INNOVATION OVERLOAD?

Byoung-Hyoun Hwang, Hugh Hoikwang Kim and Kai Wu*

This Draft: January 2020

We conjecture that some individuals are becoming increasingly overwhelmed by the myriad of new ideas and products proposed to them each year. As a result, even minor obstacles to understanding an innovation causes such innovation to not be adopted. To test this idea, we turn to ideas proposed in scientific journal articles. Consistent with our conjecture, we find that how difficult to read an article is strongly negatively predicts its subsequent number of citations. This predictability has strengthened as journals increased the number of articles published in a given year. Our evidence extends to patent descriptions and patent forward citations.

JEL Classification: A10, M30.

Keywords: Innovation Diffusion, Bounded Rationality.

---

* Hwang is affiliated with the Cornell SC Johnson College of Business, Dyson School of Applied Economics and Management, Cornell University, Warren Hall, Ithaca, NY 14853. Kim is affiliated with the Darla Moore School of Business, University of South Carolina, 1014 Greene Street, Columbia, SC 29208. Wu is affiliated with the Central University of Finance and Economics, Beijing, China. E-mail: bhwang@cornell.edu, hugh.kim@moore.sc.edu and wukai8759@cufe.edu.cn. The authors would like to thank Richard Wu, who provided excellent research assistance and Alex Butler, Clifton Green, David Hirshleifer, Soo Kim and Siew Hong Teoh for helpful comments.

1. Introduction

Innovations are an important source of a country's economic growth and, by some accounts, even the most important driver (Rosenberg 2004, OECD 2015). Factors that have been argued to stimulate the creation of innovations include, among others, sources of financing available, managerial incentive structure, behavioral traits of managers, product market competition, and regulatory environment (Hall and Lerner 2010; Ederer and Manson 2011; Kerr and Nanda 2015; He and Tian 2018).

Stimulating innovation is only part of the equation however as any innovation – no matter how useful – can only have economic impact when adopted widely. A long line of work rooted in history, sociology and economics examines the factors that potentially hinder the diffusion of seemingly useful products and technologies (surveyed in Rogers (1995) and Hall (2005)).

In this paper, we point to a possibly critical conundrum. The number of new ideas, products and technologies proposed each year has reached an unprecedented level. For instance, Ulrich's Web Directory notes that there are "*about 33,100 active scholarly peer-reviewed English-language journals in 2018, collectively publishing some 3 million articles a year*" (The STM Report 2018, page 25). The report notes further that the annual number of journals and journal articles has been growing more than exponentially. Similarly, the US Patent and Trademark Office (USPTO) notes that it granted 339,992 patents in 2018. As with scientific journal articles, the annual number of patents granted has been rising more than exponentially.

People have limited capacity for gathering and processing information. With the number of innovations to be considered each year so high, it appears plausible that at least some potential adopters are feeling increasingly overwhelmed. In this study, we speculate that, in such a world of "innovation overload," at least some individuals stop learning about an innovation if there are even minor inconveniences in understanding an innovation. This in turn hampers the diffusion of potentially important ideas, products and technologies. We hereafter refer to this possibility as the "innovation overload perspective."

To assess the validity and relevance of the innovation overload perspective, we turn to innovations proposed in scientific journal articles. Scientific journal articles represent the primary forum through which advances in the sciences are reported and discussed. As we describe in Section 3, the construction of many of our variables is labor-intensive. We therefore cannot consider all articles published in all scientific journals. Instead, we focus our analysis on journal articles published in the field of financial economics (or, simply, finance).[1] We do not believe the observations we make in this paper are particular to finance, however, and we later also provide descriptive statistics for articles published in the fields of economics and other areas of business (accounting, management, marketing, operations and information systems).

We download all papers that were published in the *Journal of Finance*, the *Journal of Financial Economics*, and *The Review of Financial Studies* from 2005 through 2014. As our measure of innovation diffusion we use the number of Google Scholar citations that the relevant article garners as of September 2016.

As an example for a minor inconvenience in understanding an innovation, we consider the readability of the article in which the innovation is couched. In particular, we apply a copy-editing software to each journal article and search for the presence of writing faults that prior literature argues lowers the readability of a text. We then construct our readability measure as the number of writing faults per one hundred words. Writing faults include the use of passive verbs, hidden verbs, complex words, abstract words, overused words and clichés, legal words, wordy phrases, overwriting, foreign words and long sentences. Section 3 describes and presents examples of each writing fault. Section 3 also discusses the relevant literature.

To assess the validity of our measure, we randomly assign finance PhD students the introduction sections of articles that, as per our measure, earn "low readability" scores. We repeat this procedure with introduction sections of articles that earn "high readability" scores. We find that students largely agree with

[1] We choose finance because we, the authors of this study, all work in finance departments and, as such, are familiar with the publication process in the finance area.

the output generated by our readability measure, as they perceive introductions with low readability scores to be significantly more difficult to read than those with high readability scores.

When relating the readability of scientific journal articles to their subsequent number of citations within a regression framework, we find that papers of lower readability receive substantially fewer citations than papers of higher readability. The economic significance of our predictability is substantial: our estimates imply that a one standard deviation increase in the number of writing faults per one hundred words comes with 7% fewer citations. Our results easily survive the inclusion of various controls, such as years-since-publication, author affiliation, number of conference and seminar presentations, whether the paper won a best-paper award, number of authors, whether the paper is a theory paper, title length, subfield of the paper (e.g., "Portfolio choice and investment decisions," "Bankruptcy and litigation") and journal- and Journal-of-Economic-Literature (JEL) fixed effects.

In further tests, we explore whether "more-inconvenient-to-learn" innovations simply take longer to become adopted, but, in the end, achieve the same level of adoption as their "less-inconvenient-to-learn" counterparts, or, whether the number of adoptions is permanently lower. In short, our results suggest that more-inconvenient-to-learn innovations never catch up. In fact, the gap in the subsequent number of citations between high- and low-readability papers widens over time.

Overall, our results are consistent with the notion that at least some potential adopters are attention- and time-constrained. As a result, even minor inconveniences in learning about an innovation hampers an innovation's level of adoption.

An alternative perspective of our results, hereafter referred to as "quality perspective," is that "high-quality innovators" not only make discoveries of greater impact, but they also communicate them more clearly. Relatedly, innovations that are inherently more complex may simply have to be couched in language that is more difficult to process ("complexity perspective").

Despite our controls, we cannot rule out the quality- or the complexity perspective. In general, we would like to emphasize that our evidence is not causal and should thus be interpreted with some caution. At the same time, we note that the correlations between our measure of readability and indicators, which likely capture the quality or complexity of an innovation, such as winning a prestigious research award or being a theory paper, is virtually zero.

In further analyses, we examine whether our readability effect becomes stronger in the second half of our sample period. The annual number of articles published in finance journals is noticeably higher in the second half of our sample period. Consequently, the innovation overload perspective predicts that our readability effect be stronger in the second half of our sample period. The quality- and the complexity perspectives do not share this prediction. Consistent with the innovation overload perspective, we find that the magnitude of our readability effect roughly doubles in the second half of our sample period.

In our final test, we extend our analysis to patents. Patents are described mostly in sketches, but still consider a non-material amount of text. We construct a representative sample of patents and relate the readability of patent descriptions to their number of forward citations within a regression framework. Similar to our findings based on innovations proposed in scientific journal articles, we find that the readability of a patent description strongly predicts a patent's number of forward citations. As with scientific journal articles, this predictability strengthens substantially as the number of patents to be considered each year increases.

Our study speaks to a few of lines of research. Our primary contribution is to the innovation literature. Given the importance of innovations to economic growth, a large literature seeks to explain what factors prevent potentially useful innovations from being adopted. Here, we augment prior empirical work and provide evidence that, even among our set of innovations evaluated by highly-trained experts, something as seemingly frivolous as the readability of an article strongly predicts its level of adoption. One possible view of the evidence is that at least some potential adopters are overwhelmed by the myriad of new ideas, products and technologies proposed to them such that even minor obstacles to understanding an innovation causes an innovation to not be adopted.

Our findings also relate to the emerging literature on "social finance," which tries to understand what factors make some investment ideas spread more easily than others (Hirshleifer 2015; Shiller 2017).[2] Here, we provide evidence that ideas that are even mildly more complex travel substantially less widely than their easier-to-process counterparts.

Finally, by linking the readability of a text to the degree such text reaches its intended audience, our study also adds to the growing textual analysis literature in finance and accounting (e.g., Miller, 2010; Lehavy, Li, and Merkley, 2011; Lawrence, 2013; Loughran and McDonald, 2014; Elliott, Rennekamp and White, 2015).

2. Background, Laboratory, and Hypothesis Development

2.1. Literature on the Diffusion of Innovation

The question of what prevents a seemingly useful innovation from being adopted widely has been examined from two perspectives: (1) a historical, sociological perspective, and (2) an economic perspective. The historical, sociological perspective argues that the level of adoption depends on what an innovation's perceived benefit is, how compatible an innovation is with a potential adopter's social norm, how complex an innovation is, and whether an innovation can easily be tested and evaluated before full adoption (Rogers 1995). As few potential adopters operate in a vacuum, the diffusion of innovation also depends on certain social conditions, such as whether an innovation was first heard about via mass media or word-of-mouth and what the social structure is among potential adopters. Rogers uses the historical, sociological framework to explain why certain innovations, such as the use of contraceptives or the boiling of water prior to consumption, diffused in some groups but not in others.

[2] Our paper also relates to the literature on knowledge dissemination in the social sciences. Ellison (2002b) notes that the mean acceptance time in the "top six" economics journals has slowed considerably between 1970 and 2000. Holden (2017) makes similar observations for finance journals, but also notes that there is substantial cross-sectional variation. Spiegel (2012) and Card and DellaVigna (2013) find that, over time, the average published finance and economics paper has become much longer. Ellison (2002a) and Hirshleifer (2015) develop models in which they try to explain the above noted "complications" in the publication process and they point to a change in reviewer norms and reviewers' effort to maximize editor perception as possible explanations. While the above work assesses primarily what factors hinder an innovation from being made public, our study examines what hinders an idea from being adopted once published. Unlike the above noted features, our source of friction, lack of clear writing, is largely innovator-controlled and can easily be reduced.

In comparison, the economic perspective views the diffusion of innovation as the sum of "individual calculations that weigh the incremental benefits of adopting a new technology against the costs of change" in the presence of uncertainty (Hall 2005). The incremental benefit is, among others, a function of the availability of substitutes. For instance, Hall argues that one reason washing machines diffused slower than the radio is that for washing machines there was a substitute available in the form of handwashing your clothes. The incremental benefit is also a function of the expected network size. For instance, social media platforms such as Facebook only become useful when large crowds use such products; some seemingly useful innovations were unsuccessful because they failed to reach a critical mass (Hall 2005). Regarding the costs of adopting a new technology, the literature emphasizes that most of the cost comes not from acquiring the technology per se, but from the difficulty of learning how to use the new technology and any complementary investment required (Bresnahan, Brynjolfsson and Hitt 2002). A commonly used anecdote to illustrate this point is that personal computers were not adopted widely until Microsoft and Apple transitioned away from operating systems with command-line interface to easier-to-navigate operating systems that have a graphical user interface (Venkatesh and Brown 2001).

Other factors that the economics literature notes affect the level of adoption are how uncertain the benefits and costs of change are, the regulatory environment, the prevailing culture, and the overall market structure (Hall 2005).

2.2. Hypothesis Development and Laboratory

Our hypothesis is most closely related to what the historical, sociological perspective refers to as "complexity" and what the economic perspective refers to as the difficulty of learning how to use a new technology.

Because scientists produce hundreds of journal articles each year, staying current, even in one's "narrow" area of expertise, represents a non-trivial task. Given the limited amount of attention and time the intended audience can spend on each journal article, we speculate that at least some in the audience stop reading about a new idea if the corresponding article is filled with writing faults and, as a result, becomes difficult to process. Any new ideas described in such articles thus do not become adopted and cited widely.

Relatedly, research in psychology finds that low readability weakens readers' trust in the text and causes readers to subconsciously evaluate the source less favorably (McGlone and Tofighbakhsh 2000, Oppenheimer 2006, Alter and Oppenheimer 2008). That is, even if the audience were to learn fully about two innovations and the two innovations were of the same incremental benefit, but one innovation was more difficult to learn about, potential adopters subconsciously develop negative sentiments towards the more-difficult-to-learn innovation and perceive that innovation to be of lower incremental benefit than the innovation that was easier to learn about. Lower perceived incremental benefit, in turn, should translate to lower citation counts.

Hypothesis: *Innovations proposed in scientific journal articles that are filled with writing faults are adopted and cited less widely.*

The null hypothesis, which, ex ante, appears equally plausible to us, is that new ideas proposed in scientific journal articles are evaluated by experts, who are presumably highly trained in processing journal articles. Writing faults, which, at most, cause only minor inconveniences in learning about an innovation are too frivolous to meaningfully affect the degree of innovation diffusion. On this view, we should observe no link between our measure of readability and our measure of innovation diffusion.

3. Data and Key Variables

3.1. Measure of Innovation Diffusion

Our sample comprises 2,618 papers from 2005 through 2014, 716 of which are from the *Journal of Finance*, 1,048 of which are from the *Journal of Financial Economics*, and 854 of which are from *The Review of Financial Studies*. For each of these papers, we assess the degree of innovation diffusion by manually searching the number of citations in Google Scholar (https://scholar.google.com/intl/en/scholar/about.html) via title, authors, and year of publication. We collect the number of citations as of September 16–20 2016, which should ensure that differences in citations do not reflect differences in points of data collection. We only consider articles published from 2005 through 2014 (as opposed to articles published from 2005 through 2016) to give each article some time to diffuse. As shown in Table 1, the average paper in our sample generates 204 citations. The median number of citations is 115; the 10^{th} and 90^{th} percentiles are 25 and 465, respectively.

3.2. Measure of Readability

The development of our readability measure is couched within a growing body of work in accounting and finance that examines how financial market participants respond to disclosure documents when the text is difficult to process. Lawrence (2013) and Elliott, Rennekamp and White (2015) provide evidence that investors shun firms whose disclosure documents are difficult-to-read. Hwang and Kim (2017) go one step further and argue that the associated reduction in investor demand causes such firms to trade at substantial discounts relative to their fundamentals.

Our measure is most similar to that adopted in Hwang and Kim (2017). We save each article as a separate Microsoft Word document. We then use a program called StyleWriter, a manuscript editor that, once installed on a computer, searches Word documents for "writing faults."

The writing faults are: the use of passive verbs, hidden verbs, complex words, abstract words, overused words and clichés, legal words, wordy phrases, overwriting, foreign words and long sentences. Appendix Table 1 provides examples of each writing fault; Appendix Table 1 also provides possible corrections to each writing fault.

Our readability measure, *Readability*, is the number of occurrences of the above writing faults, scaled by the number of words and multiplied by (100) and (-1).

$$Readability = \frac{\sum_{i=1}^{10} WritingFaults_i}{\#Words} \times (100) \times (-1). \tag{1}$$

Multiplying by one hundred later helps us interpret the coefficient estimates. We multiply by negative one so that higher readability scores imply more easily readable documents.

We acknowledge alternate measures of readability, perhaps the two most popular of which are the *Fog Index* and the *Flesch-Kincaid Index*. Both measures are designed to gauge the number of years of formal education needed to comprehend a text on a first reading. The *Fog Index* is 0.4 × (Average Number of Words per Sentence + Fraction of Complex Words × 100). The *Flesch-Kincaid Index* is 0.39 × (Total Number of Words / Total Number of Sentences) + 11.8 × (Total Number of Syllables / Total Number of Words) - 15.59 (Kincaid, Fishburne, Rogers and Chissom, 1975). Other measures of readability include document length and the file size of an electronic document (e.g., Li, 2008; Lawrence, 2013; Loughran and McDonald, 2014).

All of the above measures have their place in the literature. At the same time, we conjecture that using a measure based on actual writing faults that writing classes and textbooks teach us to avoid increases the power of the analysis. Hwang and Kim (2017) provide experimental and regression-based evidence to this regard.

Table 1 shows that there is great variation in our readability measure in our sample. The 10^{th} and 90^{th} percentiles for *Readability* are -7.5 and -4.8; the mean and standard deviation are -6.13 and 1.12, respectively. The mean of -6.13 implies that, on average, there are 6.1 writing faults for every one hundred words. For reference, the *Readability* of this paper is -5.6, which puts this paper in the 75^{th} percentile.

Figure 1 shows the time-series average *Readability* for each of the three finance journals separately. *Readability* is fairly stable through time. On average, the *Journal of Financial Economics* has the highest readability score, followed by the *Journal of Finance* and *The Review of Financial Studies*. Overall, the differences in *Readability* across the finance journals are not economically meaningful, however.

Table 2 extends the across-journal comparison to all fields in business and (general) economics. As (general) economics "A-level" journals, we include the *American Economic Review*, the *Journal of Political Economy* and the *Quarterly Journal of Economics*. Our list of A-level business journals is that of the *Business School Research Rankings*^{TM} compiled by the University of Texas (Dallas) (http://jindal.utdallas.edu/the-utd-top-100-business-school-research-rankings). In addition to our three finance journals, the list includes: *Journal of Accounting and Economics*, *Journal of Accounting Research*, *The Accounting Review*, *Academy of Management Journal*, *Academy of Management Review*, *Administrative Science Quarterly*, *Journal of International Business Studies*, *Organization Science*, *Strategic Management Journal*, *Journal of Consumer Research*, *Journal of Marketing*, *Journal of Marketing Research*, *Marketing Science*, *Information Systems Research*, *MIS Quarterly*, *Journal on Computing*, *Journal of Operations Management*, *Manufacturing and Service Operations Management*, *Operations Research*, *Production and Operations Management* and *Management Science*. Since the computation of *Readability* is labor-intensive, we consider only the 1,752 articles in the above journals that were published in 2014, the last year of our sample period.

Table 2 reveals strong differences across journals and fields. Economics- and accounting journals fare relatively well with *Readability* ranging from -6.02 for the *Journal of Political Economy* to -6.64 for *The Accounting Review*. Marketing journals also fare relatively well with the quantitative journals having slightly fewer writing faults than the more psychology-based journal.

Management journals tend to have lower readability scores. The journals with the lowest readability scores, however, are in the field of Operations and Information Systems. The average article in the *Journal of Operations Management* has 8.71 writing faults per one hundred words and the average article in *Information Systems Research* has 8.33 writing faults per one hundred words.

3.3. Other Article and Author Characteristics

While innovations proposed in finance journal articles represent a relatively homogeneous pool, there remain important differences in quality as well as differences in article and author characteristics that likely affect citation counts outside of the readability channel.

In an attempt to control for these differences, we include the following variables in our regression analysis: number of years since publication as of 2016 (*Years since Publication*), a relative ranking of authors' affiliations based on publication numbers (*Affiliation Ranking*), the number of conferences and seminars the paper has been invited for presentation prior to publication (*Number of Presentations*), whether the article won a "best-paper award" (*Award Paper*), the number of authors (*Number of Authors*), whether the article is a theory paper (*Theory Paper*), the length of the title (*Length of Title*), and the number of JEL codes (*Number of JEL Codes*). In our regression analysis, we also include journal-fixed effects along with fixed effects based on the article's first JEL code.

We include *Years since Publication* since the degree of diffusion is naturally a function of how long the innovation has been around. The *Affiliation Ranking* variable measures the "prestige" of the institutions the authors of the relevant articles are affiliated with. We turn to the *Finance Research Ranking* compiled by the Arizona State University (http://apps.wpcarey.asu.edu/fin-rankings/rankings). The *Finance Research Ranking* counts for each institution the number of publications in the *Journal of Finance*, the *Journal of Financial Economics*, and *The Review of Financial Studies*; it runs from 1990 through the present. To avoid look-ahead bias, we save the Top 50 ranking as of 2004, the year before our sample begins.[3] To facilitate interpretation, the number one institution receives a score of 50, the number two institution receives a score of 49, and so on. Any institution that is not represented in the Top 50 receives a score of zero. A high score therefore indicates that the institution in question has a strong reputation. We compute the average score across the institutions with which the author or authors of the paper in question are primarily affiliated with. Taking the score of the highest ranking institution instead of the average score does not noticeably alter any of our results (results available upon request).

Number of Presentations is the number of conferences and seminars the innovation has been presented at prior to publication. Innovations that have benefitted from comments received during talks are likely of higher quality. Vice versa, high-quality innovations proposed by high-quality researchers are likely to be invited to more presentations. We expect innovations of higher quality to receive higher citation counts. In a tangential, yet related vein, conferences and seminars are primary vehicles through which researchers can draw attention to their work prior to publication, providing further rationale for why *Number of Presentations* may relate positively to citation counts.

[3] The top-ten schools in our *Affiliation Ranking* are (1) NYU, (2) the University of Pennsylvania, (3) the University of Chicago, (4) Harvard University, (5) UCLA, (6) the University of Michigan, (7) Northwestern University, (8) Duke University, (9) Columbia University and (10) MIT. The full top-fifty is available upon request.

Our *Award Paper* variable is constructed as follows: Each year, the *Journal of Finance,* the *Journal of Financial Economics* and *The Review of Financial Studies* award best-paper prizes to the best published articles in that year. Our *Award Paper* variable equals one if the relevant article won such a prize, and zero otherwise. The awards are the Amundi Smith Breeden Prize and the Brattle Group Prize for the *Journal of Finance*; the Jensen Prize and the Fama-DFA Prize for the *Journal of Financial Economics*; and the Michael J. Brennan Best Paper Award for *The Review of Financial Studies*. These awards represent some of the most prestigious prizes in the field of finance and should correlate strongly and positively with the quality of an innovation and its citation count.

We expect *Number of Authors* to positively associate with citation counts as having multiple authors may improve the quality of an innovation. *Number of Authors* may also capture network effects as co-authors likely help raise awareness of an innovation through their own personal networks.

We further differentiate between theory papers and empirical papers as theory papers, by their very nature, may have lower citation counts. *Theory Paper* equals one if the corresponding text contains the term "proof" along with either one of the following terms: "proposition," "theorem," "lemma," "corollary".

Finally, we include *Length of Title* and *Number of JEL Codes*. *Length of Title* is the length of the title in words. We speculate that innovations presented with short titles come with higher citation counts as such innovations may be broader (e.g., "*The Cross-Section of Expected Stock Returns*" versus "*The Effect of Introducing a Non-Redundant Derivative on the Volatility of Stock-Market Returns When Agents Differ in Risk Aversion*"). Shorter titles may also be more attention-grabbing. For instance, when conducting a randomized field experiment on *Seeking Alpha*, a leading investments-related social media platform, Umar (2016) finds that articles posted on *Seeking Alpha* with twice the title length receive 30% fewer page views. In an online survey, Umar also finds that participants perceive the news that "Apple *Inc.* Sees iPhone Sales Slump" to be more interesting than the news that "Apple *Incorporated* Sees iPhone Sales Slump".

The *Number of JEL Codes* variable counts the number of JEL codes for the relevant article. For papers published in the *Journal of Finance*, which do not have JEL codes, the *Number of JEL Codes* variable is set at zero. JEL codes represent subfields in economics.[4] For instance, a paper with JEL code "G11" is a paper on "Portfolio Choice and Investment Decisions"; a paper with JEL code "G33" is a paper on "Bankruptcy and Litigation." Most articles have multiple JEL codes and we conjecture that the more JEL codes an article lists, the broader the innovation proposed in such article and, consequently, the higher the citation count.

Table 3 presents a correlation matrix. Table 3 shows that, as expected, *Affiliation Ranking*, *Number of Presentations* and *Award Paper* all significantly positively correlate with *Citations*. *Theory Paper* significantly negatively correlates with *Citations*. Importantly, none of these measures reliably associate with *Readability*. That is, there are no systematic differences in the occurrence of writing faults between "high-affiliation-rank" papers and "less-high-affiliation-rank" papers, between papers that have been presented many times prior to publication and papers that have been presented fewer times, between award-winning papers and non-award winning papers, and between theory papers and non-theory papers. This finding does not lend support to the notion that innovations of greater quality or greater inherent complexity systematically come with fewer or more writing faults.

4. Experimental Evidence on the Validity of our Readability Measure

Before estimating the impact of readability, we pause briefly to assess the validity of our readability measure in an experimental setting. We randomly select twenty papers that are in the top *Readability* quartile ("high readability papers") and twenty papers that are in the bottom quartile ("low readability papers"). We then assign these papers to finance PhD students and ask them to rate the readability of the introduction section. We focus on the introduction section, as reading the entire paper would require too much time for the PhD students. Moreover, the readability of the introduction section and the readability of the full paper are highly positively correlated. In our random subsample, the average *Readability* of the

[4] The complete list of JEL codes can be found here: https://www.aeaweb.org/econlit/jelCodes.php.

introductions of high readability papers is -5.68; the average *Readability* of the introductions of low readability papers is -8.71; the difference is -3.04 (*t*-statistic = -9.01).

To ensure that our sample of papers represents all areas of finance, we adopt the following procedure: Of the twenty randomly chosen high-readability papers (low readability papers), five are from the pool of papers that are in the area of financial markets (JEL codes: G10-G19) and are purely empirical, five are from the pool of papers that are in the area of financial markets and contain a theoretical model, five are from the pool of papers that are in the areas of Financial Institutions & Services and Corporate Finance & Governance (JEL codes: G20-G39) and are purely empirical, and five are from the pool of papers that are in the areas of Financial Institutions & Services and Corporate Finance & Governance and contain a theoretical model

Our subject pool consists of twenty-one finance PhD students from the following schools: Cornell University, Emory University, Indiana University, University of Southern California, University of Washington, and Yale University.[5] Each of the forty introductions is read by three finance PhD students. We ask the following question: "*How easy to read was the introduction? The scales are 7 ("Very easy") to 1 ("Not at all easy").*"

Appendix Table 2 reports the average response for the group of twenty high-readability papers and the group of twenty low-readability papers. Because each paper is read by three students, each of the two cells contains sixty observations. Papers that are in the top quartile based on *Readability* receive an average score of 5.38. In comparison, papers that are in the bottom quartile receive an average score of 4.70. The difference is +0.68 (*t*-statistic = 2.70). Since students generally avoided the extremes and mostly assigned scores of four, five or six, the difference of +0.68 is economically meaningful. The relatively strong agreement of survey participants with the outputs generated by our readability measure helps build confidence in the validity of our measure.

[5] We are unable to match papers with survey participants based on area of expertise. Our survey participants report that 36.67% of the papers that they were assigned to read are in their area of expertise. There is no reliable difference in this fraction between the twenty high readability papers and the twenty low readability papers.

5. Readability and the Diffusion of Innovation

To quantify the effect of readability on innovation diffusion, we first estimate the following regression equation:

$$Y_i = \alpha + \beta\, Readability_i + \delta' X_i + \varepsilon_i. \quad (2)$$

where *i* indexes an article. The dependent variable is the natural logarithm of the number of citations. We take the natural logarithm since *Citations* is highly right-skewed. *X* includes our explanatory variables described above. As alluded to earlier, we also include journal fixed effects and fixed effects based on the paper's first JEL code. *T*-statistics are based on standard errors adjusted for heteroscedasticity and clustered by the first JEL code.

We present our regression results in Table 4. Depending on the set of controls, the coefficient estimate for *Readability* reported in Columns 1 through 3 ranges from 0.063 (*t*-statistic = 3.83) to 0.084 (*t*-statistic = 5.35). Our regression analysis thus indicates that, holding all else equal, one less writing fault per one hundred words is followed by ~7% more citations. This result is consistent with the notion that making it easier to learn about an innovation facilitates an innovation from being adopted, or, put in analogue form, making it more difficult to learn about an innovation hinders an innovation's level of diffusion.

One of the most established patterns in the innovation diffusion literature is that when the number of adoptions is plotted against the number of years the corresponding innovations has been present, the plot follows an S-shape (Hall 2005). That is, the rise in the number of adoptions proceeds slowly at first, accelerates as it spreads throughout the population, and then slows down again as the innovation reaches its saturation point. To test for the presence of such an S-shaped pattern in our setting, we include a squared term of *Years since Publication*. As reported in Column 3 of Table 4, the coefficient estimate for squared *Years since Publication* is negative and statistically significant suggesting that a plot of number of citations against number of years since publication, indeed, follows an S-shaped pattern.

In further tests, we assess whether readability affects the *rate* or the *degree* of innovation diffusion. That is, do more-inconvenient-to-learn innovation simply take longer to become adopted, but, in the end, achieve the same level of adoption as their less-inconvenient-to-learn counterparts, or, is the number of adoptions permanently lower for the more-inconvenient-to-learn innovations?

To speak to this question, we include an interaction term between *Readability* and *Years since Publication*. If the effect of readability on innovation diffusion is temporary and does not influence the long-term number of adoptions, the negative impact of *Readability* on the number of citations should weaken over time. In other words, as values for *Years since Publication* become larger, the magnitude of the estimated coefficient for *Readability* should become smaller. The coefficient estimate for the interaction term should therefore be negative and statistically significant.

Our results reported in Column 4 of Table 4 show that the coefficient estimate for the interaction term is not reliably negative, implying that the negative effect of low readability is permanent. In fact, since the impact of *Readability* on the percentage change in the number of citations is similar across *Years since Publication*, we can infer that the difference in raw citation numbers between high- and low-readability papers widens over time due to compounding.[6]

To further illustrate this point, Figure 2 plots the predicted number of citations of high-readability papers versus low-readability papers over time. High- and low readability papers represent papers that are above the 90th and below the 10th percentile in terms of their *Readability* score, respectively. The predicted number of citations is based on the coefficient estimates reported in Column 4 of Table 4.[7] Figure 2 shows that papers with high readability scores receive substantially more citations than papers with low readability scores and that the gap between the two groups widens over time.

In Columns 5 and 6 of Table 4, we explore whether our predictability is stronger in the second half of our sample period, i.e., from 2010 through 2014, than in the first half of our sample period, i.e., from 2005 through 2009. The annual number of articles published in our three finance journals is ~20% higher

[6] Put differently, let X be the citation number of articles with high readability and Y be the citation number of articles with low readability. Given the insignificant estimate for the interaction term, we can infer that log (X) – log(Y) = constant (>0) over time, which implies that (X)/(Y) = constant (>1) over time. This, in turn, implies that (X) – (Y) increase over time as X and Y rise, in line with what is depicted in Figure 2.

[7] The predicted number of citations at *Years since Publication* is equal to exp (5.307 + 0.133 × *Readability* at 10th or 90th percentile + 0.276 × *Years since Publication* – 0.01 × *Sqr.Years since Publication* + $\Gamma' X$), where X is a vector of the other control variables at their median values and Γ is a vector of the coefficient estimates for those controls.

in the second half of our sample period. To potential adopters, it has thus become more challenging to carefully consider each article. In line with this argument, we find that the estimate for *Readability* increases from 0.053 (*t*-statistic = 2.88) in the first half to 0.096 (*t*-statistic = 4.55) in the second half. We do not think that either of the alternative perspectives, i.e., the quality perspective and the complexity perspective, can account for the observed strengthening of the predictability.

We conclude this subsection with a few notes on the coefficient estimates for the control variables. The estimates are similar across our six columns. Here, we discuss the estimates reported in Column 3.

The number of authors on a paper strongly positively associates with the number of citations (coefficient estimate = 0.115, *t*-statistic = 5.98). As alluded to earlier, one possible explanation for this finding is that co-authors help improve the quality of the work. An alternative perspective is that co-authors help raise awareness of a paper through their own personal networks (Kerr, 2008).

The coefficient estimate for *Affiliation Ranking* is 0.009 (*t*-statistic = 8.55), suggesting that a ten-rank difference in the average author's *Affiliation Ranking* comes with 9% higher subsequent citation counts. The estimate for *Number of Presentations* is 0.024 (*t*-statistic = 7.58), suggesting that presenting the paper one more time at a conference or a university prior to publication increases subsequent citation counts by 2.4%.

Not surprisingly, receiving an award strongly and positively contributes to citation counts. Our estimate implies that, holding all else equal, award-winning papers, on average, receive 46.4% more citations than non-award papers.

Theory papers receive 45.4% fewer citations (*t*-statistic = -10.78). One interpretation is that theory papers cater to a smaller audience and thus naturally are adopted less widely. An alternative perspective is that innovations in theory papers are inherently more complex and more complex innovations diffuse less widely.

Interestingly, the length of an article's title negatively associates with the number of citations. This result is consistent with the notion that innovations described with short titles are broader and/or more attention-grabbing.

In the end, all of the coefficient estimates for the control variables agree with expectations, which helps build confidence in the usefulness of our controls and the validity of our overall empirical design.

6. Readability and the Diffusion of Innovation through Patents

Scientific journals represent only one channel through which innovation are disseminated. Patents represent another vehicle and in this section we explore whether the observations made for scientific journal articles carry over to patents.

We crawl all utility-patent-filing documents in html format from the USPTO website (http://patft.uspto.gov/netahtml/PTO/srchnum.htm). We remove all patents granted prior to 1976 because filing documents are available as high-quality text files only starting in 1976 and we require high-quality text files to construct our readability measure. We merge our data with the patents data provided by Kogan, Papanikolaou, Seru, Stoffman (2017), which runs from 1926 through 2010.[8] As the construction of our readability measure is labor-intensive, we randomly select 1% of these patents. In the end, we have 12,851 patents granted between 1976 and 2010 with data available for all our dependent and independent variables.

Our dependent variable is *Patent Citations*, which is the number of forward citations received by a patent as described in other patents' filing documents through 2010. Our independent variables are: *Patent Readability*, which is the number of writing faults in a patent description per 100 words, multiplied by (-1); *Years since Granting*, which is the number of years since a patent has been granted (as of 2010); *Economic Value of Patent*, which is the estimated value of a patent based on the stock market reaction to the corresponding patent's granting, scaled by 100; *Firm-Level Innovation Value*, which is the *Economic Value of Patent* aggregated to the firm-level, over the corresponding firm's book value, scaled by 1,000; and *Firm-Level Number of Patents*, which is the number of patents granted to the relevant firm as of 2010, scaled by 100. *Patent Citations*, *Economic Value of Patent* and *Firm-Level Number of Patents* are all from Kogan, Papanikolaou, Seru, Stoffman's (2017) dataset.

[8] https://iu.app.box.com/v/patents

Table 5 presents summary statistics for the above variables. Compared with scientific journal articles, the number of citations for patents is lower (an average of 11.62 compared with an average of 203.56 for scientific journal articles) and the number of writing faults in the patent description is higher (an average of 11.31 per 100 words compared with an average of 6.13 for scientific journal articles).

Our regressions are similar to the ones presented in Table 4 for scientific journal articles:

$$Y_i = \alpha + \beta\, Readability_i + \delta' X_i + \varepsilon_i. \qquad (3)$$

where *i* indexes a patent. The dependent variable is the natural logarithm of the number of forward patent citations. We take the natural logarithm since *Patent Citations* is highly right-skewed. *X* includes *Years since Granting*, the square of *Years since Granting*, *Economic Value of Patent*, *Firm-Level Innovation Value*, and *Firm-Level Number of Patents*. We include USPTO three-digit technology class fixed effects to account for potentially confounding effects that technology-field specific characteristics have on patent citations. *T*-statistics are based on standard errors adjusted for heteroscedasticity and clustered at the year- and technology class-level.

Our regression results presented in Table 6 suggest that patents filed with more readable descriptions receive more citations than similarly valued counterparts with less readable descriptions. Depending on the set of controls, the coefficient estimate for *Readability* reported in Columns 1 through 3 ranges from 0.011 (*t*-statistic = 1.94) to 0.034 (*t*-statistic = 4.87). Our regression analysis thus indicates that, holding all else equal, one less writing fault per one hundred words is followed by 1% to 3% more citations.

To examine whether the impact of readability on patent citation numbers has increased over time, we split our sample period in half and estimate our regression with the full set of controls separately for each subsample. The results presented in Columns 4 and 5 show that our previously found predictability is coming entirely from the second half of our sample period during which the total number of patents granted per year is 169,210 on average (compared with 80,914 in the first half of our sample period).

In sum, our results based on scientific journal articles carry over to patents and are consistent with our proposition that even minor inconveniences in learning about an innovation can hinder the diffusion of an innovation.

7. Conclusion

Our study notes that on the "supply side," there is considerable variation in readability. Some papers and patent descriptions read rather well; others suffer from numerous writing faults. Some scientists and lawyers may be overconfident and erroneously believe their writing to be superb. Others may not care about their writing. Still others may take pride in their ability to construct complex phrases and use terms such as "inter alia" and "lacuna."

On the "demand side," we propose that at least some potential adopters are feeling increasingly overwhelmed by the myriad of new ideas, products and technologies proposed to them such that even minor obstacles to understanding an innovation causes an innovation to not be adopted.

If it is true that at least some individuals and institutions are beginning to "feel some innovation overload," we face a possibly critical conundrum: The more we innovate, the harder it becomes to evaluate and fully absorb each innovation, thereby hampering any positive effect coming out of greater innovative activities. Our evidence, while not causal, corroborates the innovation overload perspective and also suggests that the economic magnitude of the effect is substantial. Given the importance of innovations to economic growth, future research may consider subjecting the innovation overload perspective to additional tests and explore possible solutions for how to overcome problems arising from such overload.

References

Alter, A.L., Oppenheimer, D.M., 2008. Easy on the mind, easy on the wallet: The effects of familiarity and fluency on currency valuation. Psychonomic Bulletin and Review 15, 985-990.

Bresnahan, T.F., Brynjolfsson, E., Hitt, L.M., 2002. Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. Quarterly Journal of Economic 117(1), 339–376.

Card, D., DellaVigna, S., 2013. Nine facts about top journals in economics. Journal of Economic Literature 51(1), 144–161.

Caselli, F., Coleman, W.J. II, 2001, The U.S. structural transformation and regional convergence: A reinterpretation, Journal of Political Economy 109(3), 584-616.

Cutts, M., 2013. Oxford guide to Plain English. Oxford University Press (4th edition).

Ederer, F., and G. Manso, 2011. Incentives for innovation: Bankruptcy, corporate governance, and compensation systems. In R. E. Litan ed.: Handbook of law, innovation, and growth (Cheltenham: Edward Elgar Publishing).

Elliot, W.B., K. Rennekamp, and B.J. White, 2015, Does concrete language in disclosures increase willingness to invest? Review of Accounting Studies 20(1), 839-865

Ellison, G., 2002a. Evolving standards for academic publishing. Journal of Political Economy 110(5), 994–1034.

Ellison, G., 2002b. The slowdown of the economics publishing process. Journal of Political Economy 110(5), 947–993.

Hall, B., 2005. Innovation and Diffusion. In J. Fagerberg, D. Mowery, R. Nelson eds: The Oxford handbook of innovation (Oxford and New York: Oxford University Press).

Hall, B., and J. Lerner, 2010, The financing of R&D and innovation. In B. Hall, and N. Rosenberg eds: Handbook of the economics of innovation (North Holland: Elsevier).

He, J., and Tian, X., 2018. Finance and Corporate Innovation: A Survey. Asia-Pacific Journal of Financial Studies 47, 165–212.

Hirshleifer, D., 2015. Editorial: Cosmetic surgery in the academic review process. Review of Financial Studies 28, 637–649.

Holden, C., 2017. Do acceptance and publication times differ across finance journals? Review of Corporate Finance Studies 6(1), 102–126.

Hwang, B.H., and Kim, H., 2017. It pays to write well. Journal of Financial Economics 124(2), 373–394.

Kerr, W., 2008. Ethnic scientific communities and international technology diffusion. Review of Economics and Statistics 90(3), 518-537.

Kerr, W., and R. Nanda, 2015. Financing innovation, Annual Review of Financial Economics 7, 445–462.

Kincaid, J.P., Fishburne, R.P., Rogers, R.L., Chissom, B.S., 1975. Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel. Research Branch Report *8-75,* Chief of Naval Technical Training.

Kogan, L. Papanikolaou, D., Seru, A., Stoffman, N., 2017. Technological innovation, resource allocation, and growth, Quarterly Journal of Economics 132(2), 665–712.

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

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

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., 2014. Measuring readability in financial disclosures. Journal of Finance 69(4), 1643–1671.

McGlone, M.S., Tofighbakhsh, J., 2000. Birds of a feather flock conjointly: Rhyme as reason in aphorisms. Psychological Science 11(5), 424–428.

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

OECD, 2015. The innovation imperative: Contributing to productivity, growth and well-Being. OECD Publishing, Paris, https://doi.org/10.1787/9789264239814-en.

Oppenheimer, D.M., 2006. Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. Applied Cognitive Psychology 20(2), 139-156.

Rogers, E., 1995. Diffusion of Innovations. New York, Free Press of Glencoe (4th edition).

Rosenberg, N., 2004, Innovation and economic growth, OECD 1–6.

Securities and Exchange Commission (SEC), 1998. A Plain English handbook: How to create clear SEC disclosure documents. Washington DC, USA.

Shiller, R.J, 2017. Narrative economics. American Economic Review 107(4), 967-1004.

Spiegel, M., 2012. Reviewing less–progressing more. Review of Financial Studies 25, 1331–1338.

Umar, T. 2016. Attention grabbers when Seeking Alpha. University of Chicago Working Paper.

Venkatesh, V., and Brown, S A., 2001. A longitudinal investigation of personal computers in homes: Adoption determinants and emerging challenges. MIS Quarterly 25(1), 71–98.

Appendix Table 1
List of Writing Faults, Examples and Possible Corrections

(1) Writing fault	(2) Example	(3) Example how to avoid the corresponding writing fault
Passive verbs	*We must re-think how our resources will be best used to provide world-class customer service*.	*We must re-think how to best use our resources to provide world-class customer service*.
Hidden verbs	*. . . to make an application for employment*.	*. . . to apply for employment*.
Complex words	*While third parties sometimes endeavor to ameliorate relationships . . .*	*While third parties sometimes try to improve relationships . . .*
Abstract words	*We need to install more output devices*.	By avoiding abstract words, writers can clarify the message they are trying to convey: *We need to install more printers*.
Overused words and Cliches	*The patient was then informed about the parameters of treatment available . . . we must more carefully study the parameters of our health care system*. and *open a can of worms, we beg to differ, wakeup call*	The former are popular terms used in a variety of settings; they can essentially mean anything a writer wants them to mean; the latter are phrases that have become devalued through overuse: *The patient was then informed about the types of treatment available . . . we must more carefully study the limitations of our health care system*.
Legal words	*forthwith*	*immediately*
Wordy phrases	*an appreciable number of, has a requirement for*	*many, requires*
Overwriting	*It is completely unnecessary.*	*It is unnecessary.*
Foreign words	*The results show a high urban crime rate, inter alia . . . our paper helps fill a lacuna in the literature . . .*	*The results show a high urban crime rate, among others . . . our paper helps fill a gap in the literature . .*
Long sentences	Regarding "long sentences," there is no objective criterion as to what constitutes a long sentence. Cutts (2013) in the Oxford Guide to Plain English, for instance, recommends an average sentence length of 15–20 words. In our study, we follow our software's definition of a long sentence, which is a sentence with more than 35 words.	

Appendix Table 2
Experimental Evidence on of the Validity and the Effectiveness of our Readability Measure

This table presents survey responses from Finance PhD students that are pertinent to the readability of scientific journal articles. We conduct the following experiment: We sort introduction sections of papers based on *Readability*. We randomly select twenty papers from the top quartile ("High Readability") and twenty papers from the bottom quartile ("Low Readability"). We assign these introductions to twenty-one PhD students and ask: "How easy to read was the introduction?" The scales range from 7 ("Very easy") to 1 ("Not at all easy")." Each introduction is read by three students, yielding a total of sixty observations in each of the two cells. We report the average score given by the students for the "High Readability" articles and the "Low Readability" articles. *T*-statistics, reported in parentheses, account for heteroscedasticity. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively.

	(1) High Readability Papers	(2) Low Readability Papers	(3) Δ High- and Low Readability Papers
"How easy to read was the introduction?" Scale: 7 ("Very easy") to 1 ("Not at all easy")	5.38	4.70	0.68*** (2.70)

Figure 1
Readability of Scientific Journal Articles in Finance over Time

This figure plots our measure of readability, *Readability*, across 2,618 scientific journal articles published in the *Journal of Finance*, the *Journal of Financial Economics*, and *The Review of Financial Studies* from 2005 through 2014. *Readability* is the number of writing faults per one hundred words, multiplied by negative one. The lines represent the average *Readability* by journal and year.

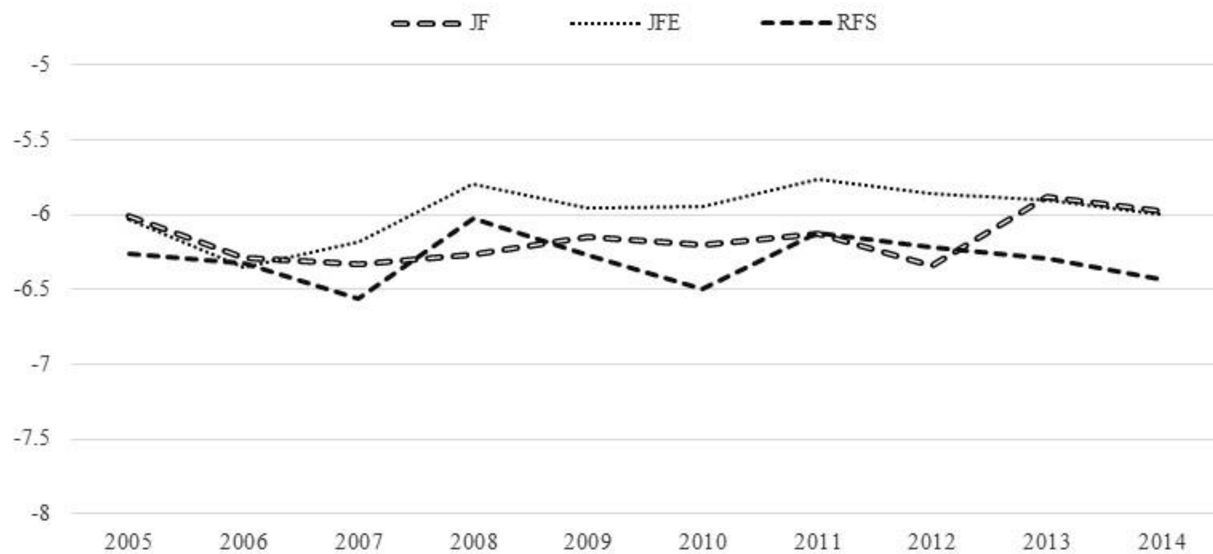

Figure 2
Readability and Predicted Number of Citations over Time

This figure plots the predicted number of citations for papers with high- and low readability since publication. The horizontal axis represents the years since publication and the vertical axis represents the predicted number of citations over years. High- and low readability papers represent papers above the 90th and below the 10th percentile in terms of their *Readability* score, respectively. The predicted number of citations is based on the regression coefficients reported in Colum 4 of Table 4. Predicted number of citations at *Years since Publication* = exp (5.307+0.133×*Readability* at 10th or 90th percentile +0.276×*Years since Publication*-0.01×*Sqr.Years since Publication*+$\Gamma' X$), where X is a vector of other control variables at their median values and Γ is a vector of estimated coefficients on the other controls. The sample includes 2,618 scientific journal articles published in the *Journal of Finance*, the *Journal of Financial Economics*, and *The Review of Financial Studies* from 2005 through 2014.

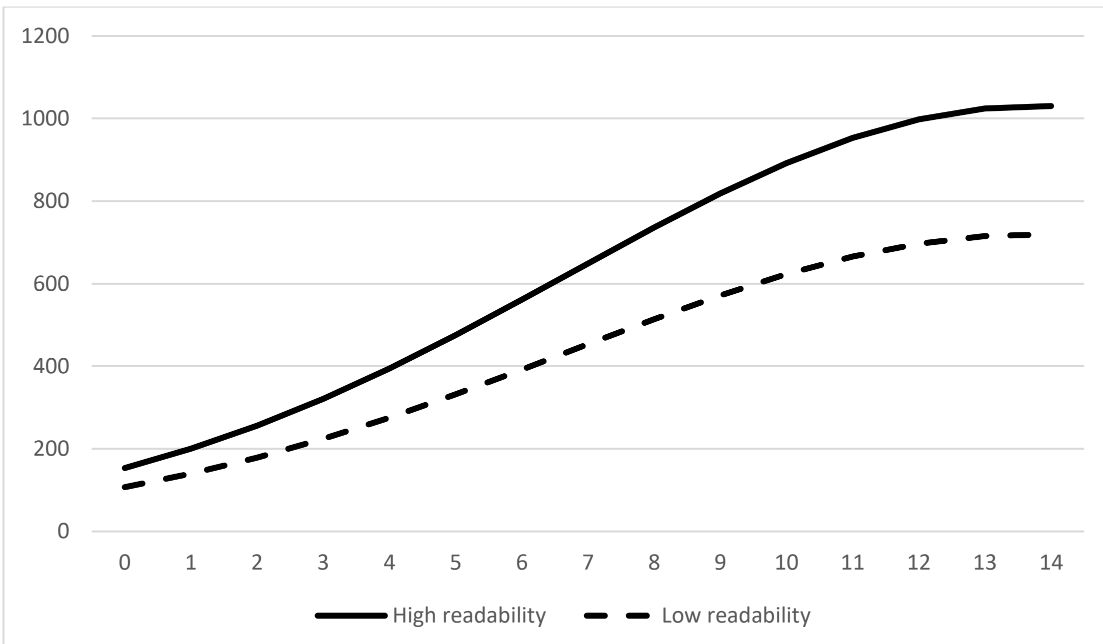

Table 1
Descriptive Statistics for Scientific Journal Articles Sample

This table presents summary statistics for our main variables in the scientific journal articles sample. The sample includes 2,618 scientific journal articles published in the *Journal of Finance*, the *Journal of Financial Economics*, and *The Review of Financial Studies* from 2005 through 2014. The *Journal of Finance* has 716 papers, the *Journal of Financial Economics* has 1,048 papers, and *The Review of Financial Studies* has 854 papers. *Citations* is the Google citations the paper receives as of September 16–20, 2016. *Readability* is the number of writing faults per one hundred words multiplied by negative one. *Years since Publication* is the number of years since publication (as of year 2016). *Number of Authors* is the number of authors listed in the paper. *Affiliation Ranking* is the average "ranking score" across the schools the authors are primarily affiliated with as detailed in Section 3.2. *Number of Presentations* is the number of conferences and seminars the paper was presented at prior to publication. Each year, the *Journal of Finance*, the *Journal of Financial Economics* and *The Review of Financial Studies* award best-paper prizes to the best published papers in that year. *Award Paper* equals one if the paper won such a prize. *Theory Paper* equals one if a paper contains the term "proof" along with either one of the following terms: "proposition," "theorem," "lemma," "corollary". *Length of Title* is the number of words in the title. *Number of JEL Codes* is the number of JEL codes listed in the paper.

	N	Mean	StDev	10^{th} Percentile	Median	90^{th} Percentile
Citations	2,618	203.56	293.14	25.00	115.00	465.00
Readability	2,618	-6.13	1.12	-7.50	-6.10	-4.80
Years since Publication	2,618	6.19	2.83	2.00	6.00	10.00
Number of Authors	2,618	2.39	0.87	1.00	2.00	3.00
Affiliation Ranking	2,618	18.00	16.40	0.00	15.67	44.50
Number of Presentations	2,618	8.68	6.50	1.00	8.00	17.00
Award Paper	2,618	0.05	0.21	0.00	0.00	0.00
Theory Paper	2,618	0.14	0.35	0.00	0.00	1.00
Length of Title	2,618	8.80	3.48	5.00	8.00	13.00
Number of JEL Codes	2,618	2.28	1.84	0.00	2.00	5.00

Table 2
Readability of Scientific Journal Articles by Business and (General) Economics Journals

This table reports the average readability score, *Readability*, across all scientific journal articles published in the relevant journal in 2014.

Readability Ranking	(1) Journal	(2) Field(s)	(3) *Readability*
1	*Journal of Finance*	Finance	-5.98
2	*Journal of Financial Economics*	Finance	-6.00
3	*Journal of Political Economy*	Economics (General)	-6.02
4	*American Economic Review*	Economics (General)	-6.02
5	*Quarterly Journal of Economics*	Economics (General)	-6.12
6	*Journal of Accounting Research*	Accounting	-6.21
7	*Journal of Accounting and Economics*	Accounting	-6.40
8	*The Review of Financial Studies*	Finance	-6.43
9	*Journal of Marketing*	Marketing	-6.55
10	*Journal of Marketing Research*	Marketing	-6.56
11	*The Accounting Review*	Accounting	-6.64
12	*Management Science*	Accounting, Finance, Management, Marketing, Operations and Information	-6.65
13	*Journal on Computing*	Operations and Information	-6.68
14	*Production and Operations Management*	Operations and Information	-6.82
15	*Operations Research*	Operations and Information	-6.92
16	*Marketing Science*	Marketing	-6.95
17	*Strategic Management Journal*	Management	-6.95
18	*Journal of Consumer Research*	Marketing	-7.19
19	*MIS Quarterly*	Operations and Information	-7.20
20	*Administrative Science Quarterly*	Management	-7.22
21	*Manufacturing and Service Operations*	Operations and Information	-7.25
22	*Academy of Management Journal*	Management	-7.39
23	*Journal of International Business Studies*	Management	-7.44
24	*Academy of Management Review*	Management	-7.70
25	*Organization Science*	Management	-7.79
26	*Information Systems Research*	Operations and Information	-8.33
27	*Journal of Operations Management*	Operations and Information	-8.71

Table 3
Correlation Matrix for Scientific Journal Articles Sample

This table presents Pearson correlation coefficients across our main independent and dependent variables. The sample includes 2,618 scientific journal articles published in the *Journal of Finance*, the *Journal of Financial Economics* and *The Review of Financial Studies* from 2005 through 2014. Correlations that are significant at the 5% level are in bold.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) *Citations*	1.000									
(2) *Readability*	**0.056**	1.000								
(3) *Years since Publication*	**0.353**	**-0.045**	1.000							
(4) *Number of Authors*	0.019	**-0.041**	**-0.047**	1.000						
(5) *Affiliation Ranking*	**0.151**	0.031	**0.039**	**-0.100**	1.000					
(6) *Number of Presentations*	**0.059**	0.004	**-0.177**	**0.080**	**0.132**	1.000				
(7) *Award Paper*	**0.149**	0.039	0.023	-0.023	**0.135**	**0.072**	1.000			
(8) *Theory Paper*	**-0.083**	-0.034	0.035	**-0.114**	**0.062**	**0.086**	-0.015	1.000		
(9) *Length of Title*	-0.024	**-0.070**	**0.127**	**0.062**	**-0.049**	**-0.083**	**-0.050**	**-0.141**	1.000	
(10) *Number of JEL Codes*	**-0.071**	0.001	-0.007	**0.099**	-0.023	-0.036	**-0.079**	**-0.044**	**0.150**	1.000

Table 4
Readability and Innovation Diffusion: Evidence based on Scientific Journal Articles

This table presents coefficient estimates from regressions of the natural logarithm of *Citations* on various article characteristics. The sample includes 2,618 scientific journal articles published in the *Journal of Finance*, the *Journal of Financial Economics*, and *The Review of Financial Studies* from 2005 through 2014. All variables are as described in Table 1. *T*-statistics are reported in parentheses and are based on standard errors adjusted for heteroscedasticity and clustered by year and the first JEL code. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively.

	(1) Published 2005-2014	(2) Published 2005-2014	(3) Published 2005-2014	(4) Published 2005-2014	(5) Published 2005-2009	(6) Published 2010-2014
Readability	0.063***	0.084***	0.072***	0.133***	0.053***	0.096***
	(3.83)	(5.35)	(4.98)	(3.40)	(2.88)	(4.55)
Years since Publication		0.192***	0.336***	0.276***	-0.311	0.498***
		(8.56)	(5.20)	(5.56)	(-1.17)	(3.07)
Sqr. Years since Publication			-0.010**	-0.010***	0.026*	-0.031*
			(-2.53)	(-2.59)	(1.77)	(-1.74)
Readability				-0.101		
× *Years since publication*				(-1.52)		
Number of Authors			0.115***	0.116***	0.094***	0.132***
			(5.98)	(6.02)	(3.34)	(5.74)
Affiliation Ranking			0.009***	0.009***	0.009***	0.008***
			(8.55)	(8.54)	(6.47)	(5.39)
Number of Presentations			0.024***	0.024***	0.031***	0.018***
			(7.58)	(7.71)	(4.78)	(5.94)
Award Paper			0.464***	0.466***	0.482***	0.385*
			(2.65)	(2.62)	(3.17)	(1.75)
Theory Paper			-0.454***	-0.457***	-0.582***	-0.317***
			(-10.78)	(-10.94)	(-10.13)	(-4.59)
Length of Title			-0.025***	-0.025***	-0.026***	-0.023***
			(-5.13)	(-4.96)	(-4.03)	(-2.87)
Number of JEL Codes			0.013	0.014	0.014	0.034
			(0.80)	(0.82)	(0.46)	(1.40)
Constant	5.307***	4.519***	4.731***	5.307***	7.082***	4.040***
	(43.33)	(16.36)	(18.92)	(43.33)	(5.78)	(10.79)
Journal FE	Yes	Yes	Yes	Yes	Yes	Yes
1st JEL Code FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.148	0.339	0.416	0.417	0.293	0.386
Number of observations	2,618	2,618	2,618	2,618	1,206	1,412

Table 5
Descriptive Statistics for Patent Sample

This table presents summary statistics for our main variables in the patent sample. We randomly select 1% from all patents granted between 1976 and 2010 that are in the patents database of Kogan, Papanikolaou, Seru and Stoffman (2017). Our final sample includes 12,851 patents. *Patent Citations* is the number of forward citations received by a patent as described in other patents' filing documents through 2010 (provided by Kogan et al. (2017)). *Patent Readability* is the number of writing faults in a patent description per 100 words multiplied by (-1). *Years since Granting* is the number of years since a patent has been granted (as of 2010). *Economic Value of Patent* is the estimated value of a patent based on the stock market reaction to the corresponding patent's granting (provided by Kogan et al. (2017), scaled by 100 in this paper). *Firm-Level Innovation Value* is the aggregate *Economic Value of Patent* at the firm-level over the corresponding firm's book value (provided by Kogan et al. (2017), scaled by 1,000 in this paper). *Firm-Level Number of Patents* is the number of patents granted to the relevant firm as of 2010 (provided by Kogan et al. (2017), scaled by 100 in this paper).

	N	Mean	StDev	10th Percentile	Median	90th Percentile
Patent Citations	12,851	11.621	23.060	0.000	5.000	28.000
Patent Readability	12,851	-11.308	2.314	-14.400	-11.100	-8.500
Years since Granting	12,851	12.191	9.466	1.000	10.000	27.000
Economic Value of Patent	12,851	0.122	0.358	0.001	0.037	0.263
Firm-level Innovation Value	12,851	3.887	9.112	0.013	0.518	11.503
Firm-level Number of Patents	12,851	7.049	9.450	0.170	3.010	19.850

Table 6

Readability and Innovation Diffusion: Evidence based on Patents

This table presents coefficient estimates from regressions of the natural logarithm of *Patent Citations* on various patent characteristics. We randomly select 1% from all patents granted between 1976 and 2010 that are in the patents database of Kogan, Papanikolaou, Seru and Stoffman (2017). Our final sample includes 12,851 patents. All variables are as described in Table 5. *T*-statistics are reported in parentheses and are based on standard errors adjusted for heteroscedasticity and clustered by year and the USPTO three-digit technology class. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively.

	(1) Granted 1976-2010	(2) Granted 1976-2010	(3) Granted 1976-2010	(4) Granted 1976-1993	(5) Granted 1994-2010
Patent Readability	0.034***	0.016***	0.011*	-0.002	0.017***
	(4.87)	(2.83)	(1.94)	(-0.17)	(3.21)
Years since Granting		0.242***	0.236***	0.016	0.312***
		(19.89)	(20.17)	(0.59)	(17.93)
Sqr. Years since Granting		-0.006***	-0.006***	-0.000	-0.009**
		(-14.13)	(-14.36)	(-0.87)	(-2.03)
Economic Value of Patent			0.082***	0.230*	0.038
			(3.74)	(1.67)	(1.18)
Firm-level Innovation Value			0.006***	0.042***	0.004***
			(3.43)	(2.62)	(2.74)
Firm-level Number of Patents			-0.006***	-0.017*	-0.006***
			(-4.39)	(-1.80)	(-3.85)
USPTO Technology Class FE	Yes	Yes	Yes	Yes	Yes
R^2	0.035	0.399	0.403	0.160	0.445
Number of observations	12,851	12,851	12,851	3,867	8,984