E., & Zhang, Y. (2022). Board surname sharing and investment efficiency: Evidence from Chinese state-owned enterprises. *Corporate Governance: An International Review*.
Jha, A. (2019). Financial reports and social capital. *Journal of Business Ethics, 155*(2), 567–596.
Jiang, X., & Yuan, Q. (2018). Institutional investors' corporate site visits and corporate innovation. *Journal of Corporate Finance, 48*, 148–168.
Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience, 32*(45), 16064–16069.
Jiang, Y., Wang, C., Li, S., & Wan, J. (2022). Do institutional investors' corporate site visits improve ESG performance? Evidence from China. *Pacific-Basin Finance Journal, 76*, Article 101884.
Kang, J. K., & Kim, J. M. (2010). Do foreign investors exhibit a corporate governance disadvantage? An information asymmetry perspective. *Journal of International Business Studies, 41*(8), 1415–1438.
Kimbourgh, M. D. (2005). The effect of conference calls on analyst and market underreaction to earnings announcements. *The Accounting Review, 80*, 189–219.
Kirk, M. P., & Markov, S. (2016). Come on over: Analyst/investor days as a disclosure medium. *The Accounting Review, 91*(6), 1725–1750.
Klein, A., Li, T., & Zhang, B. (2020). Seeking out non-public information: Sell-side analysts and the Freedom of Information Act. *The Accounting Review, 95*(1), 233–257.
Kleiner, B., Lipps, O., & Ferrez, E. (2015). Language ability and motivation among foreigners in survey responding. *Journal of Survey Statistics and Methodology, 3*, 339–360.
Kong, D., Liu, L., & Liu, S. (2020). Market information traveling on high-speed rails: The case of analyst forecasts. *Pacific-Basin Finance Journal, 61*, Article 101320.
Lameli, A., Südekum, J., Nitsch, V., & Wolf, N. (2015). Same same but different: Dialects and trade. *German Economic Review, 16*(3), 290–306.
Lee, T., & Perez, E. O. (2014). The persistent connection between language-of-interview and Latino political opinion. *Political Behavior, 36*(2), 401–425.
Li, D. C. S. (2006). Chinese as a lingua franca in Greater China. *Annual Review of Applied Linguistics, 26*, 149–176.
Li, L., Duan, Y., He, Y., & Chan, K. C. (2018). Linguistic distance and mergers and acquisitions: Evidence from China. *Pacific-Basin Finance Journal, 49*, 81–102.
Lundholm, R., Rahman, N., & Rogo, R. (2018). The foreign investor bias and its linguistic origins. *Management Science, 64*(9), 4433–4450.
Muslu, V., Mutlu, S., Radhakrishnan, S., & Tsang, A. (2019). Corporate social responsibility report narratives and analyst forecast accuracy. *Journal of Business Ethics, 154*(4), 1119–1142.
Nakagawa, M., & Sugasawa, S. (2020). Linguistic Distance and Economic Prosperity: A Cross-Country Analysis. *Working Paper*.
Novick, G. (2008). Is there a bias against telephone interviews in qualitative research? *Research in Nursing & Health, 31*, 391–398.
Paul, J., & Jefferson, F. (2019). A comparative analysis of student performance in an online vs. face-to-face environmental science course from 2009-2016. *Frontiers in Computer Science, 1*, 7.
Payne, G. T., Moore, C. B., Griffis, S. E., & Autry, C. W. (2011). Multilevel challenges and opportunities in social capital research. *Journal of Management, 37*(2), 491–520.
Pennebaker, J. W., & Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science, 10*(3), 90–93.
Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*(1), 547–577.
Portes, A. (1998). Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology, 24*, 1–24.
Preston, P. (2005). Nonverbal communication: Do you really say what you mean? *Journal of Healthcare Management, 50*(2), 83–86.
Reiman, T. (2007). *The power of body language: How to succeed in every business and social encounter*. Simon and Schuster.
Rupasingha, A., Goetz, S. J., & Freshwater, D. (2006). The production of social capital in US counties. *Journal of Socio-Economics, 35*(1), 83–101.
Soltes, E. (2014). Private interaction between firm management and sell-side analysts. *Journal of Accounting Research, 52*(1), 245–272.
Stock, J., & Watson, M. W. (2003). *Introduction to econometrics*. New York: Prentice Hall.
Sturges, J. E., & Hanrahan, K. J. (2004). Comparing telephone and face-to-face qualitative interviewing: a research note. *Qualitative Research, 4*(1), 107–118.
Tan, Y., Xiao, J., Zeng, C. C., & Zou, H. (2021). What's in a name? The valuation effect of directors' sharing of surnames. *Journal of Banking & Finance, 122*, Article 105991.
Tang, C., & van Heuven, V. J. (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua, 119*(5), 709–732.
Tenzer, H., Pudelko, M., & Harzing, A. W. (2014). The impact of language barriers on trust formation in multinational teams. *Journal of International Business Studies, 45*(5), 508–535.
Viruell-Fuentes, E. A., Morenoff, J. D., Williams, D. R., & House, J. S. (2011). Language of interview, self-rated health, and the other Latino Health Puzzle. *American Journal of Public Health, 101*(7), 1306–1313.
Vogl, S. (2013). Telephone versus face-to-face interviews: Mode effect on semistructured interviews with children. *Sociological Methodology, 43*, 133–177.
Walther, B. R., & Willis, R. H. (2013). Do investor expectations affect sell-side analysts' forecast bias and forecast accuracy? *Review of Accounting Studies, 18*(1), 207–227.
Ward, K., Gott, M., & Hoare, K. (2015). Participants' views of telephone interviews within a grounded theory study. *Journal of Advanced Nursing, 71*(12), 2775–2785.
Wenz, A., Al Baghal, T., & Gaia, A. (2021). Language proficiency among respondents: Implications for data quality in a longitudinal face-to-face survey. *Journal of Survey Statistics and Methodology, 9*(1), 73–93.
Whorf, B. L. (1956). In John (Ed.), *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge: MIT Press.
Wierzbicka, A. (2006). *English: Meaning and culture*. Oxford: Oxford University Press.
Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, 104*(19), 7780–7785.
Woolcock, M. (2001). The place of social capital in understanding social and economic outcomes. *Canadian Journal of Policy Research, 2*(1), 11–17.
Wu, W., Zhang, X., & Zhou, Z. (2022). Institutional investors' corporate site visits and pay-performance sensitivity. *Pacific-Basin Finance Journal, 76*, Article 101875.
Yang, J., Lu, J., & Xiang, C. (2020). Do disclosures of selective access improve market information acquisition fairness? Evidence from company visits in China. *Journal of Corporate Finance, 64*, Article 101631.
Yin, Y., Niu, Z., Liu, Y., & Mou, S. (2018). Firm site visits and differential information of text in analyst reports. *China Journal of Accounting Studies, 6*(3), 231–274.
Zhang, R. (2022). Language commonality and sell-side information production. *Management Science, 68*(6), 4435–4453.
Zhou, T., & Gan, J. (2022). Institutional investors' site visits and corporate social responsibility: Evidence from China. *Economic Modelling, 115*, Article 105965.