



# MIT Open Access Articles

## *Gender Differences in Scientific Communication and Their Impact on Grant Funding Decisions*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

<b>As Published</b>	10.1257/PANDP.20201043
<b>Publisher</b>	American Economic Association
<b>Version</b>	Final published version
<b>Citable link</b>	<a href="https://hdl.handle.net/1721.1/135234">https://hdl.handle.net/1721.1/135234</a>
<b>Terms of Use</b>	Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

# Gender Differences in Scientific Communication and Their Impact on Grant Funding Decisions<sup>†</sup>

By JULIAN KOLEV, YULY FUENTES-MEDEL, AND FIONA MURRAY\*

Why are women underrepresented in the fields of science, technology, and innovation? Even as women make up almost half of the US labor force and an outright majority of recent college graduates, they remain far from parity in science, technology, engineering, and math (STEM) fields (Delgado and Murray forthcoming). In contrast to broader disparities such as the gender wage gap, for which recent work has documented a range of contributing mechanisms (Goldin 2014), there has been relatively little progress in identifying underlying causes for the lack of gender inclusion in scientific fields. Building on an emerging literature on the contribution of differences in communication to the STEM gender gap (Kanze et al. 2018; Lerchenmueller, Sorenson, and Jena 2019), we implement a range of text analysis methods associated with gendered language in the context of scientific grant applications. We document consistent gender differences in scientific communication across two grant-awarding institutions: female applicants are more likely to use language with high readability, concreteness, and narrow words and are less likely to use positive words and broad words and to concentrate in a single medical subject. We then show that

institutional design is a crucial moderator in determining whether these differences lead to a significant impact on funding outcomes.

## I. Institutional Setting

Our analysis focuses on scientific communication at two major grant-awarding institutions: the Bill and Melinda Gates Foundation (henceforth Gates) and the National Institutes of Health (henceforth NIH). While both institutions share the broad goal of allocating resources to promising research ideas in health and medicine, they differ along a number of important dimensions. Most crucially for our analysis, these institutions have chosen differing approaches to the grant application process. NIH applications can be up to 20 pages long, require detailed budgets and preliminary data analysis, and may also request biographical information and reference letters from the applicant. These applications are then evaluated by a panel of experts, and successful applications for the prominent R01 grant receive, on average, over \$400,000 per year for four or five years.

By contrast, Gates applications are only two pages in length and do not require a budget, preliminary data analysis, or biographical information. Instead, Grand Challenges and Explorations (GCE) applications are reviewed anonymously and independently by multiple experts, and successful applications receive a single Phase I award of \$100,000, with the opportunity to apply for follow-on Phase II funding if early results are promising.

These differing application processes reflect underlying contrasts in institutional goals: the NIH focuses on in-depth and long-term research projects, while Gates prioritizes shorter-term evaluation of potentially high-impact ideas. This institutional variation allows us to evaluate whether there are persistent gender differences

\* Kolev: Cox School of Business, Southern Methodist University (email: [jkolev@smu.edu](mailto:jkolev@smu.edu)); Fuentes-Medel: Sloan School of Management, Massachusetts Institute of Technology (email: [yfuentes@mit.edu](mailto:yfuentes@mit.edu)); Murray: Sloan School of Management, Massachusetts Institute of Technology (email: [fmurray@mit.edu](mailto:fmurray@mit.edu)). We are grateful to the Bill and Melinda Gates Foundation for providing access to their internal data and procedures and for their overall support. We thank Pierre Azoulay, Scott Stern, Philippe Aghion, Joshua Krieger, Daniel Fehder, Annamaria Conti, Donna Ginther, and Lisa Cook for their helpful comments and suggestions.

<sup>†</sup> Go to <https://doi.org/10.1257/pandp.20201043> to visit the article page for additional materials and author disclosure statement(s).

in language across different application formats and whether these text characteristics influence each institution's ultimate funding decisions.

## II. Data and Methods

To construct our sample of grant applications, we begin with the full set of submissions to the first 17 rounds of the GCE Program at the Gates Foundation, covering 2008–2016 and focusing primarily on global health and infectious disease research. To focus our sample on applicants who are likely to also seek NIH funding, we restrict our sample to the 5,236 Gates applicants affiliated with US-based academic institutions. We then match this sample of applicants to records of successful grant applications from the NIH, for a total of 6,931 Gates grant applications and 12,589 NIH grant applications. We use the abstracts of these applications to generate a range of text analysis measures, as explained in the following section. In addition, we obtain a probabilistic measure of applicant gender by matching first names to records from the US Census Bureau and the Social Security Administration. Finally, we match our sample of applicants to publication records in Scopus and PubMed to obtain covariates including applicant career lengths, journal article publication rates, and publications in the top 10 percent of journals by impact factor.

### A. Text Analysis Measures

The field of text analysis offers a broad and growing set of methods to analyze language and communication. When selecting the metrics presented in our analysis, we prioritized those that have a documented association with gender, an established link to the scientific process, or both. All of our measures fall into the category of “bag-of-words” metrics, meaning that they ignore syntax and word order and account only for the collection of words in a given abstract.

Our first two measures replicate Lerchenmueller, Sorenson, and Jena (2019), with *Positive\_Presentation* capturing whether the abstract contains one or more of a list of 25 words associated with positive descriptions. Their findings highlight the word “novel” as the positive word with the largest difference in usage by gender, so we also track its presence via our

*Novel\_Presentation* measure. Both measures have previously been shown to be lower in the abstracts of clinical research and life sciences journal publications with female authors.

Our second measure, *Flesch\_Reading\_Ease*, is motivated by the findings in Hengel (2020), which shows that female economists tend to use clearer and more readable language in their journal articles. This measure is calculated by using the ratios of words per sentence and syllables per word, with higher scores reflecting clearer writing. Our next measure, *Concreteness*, replicates the methodology in Joshi et al. (forthcoming), which shows that across a variety of (nonacademic) contexts, women tend to use more concrete and less abstract language than men, as captured by the average Brysbaert concreteness index score of their vocabulary.

Our final three measures are meant to capture whether a scientific abstract is narrow or broad, motivated by Kolev, Fuentes-Medel, and Murray (2019), which shows that male scientists tend to use broad language while women are more likely to use narrow language to describe their research. The first of these, *MeSH\_Concentration*, first uses the National Library of Medicine's Medical Text Indexer to identify the top 25 Medical Subject Heading (MeSH) terms relevant to a given grant abstract and then calculates the Herfindahl index of these terms based on the first level (e.g., C14: cardiovascular diseases) of the MeSH tree taxonomy. While this measure focuses on broad versus narrow science, our final two measures focus on broad versus narrow language. These are in-sample metrics that calculate whether the words in an abstract appear primarily within a single grant topic category or across multiple categories, with the division based on the median intertopic variance of word use rates.

### B. Estimated Equations

Our empirical analysis begins with an evaluation of the determinants of abstract text characteristics, where we implement the following ordinary least squares specifications:

$$(1) \text{Text\_Measure}_{ij} = \alpha_0 X_{ij} + \alpha_1 \text{Female}_j + \epsilon_{ij}.$$

In equation (1), the text characteristic of grant application  $i$  from applicant  $j$  is explained by the probability that applicant  $j$  is female and a

TABLE 1—SUMMARY STATISTICS

Variable	Gates sample ( <i>N</i> = 6,931)		NIH sample ( <i>N</i> = 12,589)		<i>t</i> -test for difference in means
	Mean	SD	Mean	SD	
log(word count)	3.291	0.546	5.839	0.469	<i>p</i> < 0.0001
Positive presentation (0/1)	0.208	0.406	0.567	0.495	<i>p</i> < 0.0001
Novel presentation (0/1)	0.177	0.382	0.373	0.484	<i>p</i> < 0.0001
Flesch reading ease	16.874	23.794	13.777	17.137	<i>p</i> < 0.0001
Concreteness	2.497	0.277	2.392	0.096	<i>p</i> < 0.0001
MeSH concentration	0.256	0.108	0.126	0.054	<i>p</i> < 0.0001
Broad word count	6.796	3.455	43.700	14.175	<i>p</i> < 0.0001
Narrow word count	3.923	2.206	35.565	13.293	<i>p</i> < 0.0001
Female applicant	0.333	0.432	0.267	0.407	<i>p</i> < 0.0001

Note: The *t*-tests are based on a two-tailed, two-sample test statistic and assume unequal variance across samples.

set of control variables that include calendar time and application topic fixed effects, controls for total word count and the count of relevant words for dictionary-based metrics, and applicant publication history. Aside from our binary text measures (*Positive\_Presentation* and *Novel\_Presentation*), all dependent variables are standardized to a mean of zero and unit standard deviation. The coefficient of interest is  $\alpha_1$ , which captures the average impact of gender on the measures of scientific language in our sample.

The second portion of our empirical analysis focuses on the outcome of funding. Specifically, we estimate the probability of receiving Phase I funding versus no funding for Gates applications, and we look at the division between receiving a long-term R01 grant versus a smaller short-term R56 grant for the subset of NIH applications seeking R01 funding.<sup>1</sup> Because of significant differences in base rates of funding, we use the following logit specifications to allow for comparability between the Gates and NIH results:

$$(2) \text{Funding}_{ij} = \beta_0 X_{ij} + \beta_1 \text{Text\_Measure}_{ij} + \epsilon_{ij}.$$

In equation (2), the probability of funding for grant application *i* from applicant *j* is explained by the full set of text characteristics discussed earlier in this section and a set of

<sup>1</sup>Importantly, NIH applicants may not directly apply for R56 grants; instead, these grants are awarded to applications whose scores fall just outside the R01 funding cutoff of their respective NIH institute.

control variables. These include calendar time and application topic fixed effects, controls for total word count and the count of relevant words for dictionary-based metrics, and applicant publication history and gender. As in the previous analysis, all nonbinary text measures are standardized to facilitate comparability. The coefficient of interest is the vector  $\beta_1$ , which captures the impact of our full set of text measures on the likelihood of obtaining grant funding.

### III. Results

Our results begin with Table 1, which lists summary statistics for key variables of interest across our Gates and NIH samples. The final column indicates that across all measures, there are significant differences between our two samples. Most prominently, NIH abstracts are significantly longer than those of Gates applications, and this in turn affects many of the text analysis measures that we calculate. This supports our empirical approach of including detailed controls both for overall word count and for counts of words specific to our dictionary-based text measures.

Proceeding to our primary analysis, we report the results of the impact of gender on abstract text characteristics in Figure 1. Estimating effects independently for our Gates and NIH samples, we find that while significance levels may differ, the point estimates of the impact of a female applicant go in the same direction across all measures. We find support for all of the patterns documented in the prior literature:

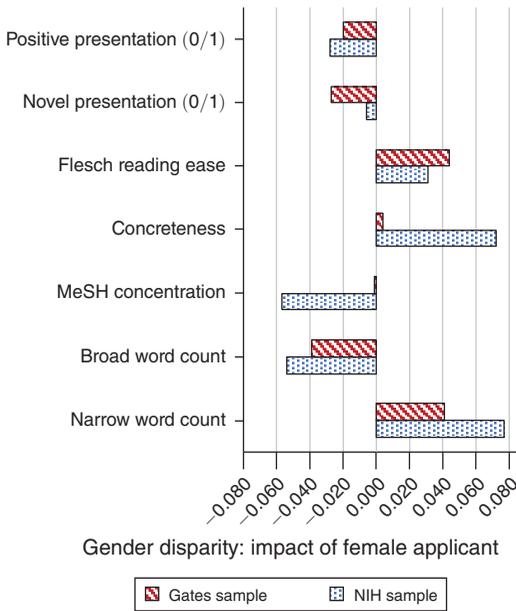


FIGURE 1. GENDER DIFFERENCES IN ABSTRACT TEXT CHARACTERISTICS

Notes: Adjusted gender differences in grant application abstract text characteristics, 2008–2016. The figure shows the impact of a female applicant on text characteristics after controlling for calendar time and application topic fixed effects, word counts, and applicant publication history. The Gates sample is based on 6,931 applications, while the NIH sample is based on 12,589 applications, all from the same population of 5,236 applicants. The first two text characteristics are binary indicators, while the remaining variables have been standardized to a mean of zero and unit standard deviation.

female applicants are less likely to present their research using positive vocabulary, they are more likely to write with high readability, and they prefer concrete language. Moving to our final three measures, we find an interesting dichotomy: even as female applicants use fewer broad words and more narrow words in their abstracts, we find that their research is characterized by lower MeSH concentrations, meaning that they cover a wider range of medical subjects in their work, at least within the NIH sample. Effect sizes are relatively small: the impact of gender ranges from approximately 0.04 to 0.08 standard deviations for our significant effects. At the same time, small differences may well lead to large effects in highly competitive contexts such as the grant applications of our setting.

TABLE 2— IMPACT OF TEXT CHARACTERISTICS ON FUNDING OUTCOMES

Variables	Gates sample (1)	NIH sample (2)
Positive presentation (0/1)	0.028 (0.232)	-0.169 (0.178)
Novel presentation (0/1)	-0.246 (0.249)	0.113 (0.173)
Flesch reading ease	-0.040 (0.052)	0.032 (0.087)
Concreteness	0.097 (0.049)	-0.038 (0.077)
MeSH concentration	0.019 (0.051)	0.256 (0.104)
Broad word count	0.116 (0.099)	0.335 (0.185)
Narrow word count	0.053 (0.074)	0.344 (0.190)
<i>Controls</i>		
Calendar time fixed effects	Yes	Yes
Application topic fixed effects	Yes	Yes
Applicant characteristics	Yes	Yes
Word count controls	Yes	Yes
Observations	6,924	5,131
Pseudo-R <sup>2</sup>	0.0740	0.218

Notes: The table shows the impact of text characteristics on the binary outcome of receiving grant funding. The dependent variable is funding (0/1). All specifications are logit models with robust standard errors clustered by applicant in parentheses. The Gates sample’s funding measure is based on Phase I funding, while the NIH sample’s funding measure is based on the difference between R01 and R56 awards. All specifications include calendar time and application topic fixed effects, word counts, applicant gender, and applicant publication history. The first two text characteristic variables are binary indicators, while the remaining variables have been standardized to a mean of zero and unit standard deviation.

Our final empirical analysis explores the impact of text characteristics on funding outcomes. In Table 2, we report the results of a multiple-regression logit model that simultaneously evaluates the impact of all text measures on grant funding outcomes. We find strong institutional differences, with no overlap in significant determinants of funding outcomes. For Gates applicants, high levels of concreteness tend to improve the odds of funding; by contrast, at the NIH, we find a strong positive impact for MeSH concentration and marginal effects for both broad and narrow words. Effect sizes for our significant

effects are quite large: a one-standard-deviation increase in concreteness leads to a 10 percent increase in the odds of receiving Gates funding, while a comparable increase in MeSH concentration leads to an almost 30 percent increase in funding odds at the NIH. These results suggest that the funding decision process is highly dependent on institutional characteristics. Indeed, our results are consistent with the NIH focus on in-depth development of longer-term projects and the Gates focus on measurable short-term impact. In both institutions, we find that grant application text characteristics can be a significant driver of funding outcomes.

#### IV. Conclusions

Using a sample of Gates and NIH grant applications submitted by a common pool of US-based academic researchers, we find that despite significant institutional variation, there are consistent gender differences in scientific communication across a wide range of abstract text characteristics. By contrast, we find starkly differing impacts of these text characteristics on funding outcomes across our institutional settings. This contrast suggests that rather than seeking to eliminate gender differences in communication, a more viable strategy to increase gender inclusion in science and technology would focus on updating institutional processes so that they are not negatively predisposed toward female scientific language.

#### REFERENCES

- Delgado, Mercedes, and Fiona Murray.** Forthcoming. "Catalysts for Gender Inclusion in Innovation: The Role of Universities and Their Top Inventors." *Science*.
- Goldin, Claudia.** 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104 (4): 1091–119.
- Hengel, Erin.** 2020. "Publishing while Female: Are Women Held to Higher Standards? Evidence from Peer Review." [http://www.erinhengel.com/research/publishing\\_female.pdf](http://www.erinhengel.com/research/publishing_female.pdf).
- Joshi, Priyanka D., Cheryl J. Wakslak, Gil Appel, and Laura Huang.** Forthcoming. "Gender Differences in Communicative Abstraction." *Journal of Personality and Social Psychology*.
- Kanze, Dana, Laura Huang, Mark A. Conley, and E. Tory Higgins.** 2018. "We Ask Men to Win and Women Not to Lose: Closing the Gender Gap in Startup Funding." *Academy of Management Journal* 61 (2): 586–614.
- Kolev, Julian, Yuly Fuentes-Medel, and Fiona Murray.** 2019. "Is Blinded Review Enough? How Gendered Outcomes Arise Even under Anonymous Evaluation." NBER Working Paper 25759.
- Lerchenmueller, Marc. J., Olav Sorenson, and Anupam B. Jena.** 2019. "Gender Differences in How Scientists Present the Importance of Their Research: Observational Study." *BMJ* 367: l6573.