Narrative R&D Disclosure and the R&D Anomaly

Abstract: Prior research documents that investors underreact to R&D expense because they have difficulty valuing innovation (Chan et al 2001; Eberhart et al. 2004; Cohen et al. 2013). This phenomenon is commonly referred to as the R&D anomaly. We extend prior research by examining how narrative R&D disclosure in 10-K filings impacts market participants' understanding of the value of innovation. We first show that narrative R&D disclosure is positively related with future R&D outcome. Despite such value-relevance, we find that R&D anomaly is magnified in intense narrative R&D disclosure firms, as reflected in larger future returns associated with current R&D expense. We further find that the impact of R&D disclosure on the R&D anomaly is more pronounced when the number of investors' 10-K views is low and when 10-K reports are less readable. Overall, our findings suggest that narrative R&D disclosure does not necessarily help investors' ability to impound information about R&D into stock prices on a timely basis. Our study has implications for regulators in that users of financial statements have difficulty processing on a timely basis the information contained not only in R&D, but also in R&D narrative disclosure.

JEL classification: G14; L22; M41
Keywords: R&D expense, Narrative disclosure, Textual analysis, EDGAR, Limited attention, Mispricing

1 Introduction

Given that corporations spend billions of dollars annually on research and development (R&D) activities, it is not surprising that R&D is the subject of numerous studies in the finance and accounting literatures. An anomalous finding from prior studies is that investors underreact to R&D expense and that this underreaction is more pronounced for firms with high levels of R&D (Chan et al. 2001; Eberhart et al. 2004; Cohen et al. 2013). In particular, current R&D expense is related to positive excess returns over the following year. This phenomenon is commonly referred to as the R&D anomaly. One explanation for the R&D anomaly is that equity investors underreact to R&D expense because, compared to the consequences of other firm activities (e.g. sales, inventory management), the extent to which R&D ultimately affects firm value is highly uncertain. Moreover, R&D is idiosyncratic, further hindering investors' ability to decipher its valuation implications.

Because disclosure has been proposed as a mechanism to alleviate information asymmetry (Healy and Palepu 2001), we propose that narrative R&D disclosure[1] might play a role in mitigating the R&D anomaly. Prior research suggests that narrative R&D disclosure conveys value-relevant information beyond that provided by income statement disclosure of R&D expense (Feldman et al. 2010; Merkley 2014). Specifically, Merkley (2014) reports short-window positive excess returns associated with narrative R&D disclosure in firms' 10-Ks. His evidence suggests that firms with

[1] We use the terms "narrative R&D disclosure" and "qualitative R&D disclosure" interchangeably throughout the paper.

high R&D expense might benefit from enhanced qualitative R&D disclosure in their 10-K reports because doing so may enhance investors' ability to impound information about R&D, thereby reducing the underreaction. Given that narrative disclosures about R&D bridge the gap between quantitative R&D disclosure (i.e. R&D expense) and investors' valuation of R&D, we expect less mispricing of R&D expense for firms with enhanced narrative R&D disclosure.

However, a plausible alternative story is that narrative R&D disclosure exacerbates the R&D anomaly. Cazier and Pfeiffer (2015) report that narrative disclosures are voluminous and complex compared to information released via earnings announcements and conference calls. Merkley (2014) shows that narrative R&D disclosures are less readable than other narrative disclosures in 10-K reports. To the extent that investors have limited attention and consequently place less weight on information that is difficult to process (Hirshleifer and Teoh 2003; You and Zhang 2009), it is possible that investors find it difficult to assess the narrative information about R&D, thereby exacerbating the underreaction to R&D expense. Moreover, Zhong (2018) finds innovative effort and output is positively related with disclosure transparency and Merkley (2014) notes narrative R&D disclosures has significantly more positive tone than other narrative disclosures. This suggests intense narrative R&D disclosures might foreshadow positive outcome of R&D, adding potential reason why investors' inability to comprehend narrative R&D disclosure might lead to *more* mispricing of R&D.

Finally, although Merkley (2014) reports that narrative R&D disclosure is positively associated with short-run returns, it is possible that the market response to narrative R&D disclosure is incomplete due to the complexity of the disclosure. To disentangle the two competing stories, we use a long-window setting to examine the relation between R&D narrative disclosure and the R&D anomaly.

Using a large sample of firms for the period of 1993-2016, we provide several new and important findings. First, similar to Zhong (2018), we show that intense narrative R&D disclosure combined with intense R&D expense is positively related to future patents, future citations, and innovation efficiency. These results are economically significant in that an increase in narrative R&D disclosure from the first quartile to the third quartile increases patents by about 21%. We then confirm the R&D anomaly reported by prior studies (Chan et al 2001; Eberhart et al. 2004; Cohen et al. 2013). More importantly, we report that investors fail to recognize information embedded in narrative R&D disclosure in 10-K reports and the R&D anomaly is magnified even after controlling for firm characteristics that may affect stock returns. We demonstrate that a long-short portfolio based on the high R&D disclosure and high R&D expense versus high R&D disclosure and low R&D expense generates an economically significant monthly alpha of 1.31% based on the Fama-MacBeth five-factor model while the comparison between low R&D disclosure and high R&D expense versus low R&D disclosure and low R&D expense yields 0.78% monthly alpha. Similarly, results from Fama-MacBeth regressions show that monthly excess returns increase by 0.40% and R&D anomaly intensifies about 0.25% more per 1% change in R&D intensity (i.e. R&D anomaly) when we move narrative R&D disclosure from the first quartile to the third quartile of its distribution. The results suggest that investors do not fully impound information embedded in narrative R&D disclosure into stock prices on a timely basis.

To gain further insight into the impact of narrative R&D disclosure on stock returns, we also examine cross-sectional variation in the relation between narrative R&D disclosure and the R&D anomaly. We find that the impact of R&D disclosure on the R&D anomaly is alleviated when the number of investors' 10-K views is high and when 10-K reports are more readable. Specifically, we partition the sample based on the number of 10-K views via EDGAR and the fog readability index of 10-K reports. We then re-run monthly Fama-MacBeth regressions for each subsample and find that the impact of narrative R&D disclosure on the R&D anomaly is alleviated when more investors access 10-K filings and when 10-K filings are more readable. Specifically, we find that for our subsample of low EDGAR views (high readability) R&D anomaly intensifies about 0.27% more per 1% increase in R&D intensity (i.e. R&D anomaly) when we move R&D disclosure from the first quartile to the third quartile, while we find statistically insignificant and minimal change in R&D anomaly under high EDGAR views (low readability) subsample.

We run a battery of robustness tests and find that our results are qualitatively similar. We continue to find similar inferences when we control for managerial ability

(Cohen et al. (2013), innovation efficiency (Hirshleifer et al. (2013)), R&D increases (Eberhart et al. (2004)), and intangible information (Daniel and Titman (2006)). We also conduct robustness checks using alternative measures of narrative R&D disclosure (i.e., different bag of words) and find similar results. We find that our inferences with regard to narrative R&D disclosure are robust to the use of DGTW returns (Daniel et al. 1997) and the Fama-French three-factor model (Fama and French 1993). Our results are robust to the inclusion of adjusted ROA, which prior research shows is negatively related to narrative R&D disclosure quantity (Merkley 2014), total number of words in the 10-K (You and Zhang 2009), and industry fixed effects.

We contribute to the literature in several ways. Our results extend research on the R&D expense anomaly (Chan et al. 2001; Cohen et al. 2013; Eberhart et al. 2004; Hirshleifer et al. 2013) by showing that qualitative R&D disclosure adds another dimension of investors' underreaction. We contribute to the literature by providing evidence that 10-K views via EDGAR and the fog index help explain investor mispricing of qualitative R&D disclosure. More broadly, our study adds to the debate about the role of supplementary disclosure in stock price anomalies (Easton and Zmijewski 1993; E. X. Li and Ramesh 2009; You and Zhang 2009). We also shed light on the debate surrounding the benefits and costs of R&D disclosure. Several studies suggest that narrative R&D disclosure conveys value relevant information beyond quantitative R&D expenditure information (Feldman et al. 2010;

Merkley 2014). Our empirical evidence suggests that disseminating narrative R&D information does not necessarily lead to fuller understanding of R&D investment, consistent with Aboody and Lev (2000) and Glaeser (2018) that managers provide more non-proprietary information when they maintain trade secrecy to withhold proprietary information.

The remainder of this paper proceeds as follows. Section 2 reviews prior relevant literature and develops our hypotheses. Section 3 explains data sources and sample selection. Section 4 presents descriptive statistics and test results. Section 5 provides results of robustness tests. Section 6 summarizes and concludes the study.

2 Literature Review and Hypotheses

2.1 Investor Response to R&D Investment

Prior studies document that R&D expense predicts positive abnormal stock returns, suggesting that equity investors underreact to R&D expense. Since Chan et al. (2001) presented empirical evidence on the relation between R&D expense and stock prices, numerous studies have replicated and analyzed abnormal returns to R&D expense (Chambers et al. 2002; Donelson and Resutek 2012; Eberhart et al. 2004; Gu 2016; Hou et al. 2015; Lev and Sougiannis 1996; D. Li 2011). The literature has provided two main explanations for the R&D anomaly: (i) mispricing and (ii) risk.

Extant research shows that limited attention and accounting conservatism are the mechanisms that drive mispricing. One stream of research argues that investors with limited attention fail to appreciate the long-term benefits of R&D because investors appear to be overly pessimistic about R&D expenditures. (Chan et al. 2001; Barber and Odean 2007; Hirshleifer and Teoh 2003) Another stream of research argues that the conservative nature of accounting misleads investors. Under conservative accounting standards, R&D expenditures are treated as periodic costs, consequently understating the current earnings and leading investors to misprice R&D expense. Both mispricing explanations regard future over-performance of R&D-intensive stocks as a correction of investors' underreaction at the time that R&D expense was reported.

The literature espousing a risk story for the R&D anomaly asserts that higher risk associated with R&D drives the over-performance of R&D-intensive firms. Leung et al. (2017) argue that technical risk, risk of obsolescence, uncertainty about expected cost to completion, and uncertainty surrounding the potential cash flow from R&D are related to investors' risk assessment (Leung et al. 2017). Advocates of risk-based explanations diminish the role of mispricing with evidence suggesting that the anomaly is attributable to non-R&D information correlated with R&D intensity (Al-Horani et al. 2003; Donelson and Resutek 2012), and with evidence of unbiased analysts' forecasts of R&D firms (Chambers et al. 2002; Donelson and Resutek 2012).

Other studies provide evidence suggesting that R&D intensity and innovation ability are two channels through which R&D impacts future performance. For example, Eberhart et al. (2004) find that a large increase in R&D predicts higher future returns. Hirshleifer et al. (2013) show that a firm-level measure of innovation efficiency, defined as future patents divided by lagged R&D capital, is positively related to future returns. Cohen et al. (2013) define innovation ability as the coefficient of sales growth regressed on R&D expense and show that high ability combined with intense R&D expense predicts future positive stock returns. These findings suggest that the channel through which R&D ultimately impacts firm value is uncertain and difficult for investors to assess, thus creating room for misvaluation.

2.2 Investor Response to 10-K filings

The investor attention framework introduced by Hirshleifer and Teoh (2003) asserts that investors place less weight on information that is difficult to process. Key elements of a 10-K report such as EPS, dividend, sales growth and other summary measures are often disclosed prior to 10-K filings through earnings announcements and conference calls, thus rendering 10-K filing as a formality (You and Zhang 2009). This may lead investors to pay less attention to the contents of the 10-K filing.

Empirical findings document investors' anemic reaction around 10-K and 10-Q filing dates (Easton and Zmijewski 1993; Foster III et al. 1983; Stice 1991). While investors' accessibility to 10-K filings has increased since the SEC adopted EDGAR, 10-K filings have become more voluminous and complicated (Cazier and Pfeiffer 2015). Furthermore, the number of investors actually accessing 10-K files is low even after EDGAR was established (Loughran and McDonald 2017).

Consistently, You and Zhang (2009) find evidence that investors tend to underreact to information contained in firms' 10-K reports. They calculate three-day abnormal stock returns around the 10-K filing date and show a significant drift following both positive and negative news over the next 12-months. They further find that the complexity of 10-K reports (measured in the number of words) is related to investor underreaction. E. X. Li and Ramesh (2009) present evidence that the market response around the 10-K filing date is restricted to 10-Ks filed near the time of earnings releases or calendar quarter-ends. You and Zhang (2011) also find evidence that investors' sluggish response to information contained in 10-Ks is more salient than information embedded in earnings announcements.

2.3 Hypotheses

Prior research documents that investors underreact to R&D expense, perhaps due to an inability to value innovation (Chan et al. 2001; Cohen et al. 2013; Eberhart et al. 2004). Prior research suggests that 10-K filings have become more voluminous and complicated. In particular, firms tend to increase narrative R&D disclosure in 10-K reports.

There are plausible competing arguments for the impact of R&D disclosure on the R&D anomaly. On the one hand, provided that narrative R&D disclosure is informative in assessing innovation, we should observe less mispricing of R&D expense for firms with more narrative R&D disclosure in an efficient market. Consistently, Merkely (2014) shows that narrative R&D disclosure in 10-K reports conveys potentially value-relevant information incremental to information embedded in R&D expense, increasing short-term analyst forecast accuracy (Merkley 2014). Merkley (2014) further finds that narrative R&D disclosure describes in-process R&D, patents, and forward-looking prospects in a relatively more positive tone than other narrative information in 10-K reports.[2]

On the other hand, R&D disclosure in 10-K filings is typically longer and more complex compared to information released via earnings announcements and conference calls prior to 10-K filings (Cazier and Pfeiffer 2015). Narrative R&D disclosure tends to be less readable and less informative than other narrative disclosures in 10-K reports (Merkley 2014) and on-going R&D projects at research (development) stages would be largely qualitative (quantitative), as shown by Jones (2007).[3] Furthermore, Glaeser (2018) indicates that managers have incentives to manipulate value-relevant information and provide more non-proprietary information in narrative disclosure. Glaeser (2018) suggests that managers maintain trade secrecy to withhold proprietary information. Thus, a plausible alternative outcome is that investors will have difficulty understanding the implications of

[2] The average tone of narrative R&D disclosure is more positive (mean = 0.21, std. dev. = 0.18) than that of narrative earnings disclosures (mean = 0.02, std. dev. = 0.12). This difference is economically significant in that the top 5% of earnings disclosure are equal to the mean of R&D disclosure in terms of tone positivity.

[3] While Merkley (2014) shows a positive relation between narrative R&D disclosure and forecast accuracy, note that forecast accuracy was measured using the most recent consensus forecast following the annual report filing. Missing from Merkley (2014) is a long-term assessment of analyst forecast characteristics and R&D narrative disclosure. Moreover, equity investors and analysts may respond differently to 10-K information. For example, Rajgopal et al. (2003) suggests that analysts have a deeper understanding of accounting information, which is in stark contrast with equity investors.

narrative R&D disclosure for firm value and in turn narrative R&D disclosure aggravates the R&D anomaly.[4] The preceding discussion suggests that whether market participants incorporate information from narrative R&D disclosure into stock prices is an empirical question. The foregoing discussion leads to our first hypothesis, stated in the null form:

***H1**: Narrative R&D* disclosure is not related to *the R&D anomaly*.

The framework of Hirshleifer and Teoh (2003) for investor mispricing predicts that the magnitude of mispricing varies with the level of inattention. That is, increased inattention (attention) exacerbates (attenuates) mispricing. A number of empirical studies support the prediction of the Hirshleifer and Teoh (2003) framework (Ali and Gurun 2009; Drake et al. 2014; Eichler 2012; Jin 2014; Madsen 2017; You and Zhang 2011). We utilize EDGAR pageviews as our measure of investor attention because investors mainly access narrative R&D disclosure in 10-K reports via EDGAR (Loughran and McDonald 2017). We then examine whether the extent of investor attention influences the mispricing of narrative R&D disclosure, if any. This discussion leads to the following hypothesis, stated in the alternative form:

***H2a**: The mispricing suggested in H1, if any, is negatively associated with 10-K page views via EDGAR.*

[4] In terms of readability (fog index), the top 3.7% of narrative earnings disclosures is equivalent to the mean of R&D disclosure, implying that narrative R&D disclosure is far less readable; this difference is statistically significant (t-value=1.69).

Prior theoretical models suggest that mispricing is amplified when investors have difficulty processing information (Bloomfield 2002; Grossman and Stiglitz 1980; Hirshleifer and Teoh 2003). The empirical literature on textual information provides evidence consistent with these theoretical models (Asay et al. 2016; Lawrence 2013; F. Li 2008; Miller 2010; Rennekamp 2012; Tan et al. 2014; You and Zhang 2009). Thus, to the extent investors find it difficult to process opaque 10-K reports, we anticipate that low readability increases the price drift associated with narrative R&D disclosure. The aforementioned discussion leads to the following hypothesis, stated in the alternative form:

H2b*: The mispricing suggested in H1, if any, is negatively associated with readability of 10-K reports.*

3 Data and Research Design

3.1 Data

Our sample begins with 230,853 firm-year observations from 1993 to 2016 for those firms with 10-K filings. We follow Merkley (2014) and obtain narrative R&D disclosure quantity from the 10-K by counting sentences that include R&D-related terminology.[5] After matching market and accounting variables from COMPUSTAT and CRSP, we delete observations with missing or zero R&D expense. To mitigate confounding effects from strategic non-disclosure of R&D expense (Koh and Reeb

[5] Merkley's (2014) dictionary of R&D-related words is available in the Appendix to his paper (http://dx.doi.org/10.2308/accr-50649.s1).

2015; Koh et al. 2017), we restrict our sample to observations with reported positive R&D expense.[6] We utilize Kogan et al.'s (2017) shared database of patents and citations. [7] We also use a shared dataset from the Notre Dame Software Repository for Accounting and Finance (SRAF) for 10-K attributes.[8,9] We obtain the Fog index for each 10-K filing from Professor Li's website.[10] Our final sample consists of 36,205 firm-year observations. Due to limited data for patent and 10-K attributes and requirements to compute *Ability* measure (Cohen et al. 2013), we employ reduced samples for R&D outcomes and sensitivity tests (27,645 and 14,347 observations (1993-2009) respectively). Similarly, for the analyses using 10-K page views and the fog index, we use reduced samples of firm-year observations of 21,342 (2002-2015) and 23,689 (1993-2011), respectively.

3.2 Research Design

To investigate whether narrative R&D disclosure contains value-relevant information, we first examine the relationship between narrative R&D disclosure and future R&D outcomes. We estimate annual fixed effect panel regressions of R&D outcomes on narrative R&D disclosure as follows:

[6] Such treatment is widely accepted in the literature on R&D anomaly (Leung et al. 2017) and narrative R&D disclosure (Merkley 2014).
[7] https://iu.app.box.com/v/patents
[8] https://sraf.nd.edu/data/
[9] EDGAR log data excludes robot downloads, leaving only human access.
[10] https://webuser.bus.umich.edu/feng

$$\log(R\&D\ outcome_{i,t+1})$$
$$= \beta_0 + \beta_1 \log(RD\_ME_{i,t}) + \beta_2 \log(R\&D\ DIS\ \ _{i,t}) + \beta_3 \log(RD\_ME_{i,t})$$
$$\times \log(R\&D\ DISC_{i,t}) + \beta_4 \log(ME_{i,t}) + \beta_5 \log(BTM_{i,t}) + \beta_6 \log(1 + Lev_{i,t})$$
$$+ \beta_7 \log(1 + Age_{i,t}) + \beta_8 Instown + \beta_9 Finance$$
$$+ \beta_{10} \log(Employment) + \beta_{11} \frac{K}{L} + \beta_{12} Foreignsale\%$$
$$+ Industry\ Fixed\ Effect + Year\ Fixed\ Effect + \varepsilon_{i,t} \qquad \mathbf{(1)}$$

Following prior research (Connolly and Hirschey 1988; Deng et al. 1999; Lerner 1994; Griliches 1998; Hall et al. 2005; Hirshleifer et al. 2013; Zhong 2018), we measure our R&D outcome with future patent, future citation, and innovation efficiency. Future patent (citation) is defined as log of one plus patents (citations) issued at year t+1(Cohen et al. 2013; Zhong 2018).[11] Following Hirshleifer et al. (2013) and Zhong (2018), innovation efficiency is defined as $Patent_{t+1}/RDC_t$, where RDC_t is defined as R&D capital from Chan et al. (2001), calculated as $R\&D_t$+0.8* $R\&D_{t-1}$+0.6* $R\&D_{t-2}$+0.4* $R\&D_{t-3}$+0.2* $R\&D_{t-4}$. To examine the impact of R&D disclosure on future R&D outcomes, we include log(*RD_ME*) and log(*R&D DISC*) in the model. Similar to prior research, we deflate R&D expense with market capitalization. To measure narrative R&D disclosure quantity, we follow Merkley (2014) and count the number of sentences with at least one R&D-related word in 10-K reports.[12] We also interact

[11] We replicate the analysis with alternative scalers, such as market value of equity, total assets, and sales, following prior literature (Connolly and Hirschey 1988; Griliches 1998; Deng et al. 1999); the use of alternative scalers does not change our main inferences (untabulated).

[12] For sensitivity, we employ alternative measures of narrative R&D disclosure and find that our results are qualitatively similar (untabulated). We discuss the sensitivity tests in detail in Section 5.

R&D expense with narrative R&D disclosure and include the term in the model. Following Cohen et al. (2013) and Zhong (2018), we also control for several variables that might affect R&D outcomes, log(*ME*), log(*BTM*), log(1+*Lev*), log(1+*Age*), *Instown, Finance,* log(*Employment*), K/L and *Foreignsale%*. In order to control for potential correlated omitted variable, we add industry and year fixed effect. We cluster standard errors by firms.

To explore whether investors efficiently recognize the implications of narrative R&D disclosure for future R&D outcomes, we estimate monthly Fama-MacBeth regressions as follows:

$$\begin{aligned} Ret_{\mathrm{i,t+1-t+2}} = {} & \beta_0 + \beta_1 \log\left(RD\_ME_{i,t}\right) + \beta_2 \log\left(R\&D\ DISC_{i,t}\right) + \beta_3 \log\left(RD\_ME_{i,t}\right) \\ & \times \log\left(R\&D\ DISC_{i,t}\right) + \beta_4 ROA_{adj} + \beta_5 \log\left(RD_{ME_{i,t}}\right) \times ROA_{adj} + \beta_6 \log\left(ME_{i,t}\right) \\ & + \beta_7 \log\left(BTM_{i,t}\right) + \beta_8 ret_{-12,-2} + \beta_9 ret_{-1} + \beta_{10} \log\left(N\_words_{i,t}\right) \\ & + \beta_{11} \log\left(RD\_ME_{i,t}\right) \times \beta_{12} \log\left(N\_words_{i,t}\right) + Industry\ Fixed\ Effec \\ & + \varepsilon_{i,t} \qquad (\mathbf{2}) \end{aligned}$$

The dependent variable is the monthly excess return from July of year t+1 to June of year t+2, defined as the raw monthly return net of the one-month Treasury bill rate. We match accounting variables of year t with returns from July of year t+1 to allow for sufficient time for which information of fiscal year t in 10-K filings is disseminated. Similar to equation (1), we use R&D expense (log(*RD_ME*)) and narrative R&D disclosure (log(*R&D DISC*)); we also interact the two terms and include the interaction in the model. We control for firm size (Banz 1981), book-to-market (Ban-Rosenberg et al. 1983; Fama and French 1992) and momentum (Carhart 1997; Jegadeesh and Titman 1993). We include adjusted ROA (*ROA_adj*) because it is known to be negatively correlated with *R&D DISC* (Merkley 2014). We also control for the number of total words in 10-K reports (*N_words*) to alleviate the concern that our result simply captures the market response to long 10-Ks reported in (You and Zhang 2009). In additional tests, we also use *raw R&D DISC/N_words* instead of *raw R&D DISC* and find that our inferences remain the same.[13] Finally, we include industry fixed effects using two digit SIC codes. To adjust for potential autocorrelation arising from Fama-Macbeth regressions, we adjust t-statistics based on Newey and West (1986) robust standard errors.

4 Results

4.1 Descriptive Statistics and Correlations

Table 1 provides the descriptive statistics used in our analyses. The mean (median) of *R&D DISC* is 98 (65), which is higher than that reported in Merkley (2014). However, note that Merkley (2014) covers the years 1996 to 2007, while our sample covers the years 1993 to 2016.[14] We find that that narrative R&D disclosure increases over time, suggesting that R&D has become more important in recent years.

---

[13] As an alternative way to control for the number of words in a 10-K, we use following specification:

$$ret_{\mathrm{i,t+1-t+2}} = \beta_0 + \beta_1 \log(RD\_ME_{i,t}) + \beta_2 \log\left(\frac{R\&D\ DISC_{i,t}}{N\_word_{i,t}}\right) + \beta_3 \log(RD\_ME_{i,t}) \times \log\left(\frac{R\&D\ DISC_{i,t}}{N\_word_{i,t}}\right) + \beta_4 ROA_{adj} + \beta_5 \log(RD\_ME_{i,t}) \times ROA_{adj} + \beta_6 \log(ME_{i,t}) + \beta_7 \log(BTM_{i,t}) + \beta_8 ret_{-1\ ,-2} + \beta_9 ret_{-1} + Industry + \varepsilon_{i,t}$$

[14] When we restrict our sample period to that of Merkley (2014), we find the distribution of *R&D DISC* is almost identical (untabulated).

We also confirm similar findings of Merkley (2014) with our measure of R&D narrative disclosure (*R&D DISC*), giving added comfort to our measure. Our statistics for measures of R&D expense and firm characteristics are similar to prior research on the R&D anomaly (Leung et al. 2017). We find that the mean of adjusted ROA is almost identical to that of Merkley (2014). We find that the mean of our ability measure is 2.6, which is comparable to that of Hirshleifer et al. (2013).

[Insert Table 1 here]

Table 2 reports Pearson correlations and Spearman correlations among the variables used in our analyses, along with *p-values*. Pearson (Spearman) correlations are presented above (below) the diagonal. Our three narrative R&D disclosure measures exhibit high correlations (0.86, 0.9, 0.85) with each other, implying that our measures capture similar underlying characteristics. As expected, we find a positive and significant relation between qualitative R&D disclosure (*R&D DISC*) and future R&D outcomes, suggesting that R&D disclosure predicts future R&D outcomes. We find that the correlations are largely consistent with prior research (Merkley 2014).

[Insert Table 2 here]

4.2 R&D Disclosure and Future R&D Outcomes

We first examine whether narrative R&D disclosure contains value-relevant information. Merkley (2014) suggests that managers release narrative R&D disclosure to provide information about R&D projects. In his content and tone analysis, Merkley (2014) finds that narrative R&D disclosure includes information such as in-process R&D and patents, and that in-process R&D is portrayed more positively than other narrative R&D disclosure. Related, Zhong (2018) finds evidence that information transparency increases managerial effort on R&D and R&D efficiency. Zhong (2018) indicates that managers in transparent firms exert higher innovative effort because transparency increases ex-post verifiability of the managerial action and filters noise from uncontrollable risks, alleviating managers' career concerns in multi-period contracts. To the extent that narrative R&D disclosure includes value-relevant information on firms' research investment and outcomes (and increases managers' innovative effort), it is likely that the quantity of narrative R&D disclosure is related to future R&D success measured with patents.

To examine the relationship between narrative R&D disclosure and R&D outcomes, we estimate model (1) discussed in section 3.2. We report the results in Table 3. Column (1) of Table 3 presents annual fixed effect panel regressions of future patents on narrative R&D disclosure and R&D expense. As expected, we find a positive relation between R&D disclosure and future R&D outcomes (coeff. est. = 0.249, t-value = 5.984). We also find that the interaction term of R&D disclosure and R&D expense is positive and significant (coeff. est. = 0.0397, t-value = 3.939), indicating that firms with greater R&D disclosure combined with high R&D expense produce more patents in the following year.[15]

---

[15] We replicate the analysis with dichotomous and quintile variables of *R&D DISC* and *R&D expense* instead of continuous variables and obtain similar results (untabulated).

To gauge the economic magnitude of this result, we compare the first and third quartiles of *R&D DISC*. The difference of log of patent between the two quartiles of *R&D DISC* is 0.19(=0.249*(log(114)-log(35))+0.0397*(log(114)-log(35))*log(0.1)), which means the raw number of patent increases about 21% (=exp(0.19)) when *R&D DISC* is moved from first quartile to the third. The results based on citation intensity in Column (2) mirror our findings in Column (1) (coeff. est. = 0.422, t-value = 6.617 for *R&D DISC*; coeff. est. = 0.0499, t-value=3.186 for *RD_ME*R&D DISC*). In Columns (3) and (4), we find similar positive and significant coefficients on both *R&D DISC* and its interaction term with *RD_ME*. Overall, this implies that firms with high *R&D DISC* exhibit more successful R&D outcomes in the following year. Also, this relation appears more salient for firms with high R&D expense.

[Insert Table 3 here]

4.3 Portfolio Returns to R&D Expense

Before assessing the impact of narrative R&D disclosure, we first confirm the R&D anomaly in our sample. We conduct a single-sorted portfolio analysis in *RD_ME* three-way sorts using the same methodology of Fama and French (1996) and Cohen et al. (2013), which partitions the sample into 30%/40%/30% (low/medium/high) groups. We form *RD_ME* three-way portfolios at the end of June each year (year t) and monthly returns are matched from July of year t+1 to June of year t+2 to ensure that investors have sufficient time to observe information about R&D expense in 10-K filings.

Table 4 reports equal-weighted monthly excess returns adjusted for the risk-free rate (EXRET), DGTW returns following Daniel et al. (1997), FF-3 alpha using the approach of three-factor model developed in Fama and French (1993), and FF-5 alpha using the approach of five-factor model developed in Fama and French (2015). In all four approaches, we find that the mean abnormal monthly return is monotonically increasing in *RD_ME*. The zero-cost long-short spread portfolio (portfolio 3-1) yields a positive and significant excess monthly return of 1.096%, a DGTW return of 0.977%, a FF-3 alpha of 0.955%, (t=3.74), and a FF-5 alpha of 1.182% (t=3.62). Collectively, we confirm that the R&D anomaly exists in our sample after adjusting for conventional asset pricing factors.

[Insert Table 4 here]

4.4 Portfolio Returns to R&D Expense and Narrative R&D Disclosures

In this section, we assess whether information in narrative 10-K disclosure mitigates or magnifies the R&D anomaly (H1) by examining excess returns to portfolios based on both R&D expense and narrative R&D disclosure. Similar to single portfolio sorts, we use the three-way methodology of Fama and French (1996) and Cohen et al. (2013). Portfolios are independently sorted on the two dimensions of *RD_ME* and *R&D DISC*, with the breakpoints of 30%/40%/30%.[16] Each portfolio is formed as intersections of *RD_ME* and *R&D DISC*. Portfolios are sorted at the end of

[16] We replicate the analysis with conditional sorts (first sort with *R&D DISC* and then sort with *RD_ME* within each portfolio) to ensure that the same number of stocks are allocated to each portfolio, and obtain qualitatively similar results (untabulated).

June each year (year t) and monthly returns are matched from July of year t+1 to June of year t+2 to ensure that investors have sufficient time to assess the information contained in 10-K filings.

Table 5, Panel A reports excess returns of portfolios independently sorted on *RD_ME* and *R&D DISC*. We report positive and significant returns across all *R&D DISC* groups. In the lowest group of *R&D DISC*, a hedge strategy that goes long in high *RD_ME* and short in low *RD_ME* yields a 0.638% monthly excess return (t-value=3.12), compared to a 1.422% excess return (t-value=3.71) for the highest group of *R&D DISC*. Our results reveal that more narrative R&D disclosure exacerbates the R&D anomaly. A long-short hedge portfolio strategy employing both *R&D DISC* and *RD_ME* yields higher abnormal returns of 32 basis points relative to that using only *RD_ME*. Taken together, these results suggest that high narrative disclosure combined with R&D expense is an important factor in predicting future excess returns.

Table 5, Panel B provides excess returns based on DGTW returns, FF-3 returns, and FF-5 returns for double sorts. The first (latter) four columns report the lowest (highest) group in *R&D DISC*, with the first, second, third and zero-cost long short portfolio sorted on *RD_ME* in each *R&D DISC* quintile. Results for each asset pricing model provides similar results. A strategy that goes long in the highest *RD_ME* and short in the lowest *RD_ME* under the highest group of *R&D DISC* portfolio generates higher returns compared to that under the lowest *R&D DISC* portfolio. The DGTW monthly return based on both *R&D DISC* and *RD_ME* is 1.218%, compared to 0.977% based only on *RD_ME*. The FF-3 excess return using both R&D expense and narrative R&D disclosure is 1.222% (t-value=3.81), compared to 0.955% (t-value=3.74) based only on *RD_ME*. Lastly, the FF-5 excess return using *RD_ME* and *R&D DISC* is 1.314% (t-value=4.00), compared to 1.187% (t-value=3.62) using *RD_ME* only.

Combined with the results for R&D outcomes in section 4.2, our findings for the two-dimension portfolio analysis suggest that investors fail to incorporate in a timely fashion the value-relevant information in narrative R&D disclosure, yielding a positive and economically significant price drift over the following year.

[Insert Table 5 here]

4.5 Return Predictability of Narrative R&D Disclosure

To further assess the predictive power of narrative R&D disclosure and its interaction with R&D expense using model (2) from section 3.2, our next set of tests employs monthly Fama and MacBeth (1973) cross-sectional regressions

Table 6 reports the Fama-Macbeth regression results. The first column replicates Chan et al. (2001) to ensure that investors' underreaction to R&D expense exists in our sample. In column (1), we report a significant and positive coefficient of 0.361 (t-value=3.667) on *RD_ME,* consistent with prior research. Column (2) provides results including *RD_ME*, *R&D DISC*, their interaction term, and control variables included in model (2). Column (3) includes industry fixed effects in addition to the model in column (2). Results for columns (2) and (3) are similar, so for brevity we restrict our discussion to the results in column (2). Consistent with our portfolio analysis in section 4.4, we find that the coefficient on *R&D DISC* (coeff. est=0.837, t-value=2.737) and the interaction term (coeff. est. = 0.216, t-value=3.714) are significant and positive, indicating that firms with high narrative R&D disclosure combined with high R&D expense generate positive abnormal returns. This result is not only statistically significant, but also economically significant. When we move *R&D DISC* from the first quartile to the third quartile of its distribution, we obtain a 0.40% monthly excess return and 0.25% more sensitivity to 1% change of R&D intensity (i.e. R&D anomaly), which is comparable to our results in table 5. Results from the monthly Fama-Macbeth regressions are consistent with our portfolio analysis in table 5. Overall, results in Tables 5 and 6 suggest that high narrative R&D disclosure exacerbates investors' underreaction to R&D expense.

[Insert Table 6 here]

4.6 Cross-sectional Variation in Return Predictability of Narrative R&D Disclosure

In this subsection, we conduct additional tests to assess reasons why narrative R&D disclosure is positively related to investors' underreaction to R&D expense. The mechanisms we examine are examine limited attention (H2a) and readability (H2b).

4.6.1 Limited Attention

Hirshleifer and Teoh (2003) suggest that increased inattention worsens market mispricing. Consistently, several studies find empirical results supporting the view that inattention exaggerates mispricing (Drake et al. 2014; Eichler 2012; Hirshleifer and Teoh 2003; Jin 2014; Madsen 2017). We proxy for investor attention using EDGAR page views because our variable of interest is narrative R&D disclosures in a 10-K, and 10-Ks are primarily accessed via EDGAR (Loughran and McDonald 2017).

Table 7 reports the results of Fama-Macbeth regressions in model (2) for the subsamples. We divide the sample into two subsamples of high and low 10-K page views via EDGAR each year. We count total page views from the filing date to June of the following year (t+1), when our portfolio formation begins. Due to the lack of EDGAR log data, this analysis limits our sample period to the period 2002 - 2015 and the number of firm-month observations to 116,464 and 117,498 for the low and high pageview subsamples, respectively.

The first two columns of Table 7 report the regression results. We find that investors' mispricing is more severe for the subsample of low EDGAR page views. For example, in Column (1), Table 7 the coefficients on narrative R&D disclosure and its interaction term with R&D expense remain significantly positive (coeff. est. = 0.852, t-value = 2.211; coeff. est. = 0.228, t-value = 2.640, respectively). This effect is economically significant. A change in *R&D DISC* from the first quartile to the third quartile (while other variables are fixed at their third quartile) leads to a 0.38%

increase in monthly excess returns and 0.27% more sensitivity to 1% change in R&D intensity (i.e. R&D anomaly). On the other hand, in the subsample of high EDGAR page views tabulated in the latter two columns, we find that not only does the coefficient of interest decrease in magnitude, but it also loses its statistical significance. (coeff. est. = -0.006, t-value = -0.016; coeff. est. = 0.0324, t-value = 0.414, respectively). Compared to the pooled sample for the period of 2002 to 2015, we find more pronounced mispricing only for the subsample of firms with lower page views.[17] Overall, results in Table 7 are consistent with investor inattention.

[Insert Table 7 here]

4.6.2 Readability (Fog)

Prior theoretical works suggest that investors react more slowly when the information contained in disclosure is difficult to extract (Bloomfield 2002; Grossman and Stiglitz 1980; Hirshleifer and Teoh 2003). For example, the textual analysis literature finds that investors' underreaction is more severe when disclosure is less readable (Asay et al. 2016; Lawrence 2013; F. Li 2008; Miller 2010; Rennekamp 2012; Tan et al. 2014; You and Zhang 2009). We therefore expect that investors' underreaction to narrative R&D disclosure is more severe for firms with less readable 10-Ks. Our sample is divided into firms with high and low 10-K readability (i.e. low and high fog index as in F. Li (2008)) in each year. Due to the data limitation to

[17] The pooled regression results in columns (1) and (3) yield 0.538 (t-value=2.312) for narrative R&D disclosure and 0.158 (t-value=2.58) for its interaction term with R&D expense. The difference between this result and that of column (2) in Table 6 is attributable to the different sample period. The results in Table 7 include shorter periods due to the constraint of EDGAR log data.

calculate the fog index, our sample is limited to the years 1993 to 2011, yielding observations of 103,841 and 105,069 for the low and high readability subsample, respectively.

The first two columns of Table 8 report the regression results for the subsample of low 10-K readability. As expected, investors appear to sluggishly respond to information contained in 10-K reports. For example, in column (1) of the low readability subsample, the coefficient on narrative R&D disclosure (coeff. est. = 0.742, t-value = 1.669) and its interaction term with R&D expense (coeff. est. = 0.207, t-value = 2.621) remain significantly positive. This effect is economically significant. An increase in *R&D DISC* from the first quartile to the third quartile (while other variables are fixed at their third quartile) leads to a 0.31% increase in monthly excess returns and 0.27% more sensitivity to 1% change in R&D intensity (i.e. R&D anomaly). In contrast, for the subsample with high readability tabulated in the latter two columns, we find that the corresponding result is insignificant. We find that the low readability subsample shows incremental mispricing, not only compared to the subsample of high readability, but also compared to the pooled sample.[18]

Collectively, our subsample analysis suggests that mispricing of narrative R&D is only appears for firms with low 10-K page views and firms with low 10-K readability, consistent with H2a and H2b, respectively.

---

[18] The pooled sample regression results for columns (1) and (3) yield abnormal returns of 0.669 (t-value=2.03) for narrative R&D disclosure and 0.167 (t-value=2.66) for its interaction with R&D expense. The difference between this result and that in Table 6, column (2) is likely due to the different sample period. The results in Table 8 include a shorter period due to constraints on fog index data.

[Insert Table 8 here]

5 Robustness Tests

5.1 Controlling for other R&D-related Effects

In this section, we examine whether our results for return predictability are subsumed by other known R&D-related effects on future returns. Specifically, we control for innovation efficiency of Hirshleifer et al. (2013), R&D increases of Eberhart et al. (2004), intangible information of Daniel and Titman (2006), and managerial ability of Cohen et al. (2013) and re-estimate the monthly Fama-Macbeth regression model in Table 6. Innovation efficiency is measured as in Table 3 and Table 9. Following Eberhart et al. (2004), we treat R&D increases as large if the following conditions are met (Eberhart et al. 2004; Cohen et al. 2013): (1) raw R&D increased by 5%, (2) the level of R&D divided by lagged assets is greater than 5%, and (3) the change in R&D divided by lagged asset is greater than 5%. *R&D* $Increase_{t-1}$ takes the value of 1 if there was a large R&D increase between year t-1 and year t. *R&D* $Increase_{t-5,t-1}$ equals 1 if there is a large R&D increase in the previous 5 years. $\log(B/M)_{t,t-5}$ refers to the change of *BTM* over the past five years, ret_{book} refers to book return, and $ret_{intangible}$ refers to intangible return, defined as in Daniel and Titman (2006). We include the control variables in Column 4, Table 6 but omit their tabulation for brevity.

Table 9 reports the results. Similar to results in Table 9, we find that after controlling for *Ability_high* and its interaction with R&D expense, the coefficient on the interaction of R&D expense and narrative R&D disclosure remains positive and significant (coeff. est. = 0.229, t-value = 3.301). The coefficient on the interaction of *Ability_high* and R&D expense is significant and positive (coeff. est. = 0.238, t-value=2.152), consistent with Cohen et al. (2013). Column (2) exhibits the regression model including *IE_high*. Although the coefficient on narrative R&D disclosure becomes insignificant (coeff. est. = 0.632, t-value=1,451), the coefficient on its interaction term with R&D expense remains positive and significant (coeff. est. = 0.172, t-value=1.918). The reduced significance is likely due to the fact that narrative R&D disclosure combined with R&D expense reflects innovation efficiency as in Table 3. Nevertheless, our measure remains incrementally significant. The third and fourth columns control for firms with R&D increases; for these specifications, we find that the coefficient on narrative R&D disclosure (coeff. est. = 0.782, t-value=2.466; 0.787, t-value=2.436) and its interaction term with R&D expense (coeff. est. = 0.218, t-value=3.454; coeff. est. = 0.216, t-value=3.395) remain positive and significant. Lastly, the fifth column presents the results after controlling for *BTM* change, book return and intangible returns. We continue to find that the coefficient on narrative R&D disclosure (coeff. est. = 0.795, t-value=2.495) and its interaction term with R&D expense (coeff. est. = 0.218, t-value=3.444) remain qualitatively unaffected. Collectively, our findings are robust to the inclusion of other known determinants of returns related to R&D.

[Insert Table 9 here]

5.2 Alternative Measures of Narrative R&D Disclosure

To test the robustness of qualitative R&D disclosure, we re-estimate monthly Fama-Macbeth regressions reported in Table 6 using three alternative measures of narrative R&D disclosure. We first use the raw measure of narrative R&D disclosure deflated by the total number of words in a firm's 10-K report, *R&D DISC/n_words*, in order to control for total number of words in 10-K filing. Second, we use the count of R&D-related words, *R&D DISC_words*. Third, we use the count of only core R&D-related words, *R&D DISC_corewords*, which are 'Research', 'R&D', and 'Research and Development'.

Table 10 reports monthly Fama-Macbeth regressions using these measures. The first two columns of Table 10 exhibit results for narrative R&D disclosure measured with *R&D DISC/n_words*. The coefficients on narrative R&D disclosure (coeff. est. = 0.759, t-value=2.913; coeff. est. = 0.718, t-value=2.823) and the interaction terms (coeff. est. = 0.193, t-value=3.742; coeff. est. = 0.193, t-value=3.774) remain positive and significant . In columns (3) and (4), we report results with *R&D DISC_words*. Again, we find that the coefficients on narrative R&D disclosure (coeff. est. = 0.635, t-value=2.513; coeff. est. = 0.554, t-value=2.19) and the interaction term with R&D expense (coeff. est. = 0.170, t-value=3.534; coeff. est. = 0.169, t-value=3.367) remain significantly positive. The last two columns of Table 11 show the results when narrative R&D disclosure is *R&D DISC_corewords*,. We find that in columns (5) and (6), narrative R&D disclosure (coeff. est. = 0.597, t-value=2.606; coeff.

est. = 0.607, t-value=2.582) and its interaction term with R&D expense (coeff. est. = 0.157, t-value=2.906; coeff. est. = 0.171, t-value=2.999) remain positive and significant. Overall, our findings remain robust to alternative measures of qualitative R&D disclosure.

[Insert Table 10 here]

6 Summary and Conclusion

Despite the importance of understanding the benefits and costs of R&D disclosure and the increased importance of qualitative disclosure, we know little about how investors respond to information contained in narrative R&D disclosures in 10-K filings. We expand prior research by examining whether market participants efficiently recognize the implications of information embedded in narrative R&D for the R&D anomaly. Using a large sample firms for the sample period of 1993-2016, we find evidence that narrative R&D disclosure increases the relation between R&D expense and future stock returns. This evidence suggests that investors fail to fully impound such information in stock prices at the time it is reported in firms' 10-Ks, increasing a positive and significant price drift for firms with high R&D expense over the next 12-month period. We further show that the adverse impact of narrative R&D disclosure on the relation between R&D expense and future stock returns is more pronounced when the number of EDGAR page views by investors is small and when 10-K reports are less readable. Our results are robust to the inclusion of different

R&D-related effects reported in prior research and alternative measures of narrative R&D disclosures.

We extend prior research on the R&D expense anomaly (Chan et al. 2001; Cohen et al. 2013; Eberhart et al. 2004; Hirshleifer et al. 2013) by showing that qualitative R&D disclosure is another dimension of investors' underreaction. We also add to the literature by providing evidence that 10-K views via EDGAR and fog index for 10-K filings play a moderating role in investor underreaction to qualitative R&D disclosure. Overall, our findings suggest that mandatory disclosure of narrative R&D information impairs investors' ability to more fully understand R&D investment.

References

Aboody, D., & Lev, B. (2000). Information asymmetry, R&D, and insider gains. *The Journal of Finance, 55*(6), 2747-2766.

Al-Horani, A., Pope, P. F., & Stark, A. W. (2003). Research and development activity and expected returns in the United Kingdom. *Review of Finance, 7*(1), 27-46.

Ali, A., & Gurun, U. G. (2009). Investor sentiment, accruals anomaly, and accruals management. *Journal of Accounting, Auditing & Finance, 24*(3), 415-431.

Asay, H. S., Elliott, W. B., & Rennekamp, K. (2016). Disclosure readability and the sensitivity of investors' valuation judgments to outside information. *The Accounting Review, 92*(4), 1-25.

Ban-Rosenberg, Reid, K., & Lanstein, R. (1983). Persuasive Evidence of Market Inefficiency, J. Portfolio Mgmt. 9 (Spr. 1985). d Louis Chan, Yasushi Hamao & Josef Lakonishok, Fundamentals and Stock Returns in Japan, 46 J. Fin. 1739 (1991). e Sanjoy Basu, The Relationship Between Earnings Yield, Market Value, and Return for NYSE Common Stocks: Further Evidence. *J. Fin. Econ., 12*, 129.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics, 9*(1), 3-18.

Barber, B. M., & Odean, T. (2007). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies, 21*(2), 785-818.

Bloomfield, R. J. (2002). The "incomplete revelation hypothesis" and financial reporting. *Accounting Horizons, 16*(3), 233-243.

Bonsall, S. B., & Miller, B. P. (2017). The impact of narrative disclosure readability on bond ratings and the cost of debt. *Review of Accounting studies, 22*(2), 608-643.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance, 52*(1), 57-82.

Cazier, R. A., & Pfeiffer, R. J. (2015). Why are 10-K filings so long? *Accounting Horizons, 30*(1), 1-21.

Chambers, D., Jennings, R., & Thompson, R. B. (2002). Excess returns to R&D-intensive firms. *Review of Accounting studies, 7*(2-3), 133-158.

Chan, L. K., Lakonishok, J., & Sougiannis, T. (2001). The stock market valuation of research and development expenditures. *The Journal of Finance, 56*(6), 2431-2456.

Cohen, L., Diether, K., & Malloy, C. (2013). Misvaluing innovation. *The Review of Financial Studies, 26*(3), 635-666.

Connolly, R. A., & Hirschey, M. (1988). Market value and patents: A Bayesian approach. *Economics Letters, 27*(1), 83-87.

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance, 52*(3), 1035-1058.

Daniel, K., & Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *The Journal of Finance, 52*(1), 1-33.

Daniel, K., & Titman, S. (2006). Market reactions to tangible and intangible information. *The Journal of Finance, 61*(4), 1605-1643.

Deng, Z., Lev, B., & Narin, F. (1999). Science and technology as predictors of stock performance. *Financial Analysts Journal, 55*(3), 20-32.

Donelson, D. C., & Resutek, R. J. (2012). The effect of R&D on future returns and earnings forecasts. *Review of Accounting studies, 17*(4), 848-876.

Drake, M. S., Guest, N. M., & Twedt, B. J. (2014). The media and mispricing: The role of the business press in the pricing of accounting information. *The Accounting Review, 89*(5), 1673-1701.

Easton, P. D., & Zmijewski, M. E. (1993). SEC form 10K/10Q reports and annual reports to shareholders: Reporting lags and squared market model prediction errors. *Journal of Accounting Research*, 113-129.

Eberhart, A. C., Maxwell, W. F., & Siddique, A. R. (2004). An examination of long-term abnormal stock returns and operating performance following R&D increases. *The Journal of Finance, 59*(2), 623-650.

Eichler, S. (2012). Limited investor attention and the mispricing of American Depositary Receipts. *Economics Letters, 115*(3), 490-492.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance, 47*(2), 427-465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1), 3-56.

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance, 51*(1), 55-84.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics, 116*(1), 1-22.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy, 81*(3), 607-636.

Feldman, R., Govindaraj, S., Livnat, J., & Segal, B. (2010). Management's tone change, post earnings announcement drift and accruals. *Review of Accounting studies, 15*(4), 915-953.

Foster III, T. W., Jenkins, D. R., & Vickrey, D. W. (1983). Additional Evidence on the Incremental Information Content of the 10-K. *Journal of Business Finance & Accounting, 10*(1), 57-66.

Glaeser, S. (2018). The effects of proprietary information on corporate disclosure and transparency: Evidence from trade secrets. *Journal of Accounting and Economics, 66*(1), 163-193.

Griliches, Z. (1998). Patent statistics as economic indicators: a survey. In *R&D and productivity: the econometric evidence* (pp. 287-343): University of Chicago Press.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review, 70*(3), 393-408.

Gu, L. (2016). Product market competition, R&D investment, and stock returns. *Journal of Financial Economics, 119*(2), 441-455.

Hirshleifer, D., Hsu, P.-H., & Li, D. (2013). Innovative efficiency and stock returns. *Journal of Financial Economics, 107*(3), 632-654.

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics, 36*(1-3), 337-386.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies, 28*(3), 650-705.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance, 48*(1), 65-91.

Jin, J. Y. (2014). Investor attention and stock mispricing. *Accounting Perspectives, 13*(2), 123-147.

Koh, P.-S., & Reeb, D. M. (2015). Missing r&d. *Journal of Accounting and Economics, 60*(1), 73-94.

Koh, P.-S., Reeb, D. M., & Zhao, W. (2017). CEO confidence and unreported R&D. *Management Science*.

Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting and Economics, 56*(1), 130-147.

Leung, W. S., Mazouz, K., & Evans, K. P. (2017). The R&D Anomaly: Risk or Mispricing? *Working paper*.

Lev, B., & Sougiannis, T. (1996). The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting and Economics, 21*(1), 107-138.

Li, D. (2011). Financial constraints, R&D investment, and stock returns. *The Review of Financial Studies, 24*(9), 2974-3007.

Li, E. X., & Ramesh, K. (2009). Market reaction surrounding the filing of periodic SEC reports. *The Accounting Review, 84*(4), 1171-1208.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics, 45*(2-3), 221-247.

Loughran, T., & McDonald, B. (2017). The use of EDGAR filings by investors. *Journal of Behavioral Finance, 18*(2), 231-248.

Madsen, J. (2017). Anticipated earnings announcements and the customer–supplier anomaly. *Journal of Accounting Research, 55*(3), 709-741.

Merkley, K. J. (2014). Narrative disclosure and earnings performance: Evidence from R&D disclosures. *The Accounting Review, 89*(2), 725-757.

Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *The Accounting Review, 85*(6), 2107-2143.

Newey, W. K., & West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. National Bureau of Economic Research Cambridge, Mass., USA.

Rajgopal, S., Shevlin, T., & Venkatachalam, M. (2003). Does the stock market fully appreciate the implications of leading indicators for future earnings? Evidence from order backlog. *Review of Accounting studies, 8*(4), 461-492.

Rennekamp, K. (2012). Processing fluency and investors' reactions to disclosure readability. *Journal of Accounting Research, 50*(5), 1319-1354.

Stice, E. K. (1991). The market reaction to 10-K and 10-Q filings and to subsequent The Wall Street Journal earnings announcements. *Accounting Review*, 42-55.

Tan, H.-T., Wang, E. Y., & Zhou, B. (2014). How does readability influence investors' judgments? Consistency of benchmark performance matters. *The Accounting Review, 90*(1), 371-393.

You, H., & Zhang, X.-j. (2009). Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting studies, 14*(4), 559-586.

You, H., & Zhang, X.-J. (2011). Limited attention and stock price drift following earnings announcements and 10-K filings. *China Finance Review International, 1*(4), 358-387.

Zhong, R. I. (2018). Transparency and firm innovation. *Journal of Accounting and Economics, 66*(1), 67-93.

Appendix A

Variable Definitions

Variable	Definition
R&D DISC	Total sentences in 10-K filings that include R&D related words. R&D related words follow the dictionary of Merkley (2014).
R&D DISC_words	Total R&D-related words in 10-K filings.
R&D DISC_corewords	Number of words of "Research", "R&D", "Research and Development" in 10-K filings.
RD_ME	R&D expense scaled by contemporary market value of equity.
R&D Capital (RDC)	$R\&D_{t-1}+0.8*R\&D_{t-2}+0.6*R\&D_{t-3}+0.4*R\&D_{t-4}+0.2*R\&D_{t-5}$ following Chan et al. (2001)
Patent_f1	Patent issued to firm i in year t+1
Citation_f1	Number of citations of patents issued to firm i in year t+1
Ability	Coefficient γ_j of 5 years (j=1,…,5) rolling regression, where a back window of 6-8 years of nonmissing data and at least half of non-zero R&D data is required following Cohen et al. (2013) $$\log\left(\frac{\text{Sales}_t}{Sales_{t-1}}\right)=\gamma_0+\gamma_j\log\left(1+R\&D_{t-j}\right)+\varepsilon_t$$
N_Words	Total number of words in 10-K report of firm i in year t
pageview	Total number of downloads of year t 10-K report of firm i counted from the filing date to the end of June of year t+1 (beginning of portfolio formation)
fog	Fog index of 10-K report of firm i in year t defined as in F. Li (2008)
ME	Market value of equity, defined as in Daniel and Titman (1997)
BTM	Book to market ratio, defined as in Daniel and Titman (1997)
ROA_adj	Adjusted ROA, defined as ROA before R&D expense and advertising expense as in Merkley (2014)
Leverage	Book value of debt divided by book value of equity
Instown	Top 5 institutional ownership in June year t

Age	Age of the firm computed from the first appearance year in COMPUSTAT
Finance	The sum of a firm's net equity issue scaled by total assets over a rolling five-year window ending in the current fiscal year.
Employment	Natural log of one plus total number of employees in thousands
K/L	Ratio computed as net property, plant and equipment scaled by total number of employees
Foreignsale%	The percentage of foreign net income to total sales during the past five years

Table 1
Descriptive Statistics

This table reports the descriptive statistics of variables used in our analyses. The sample covers the years 1993 to 2016, with 5,235 unique firms. Refer to Appendix A for variable definitions.

Variables	n	Mean	Std Dev	Q1	Median	Q3
R&D Disclosure						
R&D DISC	36,205	98.529	111.607	35	65	114
R&D DISC_words	36,205	201.217	250.655	62	127	242
R&D DISC_corewords	36,205	85.995	107.780	23	52	106
R&D Expense						
RD_ME	36,205	0.095	0.236	0.0169	0.0426	0.0999
R&D capital (Chan et al.)	36,205	260.567	1367.23	7.702	27.888	93.074
R&D Outcome						
patent_f1	27,645	15.942	111.773	0	1	4
cites_f1	27,645	173.591	1508.280	0	0	31
ability	14,347	2.680	23.267	-0.7854	0.6016	3.9169
10-K						
N_Words	29,945	44,740.560	31,083.590	26,362	38,233	54,402
pageview	21,342	284.6	756.32	41	138	293
fog	23,689	19.459	2.350	18.657	19.464	20.358
Firm Characteristic						
ME	36,158	3,495.430	18,920.400	57.008	233.007	1,036.970
BTM	34,417	0.552	0.667	0.215	0.394	0.678
ROA_adj	36,205	0.048	0.388	-0.015	0.103	0.185
leverage	36,204	0.451	16.309	0	0.095	0.504
age	36,205	16.129	14.614	5	11	22
Instown	36,205	0.365	0.357	0	0.279	0.686
finance	36,205	2.3	6.92	0	0.16	1.9
employment	35,872	6.01	22.94	0.13	0.46	2.54
K/L	35,855	56.23	83.78	15.94	30.49	59.62
Foreignsale%	36,205	-0.26	34.01	0	0	0.01

Table 2
Correlation Matrix

This table reports Pearson and Spearman correlation coefficients between the main variables with p-values below. Pearson (Spearman) correlations are presented above (below) the diagonal. Refer to Appendix A for variable definitions.

								H0: Rho=0 under Prob > |r|									
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1)*R&D DISC*		0.898	0.844	0.391	0.301	0.176	0.135	-0.136	0.368	0.129	0.184	0.066	-0.213	0.024	-0.209	-0.374	0.025
		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.084
(2)*R&D DISC_words*	0.831		0.844	0.373	0.260	0.152	0.121	-0.125	0.358	0.113	0.197	0.038	-0.235	0.077	-0.219	-0.411	0.016
	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.276
(3)*R&D DISC_corewords*	0.888	0.761		0.376	0.299	0.188	0.144	-0.137	0.276	0.111	0.109	0.076	-0.204	0.046	-0.163	-0.310	0.049
	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.001
(4)*RD_ME*	0.123	0.087	0.113		0.162	0.054	0.029	-0.070	0.038	-0.068	0.112	-0.375	0.315	-0.101	-0.091	-0.252	-0.123
	0.000	0.000	0.000		0.000	0.000	0.046	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(5)*R&D Capital*	0.136	0.100	0.171	0.005		0.691	0.622	-0.061	0.506	0.569	0.032	0.823	-0.225	0.268	0.257	0.216	0.425
	0.000	0.000	0.000	0.749		0.000	0.000	0.000	0.000	0.000	0.027	0.000	0.000	0.000	0.000	0.000	0.000
(6)*patent_f1*	0.024	0.014	0.041	-0.018	0.599		0.913	-0.013	0.332	0.380	-0.009	0.601	-0.183	0.157	0.206	0.211	0.261
	0.105	0.337	0.004	0.214	0.000		0.000	0.376	0.000	0.000	0.528	0.000	0.000	0.000	0.000	0.000	0.000
(7)*cites_f1*	0.003	-0.004	0.013	-0.013	0.445	0.690		-0.011	0.260	0.253	-0.031	0.555	-0.188	0.146	0.184	0.183	0.217
	0.862	0.806	0.362	0.379	0.000	0.000		0.435	0.000	0.000	0.031	0.000	0.000	0.000	0.000	0.000	0.000
(8)*ability*	-0.061	-0.050	-0.064	-0.025	-0.019	-0.008	-0.007		-0.020	0.016	-0.033	-0.026	0.060	-0.025	0.044	0.084	0.022
	0.000	0.001	0.000	0.088	0.201	0.577	0.629		0.177	0.270	0.025	0.075	0.000	0.080	0.003	0.000	0.127
(9)*N_Words*	0.294	0.322	0.227	-0.010	0.198	0.140	0.105	0.006		0.455	0.288	0.454	-0.104	-0.024	0.287	0.059	0.233
	0.000	0.000	0.000	0.493	0.000	0.000	0.000	0.702		0.000	0.000	0.000	0.000	0.096	0.000	0.000	0.000
(10)*pageview*	0.131	0.096	0.130	-0.036	0.348	0.238	0.134	0.043	0.327		0.078	0.573	-0.086	0.190	0.217	0.202	0.359
	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.003	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000
(11)*fog*	0.083	0.096	0.044	0.020	0.028	0.028	0.017	-0.009	0.247	0.065		-0.035	-0.039	0.028	-0.087	-0.080	0.000
	0.000	0.000	0.002	0.167	0.057	0.055	0.234	0.556	0.000	0.000		0.015	0.008	0.057	0.000	0.000	0.980
(12)*ME*	0.038	0.006	0.023	-0.066	0.604	0.471	0.345	0.005	0.132	0.330	0.003		-0.391	0.326	0.295	0.336	0.475
	0.009	0.662	0.121	0.000	0.000	0.000	0.000	0.741	0.000	0.000	0.848		0.000	0.000	0.000	0.000	0.000
(13)*BTM*	-0.129	-0.122	-0.127	0.609	-0.072	-0.056	-0.044	0.025	-0.052	-0.063	-0.037	-0.083		-0.346	-0.094	0.020	-0.051
	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.087	0.000	0.000	0.010	0.000		0.000	0.000	0.171	0.000
(14)*ROA_adj*	-0.046	-0.047	-0.028	-0.076	0.134	0.069	0.032	-0.011	0.007	0.200	0.018	0.129	-0.155		-0.123	0.031	0.233
	0.002	0.001	0.056	0.000	0.000	0.000	0.026	0.443	0.625	0.000	0.220	0.000	0.000		0.000	0.031	0.000
(15)*leverage*	0.024	0.012	0.022	0.008	0.007	0.003	0.001	0.001	0.024	0.008	-0.012	-0.003	-0.036	-0.046		0.304	0.132
	0.094	0.427	0.134	0.577	0.614	0.846	0.969	0.940	0.091	0.576	0.390	0.815	0.012	0.002		0.000	0.000
(16)*age*	-0.266	-0.252	-0.229	-0.122	0.281	0.222	0.144	0.049	0.123	0.326	-0.049	0.298	-0.022	0.077	0.029		0.088
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.122	0.000	0.042		0.000
(17)*Instown*	0.029	0.015	0.036	-0.117	0.023	-0.009	-0.033	0.029	0.150	0.325	0.015	0.036	-0.100	0.253	-0.013	0.127	
	0.044	0.313	0.013	0.000	0.114	0.520	0.024	0.046	0.000	0.000	0.298	0.013	0.000	0.000	0.353	0.000	

Table 3
Narrative R&D Disclosure and R&D outcomes

This table presents annual panel regressions of R&D outcomes, measured by future patents, citations, and innovation efficiency. Future patents (citations) are defined as natural log of one plus patent (citation) issued at year t+1. Innovation efficiency is defined as patent (citation) issued at year t+1 scaled by R&D capital (Chan et al. 2001) of year t (Hirshliefer et al. 2013). Control variables are *ME*, *BTM*, *Leverage*, *Age, instown, finance, employment, K/L,* and *foreignsales%*, following Cohen et al. (2013) and Zhong (2018). Other variables are as defined in Appendix A. The sample period is restricted to the years 1993 to 2009 due to patent data availability. In all regression, industry and year fixed effects are included. t-values and p-values are based on robust standard errors clustered at the firm-level. All continuous variables are winsorized at the extreme 1 percentile of their distributions. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels (two-tailed), respectively.

Dependent Variable	(1) log(1+$Patent_{t+1}$)	(2) log(1+$Citation_{t+1}$)	(3) log(1+$Patent_{t+1}/RDC_t$)	(4) log(1+$Citation_{t+1}/RDC_t$)
Intercept	-0.678	-0.877	0.163***	0.200
	(-1.251)	(-1.295)	(3.523)	(1.402)
log(*RD_ME*)	0.157***	0.261***	-0.425***	-0.713**
	(3.659)	(3.968)	(-4.898)	(-2.536)
log(*R&D DISC*)	0.249***	0.422***	0.00939**	0.0798***
	(5.984)	(6.617)	(2.121)	(5.458)
log(*RD_ME*)*log(*R&D DISC*)	0.0397***	0.0499***	0.0575***	0.124**
	(3.939)	(3.186)	(3.331)	(2.206)
log(*ME*)	0.412***	0.697***	-0.002	0.0695***
	(19.639)	(21.187)	(-0.762)	(5.506)
log(*BTM*)	-0.00511	0.0174	-0.00359	-0.0118
	(-0.245)	(0.512)	(-1.115)	(-0.880)
log(*leverage*)	0.0330	0.0145	-0.000557	-0.0122
	(0.891)	(0.261)	(-0.105)	(-0.607)
log(*age*)	0.239***	0.259***	-0.00171	-0.0283*
	(9.377)	(6.734)	(-0.363)	(-1.820)
Intsown	-0.277***	-0.148	0.00447	0.0581*
	(-3.751)	(-1.432)	(0.563)	(1.956)
Finance	0.000733	-0.000196	0.000693	0.00509***
	(0.466)	(-0.058)	(1.462)	(2.580)
log(*Employment*)	0.0862***	-0.000770	0.00140	-0.0255*
	(3.624)	(-0.020)	(0.303)	(-1.699)
K/L	0.000734**	0.000507	0.0000702**	0.000106
	(2.507)	(1.233)	(2.001)	(0.868)
Foreignsale%	0.0136***	0.0223***	0.00214***	0.00785***
	(5.549)	(3.827)	(5.613)	(3.781)
Industry fixed effect	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
S.E. clustered by firm	Yes	Yes	Yes	Yes
Number of observations	25,523	25,523	25,523	25,523
Adjusted R^2	0.5038	0.4166	0.0810	0.1608

Table 4
Portfolio Returns to R&D expense

This table reports *RD_ME* three-way sort portfolios with breakpoints of 30%/40%/30% sorted in June each year (t June), following the methodology of Fama and French (1996) and Cohen et al. (2013). All return variables are matched as one-year future returns from July to June of the next year (t+1 July – t+2 June). We report the equal-weighted average (monthly) excess returns adjusted by the risk-free rate (EXRET) and DGTW returns using the approach of Daniel et al. (1997). FF-3 α is the estimated intercept of the time-series regressions based on the three-factor model of Fama and French (1993) that includes size (SMB) and value (HML) factors. FF-5 α is the estimated intercept of the time-series regressions based on the five-factor model of Fama and French (2015) that includes profitability (RMW) and investment (CMA) factors in addition to three-factor model factors. We also report the zero-cost long-short portfolio that goes long in Portfolio 3 and short in Portfolio 1. The t-statistics in parenthesis are calculated based on Newey-West (1987) robust standard errors with 4 month lags. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels (two-tailed), respectively.

	Portfolio			
	1 (Low)	**2**	**3 (High)**	**3-1 (High-Low)**
EXRET				
Mean Return	0.597	1.061**	1.694***	1.096***
t-stat	(1.59)	(2.58)	(3.16)	(3.87)
DGTW				
Mean Return	-0.185**	0.190	0.792***	0.977***
t-stat	(-2.11)	(1.61)	(3.62)	(4.82)
FF-3				
α	-0.296**	0.128	0.659**	0.955***
t-stat	(-2.51)	(-0.85)	(2.11)	(3.74)
FF-5				
α	-0.016	0.408***	1.170***	1.187***
t-stat	(-0.11)	(2.78)	(3.69)	(3.62)
Number of months	288	288	288	288

Table 5
Portfolio Returns to R&D Expense and Narrative R&D Disclosure

This table reports double sorted portfolio on *RD_ME* and *R&D DISC* sorted in June each year (t June). All return variables are matched as one-year future returns from July to June next year. (t+1 July – t+2 June). Portfolios are independently sorted on two dimensions of *RD_ME* and *R&D DISC,* with breakpoints of 30%/40%/30%, following Fama and French (1996) and each portfolio is formed as an interaction of the two. **Panel A** reports the equal weighted average (monthly) excess return adjusted by the risk-free rate (EXRET) for each cell of three-way portfolio. We also report the zero-cost portfolio that goes long on Portfolio 3 and short on Portfolio 1 based on *RD_ME*. **Panel B** reports DGTW returns (Daniel et al. 1997), FF-3 α (Fama and French 1993), and FF-5 α (Fama and French 2015). The first (last) four columns of Panel B show the lowest (highest) tercile sorted on *R&D DISC*. The first (fifth), second (sixth), third (seventh) column represents low, middle, and high tercile of RD_ME under each R&D DISC tercile. The fourth (eighth) column represents long-short portfolio on *RD_ME* under each *R&D DISC* tercile. The t-statistics in parenthesis are calculated based on Newey-West (1987) robust standard errors with 4 month lags. *, **, *** p<0.1, p<0.05, and p<0.01, respectively.

Panel A:Weighted Average Excess Returns				
	RD_ME 1 (Low)	*RD_ME* 2	*RD_ME* 3 (High)	*RD_ME* 3-1 (High-Low)
R&D DISC 1 (Low)				
Mean	0.8145**	1.0372**	1.4783***	0.638**
t-Stat	(2.30)	(2.56)	(3.18)	(2.35)
R&D DISC 2				
Mean	0.5428	1.1569**	1.6348***	1.0920***
t-Stat	(0.65)	(1.99)	(3.08)	(2.88)
R&D DISC 3 (High)				
Mean	0.3117	1.0365	1.7340**	1.4223***
t-Stat	(0.24)	(2.02)	(2.36)	(3.01)
Number of months	288	288	288	288

Table 5 (Cont'd)

Panel B	Low *R&D DISC* (1 Low)				High *R&D DISC* (3 High)			
	RD_ME 1 (Low)	*RD_ME* 2 (Mid)	*RD_ME* 3 (High)	*RD_ME* 3-1 (High-Low)	*RD_ME* 1 (Low)	*RD_ME* 2 (Mid)	*RD_ME* 3 (High)	*RD_ME* 3-1 (High-Low)
DGTW								
Mean	-0.0531	0.1137	0.5194**	0.5724**	-0.3908**	0.2278	0.8275**	1.2182***
t-Stat	(-0.58)	(1.00)	(2.57)	(2.56)	(-1.93)	(-1.02)	(2.46)	(3.79)
FF-3 α								
Mean	-0.0098	0.1444	0.5445*	0.5544**	-0.6045***	0.0758	0.6176	1.2222***
t-Stat	(-0.08)	(1.08)	(1.78)	(2.70)	(-2.74)	(0.33)	(1.54)	(3.81)
FF-5 α								
Mean	-0.0168	0.1653	0.7719**	0.7887***	0.0202	0.5411***	1.3342***	1.3140***
t-Stat	(-0.11)	(1.13)	(2.28)	(3.43)	(0.10)	(2.71)	(3.42)	(4.00)
Number of months	288	288	288	288	288	288	288	288

Table 6
Fama-MacBeth regressions of monthly returns on R&D Expense and Narrative R&D disclosure

This table presents monthly Fama-MacBeth regressions of monthly excess returns on R&D expense and narrative R&D Disclosure. All return variables are matched as one year future return from July to June next year. (t+1 July – t+2 June). Control variables include adjusted ROA, total words in 10-K, size, book-to-market and momentum. Industry fixed effect includes industry dummy variables in two digit SIC code. Variables are as defined in appendix A. The sample covers the period 1993 to 2016. The t-statistics in parenthesis are calculated based on Newey-West (1987) robust standard errors with 4 month lags. All continuous variables are winsorized at the extreme 1 percentile of their distributions. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels (two-tailed), respectively.

Dependent Variable	(1) Excess Return $_{t+1}$	(2) Excess Return $_{t+1}$	(3) Excess Return $_{t+1}$
Intercept	2.076***	2.868	1.705
	(2.594)	(1.220)	(0.889)
log(*RD_ME*)	0.361***	0.502	0.240
	(3.967)	(0.856)	(0.409)
log(*R&D DISC*)		0.837***	0.813***
		(2.737)	(2.625)
log(*RD_ME*)*log(*R&D DISC*)		0.216***	0.233***
		(3.714)	(3.792)
ROA_adj (Merkley 2014)		-0.342	0.0697
		(-0.304)	(0.057)
log(*RD_ME*)**ROA_adj*		-0.567**	-0.424
		(-2.216)	(-1.450)
log(*N_Words*)		-0.371	-0.345
		(-1.482)	(-1.337)
log(*RD_ME*)*log(*N_Words*)		-0.0918	-0.0753
		(-1.536)	(-1.247)
log(*ME*)	0.0414	-0.0230	-0.00595
	(0.665)	(-0.398)	(-0.102)
log(*BTM*)	0.241*	0.192**	0.234***
	(1.856)	(2.049)	(2.935)
$ret_{-12,-2}$	-0.0911	-0.110	-0.135
	(-0.304)	(-0.374)	(-0.462)
ret_{-1}	-3.626***	-3.661***	-3.923***
	(-5.993)	(-5.902)	(-6.211)
Industry fixed effects	No	No	Yes
Number of observations	388,402	323,139	323,139
Adjusted R^2	0.0387	0.0613	0.1134

Table 7
Fama-MacBeth regressions of monthly returns on R&D Expense and Narrative R&D Disclosure: EDGAR pageview subsamples

This table presents monthly Fama-MacBeth regressions of monthly excess returns on R&D expense and narrative R&D Disclosure on two subsamples divided by EDGAR 10-K pageviews each year. Low (High) EDGAR Pageview refers to lower (higher) half of the sample in terms of EDGAR 10-K downloads, calculated from the filing date to the beginning of portfolio formation. Other variables are as in Table 6. The sample covers from 2002 to 2015. The t-statistics in parenthesis are calculated based on Newey-West (1987) robust standard errors with 3 month lags. All continuous variables are winsorized at the extreme 1 percentile of their distributions. *, **, *** p<0.1, p<0.05, and p<0.01, respectively.

	Low EDGAR Pageview		High EDGAR Pageview	
Dependent Variable	(1) Excess Return$_{t+1}$	(2) Excess Return$_{t+1}$	(3) Excess Return$_{t+1}$	(4) Excess Return$_{t+1}$
Intercept	2.336	2.290	1.858	1.270
	(0.607)	(0.855)	(0.437)	(0.346)
log(*RD_ME*)	0.675	0.910	0.625	0.100
	(0.662)	(0.837)	(0.623)	(0.093)
log(*R&D DISC*)	0.852**	0.777**	-0.00633	-0.150
	(2.211)	(2.027)	(-0.016)	(-0.414)
log(*RD_ME*)*log(*R&D DISC*)	0.228***	0.231***	0.0324	0.0212
	(2.640)	(2.623)	(0.414)	(0.276)
ROA_adj (Merkley 2014)	1.282	1.927	2.320	2.645*
	(0.931)	(1.364)	(1.510)	(1.693)
log(*RD_ME*)**ROA_adj*	-0.232	-0.0557	0.170	0.297
	(-0.717)	(-0.164)	(0.421)	(0.714)
log(*N_Words*)	-0.442	-0.588	-0.0346	0.190
	(-0.995)	(-1.242)	(-0.070)	(0.368)
log(*RD_ME*)*log(*N_Words*)	-0.135	-0.155	-0.0517	0.00116
	(-1.209)	(-1.287)	(-0.473)	(0.010)
log(*ME*)	0.00401	0.0464	0.00416	0.00592
	(0.046)	(0.528)	(0.073)	(0.101)
log(*BTM*)	0.213**	0.283***	0.186	0.242**
	(2.027)	(3.070)	(1.614)	(2.140)
$ret_{-12,-2}$	-0.704	-0.794*	-0.0367	-0.0824
	(-1.599)	(-1.808)	(-0.099)	(-0.228)
ret_{-1}	-3.136***	-3.436***	-1.703**	-1.797**
	(-4.110)	(-4.550)	(-2.279)	(-2.461)
Industry fixed effects	No	Yes	No	Yes
Number of observations	116,464	116,464	117,498	117,498
Adjusted R^2	0.0452	0.0949	0.0689	0.1439

Table 8
Fama-MacBeth regressions of monthly returns on R&D expense and narrative R&D disclosure: Readability (Fog) subsamples

This table presents monthly Fama-MacBeth regressions of monthly excess returns on R&D expense and narrative R&D Disclosure on two subsamples divided by fog index as in Li (2008). Low (high) Readability refers to higher (lower) half of the sample in terms of 10-K fog index each year. Other variables are as defined in Table 6. The sample covers from 1993 to 2011. The t-statistics in parenthesis are calculated based on Newey-West (1987) robust standard errors with 3 month lags. All continuous variables are winsorized at the extreme 1 percentile of both ends. *, **, *** p<0.1, p<0.05, and p<0.01, respectively.

	Low Readability		High Readability	
Dependent Variable	(1) Excess Return$_{t+1}$	(2) Excess Return$_{t+1}$	(3) Excess Return$_{t+1}$	(4) Excess Return$_{t+1}$
Intercept	6.914*	-0.0902	-1.748	0.292
	(1.767)	(-0.029)	(-0.397)	(0.082)
log(*RD_ME*)	1.359	1.107	-0.244	-1.145
	(1.286)	(1.006)	(-0.226)	(-0.968)
log(*R&D DISC*)	0.742*	0.725**	0.231	0.251
	(1.669)	(1.925)	(0.598)	(0.604)
log(*RD_ME*)*log(*R&D*	0.227***	0.235**	0.110	0.0949
	(2.621)	(2.322)	(1.384)	(1.029)
ROA_adj (Merkley 2014)	-1.807	-1.665	-0.674	-0.0657
	(-1.098)	(-0.965)	(-0.410)	(-0.040)
log(*RD_ME*)**ROA_adj*	-0.859**	-0.750*	-0.751*	-0.599
	(-2.069)	(-1.786)	(-1.815)	(-1.381)
log(*N_Words*)	-0.667*	-0.553	0.287	0.501
	(-1.727)	(-1.358)	(0.604)	(0.976)
log(*RD_ME*)*log(*N_Words*)	-0.166	-0.157	0.0263	0.124
	(-1.609)	(-1.414)	(0.233)	(0.973)
log(*ME*)	-0.0914	-0.103	-0.0621	-0.0457
	(-1.211)	(-1.248)	(-0.809)	(-0.603)
log(*BTM*)	0.147	0.162	0.322**	0.321**
	(1.168)	(1.298)	(2.448)	(2.364)
$ret_{-12,-2}$	-0.0603	-0.181	-0.143	-0.235
	(-0.175)	(-0.542)	(-0.389)	(-0.589)
ret_{-1}	-5.090***	-5.363***	-3.324***	-3.397***
	(-4.808)	(-5.030)	(-3.692)	(-3.474)
Industry fixed effects	No	Yes	No	Yes
Number of observations	103,841	103,841	105,069	105,069
Adjusted R^2	0.0913	0.1772	0.0941	0.1838

Table 9
Fama-MacBeth regressions of monthly returns on R&D Expense, Narrative R&D Disclosure and other controls

This table presents monthly Fama-MacBeth regressions of monthly excess returns on R&D expense, narrative R&D disclosure, and other controls. *Ability_high* equals to 1 if Ability of the observation is in the highest quintile and 0 otherwise. (Cohen et al. 2013) Control variables are *ME*, *BTM*, *Leverage*, and *Age* in log form following Cohen et al. (2013). *IE_high* equals to 1 if Innovation Efficiency of the observation is in the highest quintile and 0 otherwise (Hirshlifer et al. 2013). *R&D Increase*$_{t-1}$(*R&D Increase*$_{t-1,t-5}$) equals 1 if large R&D increases are met between year t and t-1 (t-1 and t-5). Large R&D increases are defined as is in Eberhart et al. (2004). log(B/M)$_{t,t-5}$ refers to the change of *BTM* over the past five years, ret_{book} refers to book return, and $ret_{intangible}$ refers to intangible return defined as in Daniel and Titman (2006). Other variables are similar to Table 6. The t-statistics in parenthesis are calculated based on Newey-West (1987) robust standard errors with 4 month lags. All continuous variables are winsorized at the extreme 1 percentile of both ends. *, **, *** p<0.1, p<0.05, and p<0.01, respectively.

Dependent Variable	(1) Excess Return$_{t+1}$	(2) Excess Return$_{t+1}$	(3) Excess Return$_{t+1}$	(4) Excess Return$_{t+1}$	(5) Excess Return$_{t+1}$
Intercept	-2.267	4.009	0.023	-0.085	1.080
	(-0.723)	(1.304)	(0.010)	(-0.038)	(0.485)
log(*RD_ME*)	-0.816	0.756	0.051	0.046	0.081
	(-1.069)	(1.058)	(0.089)	(0.080)	(0.142)
log(*R&D DISC*)	0.892***	0.632	0.782**	0.787**	0.795**
	(2.739)	(1.451)	(2.466)	(2.436)	(2.495)
log(*RD_ME*)*log(*R&D DISC*)	0.229***	0.169**	0.218***	0.216***	0.218***
	(3.301)	(1.918)	(3.454)	(3.395)	(3.444)
Ability_high	1.171**				
	(2.508)				
log(*RD_ME*)**Ability_high*	0.238**				
	(2.152)				
IE_high		0.149			
		(1.33)			
R&D Increase$_{t-1}$			0.140*		
			(1.856)		
R&D Increase$_{t-5,t-1}$				0.179***	
				(2.725)	
log(*B/M*)$_{t,t-5}$					0.101
					(0.143)
ret_{book}					0.035
					(1.123)
$ret_{intangible}$					-0.239**
					(-2.270)
Controls	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes
Number of observations	166,883	258,114	323,725	323,725	323,725
Adjusted R^2	0.1203	0.1121	0.1180	0.1320	0.1246

Table 10
Fama-MacBeth regressions of monthly returns on R&D expense and alternative measure of narrative R&D disclosure

This table presents monthly Fama-MacBeth regressions of monthly excess returns on R&D expense and narrative R&D disclosure with different measures of *R&D DISC*. *R&D DISC*/n_words refers to number of sentences containing R&D related words divided by total words in 10-K. *R&D DISC_words* refers to the number of total R&D related words in 10-K report. *R&D DISC_corewords* refers to the number of "Research", "R&D", and "Research and Development" appearing in 10-K report. Other variables and specifications are as in Table 6. The t-statistics in parenthesis are calculated based on Newey-West (1987) robust standard errors with 4 month lags. All continuous variables are winsorized at the extreme 1 percentile of both ends. *, **, *** p<0.1, p<0.05, and p<0.01, respectively.

R&D DISC measure	(1) *R&D DISC*/n_words	(2) *R&D DISC*/n_words	(3) *R&D DISC_words*	(4) *R&D DISC_words*	(5) *R&D DISC_corewords*	(6) *R&D DISC_corewords*
Intercept	7.209***	-1.040	2.616	1.213	1.416	0.773
	(3.612)	(-0.441)	(1.111)	(0.563)	(0.593)	(0.418)
log(*RD_ME*)	1.637***	1.615	0.479	0.185	0.187	-0.0704
	(4.353)	(4.255)	(0.846)	(0.322)	(0.316)	(-0.119)
log(*R&D DISC*)	0.759***	0.718***	0.635**	0.554**	0.597**	0.607**
	(2.913)	(2.823)	(2.513)	(2.190)	(2.606)	(2.582)
log(*RD_ME*)*log(*R&D*) *DISC*)	0.193***	0.193***	0.170***	0.169***	0.157***	0.171***
	(3.742)	(3.774)	(3.534)	(3.367)	(2.906)	(2.999)
ROA_adj (Merkley 2014)	-0.422	-0.057	-0.457	-0.0262	-0.686	-0.200
	(-0.379)	(-0.054)	(-0.402)	(-0.021)	(-0.601)	(-0.161)
log(*RD_ME*)**ROA_adj*	-0.601**	-0.493*	-0.614**	-0.467	-0.662**	-0.508*
	(-2.31)	(-1.67)	(-2.372)	(-1.569)	(-2.546)	(-1.732)
log(*N_Words*)			-0.303	-0.242	-0.0782	-0.0854
			(-1.200)	(-0.923)	(-0.320)	(-0.342)
log(*RD_ME*)*log(*N_Words*)			-0.0806	-0.0545	-0.0275	-0.00988
			(-1.413)	(-0.918)	(-0.468)	(-0.166)
log(*ME*)	-0.021	-0.018	-0.0254	-0.00975	-0.0257	-0.00675
	(-0.447)	(-0.332)	(-0.444)	(-0.170)	(-0.447)	(-0.116)
log(*BTM*)	0.191*	0.226***	0.184*	0.227***	0.174*	0.229***
	(1.898)	(2.748)	(1.854)	(2.691)	(1.828)	(2.798)
$ret_{-12,-2}$	-0.109	-0.137	-0.0919	-0.121	-0.102	-0.131
	(-0.371)	(-0.438)	(-0.316)	(-0.417)	(-0.350)	(-0.453)
ret_{-1}	-3.623***	-3.878***	-3.641***	-3.901***	-3.645***	-3.905***
	(-5.798)	(-6.073)	(-5.859)	(-6.178)	(-5.883)	(-6.182)
Industry fixed effects	No	Yes	No	Yes	No	Yes
Number of observations	322,959	322,959	323,725	323,725	323,725	323,725
Adjusted R^2	0.0623	0.1092	0.0599	0.1123	0.0604	0.1132