Publishing while female
Are women held to higher standards? Evidence from peer review.*
18 January 2022

Erin Hengel[†]

Female authors are underrepresented in top economics journals. In this paper, I investigate whether higher writing standards contribute to the problem. I find: (i) female-authored papers are 1–6 percent better written than equivalent papers by men; (ii) the gap widens during peer review; (iii) women improve their writing as they publish more papers (but men do not); (iv) female-authored papers take longer under review. Using a subjective expected utility framework, I argue that higher writing standards for women are consistent with these stylised facts. A counterfactual analysis suggests senior female economists may, as a result, write at least five percent more clearly than they otherwise would. As a final exercise, I show tentative evidence that women adapt to biased treatment in ways that may disguise it as voluntary choice.

Keywords. Gender, Discrimination, Inequality, Diversity, Peer Review

JEL codes. D8, J7, J16, A11

1 Introduction

Female academics are less likely to make tenure, take longer when they do and earn much less than their male peers (Bandiera, 2016; Ceci et al., 2014; Ginther and Kahn, 2004; Weisshaar, 2017). In economics, only a quarter to a third of assistant and associate professors—and no more than 15 percent of professors—are women (Bateman et al., 2021; Gamage et al., 2020; Lundberg and Stearns, 2019).[1]

There are a number of factors driving these outcomes. Women have smaller research networks (Ductor et al., 2021), make different career choices and face different constraints (*e.g.*, motherhood). They may also be held to tougher standards. For example, evidence suggests that their qualifications and ability are underestimated (Foschi, 1996; Grunspan et al., 2016; Moss-Racusin et al., 2012; Reuben et al., 2014); female-authored papers are evaluated more critically (Goldberg, 1968; Krawczyk and Smyk, 2016; Paludi and Bauer, 1983); when collaborating with men, women are given less credit (Heilman and Haynes, 2005; Sarsons et al., 2020).

In this paper, I investigate whether top economics journals apply similar standards to men's and women's manuscripts—and specifically, similar writing standards. In the English language, clearly written prose is better prose, all things equal. Thoughtful word choice and simple sentence structure make text easier to understand, more interesting to read and expose inconsistencies long-winded writing often hides. Journal editors tend to agree. *Econometrica* asks authors to write "crisply but clearly" and to take "the extra effort involved in revising and reworking the manuscript until it will be clear to most if not all of our readers" (*Econometrica* submission guidelines, June 2016).

To measure writing clarity, I apply five highly tested "readability" formulas to 9,117 article abstracts published between 1950–2015 in the *American Economic Review* (*AER*), *Econometrica* (*ECA*), *Journal of Political Economy* (*JPE*) and *Quarterly Journal of Economics* (*QJE*). With these data, I document several stylised facts.

First, women aren't published that often in "top-four" economics journals. The average share of female authors per paper across the entire sample is 7.5 percent. In 2015, that share was still only 15 percent; just eight percent of papers were majority female-authored and only four percent were written entirely by women. Between 2015–2017, *QJE* did not publish a single exclusively female-authored paper. In several recent years, *Econometrica* and *JPE* have not either.

Second, the female-authored abstracts that are published in these journals are 1–6 percent more readable than those by men. Women write better despite adjusting for other factors correlated with quality—including citations, author prominence, seniority and individual fixed effects—accounting for English fluency and adding editor, journal, year and primary and tertiary *JEL* category dummies and roughly controlling for how theoretical vs. empirical a paper is.

Third, the gender gap in readability is 2–3 times larger in the published version of a manuscript compared to its pre-submission version. To arrive at these estimates, I match National Bureau of Economic Research (NBER) working papers with their final published articles. Assuming authors release their manuscripts as NBER working papers at about the same time that they submit them to peer review, these results suggest that female-authored abstracts become 2–5 percent more readable while under review.

Fourth, the portion of the gap formed in peer review reversed direction in journals that blinded referees to authors' identities before the internet. Although standard errors are large and sample sizes small, this evidence tentatively suggests that blind review can mitigate the impact of gender under certain circumstances. It also points to the possibility that editorial/refereeing bias at least partially contributes to women's better writing.

Fifth, I do not find evidence that men compensate for their lower quality writing by raising quality on another dimension. More specifically, better writing by female economists could arguably compensate for some other advantage present in men's papers. But as long as men and women are equally capable researchers and similarly informed conditional on controls, the cost to both genders of imple-

---

[1]Auriol et al. (2019) find female economists are slightly better represented at European institutions.

menting their respective publication strategies should be equal—otherwise, women could reduce the cost of producing a paper while holding acceptance rates constant by adopting a strategy marginally closer to men's (or visa versa). A rough test of this hypothesis using submit-accept times from *Econometrica* and the *Review of Economic Studies* (*REStud*) suggests this isn't the case. The cost to men of revising a paper appears to be much lower than the cost to women: female-authored papers spend three to six months longer under review compared to observably equivalent male-authored papers. The effect persists across a range of specifications that account for, among other things, citations, readability, author seniority, motherhood, childbirth and field.

Finally, it does not appear that women are rewarded for their better writing. Recent evidence from a set of comparable journals suggests female-authored papers are not accepted at higher rates after conditioning on similar co-variates (Card et al., 2020).

These stylised facts suggest women spend too much time rewriting old papers and not enough time writing new papers, relative to men. The lack of a gender gap under blind review points to external factors beyond their control, but women's better writing could also be driven by internal factors such as higher risk-aversion (for a review, see Croson and Gneezy, 2009), lower confidence (see, *e.g.*, Coffman, 2014; Exley and Kessler, 2021), a tendency to update too much when faced with negative signals (Möbius et al., 2014), be more easily swayed by the opinions of others (Born et al., 2020) or exert more effort on low stakes tasks (Schlosser et al., 2019) and those which do not yield obvious benefits (Babcock et al., 2017).

To investigate these mechanisms, I model an author's decision-making process as if it were governed by the rational behaviour of women who update their beliefs about the readability thresholds they are held to as they gain experience in peer review. The intuition of the model is as follows. A gender readability gap that decreases with experience suggests that women initially overestimate referees' and editors' writing thresholds but revise their beliefs downward as they submit more papers to peer review. This pattern indicates that internal factors predominantly drive the gap. On the other hand, a gap that increases with experience suggests that women initially underestimate writing thresholds but revise their beliefs upwards as they gain a better understanding of peer review. In this case, tougher standards probably play a role in women's choices unless their extra effort is rewarded with higher acceptance rates relative to men.

The model identifies three testable conditions which can help establish whether external factors are at all important to the existence and evolution of the gender readability gap: (1) experienced women write better than equivalent men; (2) women improve their writing over time; (3) female-authored papers are accepted no more often than equivalent male-authored papers. Evidence from the pooled sample of authors suggests conditions (1) and (2) hold: on average, women's writing gradually gets better but men's does not; between authors' first and third published articles, the readability gap increases by up to 12 percent. Although my data do not identify Condition (3), female-authored papers are accepted less often than equivalent male-authored papers at a similar set of journals (Card et al., 2020).

The validity and accuracy of these results primarily rely on two critical—and strong—assumptions. First, the more experience a women gains in peer review, the fewer mistakes she makes about referees' and editors' standards. Second, male and female authors write papers that are identical with respect to topic, novelty and quality, conditional on controls. The second assumption is especially likely to be violated. Furthermore, concluding that higher standards are present requires that all three conditions hold for the same author—that is, the same woman must write better than an equivalent man, not be accepted at rates higher than he is *and* raise the quality of her writing over time.

To improve my estimates in both respects, I additionally restrict the sample to authors with three or more top-four publications and match observably similar male and female economists based on characteristics—including citations and field—that predict the topic, novelty and quality of their research. Within-person readability comparisons are used to determine if Condition (2) was satisfied for each author in a matched pair. Between-person comparisons after authors gain experience in peer review are used to establish whether Condition (1) was satisfied for the male or female member.

I find that Conditions (1) and (2) were satisfied for the same author in 65 percent of matched pairs; in two-thirds of those, the member who satisfied them was female. A counterfactual analysis suggests that higher standards mean women write, on average, 5 percent more readably than they otherwise would. I emphasise, however, that the reliability of these estimates depends on the extent to which the observables authors are matched on fully capture differences in the non-readability qualities of papers. Within each matched pair, they also require that Condition (3) is satisfied for the same author who satisfied Conditions (1) and (2), something I cannot directly test with my data. Additionally, the validity of the counterfactual analysis and the precision of its estimates rely on strong assumptions about men's and women's beliefs and the impact co-authors have on an article's readability.

I conclude by showing suggestive evidence that women navigate higher standards by altering their behaviour. Guided by the model, I compare papers pre- and post-review as authors gain experience. This allows me to tease out the direct effect of higher standards—readability changes made *in* peer review—from its "feedback" effect—readability changes made *before* peer review in anticipation of those higher standards. I find that the direct effect dominates in authors' earliest papers. In fact, there is no significant gender difference between draft readabilities in men's and women's first top publications; it emerges entirely in peer review. In later papers, however, women write well upfront; the gap chiefly materialises *before* peer review. These results further support the hypothesis that women do not initially expect higher standards and instead learn about them over time. They also suggest that women adapt to higher standards by writing their future papers more readably prior to submission.

This paper contributes to the literature in several ways. First, to the best of my knowledge, I am the first to document empirical evidence that suggests women may be held to higher standards in the peer review process (as opposed to its outcome). Higher standards have recently been established using citations as a proxy for manuscript quality (Card et al., 2020; Grossbard et al., 2021; Hengel and Moon, 2020).[2] They also align with research on employee performance reviews, teaching evaluations and online comments: women receive more abusive feedback, less credit for intelligence and creativity and are expected to be more organised, prepared and clear (see, *e.g.*, Boring, 2017; Correll and Simard, 2016; Gardiner et al., 2016; Mengel et al., 2018; Wu, 2020).

Second, this paper proposes a novel explanation for academia's "Publishing Paradox", "Leaky Pipe-line" and general promotion gap. Higher standards cause collateral damage to women's productivity: spending more time revising old research means there's less time for new research; fewer papers results in fewer promotions, possibly driving women into fairer fields.[3] They may also help explain why so few women publish entirely female-authored papers, despite being the work tenure committees give them the most credit for (Sarsons et al., 2020).

Third, my conclusions relate to a more general debate about gender differences in labour market outcomes.[4] Higher standards impose a quantity vs. quality trade-off that characterises female output in many professions—*e.g.*, doctors, real estate agents and airline pilots (for a discussion, see Hengel, 2017). Their downstream effects may contribute to several employment phenomena, including women's tendencies to concentrate in certain sectors and occupations (Blau and Kahn, 2017; Cortés and Pan, 2016), under-negotiate pay (Babcock and Laschever, 2003) and apply only to jobs they feel fully qualified for (Mohr, 2014). They may also reinforce work habits—*e.g.*, conscientiousness, tenacity

[2]Data from a field journal find female-authored manuscripts are subject to greater scrutiny and spend longer under review (Alexander et al., 2021). A review-time gap was not, however, present in a set of journals that semi-overlap with those analysed here (Card et al., 2020).

[3]See also Bright (2017) for a similar idea in the philosophy literature. This idea has also been informed by extensive research on editorial patterns (Card and DellaVigna, 2013; Card et al., 2020; Casnici et al., 2017; Clain and Leppel, 2018; Ellison, 2002), bias in editorial decisions (Abrevaya and Hamermesh, 2012; Bransch and Kvasnicka, 2017; Card and DellaVigna, 2020; Card et al., 2020) and female academics' lagging productivity and underrepresentation (Bayer and Rouse, 2016; Chari and Goldsmith-Pinkham, 2017; Ductor et al., 2021; Ginther and Kahn, 2004; Teele and Thelen, 2017).

[4]Traditional hypotheses focus on obvious discrimination (Goldin and Rouse, 2000), motherhood (Bertrand et al., 2010) and differences in behaviour (*e.g.*, Niederle and Vesterlund, 2010). Contemporary theories tend to stress inflexible working conditions (Goldin, 2014a; Goldin and Katz, 2016), preferences (for a review, see Blau and Kahn, 2017) and policy design (Antecol et al., 2018).

and diligence—that correlate with quality and connote "femininity": for example, female physicians consult longer with patients (Roter and Hall, 2004), female politicians fundraise more intensely (Jenkins, 2007), female faculty commit fewer instances of academic misconduct (Fang et al., 2013) and female lawyers make fewer ethical violations (Hatamyar and Simmons, 2004).

Fourth, this paper joins an emerging body of economic research studying how the experience and anticipation of discrimination affects choices and behaviour. Earlier theoretical work focused on the impact discrimination has on investment in education and occupational choice (see, *e.g.*, Coate and Loury, 1993; Goldin, 2014b; Lundberg and Startz, 1983). More recent empirical research explores how stereotypes negatively impact performance (Bordalo et al., 2016; Carlana, 2019; Coffman, 2014; Glover et al., 2017; Lavy and Sand, 2018). My results suggest that rational responses to discrimination can distort productivity measurement (see also Parsons et al., 2011) and blur the line between biased treatment and voluntary choice.

Finally, this paper makes a related methodological contribution. Discrimination is generally identified from the actions (*e.g.*, Bertrand and Mullainathan, 2004; Neumark et al., 1996) and/or learning processes (*e.g.*, Altonji and Pierret, 2001; Fryer et al., 2013) of those who discriminate. But repeatedly observing individuals' choices can also bring to light the bias they are exposed to. In particular, multiple choices made under changing conditions reveals information about agents' intrinsic preferences and knowledge of underlying processes. Using this information, one can isolate group differences in the observed equilibrium from those that would have occurred in a non-discriminatory counterfactual one. For example, assuming intrinsic preferences are fixed over time, earlier choices provide an upper bound on the impact they play in gender readability gaps; assuming authors update beliefs about the relationship between readability and acceptance rates means later choices are made with more accurate beliefs. Although this strategy relies on several strong assumptions, it may be useful for understanding and bounding the effect discrimination has on the long-term decision-making processes of those who experience it.

The remainder of the paper proceeds in the following order. Section 2 describes the data. In Section 3, I present the stylised facts about gender, readability and review times in top economics journals. Section 4 investigates mechanisms driving these stylised facts. Section 5 concludes.

2   Data

The data include every English-language article published with an abstract in *AER*, *ECA*, *JPE* and *QJE* between January 1950 and December 2015 (inclusive). The largest sample is from *Econometrica* which consistently published abstracts with its articles prior to 1950. *JPE* added them in the 1960s and *QJE* in 1980. *AER* came last in 1986. Errata and corrigenda are excluded, as are articles from *Papers & Proceedings* (*P&P*) issues of *AER*, unless otherwise mentioned. Appendix B displays data coverage by journal and decade.

For textual input, I use abstracts. Abstract readability is strongly positively correlated with the readability of other sections of a paper (see Figure D.2 and Hartley et al. (2003) and Plavén-Sigray et al. (2017)). Its structure is standardised in a manner optimal for computing readability scores. Many abstracts have also been converted to accurate machine readable text therefore curbing errors in transcription.

For the analysis in Section 3.3, I collected draft abstracts from NBER Technical and Working Paper Series. To match published articles with their NBER drafts, I used citation data from RePEc and searched NBER's database directly for unmatched papers authored by NBER family members. I eventually matched 1,988 NBER working papers to 1,986 published articles. (The mapping is not one-for-one because a small number of working papers were eventually published as multiple articles or combined into one.) This represents approximately one-fifth of all manuscripts in the data and a third of all manuscripts published between 1990–2015. Descriptive statistics are shown in Section 3.3.2.

The analysis in Section 3.4 compiles submit-accept times at *Econometrica* (1970–2015) and *REStud* (1976–2015), a fifth highly respected economics journal. (*AER*, *JPE* and *QJE* do not publish the

TABLE 1: Readability formulas

Score	Formula
Flesch Reading Ease	$206.84 - 1.02 \times \frac{\text{words}}{\text{sentences}} - 84.60 \times \frac{\text{syllables}}{\text{words}}$
Flesch-Kincaid	$-15.59 + 0.39 \times \frac{\text{words}}{\text{sentences}} + 11.80 \times \frac{\text{syllables}}{\text{words}}$
Gunning Fog	$0.40 \times \left(\frac{\text{words}}{\text{sentences}} + 100 \times \frac{\text{polysyllabic words}}{\text{words}} \right)$
SMOG	$3.13 + 5.71 \times \sqrt{\frac{\text{polysyllabic words}}{\text{sentences}}}$
Dale-Chall	$3.64 + 0.05 \times \frac{\text{words}}{\text{sentences}} + 15.79 \times \frac{\text{difficult words}}{\text{words}}$

Notes. Table displays formulas used to calculate readability scores. Polysyllabic words refer to words with three or more syllables; difficult words are those not found on a list of 3,000 words understood by 80 percent of fourth-grade readers (aged 9–10) (Chall and Dale, 1995).

dates manuscripts were submitted and accepted.) I obtained the data from journals' online archives or extracted it from digitised articles using the open source command utility `pdftotext`. Section 3.4 displays and discusses basic summary statistics.

In Section 4.3, I analyse readability at the author-level using both the entire sample and the sample of published articles matched with NBER working papers. To generate a panel dataset following author i over the $t \in \{1, \dots T_i\}$ papers he publishes in a top-four journal, I duplicate each article N_j times, where N_j is the number of co-authors on paper j. I then assign observation j_n article j's nth $\in \{1, \dots, N_j\}$ co-author. To account for duplicate articles, observations in relevant estimates are weighted by $1/N_j$.

To control for the impact of blinded review, I constructed a dummy variable equal to one if a paper underwent double-blind review before the internet at *AER* and *QJE*, the only two journals with an official double-blind review policy in place at some point during the time period covered by the data.[5] *QJE* employed double-blind procedures until 1 June, 2005; *AER* between 1 July, 1989 and 1 July, 2011. From 1 May 1987 to 31 May 1989, the *AER* conducted a randomised controlled trial whereby half of all submitted papers were evaluated by single-blind review; the remaining half were subjected to double-blind review (Blank, 1991). Since referees correctly identified at least one author in 45.6 percent of double-blind reviewed papers, however, only about a quarter of these manuscripts were truly blind reviewed. I therefore classify every paper published during the trial as having undergone single-blind review.

Other control variables used in the analysis include editor fixed effects, dynamic institution fixed effects, primary and tertiary *JEL* fixed effects, controls for author prominence and seniority, English fluency dummies, citation counts (asinh), and controls for motherhood and childbirth (Section 3.4, only). I additionally categorised each tertiary *JEL* code as either theory/methodology, empirical or other in order to roughly account for how theoretical vs. empirical a paper is. See Appendix C for further information on how each of these variables were calculated.

2.1 Readability scores

To measure writing clarity, I use the five most common, widely tested and reliable readability formulas for adult-level material: Flesch Reading Ease, Flesch-Kincaid, Gunning Fog, SMOG (Simple Measure of Gobbledegook) and Dale-Chall. The formulas for each are shown in Table 1. Appendix D discusses the scores in more detail and reviews the literature on their validity.

The Flesch Reading Ease formula ranks passages of text in ascending order—*i.e.*, more readable passages earn higher scores. The other four formulas generate grade levels estimating the minimum years of schooling needed to confidently understand an evaluated text; as a result, more readable pas-

---

[5]Double-blind review was likely less effective after the internet was adopted (for anecdotal evidence, see, *e.g.*, Goldberg, 2014). I therefore only evaluate the impact of blind review pre-internet, which I define as having been published before Google incorporated in 1998.
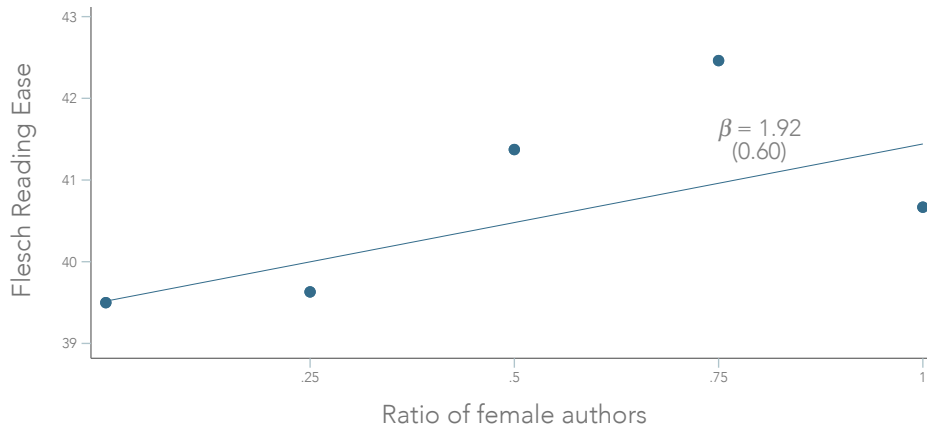
FIGURE 1: Relationship between readability and the ratio of female authors

sages earn lower scores. In order to simplify interpretation, I multiple the four grade-level scores by negative one. Thus, higher scores universally correspond to clearer writing throughout this paper.

To calculate the scores, I wrote the Python module `Textatistic`. Its code and documentation are available on GitHub; a brief description is provided in Appendix D.3. For added robustness, I also re-calculate scores and replicate most results using the `R readability` package (Appendix M).

2.2 Gender

Authors were initially assigned a gender using GenderChecker.com's database of male and female names. Three separate Mechanical Turk workers, a research assistant or I then manually verified them based on photos and other information found on faculty websites, Wikipedia articles, *etc.* In situations where the author could not be found but several people with the same first and last name were and all shared the same gender, the author was also assigned that gender. For the remaining cases, I emailed or telephoned colleagues and institutions associated with the author.

Determining the "gender" of a paper is not nearly as straightforward. For solo-authored manuscripts—of which there are 4,014 in the sample—gender corresponds to the sex of the author. As discussed in Section 3.1, however, top economics journals have collectively published just 266 by women. Only a slightly larger number were written entirely—*or even mostly*—by women.[6]

Instead, I assume an article's gender is related to its proportion of female authors. A gender readability gap—if it exists—is presumably a function of (i) the probability a passage of text was written and/or revised by a female co-author; and (ii) referees' beliefs about female authors' contributions to the writing and/or revision of a co-authored paper. Prior research suggests co-authors—regardless of seniority—share responsibility for writing and (especially) revising collaborative work (see, *e.g.*, Hart, 2000; Kumar and Ratnavelu, 2016). Thus, the intersection of (i) and (ii) is likely positively related to the ratio of female authors on a paper.

Figure 1 corroborates this hypothesis. It plots abstract readability against a paper's ratio of female authors. The slope of the regression line is positive, relatively large (1.92 points on the Flesch Reading Ease scale) and statistically significant; however the relationship is not entirely linear. In particular, it appears to be close to zero when women make up fewer than 50 percent of authors and increasing in the share of female authors only after that. For this reason, I define papers with a strict minority of female authors as male-authored; for papers with 50 percent or more female authors, I allow an

---

[6]313 papers in the sample were authored entirely by women. Women made up more than 50 percent of all authors in another 47. In 35 observations, a woman was the lead author—*i.e.*, the first author was female in a paper with authors listed non-alphabetically or in which contributions were explicitly noted.
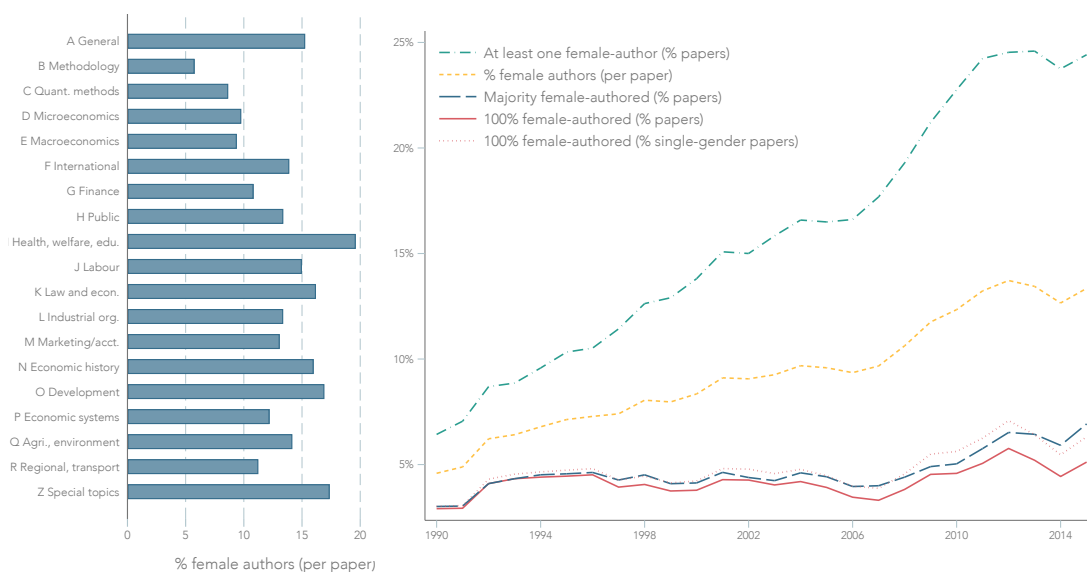
FIGURE 2: The representation of women in top economics journals

Notes. Sample 5,211 articles. Graphs illustrate the representation of female authors in articles published in top-four economics journals. Figure on the left is the average share of female authors per paper broken down by primary *JEL* category; figure on the right displays the evolution of papers' gender composition over time as five-year moving averages.

article's gender to increase linearly with its proportion of female authors. For robustness, however, I also repeat most analyses (a) on the sample of solo-authored papers, only; (b) comparing papers with a senior female co-author to entirely male-authored papers; (c) on the subset of papers authored by a single gender; (d) using a binary variable equal to one if at least one author is female; and (e) using a binary variable equal to one if at least half of all authors are female. Standard errors from (a) and (c) tend to be larger; those from (b), (d) and (e) usually similarly sized or smaller. In general, however, results do not meaningfully change (Appendix M).

3 Stylised facts

3.1 Stylised fact 1: Women are under–represented in top-four economics journals

The right-hand graph in Figure 2 illustrates women's representation in top-four economics journals over time. The number of papers these journals publish with at least one female author has been steadily increasing over the past several decades—from about 5 percent in the late 1980s to around 25 percent in 2015. Growth in the average share of female authors per paper, however, is more muted: it was about 3–4 percent in the late 1980s; by 2015 it had only increased to 15 percent. The discrepancy between these two figures is because growth in female authorship is largely thanks to an increase in the number of mixed-gendered papers authored by a strict minority of women. In fact, there has been almost no growth in the percentage of papers that are either majority or exclusively female-authored: both figures have hovered around 4–5 percent since the early 1990s. *Econometrica* publishes the fewest exclusively female-authored papers (3 percent of all papers published since 1990), *AER* the most (6 percent); *JPE* and *QJE* fall in between (4 and 5 percent, respectively). Percentages are only slightly higher (and rankings identical) for papers with a strict majority of female authors.

The left-hand graph in Figure 2 plots the average percentage of female authors per paper across primary *JEL* categories for articles published between 1990–2015. There are clear differences across fields. The average percentage of female authors per paper was lowest in *JEL* codes B (history of economic thought, methodology and heterodox approaches), C (mathematical and quantitative methods), D (microeconomics) and E (macroeconomics and monetary economics) and highest in I (health,

TABLE 2: Textual characteristics per sentence by gender

	Men	Women	Difference
No. characters	134.72	131.34	−3.37**
	(0.42)	(1.30)	(1.37)
No. words	24.15	23.38	−0.77***
	(0.08)	(0.24)	(0.25)
No. syllables	40.64	39.14	−1.50***
	(0.13)	(0.40)	(0.42)
No. polysyllabic words	4.69	4.40	−0.29***
	(0.02)	(0.06)	(0.06)
No. difficult words	9.38	9.03	−0.35***
	(0.03)	(0.10)	(0.11)
No. observations	8,268	849	9,117

Notes. Figures are means of textual characteristics (per sentence) by sex. Male-authored papers are defined as having a ratio of female authors below 50 percent; female-authored papers are those with a ratio of female authors at or above 50 percent. Last column subtracts male means from female means. Standard errors in parentheses. ***, ** and * difference statistically significant at 1%, 5% and 10%, respectively.

education and welfare), Z (other special topics) and O (economic development, innovation, technological change, and growth). Despite this variation, the percentage of female authors per paper does not exceed 20 percent in any field.

3.2 Stylised fact 2: Women's papers are more readable

Table 2 compares textual characteristics between male-authored papers (defined as having a ratio of female authors below 50 percent) and female-authored papers (defined as having a ratio of female authors at or above 50 percent). It suggests women write shorter, simpler sentences: they contain fewer characters, fewer syllables, fewer words and fewer "hard" words. Differences are highly statistically significant.

Table 3 presents results from 45 separate ordinary least squares (OLS) regressions of readability scores on the ratio of female authors (papers with fewer than 50 percent female authors are classified as male, see Section 2.2). Column (1) includes journal and editor fixed effects and controls for blind review and its interaction with the ratio of female authors on a paper. Columns (2) and (3) add journal-year interaction dummies.[7] Column (4) introduces controls for paper j's number of co-authors (N_j) and the dynamic institution effects described in Appendix C. Column (5) adds a dummy variable capturing English fluency; it also controls for article quality (citations (asinh)), co-author prominence (max. T) and seniority at the time of publication (max. t). Columns (6)–(9) are estimated on the sample of articles published after 1990. (7) includes fixed effects for primary *JEL* categories. (8) replaces it with three binary variables meant to capture how theoretical vs. empirical a paper is. (9) includes fixed effects for tertiary *JEL* categories.[8]

Results in Table 3 suggest that abstracts written by women score about 1–2 points higher on the Flesch Reading Ease scale; according to the four grade-level measures, they take about 1–5 fewer months of schooling to understand. Percentage-wise, women write about 1–4 percent better than men.

---

[7]The coefficients on the journal dummies in (2) are presented in Appendix F. Compared to *AER*, all five scores agree that *Econometrica* is harder to read; four out of five scores suggest *JPE* is, too, while *QJE* is easier.

[8]Due to small sample sizes, column (9) includes 563 articles from *AER P&P*, coded as a separate journal. Papers published in *AER P&P* are selected and edited by the American Economic Association's president-elect with the help of a Program Committee (see www.aeaweb.org for more details). *P&P* does not publish abstracts in its print version; only select years (2003 and 2011–2015) and papers were available online when I collected the data. Excluding these articles does not impact results or conclusions: coefficients are similar to those in column (9), but standard errors are somewhat higher.

TABLE 3: Gender differences in readability, article-level analysis

	1950–2015					1990–2015			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Flesch	1.25**	1.23**	1.23**	1.26**	1.49***	1.31**	1.36**	1.29**	1.94**
	(0.53)	(0.52)	(0.53)	(0.54)	(0.56)	(0.59)	(0.54)	(0.60)	(0.77)
Flesch-Kincaid	0.23*	0.22*	0.23*	0.24*	0.26*	0.34**	0.37***	0.34**	0.46***
	(0.13)	(0.13)	(0.13)	(0.13)	(0.14)	(0.15)	(0.14)	(0.15)	(0.17)
Gunning Fog	0.36**	0.36**	0.37**	0.40**	0.42***	0.51***	0.50***	0.51***	0.60***
	(0.15)	(0.15)	(0.15)	(0.15)	(0.16)	(0.18)	(0.15)	(0.18)	(0.19)
SMOG	0.25**	0.25**	0.26**	0.27**	0.29**	0.32**	0.31***	0.33**	0.40***
	(0.10)	(0.10)	(0.10)	(0.11)	(0.11)	(0.12)	(0.11)	(0.13)	(0.14)
Dale-Chall	0.08	0.08	0.08	0.08	0.09	0.11*	0.11*	0.11*	0.17**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)
No. obs.	9,117	9,117	9,117	9,117	9,117	5,211	5,211	5,211	5,774
Editor	✓	✓	✓	✓	✓	✓	✓	✓	✓
Blind	✓	✓	✓	✓	✓	✓	✓	✓	✓
Journal	✓	✓							
Year		✓							
Journal×Year			✓	✓	✓	✓	✓	✓	✓
N_j				✓	✓	✓	✓	✓	✓
Institution				✓	✓	✓	✓	✓	✓
Quality					✓[1]	✓[1]	✓[1]	✓[1]	✓[1]
Native speaker					✓	✓	✓	✓	✓
JEL (primary)							✓		
Theory/empirical								✓	
JEL (tertiary)									✓

Notes. Figures represent coefficients on female ratio from 45 separate OLS regressions of readability scores on the ratio of female authors (papers with fewer than 50 percent female authors are classified as male, see Section 2.2). (6)–(9) are estimated on the sample of papers published on or after 1990 with a primary *JEL* code; (9) includes 563 articles from *AER P&P* (see Footnote 8). Quality controls denoted by ✓[1] include citation count (asinh), max. T fixed effects (author prominence) and max. t (author seniority). Standard errors clustered on editor in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Appendix F.2 explores field in more detail; Appendix F.3 analyses readability at the author-level. Conditional on other explanatory variables, I find little evidence that field drives results in Table 3. After accounting for author-specific heterogeneity, the gender gap in readability rises to 2–6 percent.

3.3 Stylised fact 3: Women's papers improve more during peer review

3.3.1 Estimation strategy. In this section, I analyse readability changes that occurred during peer review by comparing abstracts pre- and post-review. My first estimation strategy simply regresses each paper's change in score on its gender composition. To understand it, note that the readability of a published paper depends on its earlier draft readability as well as factors that affect writing clarity any time after it was initially drafted:

$$R_{jP} = R_{jW} + \beta_{0P} + \beta_{1P} \text{ female ratio}_j + \theta_P X_{jP} + \mu_{jP} + \varepsilon_{jP}, \qquad (1)$$

where R_{jP} and R_{jW} are readability scores for working (W) and published (P) versions of paper j, respectively. β_{1P} is the coefficient of interest and reflects the particular impact female ratio$_j$ has in peer review. X_{jP} and μ_{jP} are P-specific observable and unobservable components, respectively. ε_{jP} is P's error term.

Correlation between R_{jW} and female ratio$_j$ may bias OLS estimates of β_{1P}. Equation (2) eliminates the distortion by subtracting R_{jW} from both sides of Equation (1):

$$R_{jP} - R_{jW} = \beta_{0P} + \beta_{1P} \text{ female ratio}_j + \theta_P X_{jP} + \mu_{jP} + \varepsilon_{jP}. \qquad (2)$$

Assuming zero partial correlation between female ratio$_j$ and μ_{jP}, OLS generates an unbiased estimate of β_{1P}.[9]

An alternative strategy based on Ashenfelter and Krueger (1994) separately estimates gender differences in the draft and final versions of papers using generalised least squares (GLS). The contemporaneous effect of peer review is identified post-estimation by subtracting coefficients. To implement this set-up, I combine Equation (1) with: (i) the relationship between readability scores and the gender composition of a paper before peer review; and (ii) an equation accounting for potential correlation between observable controls and version-invariant unobservables. The former is defined as:

$$R_{jW} = \beta_{0W} + \beta_{1W} \text{ female ratio}_j + \boldsymbol{\theta}_W \, \mathbf{X}_{jW} + \mu_{jW} + \varepsilon_{jW}, \tag{3}$$

where β_{1W} reflects female ratio$_j$'s impact on readability prior to peer review; $\mathbf{X}_{jW}$ and μ_{jW} are version-invariant observable and unobservable components, respectively; ε_{jW} is version W's error term. Equation (4) then defines a general structure for potential correlation between μ_{jW} and observable variables in both Equation (3) and Equation (1):

$$\mu_{jW} = \gamma + \eta \text{ female ratio}_j + \boldsymbol{\delta}_W \, \mathbf{X}_{jW} + \boldsymbol{\delta}_P \, \mathbf{X}_{jP} + \omega_j, \tag{4}$$

where ω_j is uncorrelated with female ratio$_j$, $\mathbf{X}_{jW}$ and $\mathbf{X}_{jP}$. Substituting Equation (4) into Equation (3) generates the following reduced form representation of R_{jW}:

$$R_{jW} = \widetilde{\beta}_{0W} + \widetilde{\beta}_{1W} \text{ female ratio}_j + \widetilde{\boldsymbol{\theta}}_W \, \mathbf{X}_{jW} + \boldsymbol{\delta}_P \, \mathbf{X}_{jP} + \widetilde{\varepsilon}_{jW}, \tag{5}$$

where $\widetilde{\beta}_{0W} = \beta_{0W} + \gamma$, $\widetilde{\beta}_{1W} = \beta_{1W} + \eta$, $\widetilde{\boldsymbol{\theta}}_W = \boldsymbol{\theta}_W + \boldsymbol{\delta}_W$ and $\widetilde{\varepsilon}_{jW} = \varepsilon_{jW} + \omega_j$. R_{jP}'s reduced form is similarly found by substituting Equation (5) into Equation (1):

$$R_{jP} = (\widetilde{\beta}_{0W} + \beta_{0P}) + (\widetilde{\beta}_{1W} + \beta_{1P}) \text{ female ratio}_j + \widetilde{\boldsymbol{\theta}}_W \, \mathbf{X}_{jW} + \widetilde{\boldsymbol{\theta}}_P \, \mathbf{X}_{jP} + \mu_{jP} + \widetilde{\varepsilon}_{jP}, \tag{6}$$

where $\widetilde{\boldsymbol{\theta}}_P = \boldsymbol{\theta}_P + \boldsymbol{\delta}_P$ and $\widetilde{\varepsilon}_{jP} = \widetilde{\varepsilon}_{jW} + \varepsilon_{jP}$. Equation (5) and Equation (6) are explicitly estimated via feasible GLS (FGLS). β_{1P} is identified post-estimation by subtracting reduced form coefficients. Assuming zero partial correlation between μ_{jP} and female ratio$_j$, it also generates an unbiased estimate of β_{1P}.

3.3.2   Results. Table 4 compares textual characteristics between a paper's draft and final versions in the samples of male-authored (female ratio below 50 percent) and female-authored manuscripts (female ratio at or above 50 percent). It suggests abstract text is altered during peer review. According to the first panel, draft abstracts are longer—more characters, words and sentences—and denser—more syllables, polysyllabic words and difficult words. The biggest changes are made to female-authored papers: figures in column six are 20–30 percent higher (in absolute value) than those in column three. The second panel of Table 4 suggests women's papers become more readable during peer review relative to men's. More generally, they also seem to indicate that peer review improves readability, although results are less clear for male-authored papers.

Table 5's first panel displays results from OLS estimation of Equation (1). Conditional on draft readability, published female-authored papers are more readable than published male-authored papers. Moreover, published article readability positively correlates with draft readability: coefficients on R_{jW} are about 0.8 and consistently highly significant.

Table 5's remaining columns show results from the two strategies presented in Section 3.3.1. The first strategy regresses each paper's change in score on its ratio of female authors (papers with fewer than 50 percent female authors are classified as male, see Section 2.2). As already discussed, an advantage

---

[9]Note that zero correlation between female ratio$_j$ and μ_{jP} does not preclude biased estimates of β_{1P} when μ_{jP} is correlated with other explanatory variables that are, in turn, correlated with female ratio$_j$ by some factor independent of μ_{jP}. Unbiasedness instead requires zero *partial* correlation between μ_{jP} and female ratio$_j$.

	Men			Women			Diff.-in diff.
	Working paper	Published article	Difference	Working paper	Published article	Difference	
No. sentences	6.48	5.10	−1.376***	6.76	5.06	−1.702***	−0.327**
	(0.06)	(0.04)	(0.054)	(0.15)	(0.08)	(0.139)	(0.149)
No. characters	862.64	649.74	−212.899***	905.83	636.18	−269.654***	−56.660***
	(7.19)	(4.67)	(7.162)	(18.54)	(10.35)	(18.439)	(19.632)
No. words	155.63	115.69	−39.941***	164.05	113.67	−50.386***	−10.403***
	(1.32)	(0.85)	(1.324)	(3.42)	(1.92)	(3.424)	(3.636)
No. syllables	257.06	193.33	−63.730***	268.45	187.78	−80.669***	−16.894***
	(2.15)	(1.40)	(2.136)	(5.54)	(3.09)	(5.500)	(5.852)
No. polysyllabic words	28.40	21.80	−6.603***	28.89	20.62	−8.265***	−1.648**
	(0.28)	(0.18)	(0.245)	(0.71)	(0.41)	(0.625)	(0.669)
No. difficult words	58.45	44.64	−13.815***	60.16	42.42	−17.746***	−3.923***
	(0.51)	(0.33)	(0.482)	(1.30)	(0.74)	(1.200)	(1.316)
Flesch	41.40	41.15	−0.248	42.51	43.12	0.611	0.828*
	(0.26)	(0.18)	(0.184)	(0.66)	(0.42)	(0.451)	(0.500)
Flesch-Kincaid	−13.63	−13.38	0.252***	−13.53	−13.00	0.531***	0.273**
	(0.06)	(0.05)	(0.049)	(0.15)	(0.11)	(0.122)	(0.134)
Gunning Fog	−17.28	−17.03	0.254***	−17.12	−16.58	0.545***	0.281*
	(0.07)	(0.05)	(0.055)	(0.18)	(0.13)	(0.140)	(0.149)
SMOG	−15.14	−15.00	0.145***	−15.02	−14.70	0.327***	0.179*
	(0.05)	(0.03)	(0.035)	(0.13)	(0.09)	(0.094)	(0.097)
Dale-Chall	−10.84	−10.93	−0.090***	−10.70	−10.71	−0.004	0.086**
	(0.02)	(0.02)	(0.015)	(0.06)	(0.04)	(0.037)	(0.042)
No. observations	1,716	1,716	1,716	272	272	272	1,988

Notes. Figures are means of textual characteristics by sex for NBER working papers and published articles. Male-authored papers are defined as having a ratio of female authors below 50 percent; female-authored papers are those with a ratio of female authors at or above 50 percent. Penultimate columns in each panel subtract working paper figures from published article figures for men (first panel) and women (second panel); difference-in-differences (female less male) shown in the final column. Standard errors in parentheses. ***, ** and * difference statistically significant at 1%, 5% and 10%, respectively.

of this strategy is that it more effectively removes the impact of confounding factors—*e.g.*, research field—that are constant between manuscript versions. The FGLS strategy estimates the coefficient on female ratio separately among the sample of working papers and published articles; the impact of gender on the readability gap formed during peer review is the difference between them. The advantage of this strategy is that it allows us to observe an estimate of the gap both before and after peer review.

Results from the first strategy are shown in panel two; results from the second are shown in panel three. Both strategies' estimates of the effect of gender formed during peer review are very similar (columns 4 and 10). They suggest that female-authored abstracts become 2–5 percent more readable while under review. FGLS results further indicate that the gender readability gap is 2–3 times larger in papers' published versions than it was in their pre-print versions.

Interestingly, although citations and abstract readability generally positively correlate with one another (Appendix D), the relationship between citations and the *change* in readability between draft and final versions of a paper is either negative or zero (Appendix G.1). Although I do not observe how many citations papers would have received had they not gone through peer review, these results tentatively suggest that the revisions women are asked to make during the process may not improve the general quality of their papers as proxied for by citations.

Also included in Table 5 are coefficients on the interaction between female ratio and a dummy variable equal to 1 for papers that underwent double-blind review before the internet. These estimates consistently suggest that the gender readability gap reversed direction when papers were subjected to blind review, although none are statistically significant at traditional levels. In Appendix G.3, I plot

Table 5: The impact of peer review on the gender readability gap

| | | OLS (score) | | OLS (Δ in score) | | FGLS (score) | | | | | |
| | | | | | | Working paper | | Published paper | | Difference | |
	R_{jW}	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio
Flesch	0.83***	1.47**	−2.15	1.19*	−1.75	1.68	−2.35	2.87***	−4.10*	1.19*	−1.75
	(0.02)	(0.62)	(3.84)	(0.70)	(2.18)	(1.05)	(3.37)	(1.08)	(2.16)	(0.69)	(2.14)
Flesch–Kincaid	0.74***	0.53***	−0.92	0.48**	−0.74	0.18	−0.70	0.66**	−1.44***	0.48**	−0.74
	(0.04)	(0.21)	(0.83)	(0.20)	(0.52)	(0.26)	(0.77)	(0.27)	(0.44)	(0.20)	(0.51)
Gunning Fog	0.76***	0.54**	−0.93	0.48**	−0.76	0.26	−0.71	0.74***	−1.47***	0.48**	−0.76
	(0.03)	(0.22)	(0.80)	(0.22)	(0.57)	(0.28)	(0.85)	(0.28)	(0.53)	(0.21)	(0.56)
SMOG	0.79***	0.31**	−0.59	0.27*	−0.48	0.20	−0.52	0.47***	−1.00**	0.27**	−0.48
	(0.03)	(0.15)	(0.56)	(0.14)	(0.39)	(0.17)	(0.61)	(0.18)	(0.40)	(0.14)	(0.39)
Dale–Chall	0.85***	0.18***	−0.20	0.14**	−0.20	0.21**	−0.04	0.35***	−0.24	0.14**	−0.20
	(0.02)	(0.05)	(0.20)	(0.06)	(0.13)	(0.10)	(0.23)	(0.10)	(0.20)	(0.06)	(0.13)
No. observations		1,988		1,988		1,988		1,988		1,988	
Editor effects		✓				✓		✓			
Journal×Year effects		✓		✓		✓		✓			
N_j		✓²		✓²		✓²		✓²			
Quality controls		✓²		✓²		✓²		✓²			
Native speaker		✓		✓		✓		✓			
Theory/emp. effects		✓		✓		✓		✓			

Notes. Panel one displays coefficients from estimating Equation (1) directly via OLS; standard errors clustered by editor in parentheses. Panel two displays coefficients from OLS estimation of Equation (2); standard errors clustered by year in parentheses. Panel three displays coefficients from FGLS estimation of Equation (5) and Equation (6); standard errors clustered by year and robust to cross-model correlation in parentheses. Quality controls denoted by ✓² include citation count (asinh), max. T (author prominence) and max. t (author seniority). The variable "female ratio" defines papers with a strict minority of female authors as male-authored; for papers with 50 percent or more female authors, it is the ratio of female authors on a paper (see Section 2.2 for more details). ***, **, and * statistically significant at 1%, 5% and 10%, respectively.

average residuals over time for papers published in the *AER* or *QJE* before and after they introduced (or removed) double-blind review from a regression of the differenced readability scores on the ratio of female authors. As the figures illustrate, there is a clear discontinuity in women's average unexplained changes to readability when journals switched to single-blind review (or the internet was introduced). For men, however, unexplained changes do not appear to have been substantially affected by double-blind review, conditional on included controls.

The number of manuscripts—and especially female-authored manuscripts—subjected to double-blind review is small so the coefficients on the interaction between blind review and the ratio of female authors on a paper should be interpreted with caution. Nevertheless, they do provide some (weak) indication that the gender readability gap at least partially results from factors outside of women's control, *e.g.*, editorial and refereeing bias. They may also suggest that masking authors' identities can help reduce peer review's impact on the gender readability gap—but only under certain circumstances. In Appendix G.4, I analyse the policy's post-internet impact. Gender differences are positive regardless of a journal's official review policy, suggesting that double-blind review may be effective only as long as authors are not identifiable by other means.

3.3.3 Robustness. In order to conclude that the results presented in the previous section suggest women's papers become more readable while under review, I assume that NBER working papers are not generally released before their authors submit them for peer review.[10] As Appendix G.5 illustrates, this appears to be the case: most manuscripts—and especially most female-authored manuscripts—are submitted to peer review at the same time or before being released as NBER Working Papers.

Another concern is that gender differences in how authors conform to abstract word limits may bias results in Table 5. To investigate this possibility, I exclude the 642 observations—about 40 percent of the sample—with NBER abstracts longer than the official word limit of the journals in which they were eventually published. Results are presented in Appendix G.6. Coefficient magnitudes are similar to those in Table 5; standard errors are somewhat larger.

Finally, in an effort to maximise sample sizes, I do not control for field. Although estimates in the second panel arguably implicitly account for field already, I additionally replicate Table 5 with fixed effects for primary *JEL* categories. Results are shown in Appendix G.2. Adding *JEL* fixed effects slightly increases standard errors; they otherwise make little difference.

3.4 Stylised fact 4: Women's papers spend longer under review

> "Writing simply and directly only looks easy" (Kimble, 1994, p. 53).

Good writing takes time (Hartvigsen, 1981; Kroll, 1990): skilled writers spend longer contemplating a writing assignment, brainstorming and editing; they also write fewer words per minute and produce more drafts (Faigley and Witte, 1981; Stallard, 1974). As a consequence, higher writing standards—and, indeed, higher standards applied more generally (see, *e.g.*, Card et al., 2020; Hengel and Moon, 2020)—should result in female authors spending longer in peer review, all things equal.

On the other hand, better writing by female economists could perfectly offset some other advantage present in men's papers, conditional on quality. In this case, the time-cost of publishing a paper will instead be gender neutral—since if it weren't, women could reduce their time spent in review by adopting a strategy marginally closer to men's (or visa versa).

To formalise this idea, consider male and female researchers who use strategies $x_m, x_f \in \mathcal{X}$ to produce papers of identical quality $Q \in \mathcal{Q}$. Let q represent the function mapping $\mathcal{X}$ onto $\mathcal{Q}$ and define $q^{-1}(Q)$ as the set of strategies in $\mathcal{X}$ that achieve the same Q.

---

[10]Concluding that the gender readability gap is *caused* by peer review requires making the additional assumption that authors do not make post-submission changes to their papers unless requested by referees.
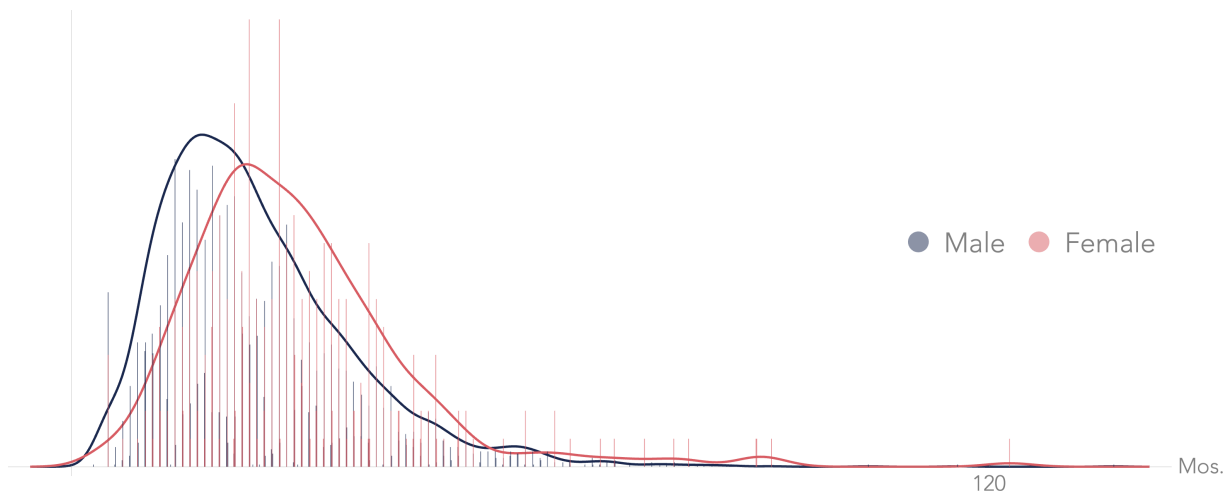
FIGURE 3: Distribution of review times at *Econometrica* and *REStud*

Notes. Sample 4,435 articles. Histograms of time spent under review for papers published in *Econometrica* and *REStud*. Male-authored papers are defined as having a ratio of female authors below 50 percent; female-authored papers are those with a ratio of female authors at or above 50 percent.

If men and women are held to identical standards in peer review, then both will accrue identical rewards, conditional on Q, *i.e.*,

$$a_m(x_m, Q) = a_f(x_f, Q) = a(x, Q), \tag{7}$$

where x is any strategy in $q^{-1}(Q)$ and $a_g(x_g, Q)$ is the acceptance rate for gender $g \in m, f$ given strategy x_g and quality Q.[11]

If men and women are also equally capable researchers, then neither side should have to exert more effort, conditional on acceptance rate (and, hence, Q)—*i.e.*, given Equation (7), there must exist some $\hat{x}_m, \hat{x}_f \in q^{-1}(Q)$ such that

$$c_m(\hat{x}_m) = c_f(\hat{x}_f), \tag{8}$$

where $c_g(x_g)$ is the cost to gender g of implementing the strategy x_g. In the absence of higher standards, Equation (8) implies that men's and women's time-cost of review should be equal, conditional on Q.

Men's and women's time-cost of review does not appear to be equal. Figure 3 displays histograms of time (in months) between dates male- (defined as having a ratio of female authors below 50 percent) and female-authored papers (defined as having a ratio of female authors above 50 percent) are first submitted to and their final revisions received by the editorial offices of *Econometria* and *REStud*. Women's review times disproportionately cluster above the mean: their articles are five times more likely to experience delays above the 75th percentile than they are to enjoy speedy revisions below the 25th.

---

[11]Higher standards come from accepting male-authored papers more often than female-authored papers, conditional on Q—*i.e.*, $a_m(x, Q) > a_f(x, Q)$—rewarding men's strategies more than women's strategies even though they both generate identical Q—*i.e.*, $a(x_m, Q) > a(x_f, Q)$—or both.

14

3.4.1 Estimation strategy and results. For more precision on gender differences in the time-cost of review—and in order to condition explicitly on quality—I build on a model by Ellison (2002):

$$
\begin{aligned}
\text{revision duration}_j = {} & \beta_0 + \beta_1 \text{ female ratio}_j + \beta_2 \text{ mother}_j + \beta_3 \text{ birth}_j \\
& + \beta_4 \text{ max } t_j + \beta_5 \text{ no. pages}_j + \beta_6 N_j + \beta_7 \text{ order}_j \\
& + \beta_8 \text{ citations}_j + \beta_9 \text{ flesch}_j + \beta_{10} \text{ theory}_j + \beta_{11} \text{ empirical}_j \\
& + \beta_{12} \text{ other}_j + \theta X_j + \varepsilon_j,
\end{aligned}
\tag{9}
$$

where female ratio$_j$ is the ratio of female authors on paper j (papers with fewer than 50 percent female authors are classified as male, see Section 2.2), mother$_j$ and birth$_j$ are binary variables equal to 1 if article j's authors were all mothers to children younger than five and gave birth, respectively, at some point during peer review, max $. t_j$ is the number of prior papers published in a top-five economics journal by article j's most prolific co-author, no. pages$_j$ refers to the page length of the published article, order$_j$ is the order in which article j appeared in an issue, citations$_j$ are the asinh-transformed number of subsequent papers citing j, flesch$_j$ is its Flesch Reading Ease score, the dummy variables theory$_j$, empirical$_j$ and other$_j$ account for how theoretical vs. empirical a paper is, and X_j captures additional fixed effects.

I first estimate Equation (9) on data from *Econometrica*. I then re-estimate it excluding readability, motherhood and childbirth controls—which I do not have for papers published in *REStud*—on the entire sample and each journal separately.

Table 6 displays results for *Econometrica*. All models include editor, acceptance year and institution fixed effects.[12] Column (1) does not control for motherhood or childbirth; (2) drops papers authored entirely by women who had children younger than five and/or gave birth during peer review; (3) controls for motherhood but not childbirth; (4) controls for childbirth but not motherhood; (5) controls for both childbirth and motherhood; (6) and (7) restrict the sample to papers published after 1990; (7) includes fixed effects for primary *JEL* categories.

Every paper published in *Econometrica* undergoes extensive review, but the consistently large and highly significant coefficient on female ratio suggests women bear the brunt of it. The average male-authored paper takes about 18.5 months to complete all revisions; papers by women need almost seven months longer.

Results pooling data from both journals and on each alone without readability, motherhood and childbirth controls are shown in Table 7. Estimates from *Econometrica* (columns one and four) coincide with those shown in Table 6. Women take 2–4 months longer in review at *REStud* (columns two and five). When observations from both journals are combined, female-authored papers take, on average, 3–6 months longer in peer review (columns three and six).

Remaining coefficients in Table 6 and Table 7 largely correspond to earlier estimates by Ellison (2002). Longer papers take more time to review, as do papers with more co-authors and (generally) those that appear earlier in an issue. Authors with an established publication history, highly cited papers and more readable papers enjoy faster reviews, although the latter effects are only noisily estimated. Papers classified as empirical take longer in review; papers classified as other spend less time under review. Finally, giving birth slows down review, but having a young child appears to have the opposite effect, at least in this particular sample.[13]

Appendix H.3 re-estimates column (5) in Table 6 and the third column of Table 7 using a quantile regression model. Appendix H.2 replicates Table 6, column (5) altering the age-threshold on mother$_j$. The gender gap is positive and significant across the entire distribution; it also does not depend on the precise definition of motherhood.

---

[12] See Appendix H.1 for results controlling for years of submission and publication, instead.

[13] This result is consistent with Ginther and Kahn (2004), who find that women with children are more productive than male and childless female doctoral recipients 10 years after receiving their Ph.D. I would interpret it with caution, however, given (i) counter-intuitive results, (ii) obtaining an unbiased estimate of β_2 was *not* this study's objective and (iii) mother$_j$ equals one for only a small number of articles in the sample.

TABLE 6: Revision duration at *Econometrica*, full control set

	1970–2015					1990–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female ratio	5.275***	6.825***	6.787***	5.645**	6.817***	9.910***	9.884***
	(1.880)	(2.294)	(2.294)	(2.203)	(2.288)	(3.036)	(2.963)
Max. t	−0.145***	−0.147***	−0.146***	−0.146***	−0.145***	−0.132***	−0.139***
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.044)	(0.045)
No. pages	0.197***	0.196***	0.196***	0.197***	0.195***	0.230***	0.216***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.042)	(0.045)
N_j	1.295***	1.283***	1.269***	1.284***	1.280***	1.745***	1.618***
	(0.299)	(0.289)	(0.296)	(0.292)	(0.293)	(0.405)	(0.426)
Order	0.199***	0.197***	0.195***	0.198***	0.195***	0.438***	0.471***
	(0.062)	(0.062)	(0.062)	(0.062)	(0.062)	(0.150)	(0.146)
No. citations (asinh)	−0.397*	−0.421**	−0.408**	−0.395*	−0.419**	−0.671	−0.683*
	(0.200)	(0.199)	(0.199)	(0.200)	(0.199)	(0.402)	(0.394)
Flesch	−0.018	−0.017	−0.017	−0.018	−0.016	−0.036	−0.041
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.028)	(0.030)
Theory	0.335	0.290	0.298	0.329	0.290	0.341	−0.630
	(0.976)	(0.976)	(0.980)	(0.979)	(0.976)	(1.010)	(1.385)
Empirical	2.322	2.767	2.323	2.345	2.267	2.296	1.383
	(1.808)	(1.808)	(1.796)	(1.786)	(1.808)	(1.941)	(2.165)
Other	−1.171	−1.165	−1.150	−1.159	−1.166	−0.900	−2.150**
	(0.884)	(0.847)	(0.867)	(0.879)	(0.876)	(0.900)	(0.909)
Mother			−7.360**		−12.085***	−23.612***	−24.268***
			(3.563)		(3.762)	(5.775)	(6.330)
Birth				−3.085	7.840	18.336***	18.694***
				(4.840)	(5.213)	(6.274)	(6.390)
Constant	14.583***	14.702***	14.689***	14.600***	14.714***	16.524***	18.801***
	(1.420)	(1.429)	(1.429)	(1.419)	(1.434)	(2.434)	(2.518)
R^2	0.289	0.292	0.291	0.289	0.291	0.132	0.150
No. observations	2,623	2,608	2,623	2,623	2,623	1,278	1,278
Editor effects	✓	✓	✓	✓	✓	✓	✓
Year effects	✓	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓	✓
JEL (primary) effects							✓

Notes. Coefficients from OLS estimation of Equation (9); (2) excludes papers authored only by women who gave birth and/or had a child younger than five at some point during peer review; (6) and (7) exclude papers published before 1990. Year fixed effects refer to the year an article was accepted. The variable "female ratio" defines papers with a strict minority of female authors as male-authored; for papers with 50 percent or more female authors, it is the ratio of female authors on a paper (see Section 2.2 for more details). Standard errors clustered by submission year in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE 7: Revision duration at *Econometrica* and *REStud*, restricted control set

	1970–2015			1990–2015		
	Econometrica	*REStud*	*Econometrica + REStud*	*Econometrica*	*REStud*	*Econometrica + REStud*
Female ratio	5.27***	1.87	3.27***	8.23***	4.07***	5.66***
	(1.87)	(1.17)	(1.11)	(2.55)	(1.47)	(1.38)
Max. t	−0.15***	−0.09	−0.13***	−0.14***	−0.06	−0.10**
	(0.04)	(0.07)	(0.04)	(0.04)	(0.08)	(0.05)
No. pages	0.20***	0.15**	0.18***	0.22***	0.05	0.17***
	(0.03)	(0.07)	(0.03)	(0.04)	(0.07)	(0.03)
N_j	1.29***	−0.05	0.82***	1.63***	0.34	1.22***
	(0.30)	(0.49)	(0.27)	(0.44)	(0.64)	(0.36)
Order	0.21***	−0.10	0.07	0.50***	−0.02	0.18
	(0.06)	(0.08)	(0.05)	(0.14)	(0.16)	(0.12)
No. citations (asinh)	−0.39*	−0.59**	−0.43***	−0.62	−1.08**	−0.85***
	(0.20)	(0.22)	(0.15)	(0.40)	(0.41)	(0.30)
Theory	0.34	0.14	0.32	−0.52	−0.21	−0.42
	(0.98)	(1.10)	(0.64)	(1.38)	(1.37)	(0.81)
Empirical	2.30	5.55**	3.80**	1.38	6.96**	4.24***
	(1.81)	(2.39)	(1.47)	(2.17)	(2.62)	(1.50)
Other	−1.21	−2.28*	−1.48**	−2.11**	−2.43*	−1.84***
	(0.89)	(1.24)	(0.67)	(0.97)	(1.36)	(0.62)
Constant	13.91***	25.35***	17.99***	16.63***	32.16***	22.96***
	(1.43)	(2.28)	(1.06)	(2.30)	(3.28)	(1.52)
R^2	0.29	0.28	0.30	0.14	0.16	0.14
No. observations	2,623	1,812	4,435	1,278	1,069	2,347
Editor effects	✓	✓	✓	✓	✓	✓
Accepted year effects	✓	✓		✓	✓	
Journal×Accepted year effects			✓			✓
Institution effects	✓	✓	✓	✓	✓	✓
JEL (primary) effects				✓	✓	✓

Notes. Coefficients from OLS estimation of Equation (9). Third and sixth column estimates pool data from *Econometrica* and *REStud*; the other four columns were separately estimated on data from each journal. The variable "female ratio" defines papers with a strict minority of female authors as male-authored; for papers with 50 percent or more female authors, it is the ratio of female authors on a paper (see Section 2.2 for more details). Standard errors clustered by submission year in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

4 Mechanisms

4.1 *Theoretical framework*

The stylised facts documented in Section 3 suggest that, relative to men, women may spend more time rewriting old papers, thus leaving them with less time for writing new papers. In this section, I investigate two potential mechanisms that can help explain why: (i) women voluntarily write better papers—*e.g.*, because they're more sensitive to referee criticism—or (ii) better written papers are women's response to external factors they do not control—*i.e.*, higher standards imposed by referees and/or editors.

To help distinguish between (i) and (ii), I develop a simple model of an author's decision making process. It follows an author—denoted by i—who publishes several articles in prestigious academic journals over the course of his career. Each article is roughly equivalent in terms of topic, novelty and quality, but may vary on readability. Upon submission to a journal, it is refereed by a review group $s \in \Sigma$, where Σ is the (finite) set of all potential review groups and μ_i are strictly positive probability measures on Σ.

I assume s accepts i's papers if and only if $R_{it} \geq \widetilde{R}_i^s$ where R_{it} is the readability of i's tth paper and

$\widetilde{R}_i^s$ is the threshold that s applies specifically to i.[14] $\widetilde{R}_i^s$ can depend on other qualities of i's papers—*e.g.*, methodological rigour, data, originality or policy relevance. It may also reflect reviewers' objectives, idiosyncratic preferences and relative weight in determining outcomes. For example, an editor who does not care about readability and is willing to override the opinion of referees will implement a lower $\widetilde{R}_i^s$ (all else equal). I assume Σ and μ_i are known to i but $\widetilde{R}_i^s$ is not, although the process of peer review provides enough information—*e.g.*, via referee reports—for i to distinguish each $s \in \Sigma$.

i forms expectations about $\widetilde{R}_i^s$ by assigning subjective probabilities $\pi_{it}^s(R)$ to all R. He then regularly updates π_{it}^s using relevant information from his own experience in peer review and by observing others' readability choices and publication outcomes. I assume this process of learning is sufficient to ensure that $\pi_{it}^s(R_{it})$ uniformly converges on some closed interval $\mathcal{R}$ where $R_{it} \in \mathcal{R}$.[15]

Equation (10) defines i's subjective expected utility at time t from writing a paper as readable as R_{it}:

$$\int_\Sigma \Pi_{it}^s(R_{it}) u_i \, \mathrm{d}\mu_i + \phi_i(R_{it}) - c_i(R_{it}), \tag{10}$$

where $\Pi_{it}^s(R_{it})$ is the probability that $\widetilde{R}_i^s \leq R_{it}$ (*i.e.*, the cumulative sum of $\pi_{it}^s(R)$ for all $R \leq R_{it}$), u_i is the utility of having a paper accepted in a prestigious journal,[16] and ϕ_i and c_i are the satisfaction and cost, respectively, i derives from writing readable papers. ϕ_i is increasing and concave in its arguments, c_i increasing and convex—marginally higher R_{it} generates proportionally less satisfaction but needs more effort when the paper is already well written.[17]

Equation (10) incorporates a variety of factors that potentially affect authors' readability choices—editorial standards conditional on other qualities in the paper ($\widetilde{R}_i^s$); ambition (u_i); the cost of drafting and revising manuscripts (c_i); an otherwise unexplained intrinsic satisfaction from writing readable papers (ϕ_i). Poor information, overconfidence and sensitivity to criticism are not explicitly included, on the assumption that people do not *want* to be poorly informed, overconfident or excessively sensitive. These factors nevertheless enter Equation (10)—and hence influence choices—via the subjective expectations authors form about $\widetilde{R}_i^s$.

A single R_{it} cannot, therefore, establish if and to what extent i's choices are motivated by (a) intrinsic preferences and costs specific to him (u_i, ϕ_i, c_i), (b) conditional editorial standards and/or referee assignment outside his control ($\widetilde{R}_i^s$, μ_i) or (c) miscellaneous confounding factors mopped up by Π_{it}^s. Because i's intrinsic preference and cost functions are assumed to be time independent, however, observing an increase in his choice of readability at two separate t distinguishes (a) from the combined impact of (b) and (c): i may be more sensitive to criticism and he might prefer writing more clearly; nevertheless, he persistently improves the readability of his future papers relative to his past papers only when he believes that doing so boosts the probability that those future papers will be accepted. Moreover, because (c) does not reflect activities or states the author enjoys, its impact on choices declines with better information—*i.e.*, authors may miscalculate referee expectations and misconstrue their reports, but with experience they correct those mistakes. I capture this idea in Theorem 1, which is proved in Appendix A.

Theorem 1. *Consider two authors, i and k, and assume:*

Assumption 1. i choses R_{it} each t to maximise Equation (10).

Assumption 2. i and k write papers that are identical with respect to topic, novelty and quality but potentially differ on readability.

---

[14] See Hengel (2017) for a version of the model with a two stage refereeing process, where papers are either rejected or offered a revise and resubmit in the first stage and rejected or accepted in the second.

[15] Effectively, this assumption rules out systematic mistakes in beliefs that are only corrected at the limit.

[16] Authors probably care about getting their papers accepted and they may care about writing well, but their marginal utility from the intersection of the two events—*i.e.*, higher utility from writing well *only* because the paper is published in a top-four journal (as opposed to a top field journal or second-tier general interest journal)—is assumed to be negligible.

[17] See Hengel (2020) for a version of the model where c_{it} changes over time, *e.g.*, due to learning-by-doing.

Assumption 3. $\pi_{it}^s(R_{it})$ *converges uniformly on some closed interval* $\mathcal{R}$ *where* $R_{it} \in \mathcal{R}$.

When the following three conditions are satisfied, then referee assignment is biased in favour of k, referee/editor scrutiny is biased against i, or both.

Condition 1. i's papers are persistently more readable than k's papers: $R_{kt} \leq R_{it}$ for all $t > t'$ and there exists $K' > 0$ such that for no $t > t'$, $|R_{it} - R_{kt}| < K'$.

Condition 2. i's future papers are more readable than at least one of his past papers: for at least one $t'' < t'$, $R_{it''} < R_{it'}$ and there exists $K'' > 0$ such that for no $t > t'$, $|R_{it} - R_{it''}| < K''$.

Condition 3. i's papers are not accepted more often than k's papers: $\int_\Sigma \mathbf{1}_i^s(R_{it})\,\mathrm{d}\mu_i \leq \int_\Sigma \mathbf{1}_k^s(R_{kt})\,\mathrm{d}\mu_k$ for all $t > t'$, where $\mathbf{1}_i^s(R_{it})$ is an indicator function equal to 1 if $\widetilde{R}_i^s \leq R_{it}$.

Theorem 1 identifies three testable conditions that, when satisfied, provide suggestive evidence that either editors assign women "tougher" referees—*i.e.*, those with higher $\widetilde{R}_i^s$—or, on average, referees and/or editors apply higher standards to women's writing. The first condition states that the readability of i's and k's papers never converges—*i.e.*, past some point, i's papers are always more readable than k's papers by some fixed amount K'. The second condition says that i's future papers are always more readable—this time by some fixed amount K''—than at least one paper he wrote in the past. In other words, i's readability does not converge to his most poorly written paper. Finally, the third condition states that i's papers are not accepted more often than k's (on average).

The intuition behind these conditions is simple. Experience can serve as a way to complete information, so observing how women's choices change as it increases can help determine which factors are predominantly motivating those choices. More specifically, information imperfections combined with lower confidence, higher risk aversion or a tendency to update too much when faced with negative signals can mean women write more readably than otherwise equivalent men despite their papers being accepted at similar rates. In this case, the gender readability gap is primarily caused by misinformation; it should therefore decline as women's information gets better. Alternatively, women may simply prefer writing more readably than men. In this case, there should be no obvious change in the gender readability gap as women gain a better understanding of peer review. A final possibility is that the gender gap increases with experience. This pattern of behaviour indicates that women revise their beliefs *upwards* about the standards they are being held to as they learn more about those standards. Assuming women make fewer mistakes about referees' and editors' thresholds as they gain experience in peer review, this would suggest that tougher standards play a role in how they make their choices.

Of course, arriving at this conclusion requires making several strong assumptions. First, it is assumed that Equation (10) defines an author's optimal readability choice. Second, experience must indeed reduce information imperfections and asymmetries between the sexes. That is, Theorem 1's conclusion only applies if women make fewer mistakes about referees' and editors' standards as they gain more experience in peer review. (Or, at least, this statement become true at some point.) However, if women are unable to obtain the knowledge required to correct mistaken beliefs, then the gaps in women's information relative to men's will not necessarily decline—and could even increase—as women gain experience in peer review. For example, women could systematically mis-perceive a higher threshold, improve readability as a result, get accepted and then have no reason to update their (mistaken) beliefs. Moreover, this could then lead to learning so that future improvements in readability are lower cost for women than they are for men, thereby exacerbating gender differences in readability.

Third—and most critically—i's and k's papers must be identical on every dimension except readability. This assumption applies over i and k's entire lifespan and not just to a single point in time. As a result, it effectively rules out the possibility that i and k specialise over time in different dimensions of quality—*e.g.*, i on readability and k on, say, mathematical rigour—even while the general quality of their work is the same. (See Section 3.4 for an indirect test of this hypothesis.)
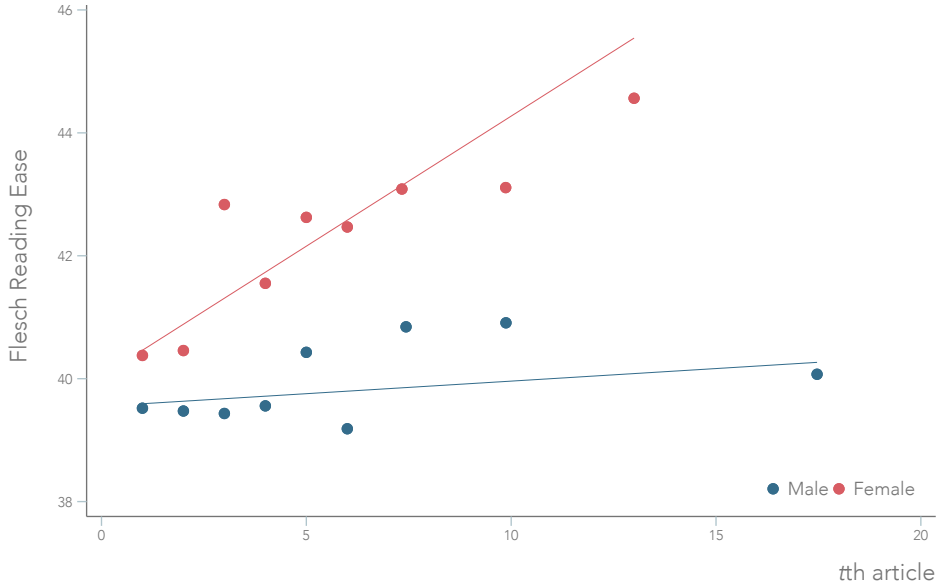
FIGURE 4: Readability of authors' tth paper

Notes. Sample 16,055 observations. Binned scatter plot of abstract readability for authors' first, second, *etc.* top-four paper.

4.2 Suggestive descriptive evidence

In this section, I show suggestive descriptive evidence that, on average, female authors satisfy Theorem 1's three conditions relative to men.

First, evidence from other studies suggests female-authored papers are not accepted more often than male-authored papers (Condition 3). A recent study of four comparable journals suggests that exclusively male- and female-authored manuscripts receive a revise and resubmit decision 8 and 6 percent of the time, respectively (Card et al., 2020). Blank (1991) found that 12.7 and 10.6 percent of male- and female-authored papers were accepted at the *AER*. A study of *JAMA*'s editorial process indicated that 44.8 percent of referees accept male-authored papers as is or if suitably revised; 29.6 percent summarily reject them. Corresponding figures for female-authored papers were 38.3 and 33.3 percent (Gilbert et al., 1994). Studies from other disciplines find female-authored papers subjected to single-blind peer review are accepted less often than would be expected by chance (Handley et al., 2015; McGillivray and De Ranieri, 2018). There appear to be no gender differences in acceptance rates to NBER's Summer Institute (Chari and Goldsmith-Pinkham, 2017). See Appendix I.1 for a table summarising these and other studies.

As for Conditions 1 and 2, women write more clearly than men (Condition 1) and their future papers are more readable than their past papers (Condition 2). As shown in Section 3.1, female-authored abstracts are 1–6 percent more readable than those by men. Figure 4 plots an author's Flesch Reading Ease score against t, where $t = 1$ for his first top-four publication, $t = 2$ for his second, *etc.* As t increases, women's readability improves whereas men's does not.[18]

Table 8 presents the marginal effect on female ratio (papers with fewer than 50 percent female authors are classified as male, see Section 2.2) for female authors (β_1) from estimating Equation (11) on subsamples of authors with $t = 1$, $t = 2$, *etc.*:

$$R_{it} = \beta_0 + \beta_1 \text{ female ratio}_{it} + \text{female ratio}_{it} \times \text{male}_i + \theta X_{it} + \varepsilon_{it}, \tag{11}$$

---

[18]As shown in Appendix I.2, women's average readability scores are 1–5 percent higher than the readability of their first papers, their latest papers 1–7 percent higher. For a man, however, his average and last papers are more poorly written than his first.

20

TABLE 8: Gender gap in readability at increasing t

	$t = 1$	$t = 2$	$t = 3$	$t = 4$–5	$t \geq 6$	All
Flesch	0.72	1.91*	4.64***	3.24	2.69	2.24***
	(0.75)	(0.99)	(1.50)	(2.31)	(2.01)	(0.74)
Flesch-Kincaid	0.07	0.23	1.04***	0.72	0.46	0.29*
	(0.19)	(0.25)	(0.25)	(0.44)	(0.40)	(0.16)
Gunning Fog	0.21	0.52*	1.29***	0.99**	0.62	0.53***
	(0.21)	(0.28)	(0.33)	(0.46)	(0.41)	(0.19)
SMOG	0.15	0.40**	0.81***	0.73*	0.42	0.42***
	(0.14)	(0.19)	(0.24)	(0.38)	(0.29)	(0.13)
Dale-Chall	0.04	0.05	0.23*	0.37*	0.42*	0.15**
	(0.07)	(0.10)	(0.13)	(0.20)	(0.23)	(0.07)
No. observations	6,874	2,827	1,675	1,906	2,773	12,008
Editor effects	✓	✓	✓	✓	✓	✓
Blind review	✓	✓	✓	✓	✓	✓
Journal×Year effects	✓	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓
Quality controls	✓[3]	✓[3]	✓[3]	✓[3]	✓[3]	✓[1]
Native speaker	✓	✓	✓	✓	✓	✓

Notes. β_1 from FGLS estimation of Equation (11). First column restricts sample to authors' first top-four publication ($t = 1$), second column to their second ($t = 2$), *etc.* Regressions weighted by $1/N_j$. Standard errors (in parentheses) adjusted for two-way clustering (editor and author) and cross-model correlation. Final column estimates from an unweighted population-averaged regression; error correlations specified by an auto-regressive process of order one and standard errors (in parentheses) adjusted for one-way clustering on author. Quality controls denoted by ✓[1] include citation count (asinh), max. T fixed effects (author prominence) and max. t (author seniority); ✓[3] includes citation count (asinh) and max. t, only. The variable "female ratio" defines papers with a strict minority of female authors as male-authored; for papers with 50 percent or more female authors, it is the ratio of female authors on a paper (see Section 2.2 for more details). ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

where R_{it} is the readability score for author i's tth top-four publication, gender enters twice, male$_i$ and female ratio$_{it}$, to account for i's sex and the sex of his co-authors, X_{it} is a vector of observable controls and ε_{it} is the error term.

All figures in Table 8 agree—women write better—but the magnitude and significance of that difference increases as t increases. Between $t = 1$ and $t = 2$, the gap marginally widens but is not significant. After that, it triples (at least); the increase is significant ($p < 0.05$) for all five scores (Appendix I.3). At higher publication counts, figures are less precisely estimated and smaller than in column 3, but still noticeably larger than estimates in columns 1 and 2.[19]

4.3 Quantifying the counterfactual

Evidence in the previous section suggests women satisfy Theorem 1's three conditions relative to men, on average. However, included controls undoubtedly fail to fully account for differences in the non-readability aspects of men's and women's papers (Assumption 2). Furthermore, concluding that higher standards are present actually requires that all three of Theorem 1's conditions hold for the same author—that is, the same woman must write better than an equivalent man, not be accepted at rates higher than he is *and* raise the quality of her writing over time.

In this section, I attempt to improve my estimates in both respects by restricting the sample to authors with three or more top-four publications. I then match observably similar male and female economists based on characteristics that predict the topic, novelty and quality of their research. Within-person readability comparisons are used to determine if Condition 2 was satisfied for each author in a matched pair. Between-person comparisons after authors gain experience in peer review are used to

---

[19]Only 40 female authors have 4–5 publications in the data; just 28 have six or more.

establish whether Condition 1 was satisfied for the male or female member. I then use these results to construct a counterfactual estimate of the impact higher standards have played in women's readability choices.

4.3.1 Theoretical strategy. To understand the counterfactual analysis, note that i's optimal choice of readability (defined in Equation (10)) can be formulated as a binary decision problem. Specifically, i initially faces the following choice: he can either ignore the impact readability has on his acceptance rate entirely and simply set R_{it} so that marginal cost equals the marginal benefit of the intrinsic satisfaction he derives from writing readable papers (*i.e.*, $R_{it} = R_i^\star$ where $R_i^\star$ solves $c_i'(R) = \phi_i'(R)$); alternatively, he can chase after higher acceptance rates and set $R_{it} > R_i^\star$. If the latter option is chosen, then i optimally minimises R conditional on acceptance rate—*i.e.*, he sets R_{it} just equal to the readability threshold of the last review group he believes will accept his paper. Thus, $R_{it} = \widetilde{R}_i^{\bar{s}} + e_{it}$, where e_{it} is the error in i's time t beliefs about $\widetilde{R}_i^{\bar{s}}$ and $\bar{s}$ is the toughest review group to accept i's papers.

Suppose i satisfies the assumptions and conditions of Theorem 1 relative to k and assume that at time t' i and k are sufficiently experienced in peer review to ensure that e_{it} and e_{kt}: (i) are on a path converging to zero (*i.e.*, both getting closer and closer to zero each time i and k go through another round of peer review); and (ii) have already converged to one another (*i.e.*, $e_{it} = e_{kt}$). When these assumptions hold, Corollary 1 suggests a conservative measure of the impact external factors have on i's time t' readability choice. It is proved in Appendix A.

Corollary 1. *Let i satisfy the assumptions and conditions of Theorem 1 relative to k and assume that for all $t \geq t'$:*

Assumption 4. $e_{it} = e_{kt}$.

Assumption 5. e_{it} *and* e_{kt} *are on a path converging to zero.*

Then Equation (12) is a conservative estimate of the impact higher standards play in i's time t' readability choice:

$$D_{ik} \equiv R_{it'} - \max\left\{R_{it''}, R_{kt'}\right\}, \tag{12}$$

where $R_{it''}$ is the readability of i's time $t'' < t'$ paper, as defined in Theorem 1, Condition 2.

D_{ik} represents a lower bound on the difference between i's time t' optimal readability score given he's held to higher standards than k (*i.e.*, $R_{it'}$) and the readability score he would have chosen had be been subject to the same standards as k (*i.e.*, $R \leq \max\{R_{kt'}, R_{it''}\}$). The intuition behind it is simple. First note that it is never optimal for i to choose an R less than $R_i^\star$. Since i already chose $R_{it''} < R_{it'}$ (Theorem 1's Condition 2), that must mean that $R_i^\star \leq R_{it''}$. For all $R > R_i^\star$, i prefers to minimise R conditional on acceptance rate, thus, i prefers any $R \in [R_{it''}, R_{it'})$ to $R_{it'}$ if it achieves the same acceptance rate as $R_{it'}$. Furthermore, from Condition 3, we know that i's acceptance rate at $R_{it'}$ is identical to k's acceptance rate at $R_{kt'}$. Suppose $R_{kt'} < R_{it''}$. If i and k were subject to identical standards, then i's acceptance rate a $R_{it'}$ would be the same as his acceptance rate at $R_{kt'}$ and therefore also the same as his acceptance rate at $R_{it''}$. Because it may be that $R_{kt'} < R_i^\star \leq R_{it''}$, i does not necessarily prefer $R_{kt'}$ to $R_{it''}$, conditional on acceptance rate. We can, however, conclude that i would, at the very least, prefer $R_{it''}$ to $R_{it'}$. Suppose $R_{it''} \leq R_{kt'}$. As before, i's acceptance rate at $R_{it'}$ is the same as k's acceptance rate at $R_{kt'}$, but it is no longer necessarily the case that i's acceptance rate at $R_{it''}$ is the same as k's acceptance rate at $R_{kt'}$. Thus, if both were subject to the same standards, then i would, at the very least, prefer $R_{kt'}$ to $R_{it'}$.

4.3.2 Estimation strategy. Implementing Corollary 1 first requires that measurement occurs at time t'—*i.e.*, a point at which authors are sufficiently experienced for Assumptions 4 and 5 to hold. I assume this point occurs at or before authors' third top-four paper. Authors with one or two top-four publications are probably tenured and well-established in their fields. By publication three, all frequently

22

referee (and some edit) prestigious economics journals. I assume this accumulated experience means: (a) equivalent authors are equally accurate about the standards they are being held to and remaining errors are no longer gender specific; and (b) those errors are getting smaller and smaller each time authors go through another round of peer review.

Additionally, in order to estimate D_{ik}, Theorem 1's Assumptions 1–3 must also hold. Most critically, Assumption 2 requires that i's and k's papers are identical with respect to topic, novelty and overall quality. I attempt to satisfy this assumption by matching every female author with three or more top-four publications to her closest male counterpart. Matches were made using a Mahalanobis procedure with the following co-variates: (1) maximum citation count over t; (2) institutional rank at $t = 1$; (3) fraction of papers published per decade; (4) fraction of papers published by each journal; and (5) number of articles per primary *JEL* category.[20] Co-variate balance pre- and post-match are shown in Appendix J.1. Appendix J.2 lists each matched pair.

Assume authors are indeed well-matched and also sufficiently experienced at $t = 3$. Then under ideal circumstances, comparing R_{i3} to R_{i1} determines the impact information (as proxied for by experience) has on readability, conditional on gender (Condition 2); comparing R_{i3} to R_{k3} determines the impact of gender, conditional on information (Condition 1). Because of co-authoring, however, circumstances are not ideal. In particular, co-authoring means that article gender is neither fixed over t conditional on i, nor is i's and k's experience—and hence information—necessarily identical at time $t = 3$. I attempt to account for this by predicting i's tth paper readability had it only been co-authored with members of i's same sex. To do so, I reconstruct i's time t readability choice at female ratio equal to 1 for women and 0 for men using errors and coefficients from OLS estimation of Equation (13) in the gender and time appropriate subsample of authors:[21]

$$\widehat{R}_{it} = \alpha_{tg_i} + \beta_{tg_i} \text{ female ratio}_{it} + \varepsilon_{it}, \tag{13}$$

where $g_i = m, f$ if i is male or female, respectively, "female ratio" defines papers with a strict minority of female authors as male-authored; for papers with 50 percent or more female authors, it is the ratio of female authors on a paper (see Section 2.2 for more details) and ε_{it} is the estimated error term. As long as ε_{it} does not partially correlate with a paper's ratio of female authors conditional on t and g_i, then $\widehat{R}_{it}$ provides an unbiased prediction of R_{it}. Regression output from Equation (13) is shown in Appendix J.3. To adjust for the degrees of freedom lost when generating $\widehat{R}_{it}$, standard errors in subsequent calculations are inflated by 1.03. Appendix J.5 presents results using the unadjusted, observed R_{it} instead of $\widehat{R}_{it}$.

4.3.3   Results. Table 9's first and second panels display means and standard deviations of D_{ik} (Equation (12)) for matched pairs where one member satisfies both Conditions 1 and 2 relative to the other member. In the first panel, the female member does; in the second, it's the male member. Male scores are subtracted from female scores, so D_{ik} is, by definition, positive in panel one and negative in panel two.

Results in Table 9 suggest that Conditions 1 and 2 were satisfied for the same author in 65 percent of matched pairs. In two-thirds of those, the member who satisfied them was female. Moreover, D_{ik}'s magnitude is (on average) 1.5 times larger in matched pairs where the female member satisfied Conditions 1 and 2 compared to pairs in which the male member did. Figure 5 emphasises this point. It displays the distribution of D_{ik}: when the evidence suggests a man is subject to higher standards, D_{ik} clusters close to zero; when it suggests the woman is, D_{ik} is far more spread out.

---

[20]Two notes on co-variate choice. First, I eschew mean, median and minimum citation counts in favour of the maximum on the assumption that an author's "quality" is principally a function of his best paper. Second, most people are at top ranked institutions by $t = 3$; by matching on institutions at $t = 1$, I try to pair authors with similar career paths.

[21]More specifically, I separately estimate Equation (13) in the following four subsamples: (i) female authors at $t = 1$; (ii) male authors at $t = 1$; (iii) female authors at $t = 3$; (iv) male authors at $t = 3$. I then generate $\widehat{R}_{it}$ using the appropriate coefficients and errors for each author: (i) $\widehat{R}_{i1} = \alpha_{1f} + \beta_{1f} + \varepsilon_{i1}$ for a female i at $t = 1$; (ii) $\widehat{R}_{i1} = \alpha_{1m} + \varepsilon_{i1}$ for a male i at $t = 1$; *etc.*

TABLE 9: D_{ik} (Corollary 1)

	Higher standards for women ($D_{ik} > 0$)			Higher standards for men ($D_{ik} < 0$)			Mean D_{ik}
	Mean	S.D.	N	Mean	S.D.	N	All obs.
Flesch	9.39	7.80	52	−4.96	5.50	27	2.92***
							(0.79)
Flesch Kincaid	1.65	1.40	63	−1.30	1.36	23	0.66***
							(0.16)
Gunning Fog	2.31	1.85	55	−1.55	1.74	24	0.79***
							(0.20)
SMOG	1.74	1.42	51	−1.00	1.12	27	0.51***
							(0.15)
Dale-Chall	0.83	0.67	61	−0.67	0.50	25	0.29***
							(0.08)

Notes. Sample 121 matched pairs. First and second panels display conditional means, standard deviations and observation counts of D_{ik} (Corollary 1) from subpopulations of matched pairs in which the woman or man, respectively, satisfies Conditions 1 and 2. Male scores are subtracted from female scores, meaning D_{ik} is, by definition, positive in panel one and negative in panel two. Third panel averages D_{ik} over all matched pairs ($D_{ik} = 0$ whenever neither member simultaneously satisfies Conditions 1 and 2). Estimates are weighted by frequency observations are used in a match; degrees-of-freedom corrected standard errors in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table 9's final panel averages D_{ik} over all observations, setting $D_{ik} = 0$ in matched pairs where neither member satisfied Conditions 1 and 2. Mean D_{ik} is consistently positive and significant indicating that higher standards are, on average, directed at women: women write about 3 points more readably on the Flesch Reading Ease scale; according to the four grade-level scores, their writing takes around 4–9 fewer months of schooling to understand than it would if they weren't subject to higher standards. Percentage-wise, these results suggest that higher standards mean women write, on average, 5 percent more readably than they otherwise would.

I emphasis, however, that these conclusions are predicated on several strong assumptions; if any are violated, then higher standards against women cannot be inferred from Table 9 and Figure 5. First, and most critically, i and k must be identical with respect to topic, novelty and overall quality. Second, at time $t = 3$ authors must have learned enough about the process of peer review for Assumptions 4 and 5 in Corollary 1 to hold. Finally, Equation (13) must accurately predict the readability of i's and k's papers as if they had been co-authored by members of their same sex. Please see Appendix J.6 where I discuss the validity and robustness of these assumptions in more detail.

4.4 Understanding how women respond to higher standards

Women can respond to higher standards in two different ways: immediately (direct effect) and pre-emptively (feedback effect). As emphasised in Section 4.1, the weight of each effect likely depends on authors' information about—hence experience with—the peer review process. In this section, I illustrate the evolution of the relative importance of each by comparing papers pre- and post-review as authors' publication counts rise. To do so, I estimate the following equation:

$$R_{itm} = \beta_0 + \beta_1 \text{ female ratio}_{it} + \beta_2 \text{ female ratio}_{it} \times t_{it} + \beta_3 t_{it} + \theta X_{it} + \varepsilon_{it}, \qquad (14)$$

where $m = W, P$ for working papers and published articles, respectively, "female ratio" defines papers with a strict minority of female authors as male-authored; for papers with 50 percent or more female authors, it is the ratio of female authors on a paper (see Section 2.2 for more details), t_{it} is author i's number of top-four papers at time t, X_{it} is a vector of observable controls and ε_{it} is the error term.

Results from estimating Equation (14) are shown in Figure 6 and Table 10. In Figure 6, hollow circles denote draft readability; solid diamonds reflect readability in the final, published versions of those same papers. Dashed lines trace readability as papers undergo peer review (direct effect) and correspond
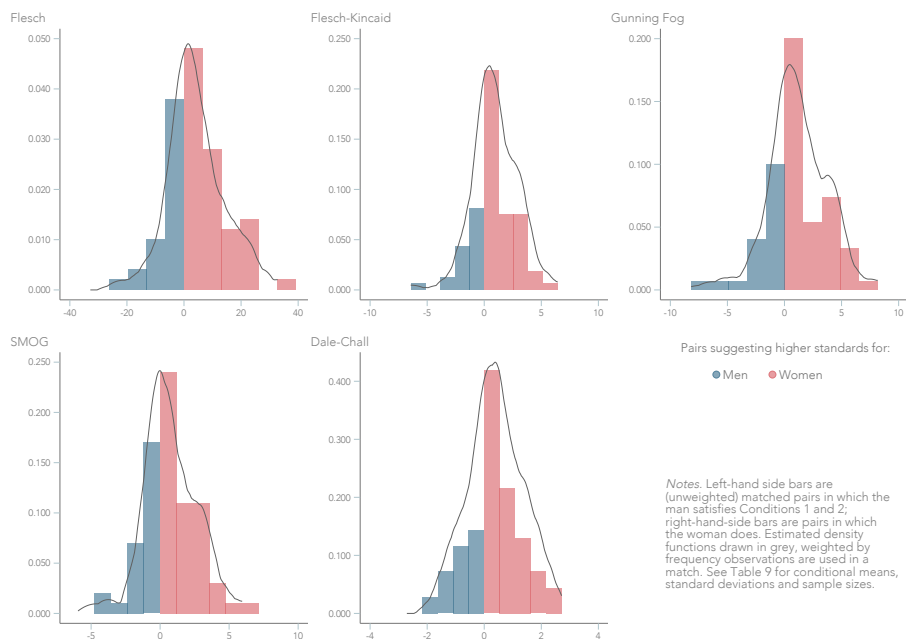
FIGURE 5: Distributions of D_{ik} (Corollary 1)

to estimates in the first panel of Table 10. Table 10's second panel shows the marginal effect of female ratio for each version of a manuscript over increasing t. Estimates for published articles correspond to the difference between female and male diamonds in Figure 6; estimates for draft papers represent feedback effects and correspond to differences between hollow circles. Difference-in-differences are shown in the final panel of Table 10.

Figure 6 and Table 10 suggest that the gender readability gap in published articles is statistically significant and relatively stable at almost every t. At $t = 1$, it is formed almost entirely during peer review; by $t \geq 4$ is is largely formed prior to submission. That is, gender differences in the direct effect of peer review start off large, positive and significant but as t increases, they gradually go away. For the feedback effect, however, the pattern is reversed. Differences in draft readability do not appear to contribute to the gender gap at $t = 1$, but this gap rises as t increases.

4.4.1 Interpretation. A number of *tentative* conclusions about the gender readability gap can be made from Figure 6 and Table 10. First, inexperienced men and women seem to make similar choices in draft readability. This suggests similar initial preferences for and beliefs about the impact of writing well. In one important sense, however, men are still better informed: the standards they believe apply actually do; junior women appear to mistakenly assume similar standards apply to them, too.

Second, experienced men *and* women seem to sacrifice time upfront in order to improve their odds in peer review. By anticipating referees' demands, authors can partially insure themselves against rejection and/or excessively long review. The price is having to spend more time revising a manuscript before submitting it. Assuming choices by senior economists express optimal trade-offs with full information, Figure 6 implies little—if any—gender differences in these preferences for insurance. Again, however, higher standards will mean that the price of that insurance is greater for women than it is for men.

Finally, Figure 6 suggests the direct effect of peer review dominates when women have less experience; the feedback effect dominates when they have more experience. This pattern of behaviour implies that women initially underestimate referees' thresholds but learn about them over time and adapt by writing their future papers more readably prior to submission. This last observation suggests inexperienced female economists go through the toughest review, conditional on acceptance. To inves-

25
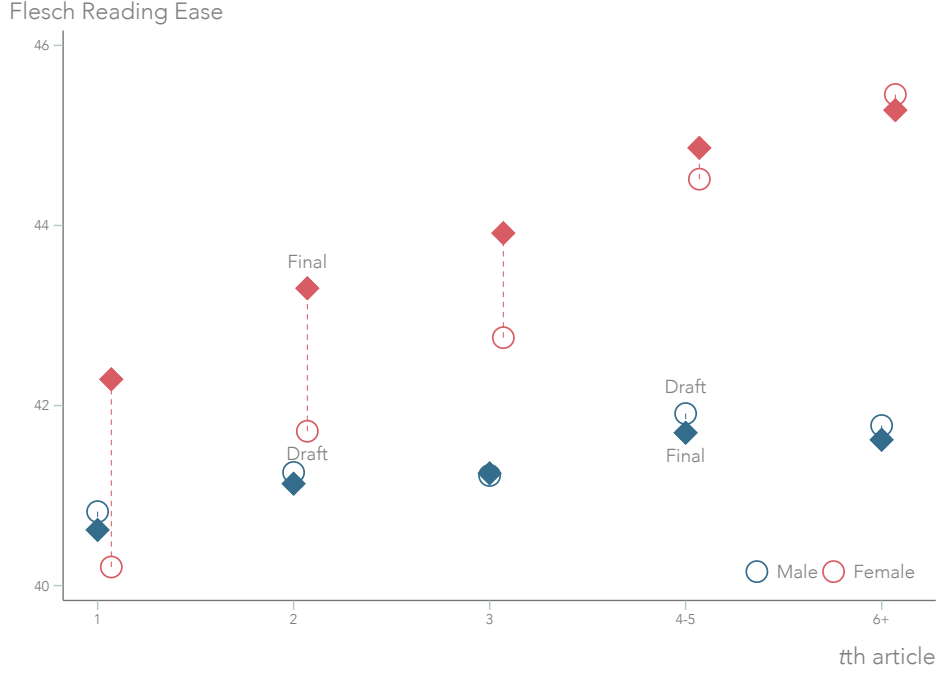
Figure 6: Readability of authors' tth paper (draft and final)

Notes. Flesch Reading Ease marginal mean scores for male and female authors' $t = 1$, $t = 2$, *etc.* top-four publications. Hollow circles denote draft readability; solid diamonds denote readability in published versions of the same papers. See Table 10 for point estimates, standard errors and further estimation details.

Table 10: Readability of authors' tth paper (draft and final)

	$t = 1$	$t = 2$	$t = 3$	$t = 4$–5	$t \geq 6$
Predicted $R_{jP} - R_{jW}$					
Women	2.09***	1.59**	1.16	0.35	−0.17
	(0.72)	(0.71)	(0.86)	(1.12)	(1.46)
Men	−0.20	−0.12	0.02	−0.21	−0.16
	(0.17)	(0.10)	(0.09)	(0.14)	(0.17)
Marginal effect of female ratio					
Published article	1.67	2.17**	2.67***	3.16***	3.66**
	(1.11)	(0.85)	(0.86)	(1.14)	(1.54)
Draft paper	−0.62	0.46	1.53**	2.60***	3.68***
	(1.31)	(0.94)	(0.75)	(0.88)	(1.23)
Diff.-in-diff.	2.29***	1.71**	1.14	0.56	−0.02
	(0.76)	(0.76)	(0.93)	(1.20)	(1.52)

Notes. Sample 4,289 observations. Panel one displays magnitude of predicted $R_{jP} - R_{jW}$ (the direct effect of peer review) for women and men over increasing t. Panel two estimates the marginal effect of an article's female ratio ($\beta_1 + \beta_2 \times t$), separately for draft papers and published articles. Figures from FGLS estimation of Equation (14), weighted by N_{it} (see Section 2). Control variables include citation count (asinh), max. T (author prominence) and max. t (author seniority), native speaker and editor and journal-year fixed effects. Standard errors clustered by editor and robust to cross-model correlation in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

tigate further, I test the impact of experience on time spent in review by re-estimating Equation (9) on sub-samples of junior and senior authors. Results are displayed in Appendix K.1. They suggest papers by junior women do indeed take longer in review; the gender gap is significantly smaller—albeit still positive—for senior women.

5   Conclusion

Most raw numerical counts suggest women produce less than men: female real estate agents list fewer homes (Seagraves and Gallimore, 2013); female lawyers bill fewer hours (Azmat and Ferrer, 2017); female physicians see fewer patients (Bloor et al., 2008); female academics write fewer papers (Ceci et al., 2014). When evaluated by narrowly defined quality measures, however, women often outperform: houses listed by female real estate agents sell for higher prices (Salter et al., 2012; Seagraves and Gallimore, 2013); female lawyers make fewer ethical violations (Hatamyar and Simmons, 2004); patients treated by female physicians are less likely to die or be readmitted to hospital (Tsugawa et al., 2017).

As I argue in this paper, female economists surpass men on another dimension: writing clarity. Abstracts written by women are 1–6 percent more readable than similar abstracts by men. They also become 2–5 percent more readable while under review when referees aren't blinded to authors' identities. However, the cost to women of revising their papers appears to be much higher than the cost to men: female-authored papers spend 3–6 months longer under review compared to observably equivalent male-authored papers. Finally, it does not appear that women are rewarded for their better writing: recent evidence from a set of comparable journals suggests female-authored papers are accepted at lower rates, conditional on quality (Card et al., 2020).

To interpret these stylised facts, I model an author's decision-making process as if it were governed by the rational behaviour of women who update their beliefs about the readability thresholds they are held to as they gain experience in peer review. I then derive three testable conditions which can help establish whether higher standards are at all important to the existence and evolution of the gender readability gap: (1) experienced women write better than equivalent men; (2) women improve their writing over time; and (3) female-authored papers are accepted no more often than equivalent male-authored papers.

On average, I find that all three conditions hold: women's writing gradually gets better but men's does not; between authors' first and third published articles, the readability gap increases by up to 12 percent; as already mentioned, female-authored papers are not accepted at higher rates after conditioning on similar co-variates (Card et al., 2020). I then conduct a counterfactual analysis that exploits within- and between-individual variation in readability among well-published economists. Its results suggest that higher standards lead women to write, on average, 5 percent more readably than they otherwise would.

I emphasis, however, that these conclusions are predicated on several strong assumptions; if any are violated then other hypotheses are also consistent with the data. For example, if matching does not fully account for differences in the non-readability aspects of men's and women's papers, then gender differences in readability may be influenced by gender differences in specialisation over time. Similarly, no control perfectly captures how empirical vs. theoretical a paper is; as a result, the gender gaps I observe may be biased by differences between fields. Additionally, the validity of the counterfactual analysis and the precision of its estimates rely on strong assumptions about men's and women's beliefs and the impact co-authors have on an article's readability.

Higher standards, wherever present, reduce women's labour market opportunities. Work that is evaluated more critically at any point in the production process will be systematically better (holding prices fixed) or systematically cheaper (holding quality fixed). This will reduce women's wages, distort measurements of their productivity and negatively impact their labour market outcomes. For example, if judges require better writing in female-authored briefs, female attorneys must charge lower fees and/or under-report hours to compete with men; billable hours and client revenue will decline, making

female lawyers appear less productive than they truly are. In academia, higher standards coupled with longer peer review times likely affect women's probability of obtaining tenure.

Unfortunately, there are no easy solutions for addressing higher standards. But least intrusive—and arguably most effective—is simple awareness and constant supervision. I hope journals are challenged to address the tougher standards they likely impose on women, open to policies that transparently monitor them and supportive of research that helps us better understand them.

University College London, Social Research Institute, UK

References

Abrevaya, J. and D. S. Hamermesh (2012). "Charity and favoritism in the field: are female economists nicer (to each other)?" *Review of Economics and Statistics* 94 (1), pp. 202–207.

Alexander, D., O. Gorelkina, and E. Hengel (2021). "Gender and the time cost of peer review". Mimeo.

Altonji, J. G. and C. R. Pierret (2001). "Employer learning and statistical discrimination". *Quarterly Journal of Economics* 116 (1), pp. 313–350.

American Economic Association (2022). *EconLit [database]*. https://www.aeaweb.org/econlit/ (last accessed January 2019). Nashville, Tennessee.

Antecol, H., K. Bedard, and J. Stearns (2018). "Equal but inequitable: who benefits from gender-neutral tenure clock stopping policies?" *American Economic Review* 108 (9), pp. 2420–2441.

Ashenfelter, O. and A. Krueger (1994). "Estimates of the economic return to schooling from a new sample of twins". *American Economic Review* 84 (5), pp. 1157–1173.

Auriol, E., G. Friebel, and S. Wilhelm (2019). "Women in European economics". Mimeo.

Azmat, G. and R. Ferrer (2017). "Gender gaps in performance: evidence from young lawyers". *Journal of Political Economy* 125 (5), pp. 1306–1355.

Babcock, L. and S. Laschever (2003). *Women Don't Ask: Negotiation and the Gender Divide*. Princeton, New Jersey: Princeton University Press.

Babcock, L. et al. (2017). "Gender differences in accepting and receiving requests for tasks with low promotability". *American Economic Review* 107 (3), pp. 714–747.

Bandiera, O. (2016). *The gender and ethnicity earnings gap at LSE*. Tech. rep. September. London School of Economics.

Bateman, V. et al. (2021). *Report on the status of women in academic economics within the UK*.

Bayer, A. and C. E. Rouse (2016). "Diversity in the economics profession: a new attack on an old problem". *Journal of Economic Perspectives* 30 (4), pp. 221–242.

Bertrand, M., C. Goldin, and L. F. Katz (2010). "Dynamics of the gender gap for young professionals in the financial and corporate sectors". *American Economic Journal: Applied Economics* 2 (3), pp. 228–255.

Bertrand, M. and S. Mullainathan (2004). "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination". *American Economic Review* 94 (4), pp. 991–1013.

Blank, R. M. (1991). "The effects of double-blind versus single-blind reviewing: experimental evidence from the *American Economic Review*". *American Economic Review* 81 (5), pp. 1041–1067.

Blau, F. D. and L. M. Kahn (2017). "The gender wage gap: extent, trends, and explanations". *Journal of Economic Literature* 55 (3), pp. 789–865.

Bloor, K., N. Freemantle, and A. Maynard (2008). "Gender and variation in activity rates of hospital consultants". *Journal of the Royal Society of Medicine* 101 (1), pp. 27–33.

Bordalo, P. et al. (2016). "Stereotypes". *Quarterly Journal of Economics* 131 (4), pp. 1753–1794.

Boring, A. (2017). "Gender biases in student evaluations of teaching". *Journal of Public Economics* 145 (Supplement C), pp. 27–41.

Born, A., E. Ranehill, and A. Sandberg (2020). "Gender and willingness to lead: does the gender composition of teams matter?" *Review of Economics and Statistics* (forthcoming).

Bransch, F. and M. Kvasnicka (2017). "Male gatekeepers gender bias in the publishing process?" IZA Discussion Paper Series, No. 11089.

Bright, L. K. (2017). "Decision theoretic model of the productivity gap". *Erkenntnis* 82 (2), pp. 421–442.

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

— (2020). "What do editors maximise? Evidence from four economics journals". *Review of Economics and Statistics* 102 (1), pp. 195–217.

Card, D. et al. (2020). "Are referees and editors in economics gender neutral?" *Quarterly Journal of Economics* 135 (1), pp. 269–327.

Carlana, M. (2019). "Implicit stereotypes: evidence from teachers' gender bias". *Quarterly Journal of Economics* 134 (3), pp. 1163–1224.

Casnici, N. et al. (2017). "Assessing peer review by gauging the fate of rejected manuscripts: the case of the *Journal of Artificial Societies and Social Simulation*". *Scientometrics* 113, pp. 533–546.

Ceci, S. J. et al. (2014). "Women in academic science: a changing landscape". *Psychological Science in the Public Interest* 15 (3), pp. 75–141.

Chall, J. S. and E. Dale (1995). *Readability Revisited: the New Dale-Chall Readability Formula*. Cambridge, Massachusetts: Brookline Books.

Chari, A. and P. Goldsmith-Pinkham (2017). "Gender representation in economics across topics and time: evidence from the NBER Summer Institute". NBER Working Paper Series, No. 23953.

Clain, S. H. and K. Leppel (2018). "Patterns in economics journal acceptances and rejections". *American Economist* 63 (1), pp. 94–109.

Clarivate (2022). *Web of Science [database]*. `https://www.webofknowledge.com/` (last accessed January 2018). London.

Coate, S. and G. C. Loury (1993). "Will affirmative-action policies eliminate negative stereotypes?" *American Economic Review* 83 (5), pp. 1220–1240.

Coffman, K. B. (2014). "Evidence on self-stereotyping and the contribution of ideas". *Quarterly Journal of Economics* 129 (4), pp. 1625–1660.

Correll, S. and C. Simard (2016). "Vague feedback is holding women back". *Harvard Business Review*. `https://hbr.org/2016/04/research-vague-feedback-is-holding-women-back` (last accessed: 4 October 2016).

Cortés, P. and J. Pan (2016). "Prevalence of long hours and women's job choices: evidence across countries and within the U.S." IZA Discussion Paper Series, No. 10225.

Croson, R. and U. Gneezy (2009). "Gender differences in preferences". *Journal of Economic Literature* 47 (2), pp. 448–474.

Ductor, L., S. Goyal, and A. Prummer (2021). "Gender and collaboration". *Review of Economics and Statistics* (forthcoming).

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

Exley, C. L. and J. B. Kessler (2021). "The gender gap in self-promotion". *Quarterly Journal of Economics* (forthcoming).

Faigley, L. and S. P. Witte (1981). "Analyzing revision". *College Composition and Communication* 32 (4), pp. 400–414.

Fang, F. C., J. W. Bennett, and A. Casadevall (2013). "Males are overrepresented among life science researchers committing scientific misconduct". *mBio* 4 (1), pp. 1–3.

Foschi, M. (1996). "Double standards in the evaluation of men and women". *Social Psychology Quarterly* 59 (3), pp. 237–254.

Fryer, R. G., D. Pager, and J. L. Spenkuch (2013). "Racial disparities in job finding and offered wages". *Journal of Law and Economics* 56 (3), pp. 633–689.

Gamage, D. K., A. Sevilla, and S. Smith (2020). "Women in economics: a UK perspective". *Oxford Review of Economic Policy* 36 (4), pp. 962–982.

Gardiner, B. et al. (2016). "The dark side of Guardian comments". *Guardian*. `https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments` (last accessed: 4 October 2016).

Gilbert, J. R., E. S. Williams, and G. D. Lundberg (1994). "Is there gender bias in *JAMA*'s peer review process?" *Journal of the American Medical Association* 272 (2), pp. 139–142.

Ginther, D. K. and S. Kahn (2004). "Women in economics: moving up or falling off the academic career ladder?" *Journal of Economic Perspectives* 18 (3), pp. 193–214.

Glover, D., A. Pallais, and W. Pariente (2017). "Discrimination as a self-fulfilling prophecy: evidence from French grocery stores". *Quarterly Journal of Economics* 132 (3), pp. 1219–1260.

Goldberg, P. (1968). "Are women prejudiced against women?" *Trans-action* 5 (5), pp. 28–30.

Goldberg, P. K. (2014). "Report of the editor: *American Economic Review*". *American Economic Review* 104 (5), pp. 621–631.

Goldin, C. (2014a). "A grand gender convergence: its last chapter". *American Economic Review* 104 (4), pp. 1091–1119.

— (2014b). "A pollution theory of discrimination: male and female differences in occupations and earnings". In: *Human Capital in History: The American Record*. Ed. by L. P. Boustan, C. Frydman, and R. A. Margo. Cambridge, Massachusetts: National Bureau of Economic Research, pp. 313–348.

Goldin, C. and L. F. Katz (2016). "A most egalitarian profession: pharmacy and the evolution of a family-friendly occupation". *Journal of Labor Economics* 34 (3), pp. 705–746.

Goldin, C. and C. Rouse (2000). "Orchestrating impartiality: the impact of 'blind' auditions on female musicians". *American Economic Review* 90 (4), pp. 715–741.

Grossbard, S., T. Yilmazer, and L. Zhang (2021). "The gender gap in citations of articles published in two demographic economics journals". *Review of Economics of the Household* 19 (3), pp. 677–697.

Grunspan, D. Z. et al. (2016). "Males under-estimate academic performance of their female peers in undergraduate biology classrooms". *PLOS ONE* 11 (2), pp. 1–16.

Handley, G. et al. (2015). "An examination of gender differences in the American Fisheries Society peer-review process". *Fisheries* 40 (9), pp. 442–451.

Hart, R. L. (2000). "Co-authorship in the academic library literature: a survey of attitudes and behaviors". *Journal of Academic Librarianship* 26 (5), pp. 339–345.

Hartley, J., J. W. Pennebaker, and C. Fox (2003). "Using new technology to assess the academic writing styles of male and female pairs and individuals". *Journal of Technical Writing and Communication* 33 (3), pp. 243–261.

Hartvigsen, M. K. (1981). "A comparative study of quality and syntactic maturity between in-class and out-of-class writing samples of freshmen at Washington State University". PhD thesis. Washington State University.

Hatamyar, P. W. and K. M. Simmons (2004). "Are women more ethical lawyers? An empirical study". *Florida State University Law Review* 31 (4), pp. 785–858.

Heilman, M. E. and M. C. Haynes (2005). "No credit where credit is due: attributional rationalization of women's success in male-female teams". *Journal of Applied Psychology* 90 (5), pp. 905–916.

Hengel, E. (2017). "Publishing while female: are women held to higher standards? Evidence from peer review". Cambridge Working Paper Economics: 1753.

— (2020). "Publishing while female. Are women held to higher standards? Evidence from peer review". Mimeo.

Hengel, E. and E. Moon (2020). "Gender and quality at top economics journals". Mimeo.

Jenkins, S. (2007). "A woman's work is never done? Fund-raising perception and effort among female state legislative candidates". *Political Research Quarterly* 60 (2), pp. 230–239.

Kimble, J. (1994). "Answering the critics of plain language". *Scribes Journal of Legal Writing* 51 (1994-1995), pp. 51–85.

Krawczyk, M. and M. Smyk (2016). "Author's gender affects rating of academic articles: evidence from an incentivized, deception-free laboratory experiment". *European Economic Review* 90, pp. 326–335.

Kroll, B. (1990). "What does time buy? ESL student performance on home versus class compositions". In: *Second Language Writing*. Ed. by B. Kroll. Cambridge: Cambridge University Press. Chap. 9, pp. 140–154.

Kumar, S. and K. Ratnavelu (2016). "Perceptions of scholars in the field of economics on co-authorship associations: evidence from an international survey". *PLoS ONE* 11 (6), pp. 1–18.

Lavy, V. and E. Sand (2018). "On the origins of gender human capital gaps: short and long term consequences of teachers' stereotypical biases". *Journal of Public Economics* 167, pp. 263–279.

Lundberg, S. J. and R. Startz (1983). "Private discrimination and social intervention in competitive labor markets". *American Economic Review* 73 (3), pp. 340–347.

Lundberg, S. J. and J. Stearns (2019). "Women in economics: stalled progress". *Journal of Economic Perspectives* 33 (1), pp. 3–22.

McGillivray, B. and E. De Ranieri (2018). "Uptake and outcome of manuscripts in *Nature* journals by review model and author characteristics". *Research Integrity and Peer Review* 3 (1), pp. 5–17.

Mengel, F., J. Sauermann, and U. Zölitz (2018). "Gender bias in teaching evaluations". *Journal of the European Economic Association* 17 (2), pp. 535–566.

Möbius, M. M. et al. (2014). "Managing self-confidence". Mimeo.

Mohr, T. S. (2014). "Why women don't apply for jobs unless they're 100% qualified". *Harvard Business Review*. https://hbr.org/2014/08/why-women-dont-apply-for-jobs-unless-theyre-100-qualified (last accessed: 16 November 2017).

Moss-Racusin, C. A. et al. (2012). "Science faculty's subtle gender biases favor male students". *Proceedings of the National Academy of Sciences* 109 (41), pp. 16474–16479.

Neumark, D., R. J. Bank, and K. D. Van Nort (1996). "Sex discrimination in restaurant hiring: an audit study". *Quarterly Journal of Economics* 111 (3), pp. 915–941.

Niederle, M. and L. Vesterlund (2010). "Explaining the gender gap in math test scores: the role of competition". *Journal of Economic Perspectives* 24 (2), pp. 129–144.

Paludi, M. A. and W. D. Bauer (1983). "Goldberg revisited: what's in an author's name". *Sex Roles* 9 (3), pp. 387–390.

Parsons, C. A. et al. (2011). "Strike three: discrimination, incentives, and evaluation". *American Economic Review* 101 (4), pp. 1410–1435.

Plavén-Sigray, P. et al. (2017). "The readability of scientific texts is decreasing over time". *eLife* 6 (e27725), pp. 1–14.

Reuben, E., P. Sapienza, and L. Zingales (2014). "How stereotypes impair women's careers in science". *Proceedings of the National Academy of Sciences* 111 (12), pp. 4403–4408.

Roter, D. L. and J. A. Hall (2004). "Physician gender and patient-centered communication: a critical review of empirical research". *Annual Review of Public Health* 25 (1), pp. 497–519.

Salter, S. P. et al. (2012). "Broker beauty and boon: a study of physical attractiveness and its effect on real estate brokers' income and productivity". *Applied Financial Economics* 22 (10), pp. 811–825.

Sarsons, H. et al. (2020). "Gender differences in recognition for group work". *Journal of Political Economy* 129 (1), pp. 101–147.

Schlosser, A., Z. Neeman, and Y. Attali (2019). "Differential performance in high versus low stakes tests: evidence from the GRE test". *Economic Journal* 129 (10), pp. 2916–2948.

Seagraves, P. and P. Gallimore (2013). "The gender gap in real estate sales: negotiation skill or agent selection?" *Real Estate Economics* 41 (3), pp. 600–631.

Stallard, C. K. (1974). "An analysis of the writing behavior of good student writers". *Research in the Teaching of English* 8 (2), pp. 206–218.

Teele, D. L. and K. Thelen (2017). "Gender in the journals: publication patterns in political science". *PS - Political Science and Politics* 50 (2), pp. 433–447.

Tsugawa, Y. et al. (2017). "Comparison of hospital mortality and readmission rates for medicare patients treated by male vs. female physicians". *JAMA Internal Medicine* 177 (2), pp. 206–213.

Weisshaar, K. (2017). "Publish and perish? An assessment of gender gaps in promotion to tenure in academia". *Social Forces* 96 (2), pp. 529–560.

Wu, A. H. (2020). "Gender bias in rumours among professionals: an identity-based interpretation". *Review of Economics and Statistics* 102 (5), pp. 867–880.

Appendices

A Proofs

Lemma 1. $\Pi_{it}^s(R_{it})$ *converges uniformly to* $\mathbf{1}_i^s(R_i)$.

Proof. By assumption, the sequence $\{\pi_{it}^s(R)\}$ converges uniformly. Because $\Pi_{it}^s(R)$ converges for some R, $\Pi_{it}^s(R)$ also converges uniformly on a closed interval containing R and on which Π_{it}^s is differentiable (Rudin, 1976, Theorem 7.17). For example, if i is rejected at time t' with readability R by review group s, then $\Pi_{it}^s(R) = 0$ for all $t > t'$ and $[R', R]$ is the closed interval on which Π_{it}^s is differentiable, where R' is some arbitrary number smaller than R. Similarly, if i is accepted at time t', then $\Pi_{it}^s(R) = 0$ for all $t > t'$ and the closed interval on which Π_{it}^s is differentiable is $[R, R']$ where this time R' is an arbitrary number greater than R. (Note that in the second case, R may equal $\widetilde{R}_i^s$.)

By the Cauchy criterion for uniform convergence (Rudin, 1976, Theorem 7.8), for every $\varepsilon > 0$ there exists an integer N such that $t \geq N$ implies

$$\left| \Pi_{it}^s(R) - \Pi_{it+1}^s(R) \right| \leq \varepsilon. \tag{A.1}$$

Because i updates Π_{it}^s based on past experience in peer review, $\Pi_{it+1}^s(R) = \mathbf{1}_i^s(R)$, where $\mathbf{1}_i^s(R)$ equals 1 if $R \geq \widetilde{R}_i^s$ and 0, otherwise. Thus, Equation (A.1) is equivalent to

$$|\Pi_{it}^s(R) - \mathbf{1}_i^s(R)| \leq \varepsilon,$$

and therefore implies that $\Pi_{it}^s(R)$ converges uniformly to $\mathbf{1}_i^s(R)$.

It remains to show that R_{it} converges. At every t, i's optimal choice of R_{it} maximises Equation (10). It is found by solving the following first order condition:

$$\int_{\Sigma} \pi_{it}^s(R_{it})\, u_i \,\mathrm{d}\mu_i = c_i'(R_{it}) - \phi_i'(R_{it}). \tag{A.2}$$

Define $\Sigma_{A_{it}}$ as the set of $s \in \Sigma$ with $\mu_i^s > 0$ and $\Pi_{it}^s(R_{it}) > 0$ (*i.e.*, the set of potential review groups that i believes has some chance of accepting his tth paper). Let $\bar{s}$ denote the review group in $\Sigma_{A_{it}}$ such that i believes $\widetilde{R}_i^{\bar{s}}$ is highest (*i.e.*, $\Pi_{it}^{\bar{s}}(R) \leq \Pi_{it}^s(R)$ for every R and all $s \in \Sigma_{A_{it}}$). Without loss of generality, Equation (A.2) is equivalent to[1]

$$\pi_{it}^{\bar{s}}(R_{it}) = \frac{c_i'(R_{it}) - \phi_i'(R_{it})}{\mu_i^{\bar{s}}\, u_i}. \tag{A.3}$$

Applying the Cauchy criterion for uniform convergence to $\pi_{it}^{\bar{s}}(R)$, for every $\varepsilon > 0$ there exists an integer N such that $t \geq N$ implies

$$\left| \pi_{it}^{\bar{s}}(R_{it}) - \pi_{it+1}^{\bar{s}}(R_{it}) \right| \leq \varepsilon \quad \text{and} \quad \left| \pi_{it+1}^{\bar{s}}(R_{it+1}) - \pi_{it+2}^{\bar{s}}(R_{it+1}) \right| \leq \varepsilon. \tag{A.4}$$

---

[1] One could think of $\Pi_{it}^{\bar{s}}(R_{it})$ as summarising i's beliefs across all or some review groups (*e.g.*, $\Pi_{it}^{\bar{s}}(R_{it}) = \int \Pi_{it}^s(R_{it})\mathrm{d}\mu_i$). Alternatively, recall that i updates his beliefs based on past experience in peer review. Thus, if i is accepted at time t by review group s, he will observe that $\widetilde{R}_i^s \leq R_{it}$. For large enough t, i will therefore know with probability 1 that $\widetilde{R}_i^s \leq R_{it}$ (and so $\pi_{it}^s(R_{it}) = 0$) for all $s \in \Sigma_{A_{it}}$ such that $s \neq \bar{s}$; the only remaining uncertainty at time t is whether $\widetilde{R}_i^{\bar{s}} \leq R_{it}$ as well.

Since i updates $\Pi_{it}^{\bar{s}}$ based on past experience in peer review, $\Pi_{it+1}^{\bar{s}}(R_{it}) = \mathbf{1}_i^{\bar{s}}(R_{it})$ and $\Pi_{it+2}^{\bar{s}}(R_{it+1}) = \mathbf{1}_i^{\bar{s}}(R_{it+1})$. Suppose R_{it} or R_{it+1} equal $\widetilde{R}_i^{\bar{s}}$. Then either i updates his belief that $\bar{s}$ is the review group in $\Sigma_{A_{it}}$ with the highest $\widetilde{R}_i^s$, or he optimally sets $R_{it'} = \widetilde{R}_i^{\bar{s}}$ for all $t' > t$, so R_{it} converges to $\widetilde{R}_i^s$, as required. Suppose neither R_{it} nor R_{it+1} equal $\widetilde{R}_i^{\bar{s}}$. Then $\pi_{it+1}^{\bar{s}}(R_{it}) = \pi_{it+2}^{\bar{s}}(R_{it+1}) = 0$; thus Equation (A.4) implies

$$\left| \pi_{it}^{\bar{s}}(R_{it}) - \pi_{it+1}^{\bar{s}}(R_{it+1}) \right| \leq \varepsilon. \tag{A.5}$$

Equation (A.5) combined with Equation (A.3) means that for every $\varepsilon > 0$ there exists an integer N such that $t \geq N$ implies

$$\left| \frac{(c_i'(R_{it}) - c_i'(R_{it+1})) - (\phi_i'(R_{it}) - \phi_i'(R_{it+1}))}{\mu_i^{\bar{s}} u_i} \right| \leq \varepsilon. \tag{A.6}$$

According to Equation (A.3), the optimal $R_i^\star$ solves $c_i'(R_i^\star) = \phi_i'(R_i^\star)$ in the absence of uncertainty. When uncertainty is present, however, the optimal R_{it} is strictly greater than $R_i^\star$. Thus, $R_{it} \geq R_i^\star$ for all t. If R_{it} were unbounded from above, then the numerator in Equation (A.6) would not converge to zero, since $c_i''(R) - \phi_i''(R) > 0$ for all $R > R^\star$. Thus, Equation (A.6) is only true for every $\varepsilon > 0$ and $t \geq N$ if the sequence $\{R_{it}\}$ converges.

Let R_i denote the limit of R_{it}. Given $\Pi_{it}^{\bar{s}}(R_{it})$ converges uniformly to $\mathbf{1}_i^{\bar{s}}(R_{it})$ and R_{it} converges to R_i, $\Pi_{it}^{\bar{s}}(R_{it})$ converges uniformly to $\mathbf{1}_i^{\bar{s}}(R_i)$, as desired. $\qquad\square$

Proof of Theorem 1. Having established that $\Pi_{it}^s(R_{it})$ converges uniformly to $\mathbf{1}_i^s(R_i)$ (Lemma 1), we prove Theorem 1 at the limit. From Condition 1 we know that $R_i > R_k$. Thus, if $\widetilde{R}_i^s = \widetilde{R}_k^s$ for all $s \in \Sigma$, then the following is also true for all $s \in \Sigma$:

$$\mathbf{1}_i^s(R_i) \geq \mathbf{1}_k^s(R_k). \tag{A.7}$$

Suppose Equation (A.7) holds as an equality for all $s \in \Sigma$ for which $\mu_i^s > 0$. From Condition 2, we know $R_i > R_{it''} > R_i^\star$. Thus, i chooses R_i over his previously optimal choice $R_{it''}$ only because it increases his probability of acceptance. Moreover, because he observes the readability choices and publication outcomes of k, he also knows $\mathbf{1}_k^s(R_k)$ for all $s \in \Sigma$. But if $\mu_i^s = \mu_k^s$ and $\mathbf{1}_i^s(R_i) = \mathbf{1}_k^s(R_k)$ for all $s \in \Sigma$ where $\mu_i^s > 0$, then i's optimal choice of readability at the limit is $R_i \leq R_k$, contradicting Condition 1.

Suppose instead that the inequality in Equation (A.7) is strict for at least one $s \in \Sigma$ for which $\mu_i^s > 0$. Assuming $\mu_i^s = \mu_k^s$ for all $s \in \Sigma$, this implies

$$\int_\Sigma \mathbf{1}_i^s(R_i) \, d\mu_i > \int_\Sigma \mathbf{1}_k^s(R_k) \, d\mu_k, \tag{A.8}$$

contradicting Condition 3. Thus, all is proved. $\qquad\square$

Proof of Corollary 1. From Condition 2 of Theorem 1 we know that $R_{it''} < R_{it'}$. Since i would never choose $R < R_i^\star$, it must be the case that $R_i^\star \leq R_{it''}$ and hence $R_i^\star < R_{it'}$. Thus, i optimally minimises R conditional on acceptance rate and therefore sets $R_{it} = \widetilde{R}_i^{\bar{s}} + e_{it}$, where e_{it} is the error in i's time t beliefs about $\widetilde{R}_i^{\bar{s}}$, $\bar{s} \in \Sigma_{A_{it'}}$ is the readability threshold of the toughest review group that i believes will still accept his paper and $\Sigma_{A_{it'}}$ is the set of review groups that i believes will accept his paper at time t'.

Suppose $R_{it''} \leq R_{kt'}$ and recall that $e_{it} = e_{kt}$ by Assumption 4. If i and k were held to identical standards, then $\widetilde{R}_i^{\bar{s}} = \widetilde{R}_k^{\bar{s}}$. Thus, i's optimal choice of readability at time t' is

$$\begin{aligned}
R &= \widetilde{R}_i^{\bar{s}} + e_{it'} \\
&= \widetilde{R}_k^{\bar{s}} + e_{kt'} \\
&\leq \max\left\{ R_k^\star, \widetilde{R}_k^{\bar{s}} + e_{kt'} \right\} \\
&= R_{kt'}.
\end{aligned}$$

Thus, if i and k were subject to identical standards, then i would not optimally set R above $R_{kt'}$. Suppose instead $R_{kt'} < R_{it''}$. Then i's optimal choice of readability at time t' is

$$
\begin{aligned}
R &= \max \left\{ R_i^\star, \widetilde{R}_i^{\bar{s}} + e_{it'} \right\} \\
&= \max \left\{ R_i^\star, \widetilde{R}_k^{\bar{s}} + e_{kt'} \right\} \\
&= \max \left\{ R_i^\star, R_{kt'} \right\} \\
&\leq \max \left\{ R_i^\star, R_{it''} \right\} \\
&= R_{it''}.
\end{aligned}
$$

Thus, if i and k were subject to identical standards, then i would not optimally set R above $R_{it''}$.

In sum, when i and k are held to the same standards, then $R_{kt'}$ is a lower bound on R when $R_{it''} \leq R_{kt'}$ and $R_{it''}$ is a lower bound on R when $R_{kt'} < R_{it''}$. As a result, $\max\{R_{it''}, R_{kt'}\}$ is a lower bound on R and so

$$
D_{ik} \equiv R_{it'} - \max\{R_{it''}, R_{kt'}\}
$$

is a lower bound on the difference between the readability choice i makes when he is subject to different standards compared to k ($R_{it'}$) and the readability choice he makes when he is subject to the same standards as k ($\max\{R_{it''}, R_{kt'}\}$). □

B   Readability data coverage

Table B.1 displays readability data coverage by journal and decade.

TABLE B.1: Article count, by journal and decade

Decade	AER	ECA	JPE	QJE	Total
1950–59		120			120
1960–69		343	184		527
1970–79		660	633	1	1,294
1980–89	180	648	562	401	1,791
1990–99	476	443	478	409	1,806
2000–09	693	519	408	413	2,033
2010–15	732	382	181	251	1,546
Total	2,081	3,115	2,446	1,475	9,117

Notes. Included is every article published between January 1950 and December 2015 for which an English abstract was found (i) on journal websites or websites of third party digital libraries or (ii) printed in the article itself. Papers published in the May issue of *AER* (*Papers & Proceedings*) are excluded. Final row and column display total article counts by journal and decade, respectively.

C   Description of control variables

Institutions. For every article I recorded authors' institutional affiliations. Individual universities in U.S. State University Systems were coded separately (*e.g.*, UCLA and UC Berkeley) but think tanks and research organisations operating under the umbrella of a single university were grouped together with that university (*e.g.*, the Cowles Foundation and Yale University). Institutions linked to multiple universities are coded as separate entities (*e.g.*, École des hautes études en sciences sociales).

In total, 1,039 different institutions were identified. For each institution, I count the number of articles in which it was listed as an affiliation in a given year and smooth the average over a five-year period. Institutions are ranked on an annual basis using this figure and then grouped to create fifteen dynamic dummy variables. Institutions ranked in positions 1–9 are assigned individual dummy variables. Those in positions 10–59 are grouped in bins of 10 to form six dummy variables. Institutions

ranked 60 or above were collectively grouped to form a final dummy variable. When multiple institutions are associated with an observation, only the dummy variable with the highest rank is used, *i.e.*, the highest-ranked institution per author when data are analysed at the author-level and the highest-ranked institution for all authors when data are analysed at the article-level.

Citations. I use article citations from Web of Science. Unless otherwise mentioned, citation counts are transformed using the inverse hyperbolic sine function (asinh).

Author prominence. I generate 37 dummy variables that group authors by their career-total top-five journal (*AER*, *Econometrica*, *JPE*, *QJE* and *REStud*) publications as of December 2015. For example, Jean Tirole forms one group (59 articles); James Heckman and Gene Grossman form another (34 articles).

Author seniority. To account for author seniority, I control for an author's number of top-five (*AER*, *Econometrica*, *JPE*, *QJE* and *REStud*) publications at the time a paper was published. For co-authored articles, only the data corresponding to the most prolific author is used.

English fluency. To account for English fluency, most regressions include a dummy variable equal to one if an article is co-authored by at least one native (or almost native) English speaker. I assume an author is "native" if he: (i) was raised in an English-speaking country; (ii) obtained all post-secondary eduction from English speaking institutions;[2] or (iii) spoke with no discernible (to me) non-native accent. This information was almost always found—by me or a research assistant—in authors' CVs, websites, Wikipedia articles, faculty bios or obituaries. In the few instances where the criteria were ambiguously satisfied—or no information was available—I asked friends and colleagues of the author or inferred English fluency from the author's first name, country of residence or surname (in that order).

Field. I create dummy variables corresponding to the 20 primary and over 700 tertiary *JEL* categories to control for subject matter. The *JEL* system was significantly revised in 1990; because exact mapping from one system to another is not possible, I collected these data only for articles published post-reform—about 60 percent of the dataset. Codes were recorded whenever found in the text of an article or on the websites where bibliographic information was scraped. Remaining articles were classified using codes from the American Economic Association's Econlit database.

I additionally categorised each tertiary *JEL* code as either theory/methodology, empirical or other. For example, C02 (mathematical methods) and D85 (network formation and analysis: theory) are classified as theory/methodology, whereas D12 (consumer economics: empirical analysis) and F14 (empirical studies of trade) are classified as empirical. Tertiary codes that are not distinctly related to empirical or theory/methodology—*e.g.*, L29 (firm objectives, organisation and behaviour: general) or O10 (economic development: general)—are classified as "other". (Papers published before 1990 are not classified in any category.) Table C.1 lists the *JEL* codes assigned to each category.

In total, I classified 99 tertiary *JEL* codes as theory/methodology, four as empirical and the remaining 756 as "other". (Given the small number of distinctly empirical codes, most "other" papers are likely empirical papers.) When combined with my 1990–2015 dataset, there are 1,764 theory/methodology articles (34 percent), 412 empirical articles (8 percent) and 4,608 "other" articles (88 percent). (Given articles can be both theory and empirical, these percentages do not sum to 100 percent.)

Editorial policy. To control for editorial policy, I recorded editor/editorial board member names from issue mastheads. *AER* and *Econometrica* employ an individual to oversee policy. *JPE* and *QJE* do

---

[2]Non-native speakers who meet this criteria have been continuously exposed to spoken and written English since age 18. This continuous exposure likely means they write as well as native English speakers. To qualify as an English-speaking institution, all courses—not just the course studied by an author—must be primarily taught in English. *E.g.*, McGill University is classified as English-speaking; University of Bonn is not (although most of its graduate economics instruction is in English).

Table C.1: Tertiary *JEL* code classification

Category	*JEL* codes
Empirical	D12 D22 F14 O47
Theory	B52 C02 C10 C11 C12 C13 C14 C15 C18 C19 C20 C21 C22 C23 C24 C25 C26 C29 C30 C31 C32 C33 C34 C35 C36 C38 C39 C40 C41 C43 C44 C45 C46 C49 C50 C51 C52 C53 C54 C55 C57 C58 C59 C60 C61 C62 C63 C65 C67 C68 C69 C70 C71 C72 C73 C78 C79 C80 C81 C82 C83 C88 C89 C90 C91 C92 C99 D01 D11 D15 D21 D40 D41 D42 D43 D44 D46 D50 D58 D85 D86 E10 E11 E12 E13 E16 E17 E19 E27 E37 E47 F11 F12 F37 F47 J64 O21 O42 R15
Other	A00 A10 A11 A12 A13 A14 A19 A20 A21 A22 A23 A29 A30 A31 A32 A33 A39 B00 B10 B11 B12 B13 B14 B15 B16 B17 B19 B20 B21 B22 B23 B24 B25 B26 B27 B29 B30 B31 B32 B40 B41 B49 B50 B51 B53 B54 B55 B59 C00 C01 C42 C87 C93 D00 D02 D04 D10 D13 D14 D16 D18 D19 D20 D23 D24 D25 D26 D29 D30 D31 D33 D39 D45 D47 D49 D51 D52 D53 D57 D59 D60 D61 D62 D63 D64 D69 D70 D71 D72 D73 D74 D78 D79 D80 D81 D82 D83 D84 D87 D89 D90 D91 E00 E01 E02 E14 E20 E21 E22 E23 E24 E25 E26 E29 E30 E31 E32 E39 E40 E41 E42 E43 E44 E49 E50 E51 E52 E58 E59 E60 E61 E62 E63 E64 E65 E66 E69 E70 E71 F00 F01 F02 F10 F13 F15 F16 F17 F18 F19 F20 F21 F22 F23 F24 F29 F30 F31 F32 F33 F34 F35 F36 F38 F39 F40 F41 F42 F43 F44 F45 F49 F50 F51 F52 F53 F54 F55 F59 F60 F61 F62 F63 F64 F65 F66 F68 F69 G00 G01 G10 G11 G12 G13 G14 G15 G17 G18 G19 G20 G21 G22 G23 G24 G28 G29 G30 G31 G32 G33 G34 G35 G38 G39 G40 G41 G50 G51 G52 G53 G59 H00 H10 H11 H12 H13 H19 H20 H21 H22 H23 H24 H25 H26 H27 H29 H30 H31 H32 H39 H40 H41 H42 H43 H44 H49 H50 H51 H52 H53 H54 H55 H56 H57 H59 H60 H61 H62 H63 H68 H69 H70 H71 H72 H73 H74 H75 H76 H77 H79 H80 H81 H82 H83 H84 H87 H89 I00 I10 I11 I12 I13 I14 I15 I18 I19 I20 I21 I22 I23 I24 I25 I26 I28 I29 I30 I31 I32 I38 I39 J00 J01 J08 J10 J11 J12 J13 J14 J15 J16 J17 J18 J19 J20 J21 J22 J23 J24 J26 J28 J29 J30 J31 J32 J33 J38 J39 J40 J41 J42 J43 J44 J45 J46 J47 J48 J49 J50 J51 J52 J53 J54 J58 J59 J60 J61 J62 J63 J65 J68 J69 J70 J71 J78 J79 J80 J81 J82 J83 J88 J89 K00 K10 K11 K12 K13 K14 K15 K16 K19 K20 K21 K22 K23 K24 K25 K29 K30 K31 K32 K33 K34 K35 K36 K37 K38 K39 K40 K41 K42 K49 L00 L10 L11 L12 L13 L14 L15 L16 L17 L19 L20 L21 L22 L23 L24 L25 L26 L29 L30 L31 L32 L33 L38 L39 L40 L41 L42 L43 L44 L49 L50 L51 L52 L53 L59 L60 L61 L62 L63 L64 L65 L66 L67 L68 L69 L70 L71 L72 L73 L74 L78 L79 L80 L81 L82 L83 L84 L85 L86 L87 L88 L89 L90 L91 L92 L93 L94 L95 L96 L97 L98 L99 M00 M10 M11 M12 M13 M14 M15 M16 M19 M20 M21 M29 M30 M31 M37 M38 M39 M40 M41 M42 M48 M49 M50 M51 M52 M53 M54 M55 M59 N00 N01 N10 N11 N12 N13 N14 N15 N16 N17 N20 N21 N22 N23 N24 N25 N26 N27 N30 N31 N32 N33 N34 N35 N36 N37 N40 N41 N42 N43 N44 N45 N46 N47 N50 N51 N52 N53 N54 N55 N56 N57 N60 N61 N62 N63 N64 N65 N66 N67 N70 N71 N72 N73 N74 N75 N76 N77 N80 N81 N82 N83 N84 N85 N86 N87 N90 N91 N92 N93 N94 N95 N96 N97 O00 O10 O11 O12 O13 O14 O15 O16 O17 O18 O19 O20 O22 O23 O24 O25 O29 O30 O31 O32 O33 O34 O35 O36 O38 O39 O40 O41 O43 O44 O49 O50 O51 O52 O53 O54 O55 O56 O57 P00 P10 P11 P12 P13 P14 P16 P17 P18 P19 P20 P21 P22 P23 P24 P25 P26 P27 P28 P29 P30 P31 P32 P33 P34 P35 P36 P37 P39 P40 P41 P42 P43 P44 P45 P46 P47 P48 P49 P50 P51 P52 P59 Q00 Q01 Q02 Q10 Q11 Q12 Q13 Q14 Q15 Q16 Q17 Q18 Q19 Q20 Q21 Q22 Q23 Q24 Q25 Q26 Q27 Q28 Q29 Q30 Q31 Q32 Q33 Q34 Q35 Q37 Q38 Q39 Q40 Q41 Q42 Q43 Q47 Q48 Q49 Q50 Q51 Q52 Q53 Q54 Q55 Q56 Q57 Q58 Q59 R00 R10 R11 R12 R13 R14 R19 R20 R21 R22 R23 R28 R29 R30 R31 R32 R33 R38 R39 R40 R41 R42 R48 R49 R50 R51 R52 R53 R58 R59 Y10 Y20 Y30 Y40 Y50 Y60 Y70 Y80 Y90 Y91 Y92 Z00 Z10 Z11 Z12 Z13 Z18 Z19 Z20 Z21 Z22 Z23 Z28 Z29 Z30 Z31 Z32 Z33 Z38 Z39

Notes. Table lists of each tertiary *JEL* code assigned to the empirical, theory/methodology and other categories.

not generally name one lead editor and instead rely on boards composed of four to five faculty members at the University of Chicago and Harvard, respectively. *REStud* is also headed by an editorial board, the size of which has been gradually increasing—from two members in the 1970s to 7–8 members today. Members are also located all over the world.

Editor controls are based on distinct lead editor/editorial boards—*i.e.*, they differ by at least one member. Among top four journals, 74 groups are formed in this manner. *REStud* adds another 34. Given the size of *Restud*'s editorial board and the fact that members serve fixed 3–4 full-year terms, editorial controls are highly correlated with year fixed effects. Moreover, unlike at *JPE* and *QJE*, editors are not located at the same institution. Thus, editor fixed effects may be less informative about editorial policy at *REStud* than they are for the other four journals.

Family commitments. To control for motherhood's impact on revision times, I recorded children's birth years for women with at least one entirely female-authored paper in *Econometrica*. I personally (and, I apologise, rather unsettlingly) gleaned this information from published profiles, CVs, acknowledgements, Wikipedia, personal websites, Facebook pages, background checks and local school district/popular extra-curricular activity websites. Exact years were recorded whenever found; otherwise, they were approximated by subtracting a child's actual or estimated age from the date the source material was posted online. In several instances, I obtained this information from acquaintances, friends and colleagues or by asking the woman directly. If an exhaustive search turned up no reference to children, I assumed the woman in question did not have any.

References

Rudin, W. (1976). *Principles of Mathematical Analysis*. 3rd ed. New York, New York: McGraw-Hill.

Publishing while female
Erin Hengel
18 January 2022

Online appendices

D   Readability scores

D.1   *Validity*

Advanced vocabulary and complicated sentences are the two strongest predictors of text difficulty (Chall and Dale, 1995). Hundreds of readability formulas exploit this relationship. The five most widely used, tested and reliable formulas for adult reading material are the Flesch Reading Ease, Flesch-Kincaid, Gunning Fog, SMOG (Simple Measure of Gobbledegook) and Dale-Chall (DuBay, 2004).

These five readability scores generally produce similar rankings: the yellow box plot in Figure D.1 summarises 169 inter-score correlations found in 26 studies; the median is 0.87. Moreover, they tend to correlate with (i) oral reading fluency, (ii) human judgement, (iii) reading comprehension tests and (iv) the cloze procedure.[1] The dark blue box plots in Figure D.1 summarise 167 correlations in 38 published cross-validation studies. (See Appendix D.4 for a list of studies included in the analysis.)

Furthermore, numerous studies have validated readability scores against surrogate measures of reading comprehension. More readable high school and college-level correspondence courses have higher completion rates (Klare and Smart, 1973). More readable academic journals enjoy larger readerships (Richardson, 1977; Swanson, 1948); their most readable articles win more awards (Sawyer et al., 2008) and are downloaded more often (Guerini et al., 2012). More readable abstracts are also (generally) cited more frequently (see Dowling et al. (2018) and McCannon (2019) and Figure D.1). They are also more likely to be published in top-five and other higher ranking journals (Marino Fages, 2020). In a blog post, Lukas Püttmann compares abstract readability to page views of VoxEU.org columns: more readable columns are viewed three percent more often (Püttmann, 2017). Evidence from other studies linking readability and citations is, however, weaker (Berninger et al., 2017; Laband and Taylor, 1992; Lei and Yan, 2016). My own data suggest a positive relationship in papers published after 1990—and particularly those published post–2000—but no relationship before that (Figure D.1).

Thanks to high predictive power and ease of use, readability formulas are widely employed in education, business and government. The U.S. Securities and Exchange Commission encourages clearer financial disclosure forms benchmarked against the Gunning Fog, Flesch-Kincaid and Flesch Reading Ease scores (Cox, 2007). The formulas have also guided readability assessments of, *inter alia*, standardised test questions (Chall et al., 1977; Chall et al., 1983), medical inserts (*e.g.*, Wallace et al., 2008), technical manuals (*e.g.*, Hussin et al., 2012; Klare and Smart, 1973), health pamphlets (*e.g.*, Foster and Rhoney, 2002; Meade and Byrd, 1989) and data security policies (Alkhurayyif and Weir, 2017).

In research, readability scores are often used to proxy for "complexity". Enke (2020) controls for language sophistication using the Flesch Reading Ease formula in a study of moral values in U.S. presidential elections. Spirling (2016) employs the same score to show that British parliamentarians simplified speeches to appeal to less educated voters in the wake of the Great Reform Act. Legal research has found that judges are more reliant on legislative history when interpreting complex legal statutes, as measured by the Flesch-Kincaid formula (Law and Zaring, 2010). In finance, the scores have linked clarity of financial communication to better firm and market financial health (Biddle et al., 2009; Jansen, 2011; Li, 2008), larger investment and trading volume (De Franco et al., 2015; Lawrence, 2013; Miller, 2010; Thörnqvist, 2015) and lower demand for—albeit higher reliability of—outside research by sell-side analysts (Lehavy et al., 2011). See also Loughran and McDonald (2016) for a review of the use of readability scores in finance and accounting research.

D.2   *Measurement error*

Readability scores fail to capture many elements relevant to reading comprehension, including grammar—*e.g.*, active vs. passive tense (Coleman, 1964; Coleman, 1965)—legibility—*e.g.*, typeface or layout—and content—*e.g.*, coherence, organisation and general appeal (Armbruster, 1984; Kemper,

---

[1]Oral reading fluency is generally measured as the number of words read aloud correctly per minute. The cloze procedure ranks passages of text according to average readers' ability to correctly guess randomly deleted words.
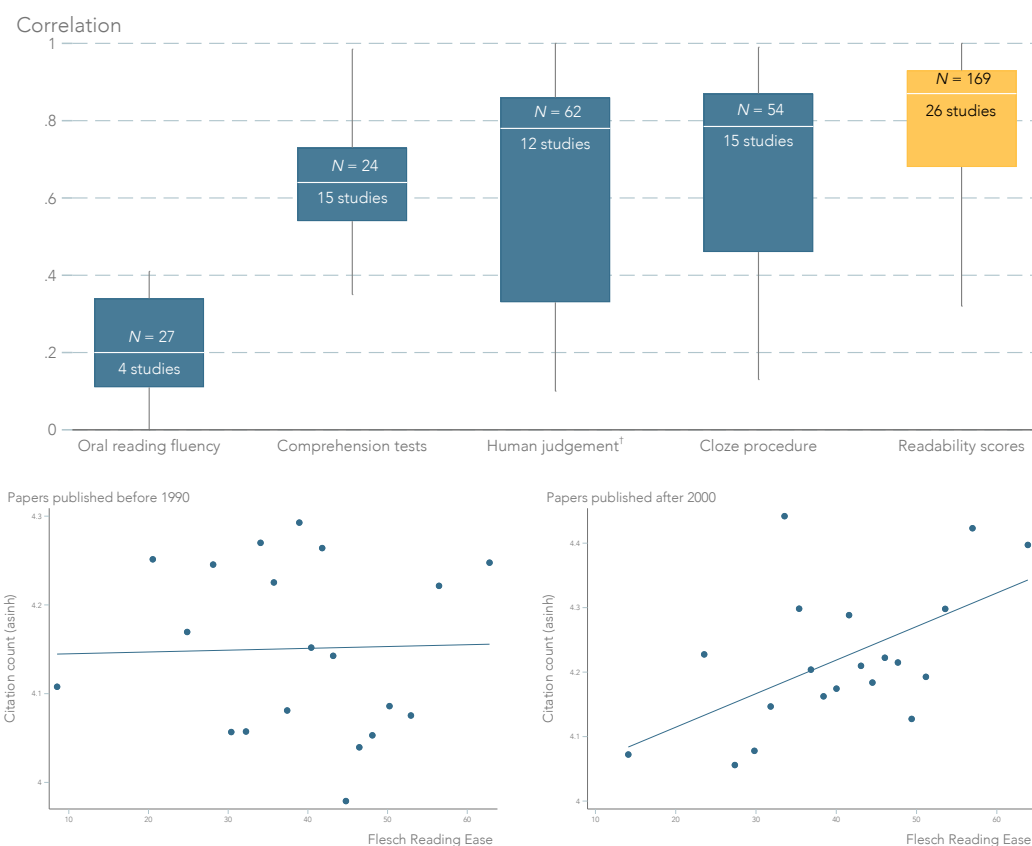
FIGURE D.1: Readability score validity

Notes. Top figure displays box plots of correlations between alternative measures of text difficulty and the Flesch Reading Ease, Flesch-Kincaid, Gunning Fog, SMOG and Dale-Chall readability scores. It includes 336 correlations found in 55 mostly peer reviewed papers. (See Appendix D.4 for the list of included studies and information on how they were selected.) Bottom figures plot abstracts' Flesch Reading Ease scores against their articles' citation counts (inverse hyperbolic sine (asinh) transformation) for the samples of top-four (excluding *AER Papers & Proceedings*) articles published before 1990 (left) and post-2000 (right). Each point represents the mean (in both dimensions) of roughly 170–180 observations. [†]Includes two studies which assessed readability using the Readability Assessment INstrument (RAIN), a comprehensive framework based on 14 variables, *e.g.*, coeherence, writing style, illustrations and typography.

1983; Kintsch and Miller, 1984; Meyer, 1982). Nevertheless, "long sentences generally correspond to complex syntactic structures, infrequent words generally refer to complex concepts, and hard texts will generally lead to harder questions about their content" (Kintsch and Miller, 1984, p. 222).

Still, readability scores' low causal power raises legitimate concerns about measurement error. As long as this error does not partially correlate with the variable of interest (gender), the analytical results I present in this paper attenuate toward zero (classical measurement error). Unfortunately, they are systematically biased in an unknown direction if it does (non-classical measurement error).

Sources of non-classical measurement error are threefold: (a) grammatical, spelling and transcription errors in the textual input; (b) errors in the estimates of vocabulary complexity and sentence length introduced by automating their calculation; or (c) embodied in the jump from using these two variables to infer readability.

Conditional on accurate calculation, readability scores combine very precise estimates of vocabulary complexity with almost perfect measures of sentence length (for a discussion, see Chall and Dale, 1995). The weighted average of these two variables is informative in much the same way that inferences about readability are. Thus, measurement error related to (c) should only shift superficial interpretation of observed gender differences—from "women are better writers" to "women use simpler words and write shorter sentences"—but leave conclusions deduced from them intact.

Nevertheless, I try to minimise measurement error from (c) by using abstracts as textual input. Ab-
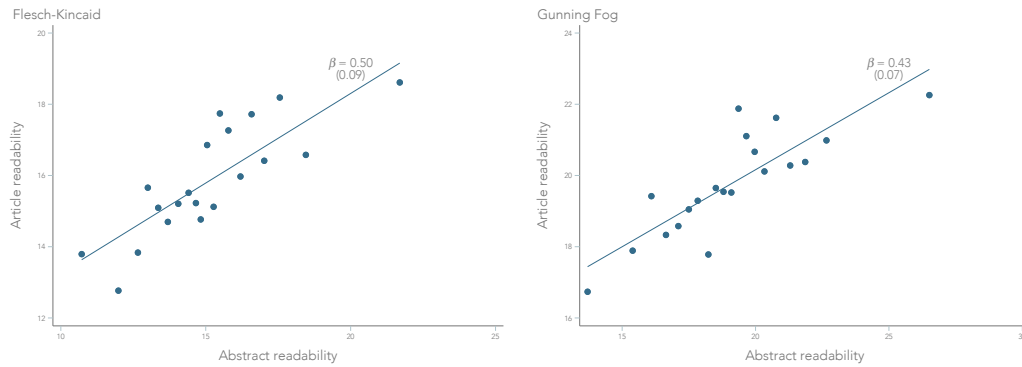
FIGURE D.2: Abstract vs. article readability

Notes. Figures plot abstract readability against the readability of a 150–200 word passage of text from the introduction of the same paper. β is the slope of the regression line (robust standard errors in parentheses). Sample only includes NBER Working Papers eventually published in a top-four economics journal with a heading explicitly titled "Introduction" (339 abstract-article pairs). Data are grouped into roughly 20 equal-sized bins; each point represents the mean (in both dimensions) of about 16–17 observations. Non-abstract text kindly provided by Henrik Kleven and Dana Scott (Kleven, 2018). Readability scores calculated using the R `readability` package.

stracts are self-contained, universally summarise the research and are the first and most frequently read part of an article (King et al., 2006). Additionally, they follow a more standardised layout compared to other parts of a manuscript: they are generally surrounded by ample whitespace and most editorial management systems anyway reproduce them in pre-formatted cover pages. These factors suggest a relatively homogenous degree of review across journals and subject matter and limit the impact that physical layout, figures and surrounding text have on readability.

Moreover, prior research suggests authors write in a stylistically consistent manner across the abstract, introduction and discussion sections of a paper. According to an analysis of published education and psychology articles, within-manuscript correlations of Flesch Reading Ease scores range from 0.64 (abstracts vs. introductions) to 0.74 (abstracts vs. discussions) (Hartley et al., 2003). Plavén-Sigray et al. (2017) also found a strong positive correlation using full text articles from several scientific journals. Figure D.2 plots abstract readability against the readability of a passage from the introduction for 339 NBER Working Papers eventually published in a top-four journal. It suggests a similarly positive relationship holds in economics, as well.[2]

In my opinion, non-classical measurement error from (a) and (b) poses a bigger concern to the identification mapped out in this paper. I have taken several steps to reduce it. First, abstract text is also ideal for calculating readability: 100–200 words containing few score-distorting features of academic writing—*e.g.*, citations, abbreviations and equations (Dale and Chall, 1948). Additionally, most abstracts have been previously converted to accurate machine-readable text by digital libraries and bibliographic databases, curbing errors in transcription.

Second, I carefully proofread the text in order to identify (and fix) remaining transcription errors,[3] eliminate non-sentence-ending full stops, and replace typesetting code—typically used to render equations—with equivalent unicode characters.[4] Readability scores were determined using the modified text.

Finally, some programs that calculate scores rely on unclear, inconsistent and possibly inaccurate algorithms to count words and syllables, identify sentence terminations and check whether a word is

---

[2]For comparison, I randomly assigned abstracts to introductions in 1,000 simulated samples. The average coefficient of correlation between abstract text readability and the readability of a passage of text from a randomly selected introduction was −0.0006 for the Gunning Fog score and 0.0007 for the Flesch-Kincaid score.

[3]*E.g.*, words in transcribed text are often inappropriately hyphenated—typically because the word was divided at the end of the line in the original text.

[4]When no exact replacement existed, characters were chosen that mimicked as much as possible the equation's original intent while maintaining the same character and word counts. (Equations in abstracts generally only occur in *Econometrica* articles published before 1980.)

on Dale-Chall's easy word list (for a discussion, see Sirico, 2007). To transparently handle these issues and eliminate ambiguity in how the scores were calculated, I wrote the Python module `Textatistic`. Its code and documentation are available on GitHub; a brief description is provided in Appendix D.3. For added robustness, I also re-calculate scores and replicate most results using the `R readability package` (Appendix L). Coefficients and standard errors are very similar to those presented in the body of the paper.

D.3 `Textatistic`

I wrote the Python module `Textatistic` to transparently calculate the readability scores in this study. The code and documentation are available on GitHub; I provide a brief description here.

To determine sentence count, the program replaces common abbreviations with their full text,[5] decimals with a zero and deletes question and exclamation marks used in an obvious, mid-sentence rhetorical manner.[6] The remaining full stops, exclamation and question marks are assumed to end a sentence and counted.

Next, hyphens are deleted from commonly hyphenated single words such as "co-author" and the rest are replaced with spaces, remaining punctuation is removed and words are split into an array based on whitespace. Word count is the length of that array.[7]

An attempt is made to match each word to one on an expanded Dale-Chall list. The count of difficult words is the number that are not found. This expanded list, available on GitHub, consists of 8,490 words. It is based on the original 3,000 words, but also includes verb tenses, comparative and superlative adjective forms, plural nouns, *etc.* It was created by first adding to the Dale-Chall list every conceivable alternate form of each word using Python's Pattern library. To eliminate nonsense words, the text of 94 English novels published online with Project Gutenberg were matched with words on the expanded list. Words not found in any of the novels were deleted.

Syllable counts are based on the C library `libhyphen`, an implementation of the hyphenation algorithm from Liang (1983). Liang (1983)'s algorithm is used by TEX's typesetting system. `libhyphen` is employed by most open source text processing software, including OpenOffice.

D.4 *Studies included in the meta analysis*

Below are the studies included in the analysis from Figure D.1, which summarises correlations between readability scores and alternative measures of reading comprehension found in other research. A few notes on the criteria for inclusion and how some correlations were determined:

- I include only documents produced for the U.S. government or published peer reviewed studies, with the exception of results from dissertations that were presented and discussed in a peer reviewed manuscript.

- I include a small number of studies with correlations between alternative readability measures and the number of words not listed on the Dale-Chall word list. In all other cases, however, correlations with only parts of a score (*e.g.*, syllables per words) are omitted.

- A few earlier studies calculated and listed various readability measures for many passages of text, but did not report coefficients of correlation between them. I manually calculated these correlations myself.

---

[5] Abbreviations which do not include full-stops are not altered. I manually replaced common abbreviations, such as "*i.e.*" and "U.S." with their abbreviated versions, sans full stops.

[6] For example, "?)." is replaced with ").".

[7] Per Chall and Dale (1995), hyphenated words count as two (or more) words.

Ardoin, S. P. et al. (2005). "Accuracy of readability estimates' predictions of CBM performance". *School Psychology Quarterly* 20 (1), pp. 1–22.

Benoit, K., K. Munger, and A. Spirling (2019). "Measuring and explaining political sophistication through textual complexity". *American Journal of Political Science* 63 (2), pp. 491–508.

Bormuth, J. R. (1966). "Readability: a new approach". *Reading Research Quarterly* 1 (3), pp. 79–132.

Brown, J. D. (1998). "An EFL readability index". *JALT Journal* 20 (2), pp. 7–36.

Carver, R. P. (1974). *Improving reading comprehension*. Tech. rep. Washington, DC: American Institutes for Research in the Behavioral Sciences.

Caylor, J. S. et al. (1973). *Methodologies for determining reading requirements of military occupational specialties*. Tech. rep. Alexandria, Virginia: Human Resources Research Organization.

Chall, J. S. and E. Dale (1995). *Readability Revisited: the New Dale-Chall Readability Formula*. Cambridge, Massachusetts: Brookline Books.

Clauson, K. A., Q. Zeng-Treitler, and S. Kandula (2010). "Readability of patient and health care professional targeted dietary supplement leaflets used for diabetes and chronic fatigue syndrome". *Journal of Alternative and Complementary Medicine* 16 (1), pp. 119–124.

Compton, D. L., A. C. Appleton, and M. K. Hosp (2004). "Exploring the relationship between text-leveling systems and reading accuracy and fluency in second-grade students who are average and poor decoders". *Learning Disabilities Research and Practice* 19 (3), pp. 176–184.

Crossley, S. A., J. Greenfield, and D. S. McNamara (2008). "Assessing text readability using cognitively based indices". *TESOL Quarterly* 42 (3), pp. 475–493.

Crossley, S. A. et al. (2017). "Predicting text comprehension, processing, and familiarity in adult readers: new approaches to readability formulas". *Discourse Processes* 54 (5-6), pp. 340–359.

Cunningham, J. W., E. H. Hiebert, and H. A. Mesmer (2018). "Investigating the validity of two widely used quantitative text tools". *Reading and Writing* 31 (4), pp. 813–833.

Dale, E. and J. S. Chall (1948). "A formula for predicting readability". *Educational Research Bulletin* 27 (1), pp. 11–20.

Dale, E. and R. W. Tyler (1934). "A study of the factors influencing the difficulty of reading materials for adults of limited reading ability". *Library Quarterly: Information, Community, Policy* 4 (3), pp. 384–412.

Entin, E. B. and G. R. Klare (1978). "Some inter-relationships of readability, cloze and multiple choice scores on a reading comprehension test". *Journal of Literacy Research* 10 (4), pp. 417–436.

Flesch, R. (1948). "A new readability yardstick". *Journal of Applied Psychology* 32 (3), pp. 221–233.

Froese, V. (1971). *Cloze readability versus the Dale-Chall formula*. Tech. rep. Winnipeg, Manitoba: University of Manitoba.

Fulcher, G. (1997). "Text difficulty and accessibility: reading formulae and expert judgement". *System* 25 (4), pp. 497–513.

Gray, W. W. and B. E. Leary (1935). *What Makes a Book Readable*. Chicago, Illinois: University of Chicago Press.

Greenfield, J. (1999). "Classic readability formulas in an EFL context: are they valid for Japanese speakers?" PhD thesis. Temple University.

— (2004). "Readability formulas for EFL". *JALT Journal* 26 (1), pp. 5–24.

Guthrie, J. T. (1972). "Learnability versus readability of texts". *Journal of Educational Research* 65 (6), pp. 273–280.

Harris, A. J. and M. D. Jacobson (1976). "Predicting twelfth graders' comprehension scores". *Journal of Reading* 20 (1), pp. 43–46.

Harwell, M. R. et al. (1996). "Evaluating statistics texts used in education". *Journal of Educational and Behavioral Statistics* 21 (1), pp. 3–34.

Hayes, D. P., L. T. Wolfer, and M. F. Wolfe (1996). "Schoolbook simplification and its relation to the decline in SAT-verbal scores". *American Educational Research Journal* 33 (2), pp. 489–508.

Hengel, E. (2017). "Publishing while female: are women held to higher standards? Evidence from peer review". Cambridge Working Paper Economics: 1753.

Hull, L. C. (1979). "Beyond readability: measuring the difficulty of technical writing". PhD thesis. Renesselaer Polytechnic Institute.

Janan, D. and D. Wray (2014). "Reassessing the accuracy and use of readability formulae". *Malaysian Journal of Learning and Instruction* 11 (1), pp. 127–145.

Jongsma, E. A. (1972). "The difficulty of children's books: librarians' judgments vs. formula estimates". *Elementary English* 49 (1), pp. 20–26.

Kanouse, D. E. et al. (1981). *Informing patients about drugs: summary report on alternative designs for prescription drug leaflets*. Tech. rep. Santa Monica, California: Rand Corporation.

Kemper, S. (1983). "Measuring the inference load of a text". *Journal of Educational Psychology* 75 (3), pp. 391–401.

Kincaid, J. P. et al. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease formula) for Navy enlisted personnel*. Tech. rep. Memphis, Tennessee: Naval Technical Training Command.

King, M. M., A. S. W. Winton, and A. D. Adkins (2003). "Assessing the readability of mental health internet brochures for children and adolescents". *Journal of Child and Family Studies* 12 (1), pp. 91–99.

Klare, G. R. (1952). "Measures of the readability of written communication: an evaluation". *Journal of Educational Psychology* 43 (7), pp. 385–399.

Klingbeil, C., M. W. Speece, and H. Schubiner (1995). "Readability of pediatric patient education materials. Current perspectives on an old problem". *Clinical Pediatrics* 34 (2), pp. 96–102.

Lee, W. D. and B. R. Belden (1966). "A cross-validation readability study of general psychology textbook material and the Dale-Chall readability formula". *Journal of Educational Research* 59 (8), pp. 369–373.

Lenzner, T. (2014). "Are readability formulas valid tools for assessing survey question difficulty?" *Sociological Methods and Research* 43 (4), pp. 677–698.

Ley, P. and T. Florio (1996). "The use of readability formulas in health care". *Psychology, Health and Medicine* 1 (1), pp. 7–28.

Lorge, I. (1948). "The Lorge and Flesch Readability Formulas: a correction". *School & Society* 67, pp. 141–142.

McLaughlin, G. H. (1969). "SMOG grading: a new readability formula". *Journal of Reading* 12 (8), pp. 639–646.

Meade, C. D. and J. C. Byrd (1989). "Patient literacy and the readability of smoking education literature". *American Journal of Public Health* 79 (2), pp. 204–206.

Meade, C. D. and C. F. Smith (1991). "Readability formulas: cautions and criteria". *Patient Education and Counseling* 17 (2), pp. 153–158.

Miller, L. R. (1974). "Predictive powers of the Flesch and Bormuth readability formulas". *International Journal of Business Communication* 11 (2), pp. 21–30.

Morris, L. A., A. Myers, and D. G. Thilman (1980). "Application of the readability concept to patient-oriented drug information". *American Journal of Health-System Pharmacy* 37 (11), pp. 1504–1509.

Powell-Smith, K. A. and K. L. Bradley-Klug (2001). "Another look at the 'C' in CBM: does it really matter if curriculum-based measurement reading probes are curriculum-based?" *Psychology in the Schools* 38 (4), pp. 299–312.

Powers, R. D., W. A. Sumner, and B. E. Kearl (1958). "A recalculation of four readability formulas". *Journal of Educational Psychology* 49 (2), pp. 99–105.

Russell, D. H. and H. R. Fea (1951). "Validity of six readability formulas as measures of juvenile fiction". *Elementary School Journal* 52 (3), pp. 136–144.

Singer, H. (1975). "The Seer Technique: a non-computational procedure for quickly estimating readability level". *Journal of Reading Behavior* 7 (3), pp. 255–267.

Singh, J. (2003). "Reading grade level and readability of printed cancer education materials". *Oncology Nursing Forum* 30 (5), pp. 867–870.

Štajner, S. et al. (2012). "What can readability measures really tell us about text complexity?" In: *Workshop on Natural Language Processing for Improving Textual Accessibility*. Istanbul, pp. 14–21.

Sullivan, R. J. (1976). "A comparison of results obtained using the cloze procedure with readability levels using the Dale-Chall formula on selected university textbooks". In: *26th Annual Meeting of the National Reading Conference*. Atlanta, Georgia.

Van Oosten, P., D. Tanghe, and V. Hoste (2010). "Towards an improved methodology for automated readability prediction". In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation*. Valletta, pp. 775–782.

Wells, J. A. (1994). "Readability of HIV/AIDS educational materials: the role of the medium of communication, target audience, and producer characteristics". *Patient Education and Counseling* 24 (3), pp. 249–259.

Woods, B., G. Moscardo, and T. Greenwood (1998). "A critical review of readability and comprehensibility tests". *Journal of Tourism Studies* 9 (2), pp. 49–61.

Zheng, J. and H. Yu (2017). "Readability formulas and user perceptions of electronic health records difficulty: a corpus study". *Journal of Medical Internet Research* 19 (3), pp. 1–15.

E   Discussion of potential alternative mechanisms

A gender readability gap exists. It's still there after including editor, journal and year effects—meaning it's probably not caused by antiquated policies and attitudes, long since overcome. The gap is unaffected by field controls, so it doesn't seem to be related to women researching topics that are easier to explain. Nor does it appear to be caused by factors correlated with gender but actually linked to authors' (or co-authors') competence as economists and fluency in English—if so, institution, native speaker and citation controls would reduce it. They do not.[8]

The gap grows between first draft and final publication and over the course of women's careers, likely precluding inborn advantage and one-off improvements in response to external circumstances unrelated to peer review. This probably also rules out gender differences in (i) biology/behaviour—*e.g.*, sensitivity to referee criticism[9]—or (ii) knowledge about referee expectations. If diligently addressing every referee concern has no apparent upside—acceptance rates are unaffected—and a very clear downside—constant redrafting takes time—even the most oversensitive, ill-informed woman would *eventually* re-examine initial beliefs and start acting like a man, no?[10] Yet this is not what we observe. The largest investments in writing well are made by female economists with greatest exposure to peer review—*i.e.*, those with the best opportunity to update their priors.

Women's papers are more likely assigned female referees (Abrevaya and Hamermesh, 2012; Gilbert et al., 1994).[11] If women are more demanding critics, clearer writing could reflect their tougher reviews.[12] Women concentrate in particular fields, so it's natural female referees more often review female-authored papers. Nevertheless, for the readability gap to exist only because of specialisation, controlling for *JEL* classification should help explain it.[13] It does not: including 20 primary or 731 tertiary *JEL* category dummies has little effect. So if referee assignment is causing the gap, it may be because journals disproportionately refer female-authored papers to the toughest critics.[14] Meaning it isn't referees who are biased—it's editors.[15]

Section 3.3 suggests a link between the gender readability gap and peer review; the evidence presented in Section 4 suggests that factors outside women's control drive it. Yet oversensitivity and/or poor information could create the former gap while *another* gender bias unconnected to peer review generates the latter. One in particular comes to mind: the feedback women receive in conferences and seminars. Perhaps experienced female economists tighten prose (before or after submission) in

---

[8]I also conducted a primitive surname analysis (see Hengel, 2016, pp. 35–36). It suggests that the female authors in my data are no more or less likely to be native English speakers.

[9]While women do appear more *internally* responsive to feedback—criticism has a bigger impact on their self-esteem—available evidence suggests they aren't any more *externally* responsive to it, *i.e.*, women and men are equally likely to change behaviour and alter performance after receiving feedback (Johnson and Helgeson, 2002; Roberts and Nolen-Hoeksema, 1989).

[10]This statement is especially relevant if the opportunity cost to women for "wasting" time on needless tasks is higher—*e.g.*, because of family responsibilities.

[11]Note that women are only a fraction of all referees—8 percent in 1986 (Blank, 1991), 10 percent in 1994 (Hamermesh, 1994) and 14 percent in 2013 (Torgler and Piatti, 2013). Abrevaya and Hamermesh (2012) report female-authored papers were only slightly more likely to be assigned a female referee between 1986–1994; matching increases between 2000–2008.

[12]It's not clear whether women's reports are more critical. A study specific to post-graduate biologists suggests yes (Borsuk et al., 2009); other studies specific to economics suggest not (Abrevaya and Hamermesh, 2012; Card et al., 2020).

[13]Specifically, men and women publishing in the same field face the same pool of referees, so controlling for that pool should reduce gender differences in readability if specialisation contributes to it.

[14]Relatedly, perhaps female-authored research is more provocative and therefore warrants more scrutiny. Yet if this explained the gap, controlling for *JEL* classification should reduce (or eliminate) it—unless women's work is systematically more provocative even among researchers in very narrow fields. There is some evidence for this hypothesis—provocative work is (presumably) highly cited work and recent female-authored papers published in top economics journals are cited more (Card et al., 2020; Hengel and Moon, 2020). Yet more provocative, cited research would probably be published at higher rates—and there is no evidence women's paper's are more frequently accepted (Ceci et al., 2014).

[15]This is a form of biased referee assignment identified in Theorem 1. It would also apply if the readability gap reflects referees' apathy for women's work. Readability is particularly relevant when interest in—and knowledge about—the topic is low (Fass and Schumacher, 1978; Klare, 1976). Thus, a gap could emerge if editors fail to assign interested and knowledgable referees to female-authored papers.

response to audience member remarks? Recent evidence suggests female speakers are indeed given a harder time (Dupas et al., 2021). Nevertheless, most conference and seminar participants are also current (or future) journal referees. Neutral peer review feedback is inconsistent with non-neutral conference/seminar feedback when originating from the same group—especially since gender neutrality is emphasised in both environments.

Perhaps women focus on writing at the expense of some other aspect of a paper due to a comparative advantage? Women's chosen publication strategy results in similar (or lower) acceptance rates and longer review times compared to the one employed by men. If men and women are equally capable researchers then writing well cannot be a comparative advantage and at the same time be strictly dominated by another strategy.[16]

In the universe of straightforward alternatives, this leaves us with one: female economists are less capable researchers. As mentioned earlier, factors correlated with gender but actually related to competency should decline when appropriate proxies are included. The sample itself is one such proxy—these are, after all, only articles published in the top four economics journals. Adding other controls—author seniority, institution, total article count, citations and published order in an issue—has no effect.[17] The gap is widest for the most productive economists and even exists among articles originally released as NBER working papers—both presumably very clear signals of merit. Indeed, contemporary female-authored papers published in a top-four economics journal are, in fact, cited more than male-authored papers (Hengel and Moon, 2020).

Yet I cannot rule out the possibility that women's work is systematically worse than men's in a way that is somehow not full captured by citations, proxies for author prominence and seniority or author-specific fixed effects—or that the female and male authors in Section 4.3.3 are not really equivalent. And if this is true, editors and referees *should* select and peruse our papers more carefully—a byproduct of which could be better written papers after-the-fact or more attractive prose compensating for structural weaknesses before it.

"Quality" is subjective; measurement, not easy. Nevertheless, attempts using citation counts and journal acceptance rates do not indicate that men's research is any better, conditional on publication: as discussed in Section 4.1, men's and women's papers are accepted at similar rates unconditional of quality. As already mentioned, recent research specific to economics suggests female-authored papers may be cited more conditional on publication (Card et al., 2020; Grossbard et al., 2021; Hengel and Moon, 2020).[18]

More complicated, multi-factor explanations could resolve inconsistencies present when each is analysed in isolation. Perhaps female economists are perfectionists, and it gets stronger with age?[19] Or, a preference for writing well coupled with unaccounted for co-author characteristics could combine to cause women's more readable papers *and* their increasing readability.[20] Alternatively, measurement error and/or co-variate controls could have interacted with gender in ways I did not anticipate.

---

[16]Assuming men and women are equally capable researchers, women would only emphasise a particular aspect of a paper at the expense of others if doing so achieved a similar outcome/effort trade-off as the one employed by men. As discussed in Section 3.4, however, the outcome/effort combination women *currently* experience appears to be strictly worse than men's.

[17]Published order in an issue was introduced as a set of indicator variables in an earlier version of this paper (Hengel, 2016, pp. 42 and 44).

[18]Nevertheless, a significant amount of research finds evidence of bias against women in the decision to cite, unconditional of publication (Dion et al., 2018; Ferber, 1986; Ferber, 1988; Koffi, 2021). This suggests that citations under-estimate the quality of female-authored work.

[19]While women score higher on maintaining order (Feingold, 1994)—a trait including organisation and perfectionism—significant differences are not universally present in all cultures (Costa et al., 2001), and differences that are present appear to decline—or even reverse—as people age (Weisberg et al., 2011).

[20]This might occur, for example, if women are excluded from male networks as t increases; consequently, senior female economists may be more likely to co-author with other women than junior female economists. Relatedly, women may have preferred to have written their $t = 1$ publication more clearly, but senior male co-authors held them back; at $t = 3$, they enjoy more freedom to achieve their desired (higher) readability by writing on their own or with other women. As I show in an earlier version of this paper, however, as t increases, women are more likely to co-author with men, while men are more likely to co-author with women (Hengel, 2016, Table 12, p. 25).

Still, no explanation matches the simplicity of biased referees and/or editors. Coherence and economy do not establish fact, but they are useful guides. This single explanation neatly accounts for all observed patterns. If reviewers apply higher standards to female-authored papers, they will be rejected more often and/or subject to tougher review. Added scrutiny should improve exposition but prolong publication. Women would internalise the rewards they receive from writing more clearly, accounting for their better writing over time.[21]

---

[21] In support of this hypothesis, existing and ongoing research suggests female workers are held to higher standards in job assessments: they are acknowledged less for creativity and technical expertise, their contributions are infrequently connected to business outcomes; guidance or praise supervisors do offer is vague and tends to under-estimate women's "potential" (Benson et al., 2021; Correll and Simard, 2016). Students display a similar bias. Data from Rate My Professors suggest female lecturers should be "helpful", "clear", "organised" and "friendly". Men, instead, are praised (and criticised) for being "smart", "humble" or "cool" (Schmidt, 2015). A study of teaching evaluations similarly finds students value preparation, organisation and clarity in female instructors; their male counterparts are considered more knowledgable, praised for their "animation" and "leadership" and given more credit for contributing to students' intellectual development (Boring, 2017).

F Section 3.2, suplemental output

F.1 Readability differences across journals

Table F.1 shows the coefficients on the journal dummies in column (2), Table 3. They compare *AER*'s readability to the readability of *Econometrica*, *JPE* and *QJE*.

TABLE F.1: Journal readability, comparisons to *AER*

	Flesch Reading Ease	Flesch-Kincaid	Gunning Fog	SMOG	Dale-Chall
Econometrica	−12.25***	−4.42***	−4.23***	−2.58***	−0.67***
	(1.92)	(0.41)	(0.47)	(0.38)	(0.16)
JPE	−5.54***	−3.98***	−3.38***	−1.80***	0.18
	(1.91)	(0.41)	(0.47)	(0.38)	(0.16)
QJE	1.51**	−0.01	0.30***	0.21***	0.27***
	(0.61)	(0.13)	(0.08)	(0.06)	(0.05)
No. observations	9,117	9,117	9,117	9,117	9,117

Notes. Figures are the estimated coefficients on the journal dummy variables from column (2) in Table 3. Each contrasts the readability of the journals in the left-hand column with the readability of *AER*. Standard errors clustered on editor in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Figure F.1 displays results from an ordinary least squares regression on the Dale-Chall score; regressors are: (i) ratio of female co-authors (papers with fewer than 50 percent female authors are classified as male, see Section 2.2); (ii) dummies for each primary *JEL* code; (iii) interactions from (i) and (ii); (iv) controls for editor, journal, year, institution and English fluency; and (v) quality controls—citation count, max. T fixed effects (author prominence) and max. t (author seniority). Codes A, B, M and P are dropped due to insufficient number of female-authored papers. (Each had fewer than 10 papers authored only by women; no paper is classified under category Y.) Due to small samples—particularly of female authors—Figure F.1 includes 563 articles from *AER Papers & Proceedings*.[22]
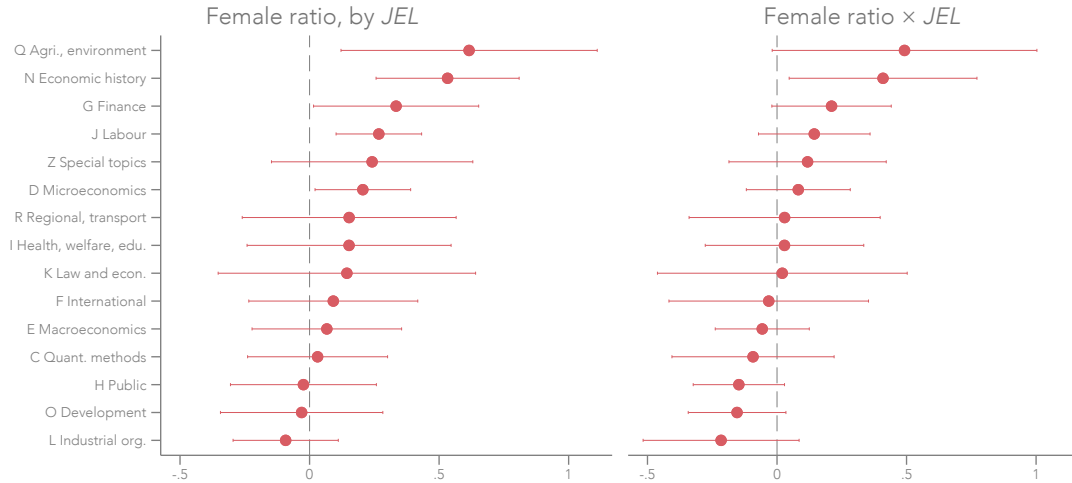
FIGURE F.1: Gender differences in readability, by *JEL* classification

Notes. Estimates from an OLS regression of:

$$R_j = \beta_0 + \beta_1 \text{female ratio}_j + \boldsymbol{\beta_2} J_j + \boldsymbol{\beta_3} \text{female ratio}_j \times J_j + \boldsymbol{\theta} X_j + \varepsilon_j,$$

where R_j is the readability score for article j; female ratio$_j$ is paper j's ratio of female authors to total authors (papers with fewer than 50 percent female authors are classified as male, see Section 2.2); J_j is a 15×1 column vector with kth entry a binary variable equal to one if article j is classified as the kth *JEL* code; X_j is a vector of editor, journal, year, institution and English language dummies, N_j (number of co-authors on paper j) and quality controls (citation count (asinh), max. T fixed effects (author prominence) and max. t (author seniority)); ε_j is the error term. Left-hand graph shows marginal effects of female ratio for each *JEL* code ($\beta_1 + \beta_3^k$). Right-hand graph displays interaction terms (β_3^k). Horizontal lines represent 90 percent confidence intervals from standard errors adjusted for clustering on editor.

Points in Figure F.1 reflect marginal effects across *JEL* classification; bars represent 90 percent confidence intervals from standard errors clustered by editor. The mean effect at observed *JEL* codes is 0.17 (standard error 0.054). This estimate coincides with results in Table 3—women's papers require about two fewer months of schooling to understand—and is highly significant.

Women earn higher marks for clarity in 12 out of 15 categories; only five are at least weakly significant: Q (Agricultural and Natural Resource Economics; Environmental and Ecological Economics), N (Economic History), G (International Economics/Finance), J (Labour Economics) and D (Microeconomics). Men may be better writers in L (Industrial Organisation), O (Economic Development, Innovation, Technological Change, and Growth) and H (Public Economics); none, however, are statistically different from zero. Figure F.1's right-hand graph displays coefficients from interacting the ratio of female co-authors with each *JEL* code. N is (weakly) significantly above the mean; remaining categories are not statistically different from the mean effect.

In general, sample sizes are small and estimates imprecise—only Labour Economics and Microeconomics contain more than 100 papers written only by women (the others average 35). Nevertheless, Figure F.1 suggests two things. First, the mostly insignificant interaction terms indicate outlier fields are probably not driving journals' gender readability gap—nor is any specific field bucking the trend. Second, the number of women in a field appears to have little effect on the size of the gap: Agriculture/

---

[22]See Hengel (2016, pp. 42–43) for a version of Figure F.1 excluding *AER Papers & Proceedings* articles.

Environment has one of the lowest concentrations of female-authored papers—but Economic History has one of the highest (Labour Economics falls between the two). Of course, Economic History papers are still overwhelmingly—as in 74 percent—penned just by men. But given the readability gap is present in subfields with both above- and below-average rates of sole female authorship, women may need to be better writers even where more of them publish.

F.3   Author-level analysis

In this appendix, I analyse readability at the author-level. To disaggregate the data, each article is duplicated N_j times, where N_j is article j's number of co-authors; observation $j_k \in \{1, \ldots, N_j\}$ is assigned article j's kth author. I then estimate the dynamic panel model in Equation (F.1):

$$R_{j_{it}} = \beta_0\, R_{it-1} + \beta_1\, \text{female ratio}_j + \beta_2\, \text{female ratio}_j \times \text{male}_i + \boldsymbol{\theta}\, \mathbf{X}_j + \alpha_i + \varepsilon_{it}. \qquad \text{(F.1)}$$

$R_{j_{it}}$ is the readability score for article j—author i's tth top-four publication; R_{it-1} is the corresponding value of author i's $t - 1$th top-four paper. Gender enters twice—the binary variable male_i and female ratio_j—to account for author i's sex and the sex of his co-authors, respectively (papers with fewer than 50 percent female authors are classified as male, see Section 2.2). $\mathbf{X}_j$ is a vector of observable controls. It includes: editor, journal, institution and English fluency dummies, controls for blind review and quality—citation count (asinh), max. T (author prominence) and max. t (author seniority)—and N_j to control for author i's proportional contribution to paper j.[23] α_i are author-specific effects and ε_{it} is an idiosyncratic error. α_i are eliminated by first-differencing. For each time period, endogeneity in the lagged dependant variable is instrumented with up to five earlier lags (Arellano and Bover, 1995; Blundell and Bond, 1998). To account for duplicate articles, the regression is weighted by $1/N_j$. Standard errors are clustered at the level of the author.

Table F.2 displays results. Rows one and two present contemporaneous marginal effects on co-authoring with women for female (β_1) and male ($\beta_1 + \beta_2$) authors, respectively. Both estimates are positive—everyone writes more clearly when collaborating with women—although statistically significant only for female authors. Marginal effects for women are up to twice as large as those shown in Table 3; they suggest women write 2–6 percent better than men.

Coefficients on the lagged dependant variables are small, suggesting readability is mostly determined contemporaneously. Nevertheless, their uniform positivity and occasional significance indicate some persistence. Table F.2's second panel reports test statistics of model fit. Tests for serial correlation indicate no model misspecification. p-values on the overall Hansen test statistic hover between 0.79–0.96, thus failing to reject that the instruments are valid; however, their very high values suggests the model may suffer from instrument proliferations which weakens the test (for a discussion, see Roodman, 2009). The p-value on the Sargan test also fails to reject that the instruments are valid at traditional significance thresholds; but although it is not vulnerable to instrument proliferation, it does require homoskedastic errors. Additional tests (available on request) suggest results are not sensitive to including the full set of (non-collapsed) instruments or to reductions in the number of instruments. Given the possibility of instrument proliferation, however, results in Table F.2 should be interpreted with caution.

---

[23]To reduce the number of instruments (and thanks to a high degree of correlation with editor fixed effects) year fixed effects are omitted. Results are similar when they are included (see, *e.g.*, Hengel, 2017, p. 17).

TABLE F.2: Gender differences in readability, author-level analysis

	Flesch Reading Ease	Flesch-Kincaid	Gunning Fog	SMOG	Dale-Chall
Female ratio for women (β_1)	2.28**	0.25	0.49**	0.37**	0.20**
	(0.90)	(0.20)	(0.24)	(0.17)	(0.09)
Female ratio for men ($\beta_1 + \beta_2$)	1.18	0.26	0.26	0.13	0.02
	(1.37)	(0.29)	(0.34)	(0.24)	(0.13)
Female ratio$\times$male (β_2)	-1.09	0.00	-0.22	-0.24	-0.18
	(1.59)	(0.34)	(0.40)	(0.28)	(0.15)
Lagged score (β_0)	0.04**	0.05**	0.03	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
No. observations	9,181	9,181	9,181	9,181	9,181
Tests of instrument validity					
Hansen test (p-value)	0.79	0.96	0.96	0.97	0.84
Sargan test (p-value)	0.36	0.11	0.31	0.16	0.36
z-test for no serial correlation					
Order 1	-18.48	-13.64	-14.90	-17.51	-18.71
Order 2	0.52	-0.43	0.13	0.45	0.10
Editor effects	✓	✓	✓	✓	✓
Blind review	✓	✓	✓	✓	✓
Journal effects	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓
Quality controls	✓^{2}	✓^{2}	✓^{2}	✓^{2}	✓^{2}
Native speaker	✓	✓	✓	✓	✓

Notes. Figures from first-differenced, IV estimation of Equation (F.1) (Arellano and Bover, 1995; Blundell and Bond, 1998) where instruments have been collapsed to create one instrument for each variable and lag distance. Female ratio (women): contemporaneous marginal effect of a paper's female co-author ratio for female authors (β_1); female ratio (men): analogous effect for male authors ($\beta_1 + \beta_2$). (The variable defines papers with a strict minority of female authors as male-authored; for papers with 50 percent or more female authors, it is the ratio of female authors on a paper. See Section 2.2 for more details.) z-statistics for first- and second-order autocorrelation in the first-differenced errors (Arellano and Bond, 1991); null hypothesis no autocorrelation. Quality controls denoted by ✓^{2} include citation count (asinh), max. T (author prominence) and max. t (author seniority). Standard errors clustered on author (in parentheses). ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

G Section 3.3, suplemental output

G.1 *Table 5, full output (first and final columns)*

Table G.1 displays coefficients from estimating Equation (1) using OLS. The first row displays coefficients on working paper score (R_{jW}); the second row shows the coefficient on female ratio (papers with fewer than 50 percent female authors are classified as male, see Section 2.2); the third rows shows the coefficient on the interaction between blind review and the ratio of female authors on a paper. (All three coefficients are also shown in the first panel of Table 5.) Remaining rows present estimated coefficients on the other (non-fixed effects) control variables: N_j (number of co-authors), max. t (author seniority), max. T (author prominence), number of citations (asinh) and dummy variables equal to one if article j is authored by at least one native English speaker or is classified as theory, empirical or other. Similarly, Table G.2 displays coefficients from estimating Equation (2). The coefficients on female ratio and its interaction with blind review correspond to estimates presented in the second panel of Table 5.

As discussed in Section 3.3.2, we do not observe the citations papers would have received had they not undergone peer review. Nevertheless, Table G.1 suggests a negative (albeit insignificant) relationship between published readability *conditional* on draft readability; Table G.2 suggests a negative (but again insignificant) relationship between citations and the change in readability that occurs *during* peer review. Thus, they tentatively suggest that the readability revisions women are asked to make during peer review may not ultimately improve the quality of their papers.[24]

---

[24]The coefficient on citations is negative **only** when controlling for R_{jW} or using the change in readability as the dependant variable. Otherwise, readability positively correlates with both working paper and published paper readability (see Appendix D.2; results for draft readability and using the specific sample from Table 5 available on request).

TABLE G.1: Table 5 (first panel), full output

	Flesch Reading Ease	Flesch-Kincaid	Gunning Fog	SMOG	Dale-Chall
R_{jW}	0.833***	0.738***	0.758***	0.786***	0.846***
	(0.020)	(0.036)	(0.034)	(0.026)	(0.016)
Female ratio	1.470**	0.531**	0.541**	0.314**	0.176***
	(0.624)	(0.206)	(0.223)	(0.149)	(0.055)
Blind review	1.764***	−0.515***	−0.833***	−0.348***	−0.229***
	(0.231)	(0.078)	(0.097)	(0.055)	(0.021)
Blind×female ratio	−2.145	−0.922	−0.931	−0.594	−0.202
	(3.841)	(0.830)	(0.799)	(0.562)	(0.195)
Max. t	−0.031	−0.007	−0.005	−0.004	−0.006**
	(0.055)	(0.014)	(0.015)	(0.009)	(0.003)
Max. T	0.031	0.007	0.006	0.004	0.003
	(0.044)	(0.009)	(0.010)	(0.007)	(0.002)
N_j	0.194	0.078	0.096	0.065	0.009
	(0.173)	(0.059)	(0.065)	(0.041)	(0.011)
No. citations (asinh)	−0.172	−0.046	−0.045	−0.034	−0.003
	(0.129)	(0.043)	(0.052)	(0.030)	(0.014)
Native speaker	−0.376	−0.043	−0.045	−0.048	−0.043
	(0.405)	(0.142)	(0.182)	(0.110)	(0.028)
Theory	−0.078	0.006	−0.075	−0.018	−0.044
	(0.531)	(0.126)	(0.118)	(0.085)	(0.038)
Empirical	0.084	0.155*	0.213***	0.093	−0.026
	(0.444)	(0.082)	(0.076)	(0.057)	(0.052)
Other	−0.594	−0.071	−0.093	−0.018	−0.065
	(1.060)	(0.281)	(0.270)	(0.178)	(0.066)
Constant	7.628***	−3.109***	−3.692***	−2.977***	−1.608***
	(1.805)	(0.488)	(0.601)	(0.443)	(0.211)
No. observations	1,988	1,988	1,988	1,988	1,988
Editor effects	✓	✓	✓	✓	✓
Year×Journal effects	✓	✓	✓	✓	✓

Notes. Coefficients from OLS regression of Equation (1). Coefficients in the first three rows correspond to the estimates presented in the first panel of Table 5. Standard errors clustered on editor (in parentheses). ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE G.2: Table 5 (second panel), full output

	Flesch Reading Ease	Flesch-Kincaid	Gunning Fog	SMOG	Dale-Chall
Female ratio	1.191*	0.484**	0.479**	0.272*	0.143**
	(0.701)	(0.204)	(0.219)	(0.140)	(0.062)
Blind review	1.953***	−0.439***	−0.679***	−0.255***	−0.181***
	(0.326)	(0.086)	(0.094)	(0.057)	(0.022)
Blind×female ratio	−1.754	−0.738	−0.758	−0.483	−0.196
	(2.182)	(0.524)	(0.567)	(0.395)	(0.129)
Max. t	−0.041	−0.006	−0.003	−0.003	−0.008**
	(0.054)	(0.016)	(0.018)	(0.012)	(0.004)
Max. T	0.039	0.007	0.006	0.004	0.005**
	(0.032)	(0.010)	(0.011)	(0.007)	(0.002)
N_j	0.219	0.077	0.090	0.065	0.011
	(0.231)	(0.062)	(0.067)	(0.044)	(0.019)
No. citations (asinh)	−0.187	−0.050	−0.057	−0.041	−0.005
	(0.228)	(0.059)	(0.070)	(0.044)	(0.015)
Native speaker	−0.403	−0.002	−0.033	−0.056	−0.028
	(0.496)	(0.163)	(0.185)	(0.111)	(0.031)
Theory	0.085	−0.010	−0.091	−0.016	−0.012
	(0.655)	(0.171)	(0.166)	(0.107)	(0.051)
Empirical	0.016	0.144	0.180	0.082	−0.023
	(0.466)	(0.110)	(0.118)	(0.077)	(0.041)
Other	−0.570	−0.077	−0.075	−0.008	−0.085
	(0.834)	(0.321)	(0.353)	(0.219)	(0.095)
Constant	0.669	0.441	0.515	0.279	0.056
	(1.394)	(0.430)	(0.462)	(0.267)	(0.127)
No. observations	1,988	1,988	1,988	1,988	1,988
Editor effects	✓	✓	✓	✓	✓
Year×Journal effects	✓	✓	✓	✓	✓

Notes. Coefficients from OLS regression of Equation (2). Coefficients in the first two rows correspond to the estimates presented in the second panel of Table 5. Standard errors clustered on editor (in parentheses). ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

As argued in Section 3.3, if field only impacts the readability of a paper when it is first drafted, then the change in readability between versions should not depend on it. For example, using the change in score as the dependent variable should wash out potential bias from, say, concepts in certain areas being easier to explain. Moreover, because FGLS estimates (shown in the final panel of Table 5) are almost identical to estimates using the change in readability as the dependent variable (shown in the second panel of Table 5), they may not suffer from substantial bias, either, despite only taking field broadly into account (via the empirical, theory and other dummies).

For added robustness, however, I replicate Table 5 but also control for primary *JEL* categories. Results are shown in Table G.3. Standard errors on the coefficients on female ratio are slightly higher in Table G.3 compared to Table 5. Otherwise, results and conclusions are very similar to those presented in Table 5.

TABLE G.3: Table 5, controlling for *JEL* category

| | | OLS (score) | | OLS (Δ in score) | | FGLS (score) | | | | | |
| | | | | | | Working paper | | Published paper | | Difference | |
	R_{jW}	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio
Flesch	0.84***	1.39**	−2.76	0.99	−2.41	2.56**	−2.23	3.55***	−4.64**	0.99	−2.41
	(0.02)	(0.61)	(4.11)	(0.71)	(2.15)	(1.14)	(3.81)	(1.11)	(2.22)	(0.69)	(2.11)
Flesch–Kincaid	0.74***	0.55***	−0.96	0.44**	−0.82	0.40	−0.55	0.84***	−1.37***	0.44**	−0.82
	(0.04)	(0.21)	(0.88)	(0.21)	(0.55)	(0.27)	(0.83)	(0.29)	(0.45)	(0.21)	(0.54)
Gunning Fog	0.76***	0.53**	−1.00	0.42*	−0.87	0.44	−0.57	0.86***	−1.44**	0.42*	−0.87
	(0.03)	(0.21)	(0.84)	(0.23)	(0.59)	(0.28)	(0.93)	(0.31)	(0.56)	(0.23)	(0.58)
SMOG	0.79***	0.29*	−0.67	0.22	−0.58	0.32*	−0.42	0.54***	−1.00**	0.22	−0.58
	(0.02)	(0.14)	(0.60)	(0.15)	(0.41)	(0.18)	(0.67)	(0.19)	(0.42)	(0.15)	(0.40)
Dale–Chall	0.86***	0.14***	−0.21	0.11	−0.18	0.25***	−0.19	0.36***	−0.37*	0.11	−0.18
	(0.02)	(0.05)	(0.22)	(0.07)	(0.14)	(0.09)	(0.26)	(0.09)	(0.20)	(0.06)	(0.13)
No. observations		1,703		1,703		1,703		1,703		1,703	
Editor effects		✓		✓		✓		✓			
Journal×Year effects		✓		✓		✓		✓			
N_j		✓²		✓²		✓²		✓²			
Quality controls		✓		✓		✓		✓			
Native speaker		✓		✓		✓		✓			
Theory/emp. effects		✓		✓		✓		✓			
JEL effects		✓		✓		✓		✓			

Notes. Columns display estimates identical to those in Table 5, except that fixed effects for primary *JEL* categories are included in all specifications. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

G.3   Blind review event study

In this section, I show a crude event study illustrating the impact of blind review on the gender readability gap formed during peer review. To implement it, I re-estimate Equation (2) without controlling for blind review, but otherwise accounting for the same factors in Table 5. Figure G.1 plots residuals from this regression for papers published in the *AER* and *QJE* 8–9 years, 6–7 years, *etc.* before they switched to or from single-blind review (or the advent of the internet) and 2–3 years, 4–5 years, *etc.* afterwards.

Most graphs in Figure G.1 suggest a discontinuity in women's unexplained changes in readability at the introduction of single-blind review (or the internet). For men, however, unexplained changes to readability appear largely unaffected by double-blind review, conditional on included controls. Figure G.1 therefore tentatively suggests that women may benefit (on average) from a policy of double-blind review, while men are less affected by it.
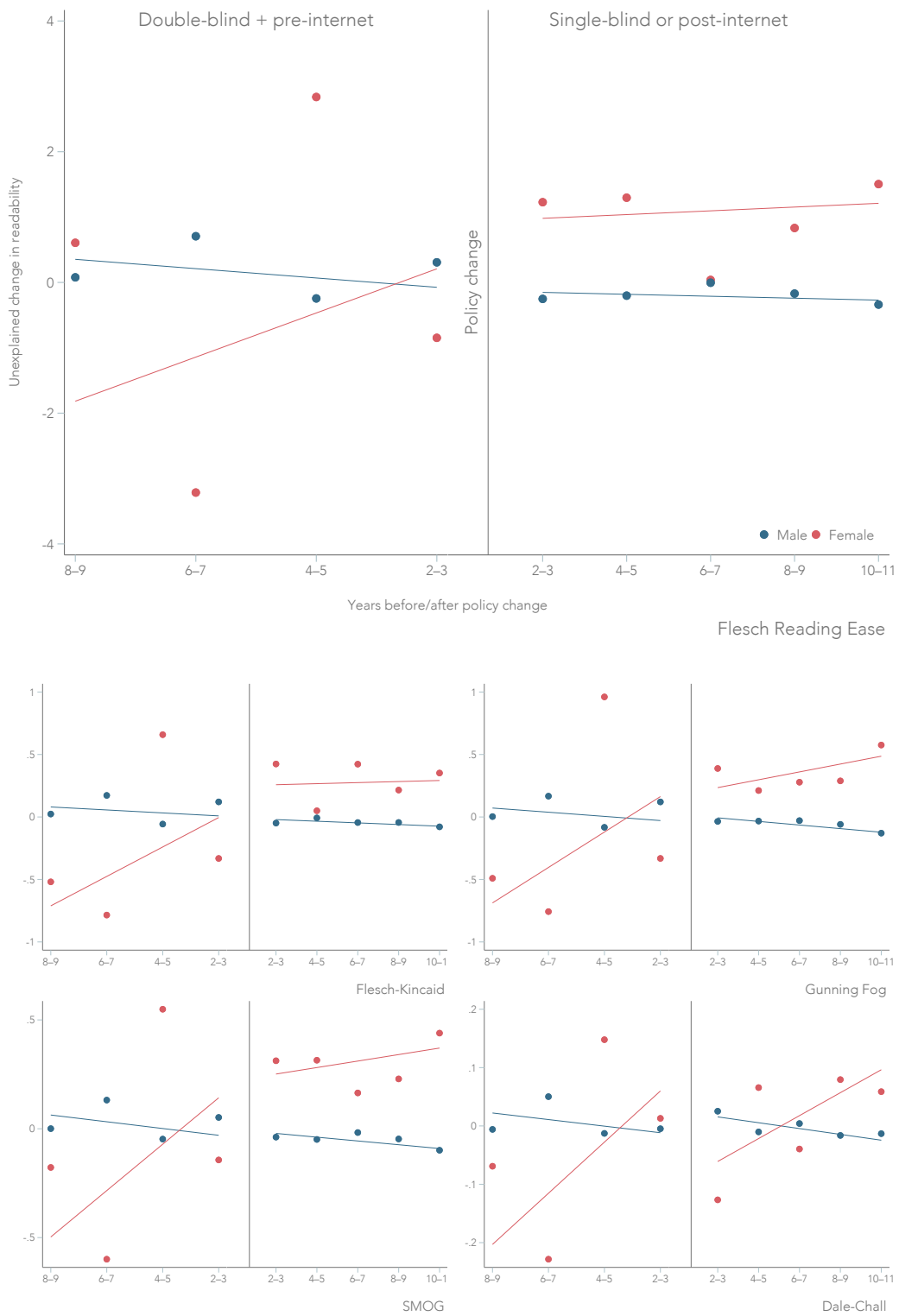
FIGURE G.1: Blind review event study

Notes. Graphs plot average residuals for papers published in the *AER* and *QJE* 8–9 years, 6–7 years, *etc.* before the journal switched from double-blind review pre-internet to single-blind review or double-blind review post-internet and 2–3 years, 4–5 years, *etc.* afterwards. Residuals generated by re-estimating Equation (2) accounting for the same controls in Table 5 with the exception of blind review.

G.4 Semi-blind review

[Table 5](#) tentatively suggests double-blind review may have successfully reduced peer review's impact on the gender readability gap *before* the internet. In this appendix, I show evidence (also tentative) suggesting that it may have been less effective *after* the internet. In particular, I re-estimated [Equation (2)](#) on the sample of articles published in or after 1998 and defined a dummy variable equal to one if a journal had in place an official policy of double-blind review at the time a paper was published. The coefficients on female ratio (papers with fewer than 50 percent female authors are classified as male, see [Section 2.2](#)) and its interaction with the redefined indicator of blind review are presented in [Table G.4](#). They do not suggest that blind review was able to alleviate the gender readability gap once referees could easily determine authors' identities by simply Googling them.

Editors knew submitting authors' identities—and therefore genders—both before and after the internet as well as under single- and double-blind review. Thus, the reversed gap in double-blind review pre-internet ([Table 5](#)) and positive gap post-internet ([Table G.4](#)) may suggest bias from referees—as opposed to editors—drives observed gender differences in readability.[25] Nevertheless, this conclusion is based on noisy (often insignificant) estimates and should therefore only be made cautiously.

TABLE G.4: The impact of double-blind review after the internet

	Flesch Reading Ease	Flesch-Kincaid	Gunning Fog	SMOG	Dale-Chall
Female ratio	1.21	0.39	0.32	0.11	0.17*
	(0.97)	(0.32)	(0.33)	(0.19)	(0.08)
Blind×fem. ratio	0.02	0.31	0.44	0.39	−0.16
	(1.74)	(0.49)	(0.57)	(0.38)	(0.17)
No. observations	1,380	1,380	1,380	1,380	1,380
Editor effects	✓	✓	✓	✓	✓
Journal×Year effects	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓
Quality controls	✓²	✓²	✓²	✓²	✓²
Native speaker	✓	✓	✓	✓	✓
Theory/emp. effects	✓	✓	✓	✓	✓

Notes. Table shows coefficients on female ratio (papers with fewer than 50 percent female authors are classified as male, see [Section 2.2](#)) and its interaction with an indicator variable equal to one if a journal had in place an official policy of double-blind review at the time a paper was published. Results from estimating [Equation (2)](#) on the sample of articles published in 1998 or afterwards.

---

[25]Many thanks to an anonymous referee for suggesting this idea.

Figure G.2 displays a histogram of the length of time between the date an author releases his draft paper as an NBER working paper and the date he submits it for peer review at *Econometrica*. Female represents papers with at least one female co-author (41 articles); male represents papers with no female co-authors (187 articles).

Figure G.2 suggests most manuscripts are submitted to peer review at the same time or *before* they are released as NBER Working Papers—and this is especially true for papers with at least one female author. This suggests that the estimates presented in Table 5 reflect gender differences in changes made to manuscripts while those manuscripts are indeed under review at the journal in which they will be eventually published.
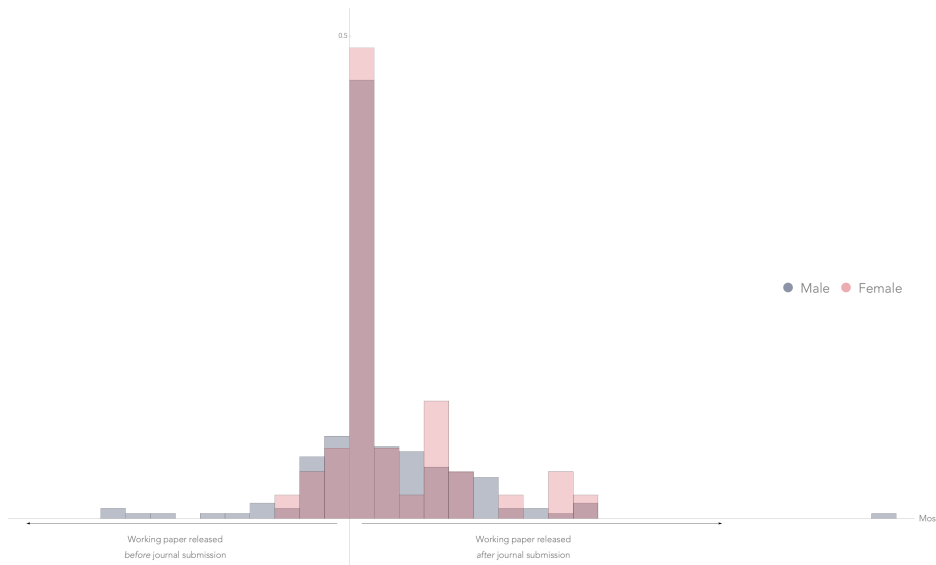
FIGURE G.2: Distribution of months between NBER release and journal submission

Notes. Sample 228 articles published in *Econometrica*. Pink represents papers with at least one female co-author (41 articles); blue are papers with no female co-authors (187 articles). Figure shows the distribution of the time difference (in months) between a paper's release as an NBER Working Paper and its submission to *Econometrica* (where it is eventually published). Observations on the right-hand-side of the y-axis were submitted to peer review first and released as working papers second; observations on the left-hand-side of the y-axis were released as working papers first and submitted to peer review second.

G.6 Abstract word limits

In Section 3.3, I argue that the gender gap in the changes in readability between draft and final versions of a paper likely occur because of the peer review process. Yet NBER working paper abstracts can be of any length while abstracts published in *Econometrica* and *AER* cannot—they are restricted to 150 and 100 words, respectively. Observed readability gaps could consequently result from gender differences in how authors conform to these limits.

To test this hypothesis, I replicated the analysis described Section 3.3.2 (and shown in Table 5) on the subset of articles with draft abstracts below the official minimum word limit of the journals in which they were eventually published. Results are shown in Table G.5. Despite dropping about 40 percent of observations, coefficient magnitudes are similar to those reported in Table 5; standard errors are somewhat larger.[26]

---

[26]Results are similar if I also include a control for the number of words in the working paper version of the abstract (available on request).

TABLE G.5: Table 5, draft abstracts below official word limits

| | OLS (score) | | | OLS (Δ in score) | | FGLS (score) | | | | | |
| | | | | | | Working paper | | Published paper | | Difference | |
	R_{jW}	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio
Flesch	0.84***	0.81	−5.38**	0.57	−5.14	1.46	−1.50	2.03	−6.64*	0.57	−5.14
	(0.02)	(1.05)	(2.72)	(0.98)	(3.86)	(1.59)	(4.12)	(1.68)	(3.77)	(0.95)	(3.75)
Flesch–Kincaid	0.74***	0.54*	−1.82***	0.59*	−1.72**	−0.17	−0.38	0.42	−2.10**	0.59*	−1.72**
	(0.05)	(0.32)	(0.52)	(0.31)	(0.85)	(0.38)	(1.12)	(0.35)	(0.88)	(0.31)	(0.83)
Gunning Fog	0.76***	0.55*	−1.77***	0.58*	−1.62	−0.10	−0.61	0.48	−2.23**	0.58*	−1.62
	(0.04)	(0.31)	(0.54)	(0.32)	(1.03)	(0.42)	(1.10)	(0.38)	(0.89)	(0.31)	(1.00)
SMOG	0.79***	0.24	−1.05**	0.24	−0.94	0.02	−0.53	0.26	−1.47**	0.24	−0.94
	(0.03)	(0.19)	(0.43)	(0.20)	(0.73)	(0.28)	(0.70)	(0.25)	(0.58)	(0.19)	(0.71)
Dale–Chall	0.84***	0.21**	−0.45***	0.18**	−0.47*	0.22*	0.13	0.40***	−0.34	0.18**	−0.47**
	(0.02)	(0.09)	(0.12)	(0.09)	(0.24)	(0.12)	(0.26)	(0.12)	(0.34)	(0.09)	(0.24)
No. observations	1,289			1,289		1,289		1,289		1,289	
Editor effects		✓		✓		✓		✓		✓	
Journal×Year effects		✓		✓		✓		✓		✓	
N_j											
Quality controls		✓²		✓²		✓²		✓²		✓²	
Native speaker		✓		✓		✓		✓		✓	
Theory/emp. effects		✓		✓		✓		✓		✓	

Notes. Estimates are identical to those in Table 5, except that the sample includes only papers with an NBER abstract below the official minimum word limit of the journal in which it was eventually published. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

H   Section 3.4, supplemental output

H.1   Alternative year fixed effects

Table H.1 and Table H.2 replicate Table 6, replacing acceptance year fixed effects with fixed effects for submission and publication years, respectively.

TABLE H.1: Table 6, submission year effects

	1970–2015					1990–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female ratio	2.777*	4.225**	4.257**	3.234**	4.273**	4.648**	4.496**
	(1.404)	(1.609)	(1.605)	(1.584)	(1.602)	(2.058)	(2.118)
Max. t	−0.105***	−0.108***	−0.106***	−0.106***	−0.105***	−0.096**	−0.101**
	(0.034)	(0.034)	(0.034)	(0.033)	(0.034)	(0.038)	(0.039)
No. pages	0.156***	0.154***	0.154***	0.155***	0.154***	0.152***	0.144***
	(0.021)	(0.020)	(0.021)	(0.021)	(0.021)	(0.025)	(0.028)
N_j	1.115***	1.093***	1.089***	1.101***	1.097***	1.035***	0.962***
	(0.246)	(0.243)	(0.246)	(0.245)	(0.244)	(0.303)	(0.329)
Order	0.113*	0.107*	0.108*	0.111*	0.108*	0.083	0.109
	(0.059)	(0.060)	(0.059)	(0.059)	(0.059)	(0.122)	(0.124)
No. citations (asinh)	−0.473***	−0.487***	−0.483***	−0.471***	−0.491***	−1.239***	−1.256***
	(0.167)	(0.168)	(0.166)	(0.166)	(0.168)	(0.286)	(0.274)
Flesch	−0.027**	−0.027**	−0.026**	−0.027**	−0.026**	−0.047**	−0.051**
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.021)	(0.022)
Theory	0.324	0.374	0.306	0.321	0.302	−0.650	−0.625
	(0.932)	(0.944)	(0.938)	(0.934)	(0.937)	(0.747)	(0.836)
Empirical	3.092**	3.476**	3.139**	3.140**	3.087**	3.174**	2.923**
	(1.375)	(1.459)	(1.378)	(1.366)	(1.377)	(1.343)	(1.426)
Other	−0.693	−0.698	−0.668	−0.678	−0.680	−0.385	−1.136*
	(0.639)	(0.636)	(0.634)	(0.639)	(0.640)	(0.630)	(0.653)
Mother			−7.149***		−10.672***	−20.550***	−20.576***
			(2.362)		(3.282)	(2.636)	(3.198)
Birth				−3.771	5.886	16.318***	16.660***
				(3.423)	(4.474)	(3.454)	(3.448)
Constant	16.512***	16.616***	16.608***	16.535***	16.620***	25.276***	26.242***
	(1.273)	(1.304)	(1.299)	(1.277)	(1.305)	(2.145)	(2.248)
R^2	0.574	0.577	0.576	0.575	0.576	0.578	0.586
No. observations	2,623	2,608	2,623	2,623	2,623	1,278	1,278
Editor effects	✓	✓	✓	✓	✓	✓	✓
Year effects	✓	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓	✓
JEL (primary) effects							✓

Notes. Estimates are identical to those in Table 6, except that submission year effects are used instead of acceptance year effects. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

	1970–2015					1990–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female ratio	5.390***	6.753***	6.751***	5.718***	6.771***	9.365***	9.429***
	(1.696)	(2.117)	(2.116)	(2.004)	(2.113)	(2.786)	(2.799)
Max. t	−0.152***	−0.156***	−0.153***	−0.153***	−0.152***	−0.144***	−0.147***
	(0.040)	(0.039)	(0.040)	(0.040)	(0.040)	(0.046)	(0.047)
No. pages	0.196***	0.196***	0.195***	0.196***	0.195***	0.226***	0.210***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.040)	(0.043)
N_j	1.292***	1.282***	1.270***	1.283***	1.279***	1.723***	1.593***
	(0.314)	(0.307)	(0.314)	(0.310)	(0.310)	(0.438)	(0.451)
Order	0.202***	0.200***	0.197***	0.200***	0.198***	0.429***	0.456***
	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)	(0.154)	(0.150)
No. citations (asinh)	−0.339*	−0.365*	−0.351*	−0.338*	−0.361*	−0.571	−0.604
	(0.199)	(0.198)	(0.198)	(0.200)	(0.199)	(0.430)	(0.410)
Flesch	−0.020	−0.019	−0.019	−0.020	−0.019	−0.043	−0.048*
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.027)	(0.028)
Theory	0.651	0.624	0.621	0.646	0.615	0.635	−0.309
	(0.967)	(0.990)	(0.971)	(0.968)	(0.970)	(1.007)	(1.371)
Empirical	2.432	2.774	2.463	2.462	2.405	2.254	1.715
	(1.822)	(1.820)	(1.812)	(1.803)	(1.820)	(1.892)	(2.130)
Other	−1.279	−1.291	−1.257	−1.267	−1.273	−1.025	−1.925**
	(0.888)	(0.860)	(0.876)	(0.886)	(0.883)	(0.896)	(0.868)
Mother			−6.599**		−10.857***	−20.476***	−20.825***
			(3.224)		(3.454)	(5.340)	(5.642)
Birth				−2.722	7.109	15.903***	15.990***
				(4.176)	(4.524)	(5.716)	(5.659)
Constant	14.460***	14.588***	14.559***	14.477***	14.580***	16.535***	18.726***
	(1.511)	(1.522)	(1.520)	(1.511)	(1.523)	(2.750)	(2.723)
R^2	0.282	0.284	0.283	0.282	0.283	0.114	0.132
No. observations	2,623	2,608	2,623	2,623	2,623	1,278	1,278
Editor effects	✓	✓	✓	✓	✓	✓	✓
Year effects	✓	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓	✓
JEL (primary) effects							✓

Notes. Estimates are identical to those in Table 6, except that publication year effects are used instead of acceptance year effects. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

H.2 Alternative thresholds for mother$_j$

Table H.3 repeats the regression presented in Table 6 column (5) using alternative age thresholds to define motherhood: mother$_j$ equals 1 if paper j's co-authors are all mothers to children younger than three (first column), four (second column), *etc.* Changing this threshold has little effect on female ratio's coefficient. The coefficients on mother$_j$ and birth$_j$ are persistently negative and positive (respectively), although magnitudes and standard errors vary. Remaining coefficients are unaffected.

TABLE H.3: Table 6 column (5), alternative thresholds for mother$_j$

	Age < 3	Age < 4	Age < 5	Age < 10	Age < 18
Female ratio	5.805**	6.130***	6.817***	6.695***	6.398**
	(2.278)	(2.277)	(2.288)	(2.330)	(2.443)
Mother	−4.823*	−9.905*	−12.085***	−9.348**	−5.244
	(2.787)	(5.107)	(3.762)	(3.856)	(3.687)
Birth	1.593	6.313	7.840	5.225	1.410
	(4.033)	(5.973)	(5.213)	(5.359)	(4.742)
Max. t	−0.146***	−0.145***	−0.145***	−0.145***	−0.145***
	(0.038)	(0.038)	(0.038)	(0.038)	(0.039)
No. pages	0.197***	0.196***	0.195***	0.196***	0.197***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
N_j	1.284***	1.286***	1.280***	1.277***	1.277***
	(0.292)	(0.292)	(0.293)	(0.292)	(0.293)
Order	0.198***	0.197***	0.195***	0.194***	0.195***
	(0.062)	(0.062)	(0.062)	(0.062)	(0.062)
No. citations (asinh)	−0.398*	−0.403**	−0.419**	−0.414**	−0.407**
	(0.200)	(0.200)	(0.199)	(0.200)	(0.200)
Flesch	−0.018	−0.017	−0.016	−0.017	−0.017
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Theory	0.331	0.299	0.290	0.294	0.309
	(0.979)	(0.974)	(0.976)	(0.980)	(0.981)
Empirical	2.341	2.293	2.267	2.279	2.297
	(1.788)	(1.803)	(1.808)	(1.805)	(1.802)
Other	−1.153	−1.152	−1.166	−1.187	−1.186
	(0.880)	(0.878)	(0.876)	(0.880)	(0.880)
Constant	14.610***	14.653***	14.714***	14.745***	14.690***
	(1.419)	(1.424)	(1.434)	(1.431)	(1.434)
R^2	0.289	0.290	0.291	0.291	0.290
No. observations	2,622	2,622	2,622	2,622	2,622
Editor effects	✓	✓	✓	✓	✓
Accepted year effects	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓

Notes. Coefficients from OLS estimation of Equation (9) at different age thresholds for mother$_j$. Column three corresponds to results presented in column (5) of Table 6. Standard errors clustered by submission year in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

H.3 Quantile regression

In Table H.4, I re-estimate gender differences in time spent under review using a quantile regression model. The first panel replicates Table 6, column (5) at the 25th, median and 75th percentiles of review times; the second panel similarly replicates the third column of Table 7.

The coefficient on female ratio is positive and significant across all three percentiles. Its magnitude is greatest in the right-tail of *Econometrica*'s distribution but is similarly sized across all percentiles when estimated using observations from both *Econometrica* and *REStud*.

TABLE H.4: Revision duration at *Econometrica* and *REStud*, quantile regression

	Econometrica			*Econometrica+REStud*		
	25th pc.	Median	75th pc.	25th pc.	Median	75th pc.
Female ratio	2.89***	5.19***	5.79***	1.99***	1.61***	2.49***
	(0.98)	(1.54)	(2.07)	(0.50)	(0.57)	(0.93)
Max. t	−0.14***	−0.17***	−0.16***	−0.13***	−0.18***	−0.16***
	(0.03)	(0.04)	(0.06)	(0.02)	(0.03)	(0.05)
No. pages	0.13***	0.21***	0.25***	0.14***	0.21***	0.25***
	(0.01)	(0.02)	(0.03)	(0.01)	(0.02)	(0.02)
N_j	0.69***	0.93***	1.26***	0.49***	0.91***	0.84***
	(0.21)	(0.24)	(0.30)	(0.14)	(0.18)	(0.21)
Order	0.08*	0.21***	0.18**	0.00	0.08**	0.05
	(0.05)	(0.04)	(0.07)	(0.03)	(0.04)	(0.04)
No. citations (asinh)	−0.39***	−0.47***	−0.59***	−0.37***	−0.37***	−0.47***
	(0.11)	(0.09)	(0.15)	(0.07)	(0.08)	(0.11)
Theory	−0.15	0.98	1.12	0.31	0.86	0.60
	(0.51)	(0.80)	(1.02)	(0.31)	(0.53)	(0.61)
Empirical	1.16	4.43	5.23*	2.15**	3.43***	4.58***
	(1.49)	(3.01)	(2.82)	(0.98)	(1.18)	(1.47)
Other	−0.56	−0.87	−1.75	−0.73	−1.06*	−1.80***
	(0.68)	(0.84)	(1.08)	(0.60)	(0.62)	(0.59)
Flesch	−0.02	−0.01	−0.03**			
	(0.01)	(0.01)	(0.02)			
Mother	−6.23	−11.33	−7.74			
	(49.95)	(12.04)	(4.93)			
Birth	6.44	7.22	9.25			
	(50.29)	(15.23)	(8.97)			
Constant	42.61***	41.19***	40.67	41.51**	38.90*	38.36**
	(3.95)	(14.72)	(25.99)	(18.84)	(21.20)	(16.11)
Pseudo R^2	0.20	0.21	0.21	0.20	0.21	0.21
No. observations	2,623	2,623	2,623	4,435	4,435	4,435
Editor effects	✓	✓	✓	✓	✓	✓
Accepted year effects	✓	✓	✓			
Journal×Accepted year effects				✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓

Notes. First panel replicates results shown in Table 6, column (5) across different percentiles of the distribution using quantile regressions; second panel similarly replicates results shown in the third column of Table 6. Robust standard errors in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

I Section 4.2, supplemental output

I.1 *Studies evaluating gender differences in acceptance rates*

TABLE I.1: Literature on gender differences in acceptance rates at peer reviewed journals

Study	Data source/context	Findings
Card et al. (2020)	*JEEA, QJE, Restat* and *REStud*	Exclusively male- and female-authored manuscripts received a revise and resubmit decision 8 and 6 percent of the time, respectively.
Blank (1991)	*AER*	12.7 and 10.6 percent of male- and female-authored papers were accepted at the *AER*, respectively.
Gilbert et al. (1994)	*JAMA*	44.8 percent of referees accepted male-authored papers as is or if suitably revised; 29.6 percent summarily rejected them. Corresponding figures for female-authored papers were 38.3 and 33.3 percent, respectively.
Handley et al. (2015)	American Fisheries Society journals	Female first-authored manuscripts were accepted 58.2 percent of the time; for male first-authored manuscripts, the corresponding figure was 62.5 percent. In no journal were female first-authored manuscripts accepted at higher rates than male first-authored manuscripts.
McGillivray and De Ranieri (2018)	25 *Nature*-branded journals	Among papers subjected to single-blind review, 23.7 percent of papers with a female corresponding author were sent out for review compared to 24.0 percent of papers with a male corresponding author. Conditional on being sent out for review, papers with a female corresponding author were accepted 44 percent of the time compared to 46 percent of papers with a male corresponding author.
Nature Neuroscience (2006)	*Nature Neuroscience*	10.9 percent of papers with a female corresponding author were accepted compared to 11.8 percent of papers with a male corresponding author.
Tregenza (2002)	Four primary research journals	There was no statistically significant overall difference in acceptance rates by gender. The average per editor acceptance rate of papers with a male first author was 40.5 percent; the average per editor acceptance rate of papers with a female first author was 34.1 percent.
Chari and Goldsmith-Pinkham (2017)	NBER Summer Institute	There was no gender difference in acceptance rates to NBER's Summer Institute.
Ceci et al. (2014)	Comprehensive overview of the literature	"When it comes to actual manuscripts submitted to actual journals, the evidence for gender fairness is unequivocal: there are no sex differences in acceptance rates." (p. 111)

Notes. Table summarises evidence on gender differences in acceptance rates at peer reviewed journals. (The list of included studies is likely not comprehensive.)

I.2 Authors' average readability scores for their first, mean and final papers

Table I.2 displays authors' average readability scores for their first, mean and final top-four papers. Grade-level scores (Flesch-Kincaid, Gunning Fog, SMOG and Dale-Chall) have been multiplied by negative one (see Section 2.1). Sample excludes authors with fewer than three publications.

As their careers advance, women do write more clearly: their average readability scores are 1–5 percent higher than the readability of their first papers; their latest papers 1–7 percent. For a man, however, his average and last paper are about as well written as his first.

TABLE I.2: Average first, mean and final top-four paper scores

	Flesch Reading Ease	Flesch-Kincaid	Gunning Fog	SMOG	Dale-Chall
Average first paper score					
Men	39.37	−13.77	−17.54	−15.35	−11.00
	(0.31)	(0.07)	(0.08)	(0.06)	(0.03)
Women	39.21	−13.81	−17.36	−15.18	−11.00
	(1.16)	(0.24)	(0.29)	(0.21)	(0.10)
Average mean paper score					
Men	39.59	−13.68	−17.41	−15.26	−11.02
	(0.19)	(0.04)	(0.05)	(0.03)	(0.02)
Women	41.23	−13.36	−16.91	−14.92	−10.91
	(0.72)	(0.15)	(0.18)	(0.13)	(0.07)
Average final paper score					
Men	39.53	−13.71	−17.40	−15.24	−11.08
	(0.33)	(0.08)	(0.09)	(0.06)	(0.03)
Women	42.06	−13.09	−16.57	−14.65	−10.90
	(1.05)	(0.21)	(0.25)	(0.18)	(0.11)
No. observations	1,675	1,675	1,675	1,675	1,675

Notes. Figures are average readability scores for authors' first, mean and last published articles. Includes only authors with three or more publications in the data. Grade-level scores have been multiplied by negative one. Standard errors in parentheses.

I.3 Table 8, tests of coefficient equality

Table I.3 tests equality of coefficients in each column of Table 8. It rejects equality between coefficients in the first and third columns at $p < 0.05$ for the Flesch Reading Ease, Flesch-Kincaid, Gunning Fog and SMOG scores.

TABLE I.3: Table 8, equality test statistics

	$t_4 = 1$ vs. 2	1 vs. 3	1 vs. 4–5	1 vs. ≥ 6	2 vs. 3
Flesch	1.05	5.67	1.16	0.81	2.16
Flesch-Kincaid	0.36	13.82	1.78	0.60	6.34
Gunning Fog	1.08	10.72	2.16	0.63	3.77
SMOG	1.44	7.94	1.87	0.65	1.93
Dale-Chall	0.01	1.64	2.75	2.48	1.00

Notes. χ^2 test statistics from Wald tests of β_1 (Equation (11)) equality across estimation results in Table 8.

J Section 4.3, supplemental output

J.1 Co-variate balance

Table J.1 compares co-variate balance pre- and post-match. The first column displays averages for the 121 female authors with at least three publications in the data. The first column of the first panel ("Pre-match means") displays corresponding averages for the 1,554 male authors with three or more publications. The first column of the second panel ("Post-match means") displays (weighted) averages for the 108 male authors matched with a female author. Gender differences are smaller post-match; t-statistics are likewise closer to zero.

TABLE J.1: Pre- and post-matching summary statistics

		Pre-match means			Post-match means		
	Women	Men	Difference	t	Men	Difference	t
$t = 1$ inst. rank	15.91	19.22	3.31	1.80	16.70	0.79	0.32
Max. citations	267.07	406.33	139.27	1.78	211.09	−55.98	−1.48
Fraction of articles per decade							
1950–59	0.00	0.01	0.01	1.57	0.00	0.00	
1960–69	0.00	0.04	0.04	2.87	0.00	0.00	−0.12
1970–79	0.01	0.11	0.09	4.72	0.02	0.00	0.25
1980–89	0.08	0.18	0.10	4.37	0.10	0.02	0.76
1990–99	0.19	0.21	0.02	1.00	0.19	0.00	0.11
2000–09	0.41	0.26	−0.15	−5.90	0.39	−0.02	−0.51
2010–15	0.31	0.20	−0.11	−4.19	0.30	−0.01	−0.15
Fraction of articles per journal							
AER	0.39	0.25	−0.14	−5.54	0.37	−0.01	−0.35
Econometrica	0.17	0.34	0.17	5.12	0.24	0.07	1.69
JPE	0.18	0.24	0.07	2.62	0.18	0.00	−0.04
QJE	0.27	0.17	−0.10	−4.79	0.21	−0.05	−1.66
Number of articles per *JEL* code							
A General	0.04	0.02	−0.02	−1.59	0.04	0.00	0.00
B Methodology	0.00	0.02	0.02	1.44	0.00	0.00	
C Quant. methods	0.64	0.81	0.17	1.03	0.57	−0.07	−0.44
D Microeconomics	1.64	1.79	0.15	0.68	1.61	−0.02	−0.12
E Macroeconomics	0.58	0.62	0.04	0.35	0.46	−0.12	−0.86
F International	0.39	0.31	−0.08	−0.85	0.29	−0.10	−0.83
G Finance	0.60	0.52	−0.07	−0.67	0.45	−0.15	−1.01
H Public	0.45	0.36	−0.10	−1.09	0.29	−0.17	−1.97
I Health, welfare, edu	0.88	0.34	−0.53	−5.40	0.52	−0.36	−2.01
J Labour	1.26	0.76	−0.49	−3.39	0.79	−0.46	−2.44
K Law and econ	0.20	0.14	−0.06	−1.14	0.14	−0.06	−0.89
L Industrial org	0.73	0.57	−0.16	−1.47	0.51	−0.21	−1.65
M Marketing/acct	0.17	0.13	−0.04	−0.93	0.12	−0.04	−0.65
N Economic history	0.29	0.14	−0.15	−2.74	0.20	−0.09	−0.96
O Development	0.86	0.52	−0.34	−2.60	0.50	−0.36	−2.21
P Economic systems	0.08	0.09	0.01	0.22	0.07	−0.02	−0.36
Q Agri., environment	0.18	0.12	−0.06	−1.20	0.09	−0.09	−1.51
R Regional, transport	0.17	0.16	−0.01	−0.16	0.06	−0.12	−2.66
Z Special topics	0.16	0.10	−0.06	−1.50	0.15	−0.01	−0.14

Notes. Sample restricted to authors with three or more publications. First panel shows pre-match summary statistics. Second panel shows post-match summary statistics. t-values for differences reported in each panel's final column.

TABLE J.2: Matched pairs

Matched pairs		Matched pairs	
Female	Male	Female	Male
Abraham, Katharine G.	Kahn, Charles M.	Kuziemko, Ilyana	Deming, David J.
Admati, Anat R.	Huang, Chi-Fu	La Ferrara, Eliana	van Soest, Arthur
Amiti, Mary	Broda, Christian	Landes, Elisabeth M.	Carlton, Dennis W.
Anderson, Siwan	Finan, Frederico	Levy, Gilat	Razin, Ronny
Ashraf, Nava	Mahajan, Aprajit	Lewis, Karen K.	Stockman, Alan C.
Athey, Susan	Haile, Philip A.	Li, Wei	Roland, Gérard
Baicker, Katherine	Shafir, Eldar	Lleras-Muney, Adriana	Kessler, Daniel P.
Bailey, Martha J.	Paserman, M. Daniele	Løken, Katrine Vellesen	Mogstad, Magne
Bandiera, Oriana	Rasul, Imran	Madrian, Brigitte C.	Lee, David S.
Barwick, Panle Jia	Winston, Clifford	Maestas, Nicole	Bettinger, Eric P.
Baxter, Marianne	Backus, David K.	Malmendier, Ulrike	Agarwal, Sumit
Bedard, Kelly	Lefgren, Lars	Matzkin, Rosa L.	Hahn, Jinyong
Bertrand, Marianne	Mullainathan, Sendhil	McConnell, Sheena	LaLonde, Robert J.
Black, Sandra E.	Kessler, Daniel P.	McGrattan, Ellen R.	Williams, Noah
Blank, Rebecca M.	Laband, David N.	Meyer, Margaret A.	Holtz-Eakin, Douglas
Boustan, Leah Platt	Abramitzky, Ran	Molinari, Francesca	Compte, Olivier
Brown, Jennifer	Vogel, Jonathan	Moser, Petra	Sunde, Uwe
Busse, Meghan R.	Zettelmeyer, Florian	Nakamura, Emi	Steinsson, Jón
Case, Anne C.	Fishman, Arthur	Ng, Serena	Hirano, Keisuke
Casella, Alessandra	Snyder, James M. (Jr.)	Niederle, Muriel	Wolfers, Justin
Chen, Xiaohong	Hahn, Jinyong	Oster, Emily	Fang, Hanming
Chen, Yan	Lange, Andreas	Pande, Rohini	Dean, Mark
Chevalier, Judith A.	Lamont, Owen A.	Paxson, Christina H.	Boldrin, Michele
Chichilnisky, Graciela	Engers, Maxim	Perrigne, Isabelle	Schmedders, Karl
Correia, Isabel	Leeper, Eric M.	Piazzesi, Monika	Schneider, Martin
Costa, Dora L.	Kahn, Matthew E.	Qian, Nancy	Ok, Efe A.
Cropper, Maureen L.	Halvorsen, Robert	Quinzii, Martine	Magill, Michael J. P.
Currie, Janet	Lavy, Victor	Ramey, Valerie A.	Bresnahan, Timothy F.
Dafny, Leemore S.	Kolstad, Jonathan T.	Reinganum, Jennifer F.	Daughety, Andrew F.
De Nardi, Mariacristina	Silverman, Dan	Reinhart, Carmen M.	Taylor, Alan M.
Demange, Gabrielle	Easley, David	Rey, Hélène	Jeanne, Olivier
Duflo, Esther	Burgess, Robin	Romer, Christina D.	Williams, John C.
Dupas, Pascaline	Urquiola, Miguel	Rose-Ackerman, Susan	Miyazaki, Hajime
Dynan, Karen E.	Ljungqvist, Lars	Rose, Nancy L.	Snyder, James M. (Jr.)
Eberly, Janice C.	Sunder, Shyam	Rosenblat, Tanya S.	Möbius, Markus M.
Eckel, Catherine C.	Dufwenberg, Martin	Rouse, Cecilia Elena	Fishman, Arthur
Edlund, Lena	Smith, Jeffrey	Sapienza, Paola	Wacziarg, Romain
Eyigungor, Burcu	Kaboski, Joseph P.	Schennach, Susanne M.	Guggenberger, Patrik
Fan, Yanqin	Sherman, Robert P.	Schmitt-Grohé, Stephanie	Leeper, Eric M.
Fernández, Raquel	Spolaore, Enrico	Schwartz, Nancy L.	Kuga, Kiyoshi
Field, Erica	Donald, Stephen G.	Shannon, Chris	Safra, Zvi
Finkelstein, Amy	Einav, Liran	Shaw, Kathryn L.	Anderson, Simon P.
Flavin, Marjorie A.	Garber, Peter M.	Spier, Kathryn E.	Ausubel, Lawrence M.
Forges, Françoise	Banks, Jeffrey S.	Stokey, Nancy L.	Smith, Bruce D.
Fortin, Nicole M.	Hyslop, Dean R.	Tenreyro, Silvana	Lloyd-Ellis, Huw
Freund, Caroline	Rose, Andrew K.	Tertilt, Michèle	Doepke, Matthias
Fuchs-Schündeln, Nicola	Woodruff, Christopher	Tesar, Linda L.	Blonigen, Bruce A.
Garfinkel, Michelle R.	Bertola, Giuseppe	Thomas, Julia K.	Khan, Aubhik
Goldberg, Pinelopi Koujianou	Levinsohn, James A.	Todd, Petra E.	Flinn, Christopher J.
Goldin, Claudia D.	Abramitzky, Ran	Vissing-Jørgensen, Annette	Veronesi, Pietro
Gopinath, Gita	Itskhoki, Oleg	Voena, Alessandra	Sunde, Uwe
Griffith, Rachel	Broda, Christian	Washington, Ebonya L.	Paserman, M. Daniele
Guerrieri, Veronica	Khan, Aubhik	White, Lucy	Yılmaz, Bilge
Hanna, Rema	Foster, Andrew D.	Whited, Toni M.	Silverman, Dan
Hastings, Justine S.	Pope, Devin G.	Williams, Heidi L.	Budish, Eric
Ho, Katherine	Kremer, Ilan	Wooders, Myrna Holtz	Gallant, A. Ronald

35

Matched pairs		Matched pairs	
Female	Male	Female	Male
Hoxby, Caroline Minter	Kessler, Daniel P.	Yariv, Leeat	Lange, Andreas
İmrohoroğlu, Ayşe	Casari, Marco	Yellen, Janet L.	Freeman, Richard B.
Jayachandran, Seema	Pop-Eleches, Cristian	Zeiler, Kathryn	Loomes, Graham
Kowalski, Amanda E.	Schrimpf, Paul	Zhuravskaya, Ekaterina	Groves, Theodore
Kranton, Rachel E.	Kosfeld, Michael		

Notes. Table lists the names of matched pairs from Section 4.3. In each panel, female members are listed first; male members second. Matches were generated in Stata using psmatch2 (Leuven and Sianesi, 2003). See Section 4.3.2 for details on the matching process.

Table J.3 displays output from time- and gender-specific regressions used to generate $\widehat{R}_{it}$ (Equation (13)).

TABLE J.3: Regression output generating $\widehat{R}_{it}$ (Equation (13))

	Women		Men	
	$t = 1$	$t = 3$	$t = 1$	$t = 3$
Flesch Reading Ease				
Female ratio	0.29	1.81	−12.65	−2.81
	(3.31)	(2.70)	(9.71)	(6.02)
Constant	39.02***	41.97***	38.19***	39.12***
	(2.47)	(1.65)	(1.13)	(1.18)
Flesch Kincaid				
Female ratio	−0.27	0.36	−0.64	0.88
	(0.68)	(0.55)	(2.12)	(1.27)
Constant	−13.63***	−13.23***	−14.03***	−14.01***
	(0.51)	(0.33)	(0.25)	(0.25)
Gunning Fog				
Female ratio	−0.40	0.66	−3.03	0.83
	(0.82)	(0.67)	(2.36)	(1.53)
Constant	−17.10***	−16.95***	−17.79***	−17.56***
	(0.62)	(0.41)	(0.27)	(0.30)
SMOG				
Female ratio	−0.22	0.43	−2.08	0.36
	(0.60)	(0.50)	(1.65)	(1.09)
Constant	−15.04***	−14.97***	−15.57***	−15.32***
	(0.45)	(0.30)	(0.19)	(0.22)
Dale-Chall				
Female ratio	−0.19	0.20	−3.49***	0.14
	(0.27)	(0.27)	(0.90)	(0.43)
Constant	−10.88***	−10.93***	−11.07***	−11.14***
	(0.20)	(0.16)	(0.10)	(0.08)

Notes. Sample restricted to matched authors. See Section 4.3.2 for details on how matches were made. Regressions weighted by the frequency observations are used in a match. Standard errors in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table J.4 and Figure J.1 replicate the analysis in Section 4.3.3 but Equation (13) controls for primary *JEL* category. $\widehat{R}_{it}$ was reconstructed at female ratio equal to 1 for women, 0 for men and for a paper classified in *JEL* categories D (microeconomics) and J (labour and demographic economics).

TABLE J.4: D_{ik} (Corollary 1), controlling for *JEL* category

	Higher standards for women ($D_{ik} > 0$)			Higher standards for men ($D_{ik} < 0$)			Mean D_{ik}
	Mean	S.D.	N	Mean	S.D.	N	All obs.
Flesch	9.83	8.15	44	−4.65	3.78	10	4.62***
							(1.60)
Flesch Kincaid	1.94	1.59	46	−1.58	1.08	6	0.99***
							(0.33)
Gunning Fog	2.56	1.99	39	−1.51	1.02	8	1.13***
							(0.39)
SMOG	1.83	1.47	37	−0.93	0.72	8	0.76***
							(0.28)
Dale-Chall	0.85	0.77	30	−0.52	0.60	5	0.28**
							(0.12)

Notes. Sample 87 matched pairs. Table displays estimates identical to those in Table 9, except that Equation (13) accounts for primary *JEL* classification dummies and $\widehat{R}_{it}$ was reconstructed at *JEL* category dummies D and J. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.
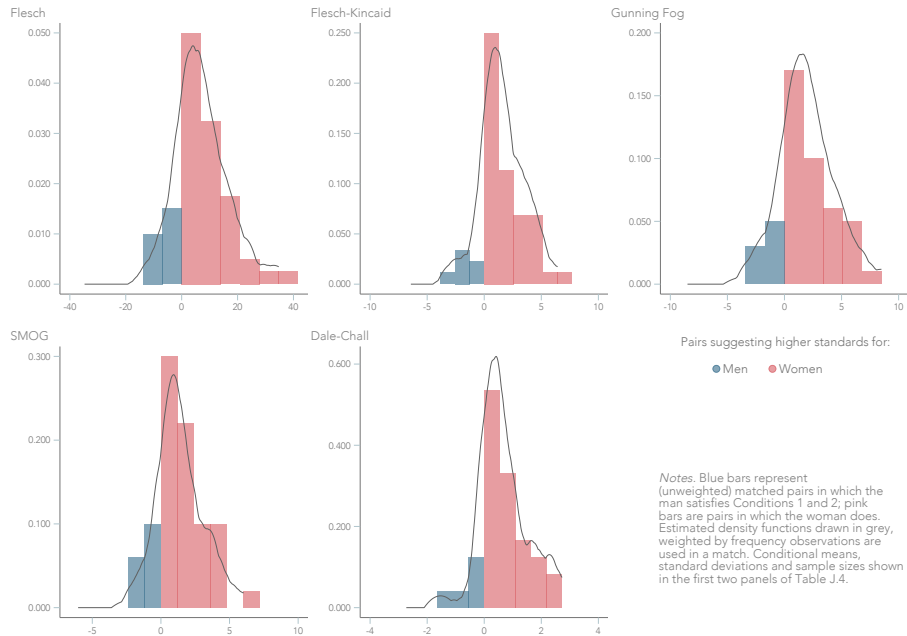
FIGURE J.1: Distributions of D_{ik} (Corollary 1), controlling for *JEL* category

Table J.5 and Figure J.2 replicate the analysis in Section 4.3.3 but Equation (13) does not adjust for the ratio of female authors on a paper.

TABLE J.5: D_{ik} (Corollary 1), without adjusting for the ratio of female authors

	Higher standards for women ($D_{ik} > 0$)			Higher standards for men ($D_{ik} < 0$)			Mean D_{ik}
	Mean	S.D.	N	Mean	S.D.	N	All obs.
Flesch	9.67	7.71	46	−5.80	5.41	26	2.47***
							(0.75)
Flesch Kincaid	1.68	1.41	51	−1.24	1.35	26	0.50***
							(0.15)
Gunning Fog	2.44	1.80	42	−1.53	1.69	27	0.55***
							(0.19)
SMOG	1.86	1.39	40	−1.05	1.08	29	0.36**
							(0.14)
Dale-Chall	0.83	0.66	50	−0.62	0.54	30	0.20***
							(0.07)

Notes. Sample 121 matched pairs. Table displays estimates identical to those in Table 9, except that R_{it} is not adjusted for the ratio of female authors on a paper. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.
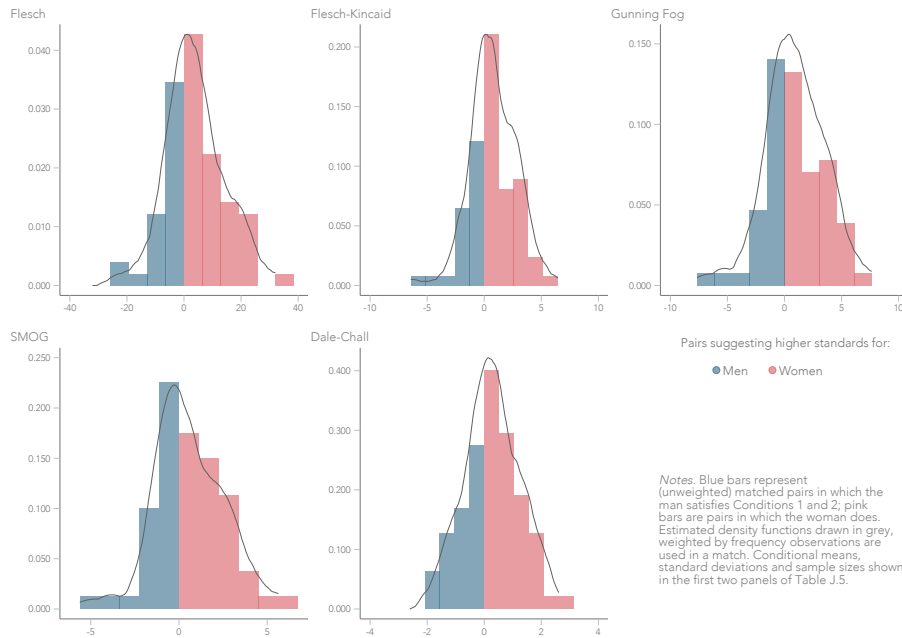
FIGURE J.2: Distributions of D_{ik} (Corollary 1), without adjusting for the ratio of female authors

39

Conclusions drawn in Section 4.3.3 are predicated on several strong assumptions. First, all results depend on match accuracy. Post-match co-variates are well balanced (see Appendix J.1). They remain well balanced—and similar to the matched population—when restricted to pairs satisfying $D_{ik} \neq 0$ (see Appendix M.1 in the August 2018 version of the paper). To facilitate further scrutiny, Appendix J.2 lists the names of economists in each pair. Matches using alternative variables (*e.g.*, minimum citation counts, mean institutional rank or fraction of articles per primary *JEL* category) and specifications (*e.g.*, propensity score matching) generate similar figures and conclusions.[27]

Additionally, authors must be sufficiently experience at t' for Assumptions 4 and 5 in Corollary 1 to hold. I assume this point occurs at or before authors' third top-four paper. Fifty percent of women with three or more top publications satisfy Conditions 1 *and* 2 when compared to equivalent men. Among them, D_{ik} is far from zero: these women write, on average, 21 percent more clearly than equivalent men with identical experience. I believe it is unlikely that half of all female economists with three top publications—plus many more second-tier publications and substantial experience refereeing and editing themselves—make mistakes of this magnitude.

To generate the counterfactual $\widehat{R}_{it}$ (Equation (13)), I assume unobserved co-author characteristics do not partially correlate with female ratio$_{it}$, conditional on i's gender and experience. To test the robustness of this assumption, Table M.30 (Appendix M.5) replicates Table 8 on exclusively, majority and senior female-authored papers. I have also repeated the analyses shown in Table 9 and Figure 5 without adjusting for female ratio$_{it}$ (Appendix J.5) and on subsets of matched pairs in which the woman's $t = 1$ and $t = 3$ papers are solo- or exclusively female-authored (16), majority female-authored (20) or at least 50 percent female-authored (76). Although sample sizes for the latter three analyses are small, they also find $D_{ik} \neq 0$ in about 70–75 percent of matched pairs; most of those (70 percent) indicate higher standards against the female member; the impact across all five scores also averages about 5 percent.

Moreover, experience appears to be the only t-varying factor driving within i changes in readability. Table 8 and additional analyses in a 2016 version of this paper (Hengel, 2016, pp. 23–24) show an identical pattern despite controlling for a large array of potential confounders. In a 2017 version, I reconstructed $\widehat{R}_{it}$ using several t-varying factors (number of co-authors, institutional rank, institutional rank of the highest ranked co-author, t for the most experienced co-author, publication year and dummies for each journal) (Hengel, 2017, pp. 30, 61); Appendix J.4 adds *JEL* classification codes to Equation (13). In Table M.30 (Appendix M.5), I restrict Table 8's analysis to solo-authored papers or those co-authored by members of the same sex. In all instances, women's readability is consistently shown to increase with t; when comparable results are estimated, they are similar to those presented in Table 9 and Figure 5.

Finally, accurate quantification requires that three additional criteria are also met. Assuming higher standards for i: (i) i's acceptance rate is no more than k's; (ii) i's draft readability is at least as high as k's; and (iii) i's draft readability at $t = 3$ is at least as high as his draft readability at $t = 1$. As already discussed in Section 4.2, (i) rules out the possibility that i is appropriately rewarded (relative to k) for writing more clearly. (ii) and (iii) eliminate situations in which women write more clearly during peer review in order to compensate for poorer writing—and consequently higher desk rejection rates—before peer review.

Unfortunately, my data do not perfectly identify acceptance rates nor do I have $t = 1$ and $t = 3$ draft readability scores for every matched pair. Nevertheless, the data I do have and prior research suggest (i)–(iii) not only hold on average, but do not exert upward bias on my estimate of D_{ik}, more generally. First, I reviewed the literature on gender neutrality in journals' acceptance rates in Section 4.2 and Appendix I.1; women are not accepted more often than men. Results and conclusions are similar when I attempt to adjust for acceptance rates explicitly by also requiring that $T_i \leq T_k$ for matched

---

[27]See Hengel (2017, pp. 30–33) for propensity score matches from a probit model performed with replacement and using a wider array of co-variates.

pairs in which i is held to higher standards relative to k (see Appendix M.4 in the August 2018 version of the paper). As shown in Section 3.3, women's draft papers are indeed more readable than men's. Section 4.4 provides further confirmation. Figure 6 plots the readability of women's and men's draft and published papers over increasing t. Women's drafts are more readable than men's drafts at $t = 3$ *and* their own drafts at $t = 1$.

K Section 4.4, supplemental output

K.1 *Experience and review times*

In Section 4.4.1 I find evidence suggesting that inexperienced female economists go through the toughest review, conditional on acceptance. To investigate further, I test the impact of experience on time spent in review by re-estimating Equation (9) on sub-samples of junior ($t = 1$) and senior ($t > 1$) authors.[28] Results are displayed in Figure K.1. They suggest papers by junior women do indeed take longer in review; the gender gap is significantly smaller—albeit still positive—for senior women.
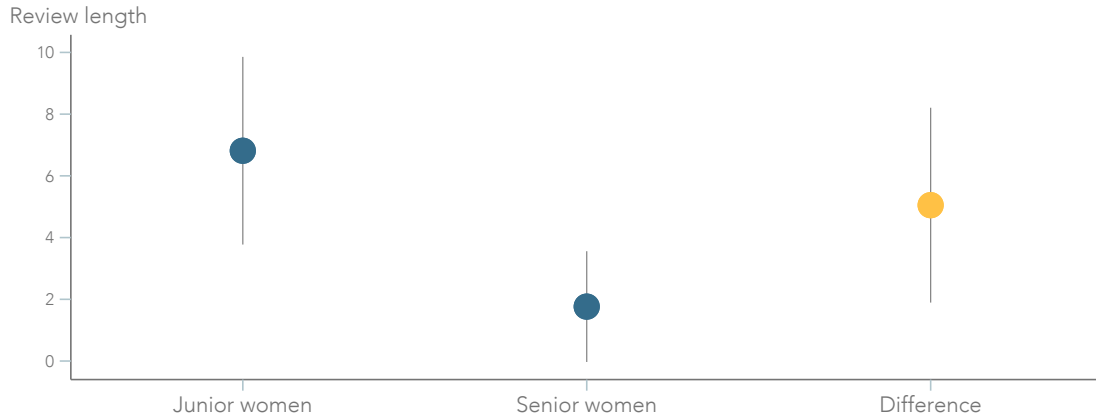
FIGURE K.1: The impact of experience on women's review times

Notes. Sample restricted to senior authors on a paper (5,007 observations). Blue dots are the coefficients on female ratio (papers with fewer than 50 percent female authors are classified as male, see Section 2.2) corresponding to separate FGLS estimations of Equation (9) on junior ($t = 1$) and senior ($t > 1$) authors, respectively. The yellow dot is their difference. Vertical grey lines correspond to 90 percent confidence intervals. Regressions weighted by $1/N_j$.

---

[28]Three notes on estimation. First, in Section 4.3, I define "experienced" as $t = 3$. However, most female-authored papers published in *Econometrica* and *REStud* are by women with no (or only one) previous top publication; only 24 have two or more previous papers and were the most senior co-author on a $t > 2$ paper. Second, to eliminate confounding by more senior co-authors, I restrict the sample to the senior authors on a paper (*i.e.*, authors satisfying max. $t = t$). (Including these observations does not substantially impact results or conclusions.) Third, because the sample includes data from *REStud*, readability, motherhood and childbirth controls are not included. See the August 2018 version of this paper for results that control for these factors (based on data from *Econometrica* alone).

L Alternative program for calculating readability scores

In this section, I replicate Table 3, Table 5, Table 8 and Table F.2 using readability scores generated by the R `readability` package, an alternative program for calculating Flesch-Kincaid, Gunning Fog and SMOG readability scores. Replications for other tables and figures presented in the paper are not shown, but will be made available on request.

`Textatistic` and `readability` employ different strategies to adapt the scores to automated calculation—*e.g.*, `readability` counts semi-colons and dashes as sentence-ending terminations; `Textatistic` does not.[29] Results appear robust to these (and other) small discrepancies: coefficients are similar to those presented in the body of the paper.

TABLE L.1: Table 3, alternative program for calculating readability

	1950–2015					1990–2015			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Flesch-Kincaid	0.23*	0.23*	0.23*	0.26**	0.28**	0.34**	0.37**	0.35**	0.49***
	(0.13)	(0.13)	(0.12)	(0.13)	(0.13)	(0.15)	(0.14)	(0.15)	(0.16)
Gunning Fog	0.37***	0.37***	0.38***	0.42***	0.43***	0.45***	0.45***	0.46***	0.54***
	(0.13)	(0.13)	(0.13)	(0.14)	(0.14)	(0.15)	(0.14)	(0.16)	(0.18)
SMOG	0.24**	0.25**	0.25**	0.27***	0.29***	0.30**	0.30***	0.31**	0.36***
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)	(0.12)	(0.13)
No. obs.	9,117	9,117	9,117	9,117	9,117	5,211	5,211	5,211	5,774
Editor	✓	✓	✓	✓	✓	✓	✓	✓	✓
Blind	✓	✓	✓	✓	✓	✓	✓	✓	✓
Journal	✓	✓							
Year		✓							
Journal×Year			✓	✓	✓	✓	✓	✓	✓
N_j				✓	✓	✓	✓	✓	✓
Institution				✓	✓	✓	✓	✓	✓
Quality				✓[1]	✓[1]	✓[1]	✓[1]	✓[1]
Native speaker					✓	✓	✓	✓	✓
JEL (primary)							✓		
Theory/empirical								✓	
JEL (tertiary)									✓

Notes. Estimates are identical to those in Table 3, except readability scores were calculated using the R `readability` program. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

---

[29] Readability scores were originally developed to be calculated by hand. Automating their calculation requires slightly adapting the algorithms. For example, all five scores define sentences as grammatically independent units of thoughts—*e.g.*, two independent clauses connected by a dash or semi-colon count as two separate sentences. Unfortunately, semi-colons and dashes are frequently used in other ways and it is difficult to programmatically distinguish between contexts.

TABLE L.2: Table 5, alternative program for calculating readability

	OLS (score)			OLS (Δ in score)		FGLS (score)					
						Working paper		Published paper		Difference	
	R_{jW}	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio
Flesch–Kincaid	0.74***	0.54***	−1.13	0.46**	−0.93	0.31	−0.76	0.77***	−1.69***	0.46**	−0.93
	(0.03)	(0.20)	(0.74)	(0.21)	(0.62)	(0.25)	(0.77)	(0.28)	(0.44)	(0.20)	(0.61)
Gunning Fog	0.77***	0.56***	−1.40	0.46**	−1.20	0.43	−0.85	0.89***	−2.05***	0.46**	−1.20
	(0.03)	(0.21)	(0.90)	(0.21)	(0.76)	(0.27)	(0.95)	(0.27)	(0.64)	(0.21)	(0.75)
SMOG	0.77***	0.39**	−1.01	0.33**	−0.86	0.28	−0.64	0.61***	−1.50***	0.33**	−0.86
	(0.02)	(0.16)	(0.65)	(0.15)	(0.55)	(0.18)	(0.71)	(0.18)	(0.49)	(0.15)	(0.54)
No. observations		1,988		1,988		1,988		1,988		1,988	
Editor effects		✓		✓		✓		✓			
Journal×Year effects		✓		✓		✓		✓			
N_j		✓		✓		✓		✓			
Quality controls		✓²		✓²		✓²		✓²			
Native speaker		✓		✓		✓		✓			
Theory/emp. effects		✓		✓		✓		✓			

Notes. Estimates are identical to those in Table 5, except readability scores were calculated using the R readability program. ***, **, and * statistically significant at 1%, 5% and 10%, respectively.

44

TABLE L.3: Table 8, alternative program for calculating readability

	$t = 1$	$t = 2$	$t = 3$	$t = 4\text{--}5$	$t \geq 6$	All
Flesch-Kincaid	0.08	0.28	1.12***	0.62	0.55	0.28*
	(0.18)	(0.27)	(0.26)	(0.49)	(0.39)	(0.17)
Gunning Fog	0.25	0.37	1.41***	0.84	0.63	0.50***
	(0.20)	(0.27)	(0.35)	(0.55)	(0.45)	(0.19)
SMOG	0.17	0.28	0.92***	0.56	0.41	0.34***
	(0.15)	(0.20)	(0.23)	(0.40)	(0.31)	(0.13)
No. observations	6,874	2,827	1,675	1,906	2,773	12,008
Editor effects	✓	✓	✓	✓	✓	✓
Blind review	✓	✓	✓	✓	✓	✓
Journal×Year effects	✓	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓
Quality controls	✓[3]	✓[3]	✓[3]	✓[3]	✓[3]	✓[1]
Native speaker	✓	✓	✓	✓	✓	✓

Notes. Estimates are identical to those in Table 8, except readability scores were calculated using the R `readability` program. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE L.4: Table F.2, alternative program for calculating readability

	Flesch–Kincaid	Gunning Fog	SMOG
Female ratio for women (β_1)	0.33	0.48**	0.30*
	(0.21)	(0.24)	(0.17)
Female ratio for men ($\beta_1 + \beta_2$)	0.38	0.50	0.32
	(0.30)	(0.34)	(0.24)
Female ratio×male (β_2)	0.05	0.02	0.02
	(0.35)	(0.40)	(0.28)
Lagged score (β_0)	0.05**	0.04*	0.04*
	(0.02)	(0.02)	(0.02)
No. observations	9,181	9,181	9,181
Tests of instrument validity			
Hansen test (*p*-value)	0.84	0.74	0.73
Sargan test (*p*-value)	0.16	0.21	0.11
z-test for no serial correlation			
Order 1	−12.48	−13.25	−16.85
Order 2	0.08	0.59	0.46
Editor effects	✓	✓	✓
Blind review	✓	✓	✓
Journal effects	✓	✓	✓
N_j	✓	✓	✓
Institution effects	✓	✓	✓
Quality controls	✓^{2}	✓^{2}	✓^{2}
Native speaker	✓	✓	✓

Notes. Estimates are identical to those in Table F.2, except readability scores were calculated using the R `readability` program. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

M Alternative proxies for article gender

In this appendix, I replicate Table 3, Table 5, Table 6, Table 7, Table 8 and Table F.2 using the alternative proxies for article gender summarised in Table M.1.

TABLE M.1: Summary of alternative proxies for female-authored papers

Proxy for female-authorship	Appendix	Description	Sample restrictions	Additional notes
Female solo-authored papers	Appendix M.1	Article gender is represented by a dummy variable equal to 1 if the paper is solo-authored by a woman and 0 if it is solo-authored by a man.	Co-authored papers are excluded in all tables except Table M.7 where they are classified as male-authored papers in order to increase within-author variation.	
Senior-female authored papers	Appendix M.2	Article gender is represented by the interaction between female ratio and a dummy variable equal to 1 if a female author had at least as many top-five papers as her co-authors at the time the paper was published.	Mixed-gendered papers without a senior female co-author are excluded in all tables except Table M.13 where they are classified as male-authored papers in order to increase within-author variation.	
Majority female-authored papers	Appendix M.3	Article gender is represented by a dummy variable equal to 1 if at least half of all authors are female.	Mixed-gendered papers with a strict minority of female authors are excluded in all tables except Table M.19 where they are classified as male-authored papers in order to increase within-author variation.	
Papers with at least one female author	Appendix M.4	Article gender is represented by a dummy variable equal to 1 if at least one author on a paper is female.		
Exclusively female-authored papers	Appendix M.5	Article gender is represented by a dummy variable equal to 1 if all authors on a paper are female.	Mixed-gendered papers are excluded in all tables except Table M.31 where they are classified as male-authored papers in order to increase within-author variation.	
Senior female-authored papers, sample of less experienced authors	Appendix M.6	Article gender is represented by the interaction between female ratio and a dummy variable equal to 1 if a female author had at least as many top-five papers as her co-authors at the time the paper was published.	Mixed-gendered papers without a senior female co-author and all papers by authors with more than three top-five papers are excluded in all tables.	Table 8 and Table F.2 are not estimated. NBER has released very few working papers in which a woman with only one or two top papers is the most senior author on a paper with more than one author. Results in Table M.33 should therefore be interpreted with caution.

48

M.1   Solo-authored

TABLE M.2: Table 3, solo-authored papers

	1950–2015					1990–2015			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Flesch	0.37	0.41	0.25	0.25	0.46	0.52	0.46	0.59	0.62
	(0.74)	(0.75)	(0.74)	(0.74)	(0.75)	(0.85)	(0.81)	(0.88)	(1.75)
Flesch-Kincaid	0.13	0.13	0.11	0.12	0.09	0.27	0.27	0.28	0.10
	(0.17)	(0.17)	(0.16)	(0.16)	(0.17)	(0.17)	(0.17)	(0.17)	(0.31)
Gunning Fog	0.21	0.22	0.19	0.21	0.18	0.40**	0.36*	0.42**	0.20
	(0.19)	(0.19)	(0.18)	(0.18)	(0.18)	(0.18)	(0.18)	(0.19)	(0.36)
SMOG	0.13	0.15	0.13	0.14	0.14	0.26*	0.22*	0.28*	0.16
	(0.14)	(0.14)	(0.13)	(0.13)	(0.13)	(0.14)	(0.13)	(0.14)	(0.26)
Dale-Chall	−0.01	−0.01	−0.02	−0.02	−0.01	0.05	0.06	0.06	0.10
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.08)	(0.12)
No. obs.	4,014	4,014	4,014	4,014	4,014	1,540	1,540	1,540	1,666
Editor	✓	✓	✓	✓	✓	✓	✓	✓	✓
Blind	✓	✓	✓	✓	✓	✓	✓	✓	✓
Journal	✓	✓							
Year		✓							
Journal×Year			✓	✓	✓	✓	✓	✓	✓
N_j				✓	✓	✓	✓	✓	✓
Institution				✓	✓	✓	✓	✓	✓
Quality					✓[1]	✓[1]	✓[1]	✓[1]	✓[1]
Native speaker					✓	✓	✓	✓	✓
JEL (primary)							✓		
Theory/empirical								✓	
JEL (tertiary)									✓

Notes. Estimates are identical to those in Table 3, except that female ratio has been replaced with a dummy variable equal to 1 if the paper is solo-authored by a woman and 0 if it is solo-authored by a man. (Co-authored papers are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.3: Table 5, solo-authored papers

| | | OLS (score) | | OLS (Δ in score) | | FGLS (score) | | | | | |
| | | | | | | Working paper | | Published paper | | Difference | |
	R_{jW}	Female	Blind× female	Female	Blind× female	Female	Blind× female	Female	Blind× female	Female	Blind× female
Flesch	0.79***	1.34	−4.31	1.19	−4.96	0.74	3.13	1.93	−1.83	1.19	−4.96
	(0.08)	(1.43)	(7.53)	(1.28)	(4.91)	(2.10)	(4.46)	(2.40)	(4.10)	(1.19)	(4.56)
Flesch–Kincaid	0.76***	0.66	−1.81	0.62	−1.78	0.17	−0.10	0.78	−1.88**	0.62*	−1.78*
	(0.14)	(0.50)	(1.34)	(0.37)	(1.14)	(0.42)	(1.04)	(0.57)	(0.94)	(0.35)	(1.06)
Gunning Fog	0.73***	0.71	−2.09	0.67	−2.19*	0.18	0.36	0.84	−1.83*	0.67*	−2.19*
	(0.11)	(0.55)	(1.50)	(0.42)	(1.27)	(0.46)	(1.17)	(0.62)	(0.95)	(0.39)	(1.18)
SMOG	0.74***	0.35	−1.32	0.31	−1.41	0.15	0.33	0.46	−1.08*	0.31	−1.41*
	(0.06)	(0.30)	(1.02)	(0.27)	(0.86)	(0.34)	(0.80)	(0.39)	(0.58)	(0.25)	(0.80)
Dale–Chall	0.86***	0.19	−0.43	0.16	−0.46	0.17	0.18	0.34**	−0.28	0.16	−0.46*
	(0.04)	(0.15)	(0.52)	(0.16)	(0.28)	(0.15)	(0.36)	(0.16)	(0.38)	(0.15)	(0.26)
No. observations		434		434		434		434		434	
Editor effects		✓		✓		✓		✓			
Journal×Year effects		✓		✓		✓		✓			
N_j		✓²		✓²		✓²		✓²			
Quality controls		✓		✓		✓		✓			
Native speaker		✓		✓		✓		✓			
Theory/emp. effects											

Notes. Estimates are identical to those in Table 5, except that female ratio has been replaced with a dummy variable equal to 1 if the paper is solo-authored by a woman and 0 if it is solo-authored by a man. (Co-authored papers are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

50

TABLE M.4: Table 6, solo-authored papers

	1970–2015					1990–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Solo female	7.313**	10.433**	10.220**	8.068**	10.354**	13.216**	14.608**
	(2.964)	(4.088)	(4.069)	(3.602)	(4.068)	(6.035)	(5.767)
Max. t	−0.311***	−0.309***	−0.308***	−0.312***	−0.305***	−0.284***	−0.279**
	(0.080)	(0.079)	(0.080)	(0.081)	(0.079)	(0.102)	(0.109)
No. pages	0.198***	0.196***	0.193***	0.197***	0.192***	0.267***	0.264***
	(0.041)	(0.041)	(0.040)	(0.041)	(0.040)	(0.063)	(0.066)
Order	0.175**	0.179**	0.170**	0.172**	0.173**	0.610*	0.479
	(0.082)	(0.082)	(0.081)	(0.081)	(0.081)	(0.307)	(0.299)
No. citations (asinh)	−0.693***	−0.745***	−0.702***	−0.686***	−0.724***	−1.150*	−1.065*
	(0.180)	(0.174)	(0.178)	(0.182)	(0.177)	(0.584)	(0.535)
Flesch	−0.009	−0.009	−0.008	−0.009	−0.006	−0.001	0.008
	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)	(0.058)	(0.063)
Theory	1.859	1.807	1.751	1.819	1.772	2.143	−0.305
	(1.312)	(1.338)	(1.347)	(1.331)	(1.344)	(1.441)	(2.554)
Empirical	−8.025**	−7.865*	−8.537**	−8.113**	−8.661**	−9.072**	−10.639**
	(3.325)	(4.064)	(3.410)	(3.303)	(3.474)	(3.854)	(4.146)
Other	1.451	1.272	1.452	1.465	1.421	1.831	2.064
	(1.918)	(1.863)	(1.889)	(1.904)	(1.916)	(2.073)	(3.000)
Mother			−9.050**		−14.724***	−31.324***	−32.020***
			(4.455)		(5.315)	(9.349)	(9.563)
Birth				−3.945	8.818*	26.982***	28.360***
				(4.821)	(5.044)	(6.856)	(7.206)
Constant	13.932***	14.287***	14.045***	13.947***	14.084***	16.312***	17.769***
	(2.554)	(2.453)	(2.515)	(2.543)	(2.519)	(5.577)	(5.379)
R^2	0.344	0.349	0.348	0.345	0.349	0.227	0.260
No. observations	1,224	1,210	1,224	1,224	1,224	418	418
Editor effects	✓	✓	✓	✓	✓	✓	✓
Year effects	✓	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓	✓
JEL (primary) effects							✓

Notes. Estimates are identical to those in Table 6, except that female ratio has been replaced with a dummy variable equal to 1 if the paper is solo-authored by a woman and 0 if it is solo-authored by a man. (Co-authored papers are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.5: Table 7, solo-authored papers

	1970–2015			1990–2015		
	Econometrica	REStud	Econometrica + REStud	Econometrica	REStud	Econometrica + REStud
Solo female	7.31**	2.40	4.07**	12.59***	5.19**	7.00***
	(2.96)	(1.72)	(1.57)	(4.51)	(2.50)	(2.21)
Max. t	−0.31***	−0.33***	−0.33***	−0.27**	−0.39***	−0.34***
	(0.08)	(0.09)	(0.07)	(0.11)	(0.12)	(0.07)
No. pages	0.20***	0.26**	0.21***	0.26***	0.11	0.22***
	(0.04)	(0.10)	(0.04)	(0.07)	(0.16)	(0.05)
Order	0.18**	−0.06	0.04	0.48	−0.04	0.24
	(0.08)	(0.09)	(0.06)	(0.30)	(0.24)	(0.21)
No. citations (asinh)	−0.69***	−0.32	−0.49***	−0.93	−1.08	−1.13**
	(0.18)	(0.30)	(0.16)	(0.57)	(0.79)	(0.48)
Theory	1.86	−1.21	0.37	−0.16	−1.46	−0.52
	(1.31)	(1.53)	(1.06)	(2.51)	(1.87)	(1.44)
Empirical	−8.05**	3.78	−1.40	−10.35**	4.56	−1.12
	(3.32)	(4.01)	(2.70)	(3.96)	(5.40)	(3.20)
Other	1.41	−2.29	−0.18	2.23	−1.11	1.05
	(1.92)	(3.11)	(1.64)	(3.12)	(3.54)	(1.73)
Constant	13.58***	21.78***	17.12***	17.26***	31.74***	23.37***
	(2.25)	(4.30)	(1.96)	(4.77)	(6.83)	(3.66)
R^2	0.34	0.36	0.36	0.25	0.31	0.25
No. observations	1,224	851	2,075	418	375	793
Editor effects	✓	✓	✓	✓	✓	✓
Accepted year effects	✓	✓		✓	✓	
Journal×Accepted year effects			✓			✓
Institution effects	✓	✓	✓	✓	✓	✓
JEL (primary) effects				✓	✓	✓

Notes. Estimates are identical to those in Table 7, except that female ratio has been replaced with a dummy variable equal to 1 if the paper is solo-authored by a woman and 0 if it is solo-authored by a man. (Co-authored papers are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.6: Table 8, solo-authored papers

	$t = 1$	$t = 2$	$t = 3\text{–}5$	$t \geq 6$	All
Flesch	−0.31	0.06	5.13*	7.07	0.46
	(1.08)	(1.74)	(2.73)	(4.99)	(0.89)
Flesch-Kincaid	−0.05	−0.01	1.21**	1.76***	0.09
	(0.27)	(0.41)	(0.61)	(0.67)	(0.21)
Gunning Fog	0.04	−0.04	1.48**	2.21**	0.18
	(0.30)	(0.49)	(0.70)	(0.86)	(0.24)
SMOG	0.04	−0.06	1.03*	1.52**	0.14
	(0.20)	(0.33)	(0.55)	(0.74)	(0.16)
Dale-Chall	−0.11	0.19	0.42*	0.80**	−0.01
	(0.11)	(0.16)	(0.24)	(0.39)	(0.08)
No. observations	2,025	758	772	459	4,014
Editor effects	✓	✓	✓	✓	✓
Blind review	✓	✓	✓	✓	✓
Journal×Year effects	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓
Quality controls	✓[3]	✓[3]	✓[3]	✓[3]	✓[1]
Native speaker	✓	✓	✓	✓	✓

Notes. Estimates are similar to those in Table 8, except that female ratio has been replaced with a dummy variable equal to 1 if the paper is solo-authored by a woman and 0 if it is solo-authored by a man. (Co-authored papers are excluded.) Due to small sample sizes, columns $t = 3$ and $t = 4\text{–}5$ have been combined and estimates are clustered on author, only. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.7: Table F.2, solo-authored papers

	Flesch Reading Ease	Flesch–Kincaid	Gunning Fog	SMOG	Dale-Chall
Female ratio for women (β_1)	1.95	0.15	0.38	0.33	0.28**
	(1.37)	(0.30)	(0.36)	(0.26)	(0.12)
Lagged score (β_0)	0.04*	0.05**	0.03	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
No. observations	9,133	9,133	9,133	9,133	9,133
Tests of instrument validity					
Hansen test (p-value)	0.81	0.97	0.96	0.96	0.86
Sargan test (p-value)	0.39	0.13	0.32	0.16	0.36
z–test for no serial correlation					
Order 1	−18.42	−13.59	−14.84	−17.46	−18.54
Order 2	0.46	−0.44	0.12	0.45	0.12
Editor effects	✓	✓	✓	✓	✓
Blind review	✓	✓	✓	✓	✓
Journal effects	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓
Quality controls	✓^{2}	✓^{2}	✓^{2}	✓^{2}	✓^{2}
Native speaker	✓	✓	✓	✓	✓

Notes. Estimates are identical to those in Table F.2, except that female ratio has been replaced with a dummy variable equal to 1 if the paper is solo-authored by a woman and 0 if it is solo-authored by a man. (Co-authored mixed-gendered papers are included and classified as male.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.8: Table 3, senior female author

	1950–2015					1990–2015			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Flesch	0.77	0.75	0.72	0.79	1.05	0.99	1.04	0.97	1.20
	(0.65)	(0.65)	(0.66)	(0.66)	(0.66)	(0.70)	(0.65)	(0.70)	(0.91)
Flesch-Kincaid	0.17	0.16	0.16	0.19	0.21	0.30**	0.34**	0.31**	0.32*
	(0.13)	(0.13)	(0.13)	(0.13)	(0.14)	(0.14)	(0.13)	(0.14)	(0.18)
Gunning Fog	0.31**	0.31**	0.32**	0.36**	0.38**	0.51***	0.51***	0.51***	0.45**
	(0.14)	(0.14)	(0.15)	(0.15)	(0.16)	(0.16)	(0.15)	(0.17)	(0.20)
SMOG	0.21*	0.22*	0.22*	0.24**	0.26**	0.33**	0.32***	0.33**	0.30*
	(0.11)	(0.11)	(0.11)	(0.12)	(0.12)	(0.13)	(0.11)	(0.13)	(0.15)
Dale-Chall	0.07	0.06	0.06	0.06	0.08	0.13*	0.12*	0.12*	0.14*
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)	(0.08)
No. obs.	8,396	8,396	8,396	8,396	8,396	4,581	4,581	4,581	4,991
Editor	✓	✓	✓	✓	✓	✓	✓	✓	✓
Blind	✓	✓	✓	✓	✓	✓	✓	✓	✓
Journal	✓	✓							
Year		✓							
Journal×Year			✓	✓	✓	✓	✓	✓	✓
N_j				✓	✓	✓	✓	✓	✓
Institution			✓	✓	✓	✓	✓	✓	✓
Quality				✓[1]	✓[1]	✓[1]	✓[1]	✓[1]	✓[1]
Native speaker				✓	✓	✓	✓	✓	✓
JEL (primary)							✓		
Theory/empirical								✓	
JEL (tertiary)									✓

Notes. Estimates are identical to those in Table 3, except that female ratio has been replaced with the interaction between female ratio and a dummy variable equal to 1 if a female author had strictly more top-five papers as her co-authors at the time the paper was published. (Mixed-gendered papers with a senior male co-author are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.9: Table 5, senior female author

| | OLS (score) | | | OLS (Δ in score) | | FGLS (score) | | | | | |
| | | | | | | Working paper | | Published paper | | Difference | |
	R_{jW}	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio	Female ratio	Blind× fem. ratio
Flesch	0.83***	1.28	−2.17	0.93	−2.05	2.10*	−0.72	3.03**	−2.77	0.93	−2.05
	(0.03)	(0.79)	(5.85)	(0.87)	(2.93)	(1.17)	(4.04)	(1.37)	(2.64)	(0.85)	(2.87)
Flesch–Kincaid	0.74***	0.53**	−0.96	0.45*	−0.82	0.31	−0.51	0.76**	−1.33**	0.45*	−0.82
	(0.05)	(0.22)	(1.14)	(0.27)	(0.77)	(0.20)	(0.90)	(0.32)	(0.60)	(0.26)	(0.75)
Gunning Fog	0.76***	0.59***	−1.13	0.48*	−1.03	0.45**	−0.40	0.93***	−1.43**	0.48*	−1.03
	(0.04)	(0.21)	(1.19)	(0.29)	(0.82)	(0.22)	(1.01)	(0.33)	(0.65)	(0.28)	(0.80)
SMOG	0.78***	0.33**	−0.77	0.25	−0.72	0.36**	−0.24	0.61***	−0.96**	0.25	−0.72
	(0.03)	(0.13)	(0.84)	(0.18)	(0.55)	(0.18)	(0.74)	(0.22)	(0.45)	(0.17)	(0.54)
Dale–Chall	0.84***	0.12	−0.21	0.07	−0.16	0.30**	−0.27	0.37**	−0.43*	0.07	−0.16
	(0.02)	(0.09)	(0.38)	(0.08)	(0.14)	(0.14)	(0.28)	(0.15)	(0.24)	(0.08)	(0.14)
No. observations		1,701		1,701		1,701		1,701		1,701	
Editor effects		✓		✓		✓		✓			
Journal×Year effects		✓		✓		✓		✓			
N_j		✓²		✓²		✓²		✓²			
Quality controls		✓		✓		✓		✓			
Native speaker		✓		✓		✓		✓			
Theory/emp. effects											

Notes. Estimates are identical to those in Table 5, except that female ratio has been replaced with the interaction between female ratio and a dummy variable equal to 1 if a female author had strictly more top-five papers as her co-authors at the time the paper was published. (Mixed-gendered papers with a senior male co-author are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

	1970–2015					1990–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Senior female	6.229**	8.878**	8.838**	6.932**	8.879**	11.209**	11.524**
	(2.614)	(3.500)	(3.502)	(3.127)	(3.499)	(4.746)	(4.604)
Max. t	−0.177***	−0.176***	−0.174***	−0.177***	−0.173***	−0.161***	−0.157**
	(0.045)	(0.046)	(0.046)	(0.046)	(0.046)	(0.057)	(0.058)
No. pages	0.208***	0.206***	0.205***	0.207***	0.205***	0.245***	0.232***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.041)	(0.042)
N_j	1.227***	1.239***	1.221***	1.220***	1.231***	1.504***	1.312***
	(0.322)	(0.318)	(0.324)	(0.320)	(0.321)	(0.462)	(0.462)
Order	0.186***	0.185***	0.182***	0.185***	0.183***	0.435***	0.446***
	(0.062)	(0.062)	(0.062)	(0.062)	(0.062)	(0.156)	(0.160)
No. citations (asinh)	−0.393*	−0.418**	−0.403*	−0.390*	−0.413**	−0.657	−0.584
	(0.201)	(0.201)	(0.200)	(0.201)	(0.201)	(0.441)	(0.448)
Flesch	−0.015	−0.014	−0.014	−0.015	−0.013	−0.030	−0.033
	(0.014)	(0.015)	(0.014)	(0.014)	(0.014)	(0.031)	(0.031)
Theory	0.344	0.293	0.315	0.341	0.307	0.407	−0.373
	(0.863)	(0.862)	(0.872)	(0.869)	(0.866)	(0.882)	(1.404)
Empirical	1.250	1.521	1.115	1.247	1.053	1.125	0.694
	(1.731)	(1.814)	(1.765)	(1.724)	(1.776)	(1.934)	(2.239)
Other	−0.921	−0.921	−0.888	−0.904	−0.902	−0.648	−1.614
	(0.602)	(0.574)	(0.588)	(0.594)	(0.595)	(0.632)	(1.198)
Mother			−8.934**		−13.380***	−23.080***	−24.496***
			(4.282)		(4.457)	(6.176)	(6.828)
Birth				−4.118	7.365	16.737***	18.110***
				(5.022)	(4.684)	(5.011)	(4.977)
Constant	14.269***	14.340***	14.313***	14.269***	14.336***	16.364***	17.974***
	(1.394)	(1.410)	(1.406)	(1.397)	(1.408)	(2.760)	(2.901)
R^2	0.290	0.293	0.292	0.290	0.292	0.134	0.150
No. observations	2,461	2,446	2,461	2,461	2,461	1,154	1,154
Editor effects	✓	✓	✓	✓	✓	✓	✓
Year effects	✓	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓	✓
JEL (primary) effects							✓

Notes. Estimates are identical to those in Table 6, except that female ratio has been replaced with the interaction between female ratio and a dummy variable equal to 1 if a female author had strictly more top-five papers as her co-authors at the time the paper was published. (Mixed-gendered papers with a senior male co-author are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.11: Table 7, senior female author

	1970–2015			1990–2015		
	Econometrica	*REStud*	*Econometrica + REStud*	*Econometrica*	*REStud*	*Econometrica + REStud*
Senior female	6.24**	2.22	3.78***	9.03**	4.51**	5.78***
	(2.61)	(1.51)	(1.32)	(3.65)	(2.10)	(1.80)
Max. t	−0.18***	−0.10	−0.15***	−0.16**	−0.06	−0.12**
	(0.05)	(0.07)	(0.04)	(0.06)	(0.09)	(0.05)
No. pages	0.21***	0.14*	0.19***	0.24***	0.02	0.18***
	(0.03)	(0.08)	(0.03)	(0.04)	(0.09)	(0.04)
N_j	1.22***	0.14	0.84***	1.29**	0.84	1.14***
	(0.32)	(0.54)	(0.30)	(0.47)	(0.68)	(0.41)
Order	0.19***	−0.05	0.09	0.47***	0.07	0.22*
	(0.06)	(0.08)	(0.05)	(0.16)	(0.17)	(0.12)
No. citations (asinh)	−0.39*	−0.49**	−0.40**	−0.53	−1.00**	−0.79**
	(0.20)	(0.21)	(0.15)	(0.46)	(0.45)	(0.33)
Theory	0.35	−0.15	0.24	−0.21	−1.02	−0.54
	(0.86)	(1.10)	(0.63)	(1.39)	(1.35)	(0.72)
Empirical	1.21	7.38**	4.14**	0.81	9.21**	4.69**
	(1.73)	(3.08)	(1.74)	(2.23)	(3.45)	(1.86)
Other	−0.95	−1.85	−1.19**	−1.55	−2.42	−1.56**
	(0.61)	(1.18)	(0.50)	(1.26)	(1.50)	(0.69)
Constant	13.68***	23.91***	17.24***	16.18***	31.35***	22.32***
	(1.33)	(2.17)	(1.07)	(2.65)	(3.41)	(1.75)
R^2	0.29	0.29	0.30	0.14	0.18	0.15
No. observations	2,461	1,653	4,114	1,154	932	2,086
Editor effects	✓	✓	✓	✓	✓	✓
Accepted year effects	✓	✓		✓	✓	
Journal×Accepted year effects			✓			✓
Institution effects	✓	✓	✓	✓	✓	✓
JEL (primary) effects				✓	✓	✓

Notes. Estimates are identical to those in Table 7, except that female ratio has been replaced with the interaction between female ratio and a dummy variable equal to 1 if a female author had strictly more top-five papers as her co-authors at the time the paper was published. (Mixed-gendered papers with a senior male co-author are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.12: Table 8, senior female author

	$t = 1$	$t = 2$	$t = 3$	$t = 4$–5	$t \geq 6$	All
Flesch	−0.14	0.78	4.13***	2.90	2.52	2.87***
	(0.91)	(1.27)	(1.45)	(2.21)	(2.26)	(1.00)
Flesch-Kincaid	−0.08	0.12	0.94***	0.53	0.51	0.53**
	(0.20)	(0.27)	(0.33)	(0.48)	(0.45)	(0.23)
Gunning Fog	0.03	0.43	1.35***	0.99*	0.73	0.89***
	(0.23)	(0.29)	(0.41)	(0.52)	(0.46)	(0.28)
SMOG	0.02	0.30	0.86***	0.80*	0.49	0.64***
	(0.17)	(0.21)	(0.28)	(0.42)	(0.34)	(0.19)
Dale-Chall	0.00	0.02	0.27**	0.36	0.50**	0.29***
	(0.09)	(0.12)	(0.13)	(0.23)	(0.23)	(0.10)
No. observations	6,017	2,536	1,489	1,688	2,427	8,361
Editor effects	✓	✓	✓	✓	✓	✓
Blind review	✓	✓	✓	✓	✓	✓
Journal×Year effects	✓	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓
Quality controls	✓[3]	✓[3]	✓[3]	✓[3]	✓[3]	✓[1]
Native speaker	✓	✓	✓	✓	✓	✓

Notes. Estimates are identical to those in Table 8, except that female ratio has been replaced with the interaction between female ratio and a dummy variable equal to 1 if a female author had strictly more top-five papers as her co-authors at the time the paper was published. (Mixed-gendered papers with a senior male co-author are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.13: Table F.2, senior female author

	Flesch Reading Ease	Flesch-Kincaid	Gunning Fog	SMOG	Dale-Chall
Female ratio for women (β_1)	1.76*	0.20	0.53**	0.40**	0.18*
	(1.02)	(0.22)	(0.25)	(0.18)	(0.10)
Female ratio for men ($\beta_1 + \beta_2$)	6.79**	1.18*	1.40*	0.96*	0.12
	(3.27)	(0.70)	(0.80)	(0.57)	(0.31)
Female ratio×male (β_2)	5.03	0.98	0.87	0.56	−0.06
	(3.39)	(0.72)	(0.83)	(0.59)	(0.32)
Lagged score (β_0)	0.04**	0.05**	0.03	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
No. observations	9,181	9,181	9,181	9,181	9,181
Tests of instrument validity					
Hansen test (*p*-value)	0.80	0.97	0.96	0.97	0.84
Sargan test (*p*-value)	0.37	0.12	0.32	0.16	0.35
z–test for no serial correlation					
Order 1	−18.47	−13.63	−14.88	−17.49	−18.71
Order 2	0.49	−0.45	0.11	0.45	0.07
Editor effects	✓	✓	✓	✓	✓
Blind review	✓	✓	✓	✓	✓
Journal effects	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓
Quality controls	✓²	✓²	✓²	✓²	✓²
Native speaker	✓	✓	✓	✓	✓

Notes. Estimates are identical to those in Table F.2, except that female ratio has been replaced with the interaction between female ratio and a dummy variable equal to 1 if a female author had at least as many top-five papers as her co-authors at the time the paper was published. (Mixed-gendered papers without a senior female co-author are included and classified as entirely male-authored.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

M.3 Majority female–authored

TABLE M.14: Table 3, majority female-authored

	1950–2015					1990–2015			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Flesch	1.05***	1.04***	1.06***	1.07***	1.22***	1.03**	1.08***	1.02**	1.32**
	(0.36)	(0.36)	(0.37)	(0.37)	(0.37)	(0.38)	(0.35)	(0.39)	(0.49)
Flesch-Kincaid	0.19**	0.18**	0.19**	0.20**	0.21**	0.25**	0.28***	0.26**	0.30**
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)	(0.09)	(0.10)	(0.11)
Gunning Fog	0.29***	0.29***	0.31***	0.31***	0.32***	0.36***	0.36***	0.36***	0.39***
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.12)	(0.11)	(0.13)	(0.13)
SMOG	0.20***	0.20***	0.21***	0.21***	0.22***	0.23***	0.23***	0.23***	0.27***
	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.08)	(0.07)	(0.09)	(0.09)
Dale-Chall	0.06*	0.06*	0.06*	0.07*	0.07*	0.07	0.07	0.07	0.10*
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.06)
No. obs.	8,800	8,800	8,800	8,800	8,800	4,913	4,913	4,913	5,403
Editor	✓	✓	✓	✓	✓	✓	✓	✓	✓
Blind	✓	✓	✓	✓	✓	✓	✓	✓	✓
Journal	✓	✓							
Year		✓							
Journal×Year			✓	✓	✓	✓	✓	✓	✓
N_j				✓	✓	✓	✓	✓	✓
Institution				✓	✓	✓	✓	✓	✓
Quality					✓[1]	✓[1]	✓[1]	✓[1]	✓[1]
Native speaker					✓	✓	✓	✓	✓
JEL (primary)							✓		
Theory/empirical								✓	
JEL (tertiary)									✓

Notes. Estimates are identical to those in Table 3, except that female ratio has been replaced with a dummy variable equal to 1 if a weak majority (50% or more) of authors are female. (Papers with a minority—but positive—number of female authors are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.15: Table 5, majority female-authored

| | OLS (score) | | | OLS (Δ in score) | | FGLS (score) | | | | | |
| | | | | | | Working paper | | Published paper | | Difference | |
	R_{jW}	Female	Blind× female	Female	Blind× female	Female	Blind× female	Female	Blind× female	Female	Blind× female
Flesch	0.83***	1.21***	−1.11	1.02**	−0.84	1.18*	−1.58	2.20***	−2.42	1.02**	−0.84
	(0.02)	(0.42)	(1.90)	(0.42)	(1.16)	(0.68)	(2.11)	(0.71)	(1.63)	(0.41)	(1.14)
Flesch–Kincaid	0.74***	0.40***	−0.48	0.36***	−0.36	0.17	−0.46	0.53***	−0.82**	0.36***	−0.36
	(0.04)	(0.15)	(0.52)	(0.12)	(0.32)	(0.17)	(0.49)	(0.16)	(0.36)	(0.12)	(0.32)
Gunning Fog	0.76***	0.40**	−0.42	0.35***	−0.30	0.21	−0.51	0.57***	−0.81*	0.35***	−0.30
	(0.04)	(0.17)	(0.50)	(0.13)	(0.37)	(0.18)	(0.54)	(0.18)	(0.45)	(0.13)	(0.36)
SMOG	0.79***	0.25**	−0.28	0.22**	−0.20	0.13	−0.36	0.36***	−0.56*	0.22***	−0.20
	(0.03)	(0.11)	(0.33)	(0.09)	(0.24)	(0.11)	(0.38)	(0.12)	(0.32)	(0.09)	(0.24)
Dale–Chall	0.85***	0.14***	−0.12	0.12***	−0.12	0.13**	0.04	0.25***	−0.09	0.12***	−0.12
	(0.02)	(0.03)	(0.07)	(0.04)	(0.09)	(0.05)	(0.16)	(0.06)	(0.16)	(0.03)	(0.08)
No. observations		1,840		1,840		1,840		1,840		1,840	
Editor effects		✓		✓		✓		✓			
Journal×Year effects		✓		✓		✓		✓			
N_j		✓²		✓²		✓²		✓²			
Quality controls		✓		✓		✓		✓			
Native speaker		✓		✓		✓		✓			
Theory/emp. effects		✓		✓		✓		✓			

Notes. Estimates are identical to those in Table 5, except that female ratio has been replaced with a dummy variable equal to 1 if a weak majority (50% or more) of authors are female. (Papers with a minority—but positive—number of female authors are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.16: Table 6, majority female-authored

	1970–2015					1990–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Majority female	3.316***	3.710***	3.691***	3.334**	3.708***	5.840***	5.767***
	(1.175)	(1.319)	(1.318)	(1.292)	(1.316)	(1.730)	(1.759)
Max. t	−0.182***	−0.185***	−0.183***	−0.182***	−0.182***	−0.174***	−0.170***
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.045)	(0.047)
No. pages	0.201***	0.202***	0.201***	0.201***	0.201***	0.239***	0.224***
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.045)	(0.048)
N_j	1.222***	1.214***	1.196***	1.220***	1.206***	1.629***	1.456***
	(0.319)	(0.311)	(0.317)	(0.310)	(0.313)	(0.423)	(0.457)
Order	0.196***	0.195***	0.193***	0.195***	0.193***	0.444***	0.463***
	(0.058)	(0.059)	(0.059)	(0.059)	(0.059)	(0.140)	(0.140)
No. citations (asinh)	−0.410**	−0.430**	−0.416**	−0.410**	−0.427**	−0.712*	−0.672
	(0.202)	(0.201)	(0.201)	(0.202)	(0.201)	(0.417)	(0.411)
Flesch	−0.014	−0.013	−0.013	−0.014	−0.013	−0.032	−0.038
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.030)	(0.030)
Theory	0.450	0.411	0.424	0.449	0.416	0.454	−0.420
	(0.894)	(0.890)	(0.892)	(0.892)	(0.889)	(0.929)	(1.445)
Empirical	1.665	2.128	1.697	1.669	1.633	1.877	1.205
	(1.615)	(1.663)	(1.600)	(1.602)	(1.614)	(1.729)	(2.117)
Other	−0.834	−0.870	−0.836	−0.833	−0.852	−0.579	−1.883*
	(0.797)	(0.768)	(0.785)	(0.798)	(0.794)	(0.813)	(1.005)
Mother			−3.776		−8.387***	−17.688***	−18.422***
			(2.874)		(2.964)	(3.874)	(4.596)
Birth				−0.315	7.675	17.104***	18.268***
				(4.201)	(4.813)	(5.324)	(5.266)
Constant	14.352***	14.476***	14.447***	14.356***	14.474***	16.545***	18.686***
	(1.472)	(1.473)	(1.472)	(1.468)	(1.476)	(2.663)	(2.742)
R^2	0.288	0.290	0.288	0.288	0.289	0.124	0.143
No. observations	2,544	2,529	2,544	2,544	2,544	1,211	1,211
Editor effects	✓	✓	✓	✓	✓	✓	✓
Year effects	✓	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓	✓
JEL (primary) effects							✓

Notes. Estimates are identical to those in Table 6, except that female ratio has been replaced with a dummy variable equal to 1 if a weak majority (50% or more) of authors are female. (Papers with a minority—but positive—number of female authors are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.17: Table 7, majority female-authored

	1970–2015			1990–2015		
	Econometrica	*REStud*	*Econometrica + REStud*	*Econometrica*	*REStud*	*Econometrica + REStud*
Majority female	3.31***	1.03	1.98**	5.44***	2.35**	3.64***
	(1.17)	(0.84)	(0.76)	(1.58)	(0.96)	(0.90)
Max. t	−0.18***	−0.07	−0.14***	−0.17***	−0.04	−0.11**
	(0.04)	(0.07)	(0.04)	(0.05)	(0.08)	(0.05)
No. pages	0.20***	0.14*	0.18***	0.23***	0.03	0.17***
	(0.03)	(0.07)	(0.03)	(0.05)	(0.08)	(0.04)
N_j	1.22***	−0.16	0.73**	1.45***	0.28	1.07***
	(0.32)	(0.49)	(0.28)	(0.48)	(0.66)	(0.39)
Order	0.20***	−0.10	0.07	0.48***	−0.01	0.19
	(0.06)	(0.08)	(0.05)	(0.14)	(0.16)	(0.12)
No. citations (asinh)	−0.41**	−0.56**	−0.43***	−0.62	−1.12**	−0.86**
	(0.20)	(0.22)	(0.16)	(0.42)	(0.46)	(0.33)
Theory	0.46	0.24	0.47	−0.32	−0.31	−0.36
	(0.89)	(1.16)	(0.64)	(1.45)	(1.41)	(0.84)
Empirical	1.63	6.10**	3.80**	1.17	7.60**	4.25**
	(1.62)	(2.70)	(1.63)	(2.10)	(2.99)	(1.71)
Other	−0.86	−2.21*	−1.18*	−1.81*	−2.12	−1.33**
	(0.80)	(1.30)	(0.62)	(1.04)	(1.34)	(0.61)
Constant	13.82***	25.33***	17.79***	16.77***	32.54***	22.86***
	(1.42)	(2.43)	(1.07)	(2.42)	(3.51)	(1.56)
R^2	0.29	0.28	0.30	0.14	0.17	0.14
No. observations	2,544	1,745	4,289	1,211	1,003	2,214
Editor effects	✓	✓	✓	✓	✓	✓
Accepted year effects	✓	✓		✓	✓	
Journal×Accepted year effects			✓			✓
Institution effects	✓	✓	✓	✓	✓	✓
JEL (primary) effects				✓	✓	✓

Notes. Estimates are identical to those in Table 7, except that female ratio has been replaced with a dummy variable equal to 1 if a weak majority (50% or more) of authors are female. (Papers with a minority—but positive—number of female authors are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.18: Table 8, majority female-authored

	$t = 1$	$t = 2$	$t = 3$	$t = 4$–5	$t \geq 6$	All
Flesch	0.73	1.32*	3.56***	2.08	2.14*	2.19***
	(0.55)	(0.70)	(1.15)	(1.58)	(1.30)	(0.65)
Flesch-Kincaid	0.08	0.10	0.75***	0.60**	0.36	0.36***
	(0.14)	(0.18)	(0.19)	(0.27)	(0.31)	(0.14)
Gunning Fog	0.21	0.28	0.93***	0.72**	0.45	0.50***
	(0.15)	(0.21)	(0.25)	(0.29)	(0.35)	(0.17)
SMOG	0.15	0.22	0.57***	0.49**	0.30	0.36***
	(0.11)	(0.14)	(0.18)	(0.24)	(0.25)	(0.12)
Dale-Chall	0.04	0.01	0.15	0.21*	0.26	0.13**
	(0.05)	(0.07)	(0.10)	(0.13)	(0.17)	(0.06)
No. observations	6,401	2,680	1,557	1,777	2,577	9,589
Editor effects	✓	✓	✓	✓	✓	✓
Blind review	✓	✓	✓	✓	✓	✓
Journal×Year effects	✓	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓
Quality controls	✓[3]	✓[3]	✓[3]	✓[3]	✓[3]	✓[1]
Native speaker	✓	✓	✓	✓	✓	✓

Notes. Estimates are identical to those in Table 8, except that female ratio has been replaced with a dummy variable equal to 1 if a weak majority (50% or more) of authors are female. (Papers with a minority—but positive—number of female authors are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.19: Table F.2, majority female-authored

	Flesch Reading Ease	Flesch-Kincaid	Gunning Fog	SMOG	Dale-Chall
Female ratio for women (β_1)	1.68**	0.22	0.32*	0.22*	0.10
	(0.69)	(0.16)	(0.19)	(0.13)	(0.07)
Female ratio for men ($\beta_1 + \beta_2$)	0.59	0.13	0.12	0.06	0.01
	(0.72)	(0.15)	(0.18)	(0.13)	(0.07)
Female ratio×male (β_2)	−1.08	−0.10	−0.20	−0.16	−0.09
	(0.96)	(0.21)	(0.25)	(0.17)	(0.10)
Lagged score (β_0)	0.04**	0.05**	0.03	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
No. observations	9,181	9,181	9,181	9,181	9,181
Tests of instrument validity					
Hansen test (*p*-value)	0.79	0.97	0.96	0.97	0.84
Sargan test (*p*-value)	0.36	0.11	0.31	0.16	0.36
z–test for no serial correlation					
Order 1	−18.48	−13.65	−14.90	−17.51	−18.71
Order 2	0.51	−0.42	0.12	0.44	0.09
Editor effects	✓	✓	✓	✓	✓
Blind review	✓	✓	✓	✓	✓
Journal effects	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓
Quality controls	✓²	✓²	✓²	✓²	✓²
Native speaker	✓	✓	✓	✓	✓

Notes. Estimates are identical to those in Table F.2, except that female ratio has been replaced with a dummy variable equal to 1 if a weak majority (50% or more) of authors are female. (Mixed-gendered papers with a strict minority of female authors are included and classified as male-authored papers.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

M.4   At least one female author

TABLE M.20: Table 3, at least one female author

	1950–2015					1990–2015			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Flesch	0.51	0.51	0.52	0.52	0.63	0.36	0.34	0.33	0.38
	(0.43)	(0.43)	(0.45)	(0.43)	(0.40)	(0.42)	(0.39)	(0.41)	(0.49)
Flesch-Kincaid	0.11	0.12	0.12	0.12	0.13	0.13	0.14*	0.13	0.12
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
Gunning Fog	0.21*	0.21*	0.22**	0.21**	0.21**	0.20*	0.19*	0.21*	0.16
	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)	(0.11)	(0.10)	(0.11)	(0.11)
SMOG	0.14*	0.14*	0.15*	0.14*	0.15**	0.13*	0.12*	0.13*	0.11
	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)	(0.08)	(0.07)	(0.08)	(0.08)
Dale-Chall	0.05	0.05	0.06	0.06	0.07*	0.06	0.06	0.06	0.06
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
No. obs.	9,117	9,117	9,117	9,117	9,117	5,211	5,211	5,211	5,774
Editor	✓	✓	✓	✓	✓	✓	✓	✓	✓
Blind	✓	✓	✓	✓	✓	✓	✓	✓	✓
Journal	✓	✓							
Year		✓							
Journal×Year			✓	✓	✓	✓	✓	✓	✓
N_j				✓	✓	✓	✓	✓	✓
Institution				✓	✓	✓	✓	✓	✓
Quality					✓[1]	✓[1]	✓[1]	✓[1]	✓[1]
Native speaker					✓	✓	✓	✓	✓
JEL (primary)							✓		
Theory/empirical								✓	
JEL (tertiary)									✓

Notes. Estimates are identical to those in Table 3, except that female ratio has been replaced with a dummy variable equal to 1 if at least one author on a paper is female. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.21: Table 5, at least one female author

| | OLS (score) | | | OLS (Δ in score) | | FGLS (score) | | | | | |
| | | | | | | Working paper | | Published paper | | Difference | |
	R_{jW}	Female	Blind× female	Female	Blind× female	Female	Blind× female	Female	Blind× female	Female	Blind× female
Flesch	0.83***	0.56*	−0.37	0.36	−0.34	1.18*	−0.21	1.55**	−0.55	0.36	−0.34
	(0.02)	(0.33)	(1.45)	(0.27)	(1.06)	(0.61)	(2.03)	(0.67)	(1.58)	(0.27)	(1.04)
Flesch–Kincaid	0.74***	0.24**	−0.25	0.17*	−0.18	0.26*	−0.30	0.43***	−0.47	0.17*	−0.18
	(0.04)	(0.11)	(0.40)	(0.09)	(0.30)	(0.13)	(0.50)	(0.15)	(0.41)	(0.09)	(0.29)
Gunning Fog	0.76***	0.22*	−0.17	0.14	−0.09	0.32**	−0.30	0.46***	−0.39	0.14	−0.09
	(0.03)	(0.12)	(0.39)	(0.10)	(0.35)	(0.14)	(0.55)	(0.17)	(0.46)	(0.09)	(0.34)
SMOG	0.79***	0.14*	−0.14	0.10	−0.10	0.20**	−0.20	0.30***	−0.30	0.10	−0.10
	(0.03)	(0.08)	(0.26)	(0.06)	(0.23)	(0.09)	(0.39)	(0.11)	(0.32)	(0.06)	(0.23)
Dale–Chall	0.85***	0.10***	−0.10	0.07***	−0.10	0.16***	0.00	0.23***	−0.10	0.07***	−0.10
	(0.02)	(0.03)	(0.06)	(0.02)	(0.08)	(0.05)	(0.17)	(0.06)	(0.16)	(0.02)	(0.08)
No. observations	1,988			1,988		1,988		1,988		1,988	
Editor effects	✓			✓		✓		✓			
Journal×Year effects	✓			✓		✓		✓			
N_j	✓²			✓²		✓²		✓²			
Quality controls	✓			✓		✓		✓			
Native speaker	✓			✓		✓		✓			
Theory/emp. effects											

Notes. Estimates are identical to those in Table 5, except that female ratio has been replaced with a dummy variable equal to 1 if at least one author on a paper is female. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

68

TABLE M.22: Table 6, at least one female author

	1970–2015					1990–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1+ female	2.671***	2.941***	2.930***	2.682**	2.945***	4.252***	4.223***
	(0.955)	(1.029)	(1.032)	(1.027)	(1.031)	(1.249)	(1.227)
Max. t	−0.153***	−0.156***	−0.155***	−0.153***	−0.153***	−0.146***	−0.151***
	(0.037)	(0.037)	(0.036)	(0.037)	(0.037)	(0.042)	(0.044)
No. pages	0.198***	0.198***	0.197***	0.198***	0.197***	0.234***	0.220***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.041)	(0.044)
N_j	1.041***	1.012***	1.000***	1.039***	1.009***	1.186***	1.072**
	(0.308)	(0.299)	(0.304)	(0.297)	(0.301)	(0.421)	(0.454)
Order	0.199***	0.199***	0.197***	0.199***	0.197***	0.446***	0.478***
	(0.061)	(0.061)	(0.061)	(0.061)	(0.061)	(0.146)	(0.143)
No. citations (asinh)	−0.396*	−0.414**	−0.400**	−0.396*	−0.411**	−0.649	−0.663
	(0.199)	(0.199)	(0.199)	(0.199)	(0.199)	(0.405)	(0.396)
Flesch	−0.016	−0.016	−0.016	−0.016	−0.015	−0.033	−0.037
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.027)	(0.029)
Theory	0.385	0.361	0.369	0.385	0.361	0.430	−0.667
	(0.957)	(0.950)	(0.954)	(0.957)	(0.951)	(0.975)	(1.381)
Empirical	2.381	2.849	2.400	2.384	2.344	2.401	1.348
	(1.751)	(1.738)	(1.731)	(1.738)	(1.744)	(1.849)	(2.143)
Other	−1.223	−1.242	−1.225	−1.223	−1.242	−1.014	−2.261**
	(0.865)	(0.832)	(0.852)	(0.866)	(0.861)	(0.875)	(0.962)
Mother			−3.656		−8.296**	−18.203***	−18.801***
			(2.794)		(3.141)	(4.580)	(5.140)
Birth				−0.282	7.722	18.014***	18.469***
				(4.189)	(5.192)	(6.087)	(6.177)
Constant	14.973***	15.097***	15.082***	14.978***	15.108***	17.252***	19.547***
	(1.397)	(1.400)	(1.400)	(1.394)	(1.404)	(2.390)	(2.508)
R^2	0.288	0.290	0.288	0.288	0.289	0.126	0.145
No. observations	2,623	2,608	2,623	2,623	2,623	1,278	1,278
Editor effects	✓	✓	✓	✓	✓	✓	✓
Year effects	✓	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓	✓
JEL (primary) effects							✓

Notes. Estimates are identical to those in Table 6, except that female ratio has been replaced with a dummy variable equal to 1 if at least one author on a paper is female. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.23: Table 7, at least one female author

	1970–2015			1990–2015		
	Econometrica	REStud	Econometrica + REStud	Econometrica	REStud	Econometrica + REStud
1+ female	2.68***	0.80	1.64***	4.06***	1.56*	2.71***
	(0.95)	(0.72)	(0.60)	(1.14)	(0.82)	(0.70)
Max. t	−0.15***	−0.10	−0.13***	−0.15***	−0.06	−0.11**
	(0.04)	(0.07)	(0.04)	(0.04)	(0.08)	(0.05)
No. pages	0.20***	0.15**	0.19***	0.22***	0.06	0.17***
	(0.03)	(0.07)	(0.03)	(0.04)	(0.07)	(0.03)
N_j	1.04***	−0.16	0.64**	1.11**	0.03	0.80**
	(0.31)	(0.49)	(0.27)	(0.47)	(0.64)	(0.38)
Order	0.20***	−0.09	0.08	0.50***	−0.02	0.19
	(0.06)	(0.08)	(0.05)	(0.14)	(0.16)	(0.12)
No. citations (asinh)	−0.39*	−0.59**	−0.43***	−0.62	−1.11**	−0.86***
	(0.20)	(0.22)	(0.15)	(0.40)	(0.41)	(0.30)
Theory	0.39	0.12	0.32	−0.58	−0.23	−0.46
	(0.96)	(1.11)	(0.64)	(1.38)	(1.38)	(0.83)
Empirical	2.36	5.51**	3.77**	1.36	6.86**	4.16**
	(1.75)	(2.42)	(1.48)	(2.13)	(2.66)	(1.52)
Other	−1.25	−2.27*	−1.50**	−2.20**	−2.37*	−1.83***
	(0.87)	(1.24)	(0.67)	(1.00)	(1.36)	(0.61)
Constant	14.35***	25.51***	18.28***	17.65***	32.69***	23.67***
	(1.40)	(2.30)	(1.07)	(2.30)	(3.26)	(1.52)
R^2	0.29	0.28	0.29	0.14	0.16	0.14
No. observations	2,623	1,812	4,435	1,278	1,069	2,347
Editor effects	✓	✓	✓	✓	✓	✓
Accepted year effects	✓	✓		✓	✓	
Journal×Accepted year effects			✓			✓
Institution effects	✓	✓	✓	✓	✓	✓
JEL (primary) effects				✓	✓	✓

Notes. Estimates are identical to those in Table 7, except that female ratio has been replaced with a dummy variable equal to 1 if at least one author on a paper is female. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.24: Table 8, at least one female author

	$t = 1$	$t = 2$	$t = 3$	$t = 4–5$	$t \geq 6$	All
Flesch	0.54	−0.05	2.96***	1.87	0.83	0.83
	(0.45)	(0.59)	(0.91)	(1.19)	(1.37)	(0.58)
Flesch-Kincaid	0.08	−0.16	0.62***	0.53**	0.15	0.08
	(0.10)	(0.17)	(0.14)	(0.24)	(0.30)	(0.12)
Gunning Fog	0.18	0.00	0.71***	0.68**	0.18	0.19
	(0.12)	(0.20)	(0.18)	(0.28)	(0.36)	(0.15)
SMOG	0.13	0.04	0.45***	0.45**	0.17	0.16
	(0.09)	(0.13)	(0.13)	(0.21)	(0.26)	(0.10)
Dale-Chall	0.05	−0.02	0.18**	0.20**	0.11	0.07
	(0.04)	(0.06)	(0.07)	(0.10)	(0.14)	(0.05)
No. observations	6,874	2,827	1,675	1,906	2,773	12,008
Editor effects	✓	✓	✓	✓	✓	✓
Blind review	✓	✓	✓	✓	✓	✓
Journal×Year effects	✓	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓
Quality controls	✓[3]	✓[3]	✓[3]	✓[3]	✓[3]	✓[1]
Native speaker	✓	✓	✓	✓	✓	✓

Notes. Estimates are identical to those in Table 8, except that female ratio has been replaced with a dummy variable equal to 1 if at least one author on a paper is female. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

71

TABLE M.25: Table F.2, at least one female author

	Flesch Reading Ease	Flesch– Kincaid	Gunning Fog	SMOG	Dale– Chall
Female ratio for women (β_1)	1.00	0.13	0.22	0.16	0.10*
	(0.67)	(0.14)	(0.17)	(0.12)	(0.06)
Female ratio for men ($\beta_1 + \beta_2$)	−0.08	0.00	0.02	0.02	0.04
	(0.56)	(0.12)	(0.14)	(0.10)	(0.05)
Female ratio×male (β_2)	−1.08	−0.13	−0.20	−0.14	−0.06
	(0.80)	(0.17)	(0.20)	(0.15)	(0.07)
Lagged score (β_0)	0.04**	0.05**	0.03	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
No. observations	9,181	9,181	9,181	9,181	9,181
Tests of instrument validity					
Hansen test (*p*-value)	0.81	0.96	0.96	0.97	0.83
Sargan test (*p*-value)	0.37	0.12	0.31	0.16	0.34
z–test for no serial correlation					
Order 1	−18.48	−13.64	−14.90	−17.51	−18.71
Order 2	0.46	−0.45	0.10	0.43	0.11
Editor effects	✓	✓	✓	✓	✓
Blind review	✓	✓	✓	✓	✓
Journal effects	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓
Quality controls	✓^{2}	✓^{2}	✓^{2}	✓^{2}	✓^{2}
Native speaker	✓	✓	✓	✓	✓

Notes. Estimates are identical to those in Table F.2, except that female ratio has been replaced with a dummy variable equal to 1 if at least one author on a paper is female. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.26: Table 3, 100% female-authored

	1950–2015					1990–2015			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Flesch	0.51	0.48	0.44	0.52	0.77	0.70	0.82	0.67	1.25
	(0.64)	(0.63)	(0.64)	(0.65)	(0.66)	(0.68)	(0.66)	(0.69)	(0.96)
Flesch-Kincaid	0.13	0.12	0.11	0.15	0.17	0.27*	0.31**	0.26*	0.34*
	(0.14)	(0.14)	(0.14)	(0.14)	(0.15)	(0.15)	(0.14)	(0.15)	(0.19)
Gunning Fog	0.24	0.24	0.25	0.30*	0.32*	0.46**	0.48***	0.46**	0.48**
	(0.15)	(0.16)	(0.16)	(0.16)	(0.17)	(0.17)	(0.16)	(0.18)	(0.21)
SMOG	0.17	0.17	0.18	0.20	0.22*	0.30**	0.30**	0.29**	0.32**
	(0.11)	(0.11)	(0.11)	(0.12)	(0.12)	(0.13)	(0.12)	(0.13)	(0.16)
Dale-Chall	0.06	0.05	0.05	0.05	0.07	0.12*	0.13**	0.12*	0.17**
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)
No. obs.	8,264	8,264	8,264	8,264	8,264	4,455	4,455	4,455	4,840
Editor	✓	✓	✓	✓	✓	✓	✓	✓	✓
Blind	✓	✓	✓	✓	✓	✓	✓	✓	✓
Journal	✓	✓							
Year		✓							
Journal×Year			✓	✓	✓	✓	✓	✓	✓
N_j				✓	✓	✓	✓	✓	✓
Institution				✓	✓	✓	✓	✓	✓
Quality				✓[1]	✓[1]	✓[1]	✓[1]	✓[1]	✓[1]
Native speaker					✓	✓	✓	✓	✓
JEL (primary)							✓		
Theory/empirical								✓	
JEL (tertiary)									✓

Notes. Estimates are identical to those in Table 3, except that female ratio has been replaced with a dummy variable equal to 1 if all authors on a paper are female. (Papers written by authors of both genders are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.27: Table 5, 100% female-authored

| | | OLS (score) | | OLS (Δ in score) | | FGLS (score) | | | | | |
| | | | | | | Working paper | | Published paper | | Difference | |
	R_{jW}	Female	Blind× female	Female	Blind× female	Female	Blind× female	Female	Blind× female	Female	Blind× female
Flesch	0.84***	0.10	−1.87	−0.16	−1.70	1.61	−1.06	1.44	−2.76	−0.16	−1.70
	(0.03)	(0.91)	(5.88)	(0.99)	(3.57)	(1.44)	(4.30)	(1.69)	(2.95)	(0.97)	(3.50)
Flesch–Kincaid	0.75***	0.30	−1.04	0.26	−0.91	0.17	−0.50	0.43	−1.41**	0.26	−0.91
	(0.05)	(0.29)	(1.08)	(0.32)	(0.90)	(0.25)	(0.94)	(0.39)	(0.65)	(0.31)	(0.88)
Gunning Fog	0.76***	0.34	−1.17	0.27	−1.08	0.27	−0.39	0.54	−1.47**	0.27	−1.08
	(0.04)	(0.28)	(1.10)	(0.33)	(0.94)	(0.29)	(1.08)	(0.40)	(0.67)	(0.32)	(0.92)
SMOG	0.78***	0.14	−0.74	0.08	−0.67	0.26	−0.29	0.34	−0.96**	0.08	−0.67
	(0.03)	(0.16)	(0.80)	(0.20)	(0.62)	(0.24)	(0.80)	(0.27)	(0.46)	(0.19)	(0.60)
Dale–Chall	0.84***	0.08	−0.24	0.03	−0.19	0.28*	−0.27	0.31**	−0.46*	0.03	−0.19
	(0.02)	(0.09)	(0.36)	(0.09)	(0.16)	(0.16)	(0.33)	(0.14)	(0.28)	(0.09)	(0.16)
No. observations		1,633		1,633		1,633		1,633		1,633	
Editor effects		✓		✓		✓		✓			
Journal×Year effects		✓		✓		✓		✓			
N_j											
Quality controls		✓²		✓²		✓²		✓²			
Native speaker		✓		✓		✓		✓			
Theory/emp. effects		✓		✓		✓		✓			

Notes. Estimates are identical to those in Table 5, except that female ratio has been replaced with a dummy variable equal to 1 if all authors on a paper are female. (Papers written by authors of both genders are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.28: Table 6, 100% female-authored

	1970–2015					1990–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Exclusively female	6.297**	9.288**	9.226**	7.039**	9.273**	11.540**	11.950**
	(2.633)	(3.657)	(3.664)	(3.205)	(3.660)	(5.117)	(5.039)
Max. t	−0.186***	−0.186***	−0.184***	−0.186***	−0.183***	−0.173***	−0.167***
	(0.045)	(0.046)	(0.045)	(0.045)	(0.045)	(0.057)	(0.058)
No. pages	0.211***	0.210***	0.208***	0.210***	0.208***	0.251***	0.237***
	(0.027)	(0.026)	(0.026)	(0.027)	(0.026)	(0.038)	(0.039)
N_j	1.288***	1.315***	1.295***	1.284***	1.306***	1.634***	1.460***
	(0.321)	(0.316)	(0.323)	(0.320)	(0.319)	(0.458)	(0.443)
Order	0.174***	0.173***	0.171***	0.173***	0.171***	0.411**	0.424**
	(0.062)	(0.062)	(0.061)	(0.062)	(0.061)	(0.157)	(0.160)
No. citations (asinh)	−0.420**	−0.446**	−0.430**	−0.417**	−0.442**	−0.758*	−0.682
	(0.198)	(0.198)	(0.197)	(0.198)	(0.197)	(0.423)	(0.429)
Flesch	−0.016	−0.015	−0.014	−0.016	−0.014	−0.033	−0.036
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.034)	(0.034)
Theory	0.374	0.313	0.334	0.368	0.325	0.420	−0.077
	(0.848)	(0.849)	(0.858)	(0.854)	(0.852)	(0.881)	(1.424)
Empirical	1.271	1.503	1.089	1.257	1.023	1.109	0.642
	(1.733)	(1.833)	(1.782)	(1.731)	(1.794)	(1.960)	(2.259)
Other	−0.952	−0.965	−0.926	−0.938	−0.941	−0.693	−1.885
	(0.624)	(0.595)	(0.608)	(0.615)	(0.614)	(0.651)	(1.187)
Mother			−9.335**		−14.066***	−24.050***	−25.513***
			(4.480)		(4.699)	(6.430)	(7.104)
Birth				−4.042	7.829	17.797***	19.159***
				(5.169)	(4.866)	(5.216)	(5.174)
Constant	14.349***	14.426***	14.391***	14.347***	14.415***	16.757***	18.288***
	(1.385)	(1.407)	(1.399)	(1.389)	(1.401)	(2.837)	(2.973)
R^2	0.292	0.296	0.294	0.292	0.294	0.139	0.155
No. observations	2,444	2,429	2,444	2,444	2,444	1,140	1,140
Editor effects	✓	✓	✓	✓	✓	✓	✓
Year effects	✓	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓	✓
JEL (primary) effects							✓

Notes. Estimates are identical to those in Table 6, except that female ratio has been replaced with a dummy variable equal to 1 if all authors on a paper are female. (Papers written by authors of both genders are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.29: Table 7, 100% female-authored

| | 1970–2015 | | | 1990–2015 | | |
	Econometrica	REStud	Econometrica + REStud	Econometrica	REStud	Econometrica + REStud
Exclusively female	6.30**	2.68*	4.12***	9.13**	5.39**	6.36***
	(2.63)	(1.55)	(1.40)	(3.80)	(2.27)	(1.95)
Max. t	−0.19***	−0.10	−0.15***	−0.17***	−0.06	−0.13**
	(0.04)	(0.07)	(0.04)	(0.06)	(0.09)	(0.05)
No. pages	0.21***	0.13*	0.19***	0.24***	0.02	0.18***
	(0.03)	(0.08)	(0.03)	(0.04)	(0.09)	(0.04)
N_j	1.28***	0.14	0.87***	1.42***	0.89	1.21***
	(0.32)	(0.51)	(0.28)	(0.45)	(0.64)	(0.37)
Order	0.18***	−0.05	0.08	0.45***	0.09	0.22*
	(0.06)	(0.08)	(0.05)	(0.16)	(0.17)	(0.12)
No. citations (asinh)	−0.42**	−0.43*	−0.40**	−0.64	−0.86*	−0.79**
	(0.20)	(0.22)	(0.15)	(0.44)	(0.46)	(0.34)
Theory	0.38	0.00	0.32	0.08	−0.74	−0.32
	(0.85)	(1.14)	(0.61)	(1.41)	(1.37)	(0.71)
Empirical	1.23	7.79**	4.32**	0.80	9.76**	4.89**
	(1.74)	(3.22)	(1.78)	(2.24)	(3.63)	(1.89)
Other	−0.98	−1.53	−1.12**	−1.80	−2.19	−1.59**
	(0.63)	(1.17)	(0.50)	(1.26)	(1.52)	(0.65)
Constant	13.75***	23.43***	17.12***	16.40***	30.51***	22.09***
	(1.34)	(2.15)	(1.06)	(2.75)	(3.36)	(1.80)
R^2	0.29	0.29	0.30	0.15	0.18	0.16
No. observations	2,444	1,636	4,080	1,140	915	2,055
Editor effects	✓	✓	✓	✓	✓	✓
Accepted year effects	✓	✓		✓	✓	
Journal×Accepted year effects			✓			✓
Institution effects	✓	✓	✓	✓	✓	✓
JEL (primary) effects				✓	✓	✓

Notes. Estimates are identical to those in Table 7, except that female ratio has been replaced with a dummy variable equal to 1 if all authors on a paper are female. (Papers written by authors of both genders are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.30: Table 8, 100% female-authored

	$t = 1$	$t = 2$	$t = 3$	$t = 4\text{–}5$	$t \geq 6$	All
Flesch	−0.26	1.00	4.74***	2.45	1.45	2.65*
	(0.90)	(1.35)	(1.50)	(2.71)	(3.27)	(1.36)
Flesch-Kincaid	−0.07	0.22	1.20***	0.33	0.36	0.39
	(0.21)	(0.27)	(0.35)	(0.58)	(0.56)	(0.34)
Gunning Fog	0.00	0.53*	1.68***	0.67	0.76	0.65*
	(0.24)	(0.31)	(0.43)	(0.66)	(0.51)	(0.39)
SMOG	−0.01	0.38*	1.09***	0.62	0.50*	0.49*
	(0.17)	(0.22)	(0.29)	(0.55)	(0.30)	(0.26)
Dale-Chall	−0.02	0.06	0.37**	0.53*	0.61***	0.18
	(0.09)	(0.12)	(0.15)	(0.31)	(0.21)	(0.14)
No. observations	5,886	2,451	1,453	1,642	2,384	8,084
Editor effects	✓	✓	✓	✓	✓	✓
Blind review	✓	✓	✓	✓	✓	✓
Journal×Year effects	✓	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓
Quality controls	✓[3]	✓[3]	✓[3]	✓[3]	✓[3]	✓[1]
Native speaker	✓	✓	✓	✓	✓	✓

Notes. Estimates are identical to those in Table 8, except that female ratio has been replaced with a dummy variable equal to 1 if all authors on a paper are female. (Papers written by authors of both genders are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.31: Table F.2, 100% female-authored

	Flesch Reading Ease	Flesch–Kincaid	Gunning Fog	SMOG	Dale-Chall
Female ratio for women (β_1)	1.57	0.06	0.35	0.32	0.30***
	(1.02)	(0.22)	(0.27)	(0.20)	(0.10)
Lagged score (β_0)	0.04**	0.05**	0.03	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
No. observations	9,181	9,181	9,181	9,181	9,181
Tests of instrument validity					
Hansen test (p-value)	0.79	0.97	0.96	0.97	0.84
Sargan test (p-value)	0.36	0.11	0.31	0.16	0.36
z–test for no serial correlation					
Order 1	−18.47	−13.64	−14.89	−17.50	−18.71
Order 2	0.48	−0.45	0.10	0.44	0.09
Editor effects	✓	✓	✓	✓	✓
Blind review	✓	✓	✓	✓	✓
Journal effects	✓	✓	✓	✓	✓
N_j	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓
Quality controls	✓²	✓²	✓²	✓²	✓²
Native speaker	✓	✓	✓	✓	✓

Notes. Estimates are identical to those in Table F.2, except that female ratio has been replaced with a dummy variable equal to 1 if all authors on a paper are female. (Co-authored mixed-gendered papers are included and classified as male.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

M.6 Senior female author, sample of less experienced authors

TABLE M.32: Table 3, senior female author ($t \leq 3$)

	1950–2015					1990–2015			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Flesch	0.52	0.52	0.42	0.54	0.73	0.72	0.80	0.70	0.69
	(0.68)	(0.69)	(0.70)	(0.72)	(0.76)	(0.85)	(0.82)	(0.85)	(1.03)
Flesch-Kincaid	0.09	0.09	0.07	0.11	0.14	0.24	0.27	0.24	0.21
	(0.15)	(0.15)	(0.15)	(0.16)	(0.17)	(0.18)	(0.17)	(0.18)	(0.21)
Gunning Fog	0.22	0.23	0.22	0.28	0.31*	0.46**	0.46**	0.46**	0.34
	(0.16)	(0.16)	(0.16)	(0.17)	(0.18)	(0.20)	(0.19)	(0.20)	(0.24)
SMOG	0.15	0.16	0.15	0.19	0.21	0.30*	0.28*	0.30*	0.22
	(0.12)	(0.12)	(0.12)	(0.13)	(0.13)	(0.15)	(0.14)	(0.15)	(0.17)
Dale-Chall	0.02	0.02	0.01	0.01	0.03	0.08	0.08	0.08	0.10
	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
No. obs.	6,762	6,762	6,762	6,762	6,762	3,521	3,521	3,521	3,858
Editor	✓	✓	✓	✓	✓	✓	✓	✓	✓
Blind	✓	✓	✓	✓	✓	✓	✓	✓	✓
Journal	✓	✓							
Year		✓							
Journal×Year			✓	✓	✓	✓	✓	✓	✓
N_j			✓	✓	✓	✓	✓	✓	✓
Institution			✓	✓	✓	✓	✓	✓	✓
Quality				✓[1]	✓[1]	✓[1]	✓[1]	✓[1]	
Native speaker				✓	✓	✓	✓	✓	
JEL (primary)							✓		
Theory/empirical								✓	
JEL (tertiary)									✓

Notes. Estimates are identical to those in Table 3, except that only papers by less experienced authors (defined as having three or fewer previous top-five articles) are included in the sample and female ratio has been replaced with the interaction between female ratio and a dummy variable equal to 1 if a female author had strictly more top-five papers as her co-authors at the time the paper was published. (Mixed-gendered papers with a senior male co-author are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.33: Table 5, senior female author ($t \leq 3$)

| | | OLS (score) | | OLS (Δ in score) | | FGLS (score) | | | | | |
| | | | | | | Working paper | | Published paper | | Difference | |
	R_{jW}	Female	Blind× female	Female	Blind× female	Female	Blind× female	Female	Blind× female	Female	Blind× female
Flesch	0.77***	1.35	1.15	1.44	0.81	−0.41	1.52	1.03	2.33	1.44	0.81
	(0.04)	(1.39)	(2.11)	(1.38)	(2.59)	(1.88)	(5.11)	(2.35)	(4.18)	(1.30)	(2.44)
Flesch–Kincaid	0.69***	0.70*	−0.50	0.76*	−0.47	−0.18	−0.09	0.58	−0.56	0.76**	−0.47
	(0.07)	(0.39)	(0.46)	(0.39)	(0.79)	(0.36)	(1.24)	(0.52)	(0.94)	(0.37)	(0.75)
Gunning Fog	0.72***	0.72*	−0.66	0.80*	−0.62	−0.29	−0.15	0.51	−0.77	0.80*	−0.62
	(0.06)	(0.41)	(0.48)	(0.44)	(0.75)	(0.41)	(1.30)	(0.57)	(1.01)	(0.41)	(0.71)
SMOG	0.75***	0.41	−0.34	0.45	−0.27	−0.15	−0.26	0.30	−0.53	0.45*	−0.27
	(0.05)	(0.26)	(0.28)	(0.28)	(0.48)	(0.30)	(0.90)	(0.39)	(0.69)	(0.27)	(0.45)
Dale–Chall	0.81***	0.15	−0.11	0.14	−0.08	0.06	−0.16	0.19	−0.24	0.14	−0.08
	(0.03)	(0.13)	(0.25)	(0.13)	(0.15)	(0.20)	(0.48)	(0.24)	(0.46)	(0.13)	(0.14)
No. observations	524			524		524		524		524	
Editor effects	✓			✓		✓		✓			
Journal×Year effects	✓			✓		✓		✓			
N_j	✓²			✓²		✓²		✓²			
Native speaker	✓			✓		✓		✓			
Theory/emp. effects	✓			✓		✓		✓			

Notes. Estimates are identical to those in Table 5, except that only papers by less experienced authors (defined as having three or fewer previous top–five articles) are included in the sample and female ratio has been replaced with the interaction between female ratio and a dummy variable equal to 1 if a female author had strictly more top–five papers as her co–authors at the time the paper was published. (Mixed–gendered papers with a senior male co–author are excluded.) *** , ** , and * statistically significant at 1%, 5% and 10%, respectively.

80

TABLE M.34: Table 6, senior female author ($t \leq 3$)

	1970–2015					1990–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sr. female ($t \leq 3$)	4.511	7.376*	7.413*	5.613	7.436*	9.274	8.803
	(2.977)	(3.857)	(3.848)	(3.450)	(3.848)	(5.588)	(5.576)
Max. t	−0.129**	−0.129*	−0.126*	−0.129*	−0.125*	−0.134	−0.134
	(0.064)	(0.065)	(0.065)	(0.064)	(0.065)	(0.082)	(0.085)
No. pages	0.206***	0.204***	0.203***	0.205***	0.203***	0.234***	0.214***
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.045)	(0.044)
N_j	0.840**	0.850**	0.844**	0.834**	0.848**	0.987	0.605
	(0.346)	(0.344)	(0.346)	(0.345)	(0.345)	(0.595)	(0.656)
Order	0.106	0.103	0.103	0.105	0.103	0.287	0.284
	(0.072)	(0.071)	(0.071)	(0.072)	(0.071)	(0.173)	(0.182)
No. citations (asinh)	−0.243	−0.268	−0.262	−0.240	−0.269	−0.358	−0.321
	(0.228)	(0.228)	(0.228)	(0.229)	(0.228)	(0.524)	(0.528)
Flesch	−0.007	−0.005	−0.005	−0.007	−0.005	−0.028	−0.041
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.031)	(0.029)
Theory	0.430	0.304	0.375	0.426	0.365	0.435	0.975
	(1.066)	(1.085)	(1.071)	(1.074)	(1.069)	(1.121)	(1.669)
Empirical	1.520	1.950	1.357	1.581	1.283	1.265	0.247
	(2.483)	(2.528)	(2.488)	(2.434)	(2.486)	(2.658)	(2.848)
Other	−2.065**	−2.145**	−2.037**	−2.030**	−2.052**	−1.824*	−3.678**
	(0.987)	(0.971)	(0.979)	(0.981)	(0.983)	(1.066)	(1.511)
Mother			−10.914**		−13.443***	−22.558***	−24.111***
			(4.601)		(4.983)	(6.915)	(7.936)
Birth				−7.321	4.319	11.280**	14.061**
				(4.779)	(4.037)	(5.280)	(5.601)
Constant	14.783***	14.938***	14.877***	14.784***	14.899***	17.474***	20.103***
	(1.694)	(1.707)	(1.709)	(1.706)	(1.710)	(3.171)	(3.513)
R^2	0.300	0.305	0.303	0.301	0.304	0.137	0.160
No. observations	1,901	1,889	1,901	1,901	1,901	843	843
Editor effects	✓	✓	✓	✓	✓	✓	✓
Year effects	✓	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓	✓
JEL (primary) effects							✓

Notes. Estimates are identical to those in Table 6, except that only papers by less experienced authors (defined as having three or fewer previous top-five articles) are included in the sample and female ratio has been replaced with the interaction between female ratio and a dummy variable equal to 1 if a female author had strictly more top-five papers as her co-authors at the time the paper was published. (Mixed-gendered papers with a senior male co-author are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

TABLE M.35: Table 7, senior female author ($t \leq 3$)

	1970–2015			1990–2015		
	Econometrica	*REStud*	*Econometrica + REStud*	*Econometrica*	*REStud*	*Econometrica + REStud*
Sr. female ($t \leq 3$)	4.52	2.67	3.38**	5.77	5.06**	5.03**
	(2.97)	(1.65)	(1.52)	(4.56)	(2.30)	(2.21)
Max. t	−0.13**	−0.01	−0.08	−0.13	0.05	−0.06
	(0.06)	(0.11)	(0.06)	(0.09)	(0.14)	(0.07)
No. pages	0.21***	0.18**	0.20***	0.22***	0.06	0.18***
	(0.03)	(0.09)	(0.03)	(0.04)	(0.10)	(0.04)
N_j	0.84**	−0.69	0.24	0.55	−0.23	0.28
	(0.35)	(0.56)	(0.31)	(0.66)	(0.73)	(0.49)
Order	0.11	−0.01	0.07	0.32	0.08	0.16
	(0.07)	(0.08)	(0.05)	(0.19)	(0.15)	(0.12)
No. citations (asinh)	−0.24	−0.59**	−0.35**	−0.24	−1.33**	−0.82**
	(0.23)	(0.22)	(0.16)	(0.54)	(0.50)	(0.35)
Theory	0.43	−0.32	0.12	1.08	−1.05	−0.25
	(1.07)	(1.09)	(0.81)	(1.68)	(1.53)	(1.05)
Empirical	1.51	4.48*	2.80	0.30	6.85**	3.33*
	(2.49)	(2.31)	(1.71)	(2.97)	(2.52)	(1.84)
Other	−2.07**	−2.01	−2.02**	−3.67**	−2.12	−2.66**
	(0.99)	(1.74)	(0.78)	(1.56)	(1.89)	(0.99)
Constant	14.53***	24.45***	18.22***	17.89***	33.18***	24.82***
	(1.67)	(3.01)	(1.33)	(3.38)	(4.31)	(2.06)
R^2	0.30	0.32	0.32	0.15	0.20	0.16
No. observations	1,901	1,324	3,225	843	730	1,573
Editor effects	✓	✓	✓	✓	✓	✓
Accepted year effects	✓	✓		✓	✓	
Journal×Accepted year effects			✓			✓
Institution effects	✓	✓	✓	✓	✓	✓
JEL (primary) effects				✓	✓	✓

Notes. Estimates are identical to those in Table 7, except that only papers by less experienced authors (defined as having three or fewer previous top-five articles) are included in the sample and female ratio has been replaced with the interaction between female ratio and a dummy variable equal to 1 if a female author had strictly more top-five papers as her co-authors at the time the paper was published. (Mixed-gendered papers with a senior male co-author are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

References

Abrevaya, J. and D. S. Hamermesh (2012). "Charity and favoritism in the field: are female economists nicer (to each other)?" *Review of Economics and Statistics* 94 (1), pp. 202–207.

Alkhurayyif, Y. and G. R. S. Weir (2017). "Readability as a basis for information security policy assessment". In: *Seventh International Conference on Emerging Security Technologies*. Canterbury, pp. 114–121.

American Economic Association (2022). *EconLit [database]*. https://www.aeaweb.org/econlit/ (last accessed January 2019). Nashville, Tennessee.

Arellano, M. and S. Bond (1991). "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations". *Review of Economic Studies* 58 (2), pp. 277–297.

Arellano, M. and O. Bover (1995). "Another look at the instrumental variable estimation of error-components models". *Journal of Econometrics* 68 (1), pp. 29–51.

Armbruster, B. B. (1984). "The problem of inconsiderate text". In: *Comprehension Instruction: Perspectives and Suggestions*. Ed. by G. G. Duffy, L. R. Roehler, and J. Mason. New York, New York: Longman, pp. 202–217.

Benson, A., D. Li, and K. Shue (2021). "'Potential' and the gender promotion gap". Mimeo.

Berninger, M. et al. (2017). "Confused but convinced: article complexity and the number of citations". Mimeo.

Biddle, G. C., G. Hilary, and R. S. Verdi (2009). "How does financial reporting quality relate to investment efficiency?" *Journal of Accounting and Economics* 48 (2-3), pp. 112–131.

Blank, R. M. (1991). "The effects of double-blind versus single-blind reviewing: experimental evidence from the *American Economic Review*". *American Economic Review* 81 (5), pp. 1041–1067.

Blundell, R. and S. Bond (1998). "Initial conditions and moment restrictions in dynamic panel data models". *Journal of Econometrics* 87 (1), pp. 115–143.

Boring, A. (2017). "Gender biases in student evaluations of teaching". *Journal of Public Economics* 145 (Supplement C), pp. 27–41.

Borsuk, R. M. et al. (2009). "To name or not to name: the effect of changing author gender on peer review". *BioScience* 59 (11), pp. 985–989.

Card, D. et al. (2020). "Are referees and editors in economics gender neutral?" *Quarterly Journal of Economics* 135 (1), pp. 269–327.

Ceci, S. J. et al. (2014). "Women in academic science: a changing landscape". *Psychological Science in the Public Interest* 15 (3), pp. 75–141.

Chall, J. S., S. S. Conard, and S. H. Harris (1977). *An analysis of textbooks in relation to declining sat scores*. Tech. rep. Prepared for the Advisory Panel on the Scholastic Aptitude Test Score Decline. Princeton, New Jersey.

Chall, J. S., A. Freeman, and B. Levy (1983). "Minimum competency testing of reading: an analysis of eight tests designed for grade 11". In: *The Courts, Validity, and Minimum Competency Testing*. Ed. by G. F. Madaus. Boston, Massachussetts: Kluwer-Nijhoff. Chap. 10, pp. 197–208.

Chall, J. S. and E. Dale (1995). *Readability Revisited: the New Dale-Chall Readability Formula*. Cambridge, Massachusetts: Brookline Books.

Chari, A. and P. Goldsmith-Pinkham (2017). "Gender representation in economics across topics and time: evidence from the NBER Summer Institute". NBER Working Paper Series, No. 23953.

Clarivate (2022). *Web of Science [database]*. https://www.webofknowledge.com/ (last accessed January 2018). London.

Coleman, E. B. (1964). "The comprehensibility of several grammatical transformations". *Journal of Applied Psychology* 48 (3), pp. 186–190.

— (1965). "Learning of prose written in four grammatical transformations". *Journal of Applied Psychology* 49 (5), pp. 332–341.

Correll, S. and C. Simard (2016). "Vague feedback is holding women back". *Harvard Business Review*. https://hbr.org/2016/04/research-vague-feedback-is-holding-women-back (last accessed: 4 October 2016).

Costa, P. T., A. Terracciano, and R. R. McCrae (2001). "Gender differences in personality traits across cultures: robust and surprising findings". *Journal of Personality and Social Psychology* 81 (2), pp. 322–331.

Cox, C. (2007). *Closing remarks to the second annual corporate governance summit*. Delivered at USC Marshall School of Business, Los Angeles, California, 23 March.

Dale, E. and J. S. Chall (1948). "A formula for predicting readability". *Educational Research Bulletin* 27 (1), pp. 11–20.

De Franco, G. et al. (2015). "Analyst report readability". *Contemporary Accounting Research* 32 (1), pp. 76–104.

Dion, M. L., J. L. Sumner, and S. M. L. Mitchell (2018). "Gendered citation patterns across political science and social science methodology fields". *Political Analysis* 26 (3), pp. 312–327.

Dowling, M., H. Hammami, and O. Zreik (2018). "Easy to read, easy to cite?" *Economics Letters* 173, pp. 100–103.

DuBay, W. H. (2004). *The Principles of Readability*. Costa Mesa, California: Impact Information.

Dupas, P. et al. (2021). "Gender and the dynamics of economics seminars". NBER Working Paper Series, No. 28494.

Enke, B. (2020). "Moral values and voting: Trump and beyond". *Journal of Political Economy* 128 (10), pp. 3679–3729.

Fass, W. and G. M. Schumacher (1978). "Effects of motivation, subject activity, and readability on the retention of prose materials". *Journal of Educational Psychology* 70 (5), pp. 803–807.

Feingold, A. (1994). "Gender differences in personality: a meta-analysis". *Psychological Bulletin* 116 (3), pp. 429–456.

Ferber, M. A. (1986). "Citations: are they an objective measure of scholarly merit?" *Signs: Journal of Women in Culture and Society* 11 (2), pp. 381–389.

— (1988). "Citations and networking". *Gender and Society* 2 (1), pp. 82–89.

Foster, D. R. and D. H. Rhoney (2002). "Readability of printed patient information for epileptic patients". *Annals of Pharmacotherapy* 36 (12), pp. 1856–1861.

Gilbert, J. R., E. S. Williams, and G. D. Lundberg (1994). "Is there gender bias in *JAMA*'s peer review process?" *Journal of the American Medical Association* 272 (2), pp. 139–142.

Grossbard, S., T. Yilmazer, and L. Zhang (2021). "The gender gap in citations of articles published in two demographic economics journals". *Review of Economics of the Household* 19 (3), pp. 677–697.

Guerini, M., A. Pepe, and B. Lepri (2012). "Do linguistic style and readability of scientific abstracts affect their virality?" In: *Proceedings of the Sixth International AAAI Conference of Weblogs and Social Media*. Dublin, pp. 475–478.

Hamermesh, D. S. (1994). "Facts and myths about refereeing". *Journal of Economic Perspectives* 8 (1), pp. 153–163.

Handley, G. et al. (2015). "An examination of gender differences in the American Fisheries Society peer-review process". *Fisheries* 40 (9), pp. 442–451.

Hartley, J., J. W. Pennebaker, and C. Fox (2003). "Using new technology to assess the academic writing styles of male and female pairs and individuals". *Journal of Technical Writing and Communication* 33 (3), pp. 243–261.

Hengel, E. (2016). "Publishing while female: gender differences in peer review scrutiny". Mimeo.

— (2017). "Publishing while female: are women held to higher standards? Evidence from peer review". Cambridge Working Paper Economics: 1753.

Hengel, E. and E. Moon (2020). "Gender and quality at top economics journals". Mimeo.

Hussin, M. F. et al. (2012). "The readability of transmission line characteristics lab manual". In: *IEEE Control and System Graduate Research Colloquium*. Shah Alam, Selangor, pp. 398–401.

Jansen, D. J. (2011). "Does the clarity of central bank communication affect volatility in financial markets? Evidence from Humphrey-Hawkins testimonies". *Contemporary Economic Policy* 29 (4), pp. 494–509.

Johnson, M. and V. S. Helgeson (2002). "Sex differences in response to evaluative feedback: a field study". *Psychology of Women Quarterly* 26 (3), pp. 242–251.

Kemper, S. (1983). "Measuring the inference load of a text". *Journal of Educational Psychology* 75 (3), pp. 391–401.

King, D. W., C. Tenopir, and M. Clarke (2006). "Measuring total reading of journal articles". *D-Lib Magazine* 12 (10), pp. 1082–9873.

Kintsch, W. and J. R. Miller (1984). "Readability: a view from cognitive psychology". In: *Understanding Reading Comprehension*. Ed. by J. Flood. Newark, Delaware: International Reading Association, pp. 220–232.

Klare, G. R. (1976). "Judging readability". *Instructional Science* 5 (1), pp. 55–61.

Klare, G. R. and K. L. Smart (1973). "Analysis of the readability level of selected USAFI instructional materials". *Journal of Educational Research* 67 (4), p. 176.

Kleven, H. J. (2018). *Language trends in public economics*. https://www.henrikkleven.com/uploads/3/7/3/1/37310663/languagetrends_slides_kleven.pdf (last accessed: 2 December 2018).

Kleven, H. J. and D. Scott (2018). *Language trends in public economics [database]*.

Koffi, M. (2021). "Innovative ideas and gender inequality". Mimeo.

Laband, D. N. and C. N. Taylor (1992). "The impact of bad writing in economics". *Economic Inquiry* 30 (4), pp. 673–688.

Law, D. S. and D. Zaring (2010). "Law versus ideology: the Supreme Court and the use of legislative history". *William and Mary Law Review* 51 (5), pp. 1653–1747.

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

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

Lei, L. and S. Yan (2016). "Readability and citations in information science: evidence from abstracts and articles of four journals (2003–2012)". *Scientometrics* 108 (3), pp. 1155–1169.

Leuven, E. and B. Sianesi (2003). *Psmatch2: stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing*. http://ideas.repec.org/c/boc/bocode/s432001.html (last accessed: 20 March 2018).

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

Liang, F. M. (1983). "Word hy-phen-a-tion by com-put-er". PhD thesis. Stanford University.

Loughran, T. and B. McDonald (2016). "Textual analysis in accounting and finance: a survey". *Journal of Accounting Research* 54 (4), pp. 1187–1230.

Marino Fages, D. (2020). "Write better, publish better". *Scientometrics* 122 (3), pp. 1671–1681.

McCannon, B. C. (2019). "Readability and research impact". *Economics Letters* 180 (July), pp. 76–79.

McGillivray, B. and E. De Ranieri (2018). "Uptake and outcome of manuscripts in *Nature* journals by review model and author characteristics". *Research Integrity and Peer Review* 3 (1), pp. 5–17.

Meade, C. D. and J. C. Byrd (1989). "Patient literacy and the readability of smoking education literature". *American Journal of Public Health* 79 (2), pp. 204–206.

Meyer, B. J. F. (1982). "Reading research and the composition teacher: the importance of plans". *College Composition and Communication* 33 (1), pp. 37–49.

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

Nature Neuroscience (2006). "Women in neuroscience: a numbers game". *Nature Neuroscience* 9 (7), p. 853.

Plavén-Sigray, P. et al. (2017). "The readability of scientific texts is decreasing over time". *eLife* 6 (e27725), pp. 1–14.

Püttmann, L. (2017). *VoxEU gobbledygook*. `http://lukaspuettmann.com/2017/12/09/voxeu-gobbledygook/` (last accessed: 19 May 2018).

Richardson, J. V. (1977). "Readability and readership of journals in library science". *Journal of Academic Librarianship* 3 (1), pp. 20–22.

Roberts, T.-A. and S. Nolen-Hoeksema (1989). "Sex differences in reactions to evaluative feedback". *Sex Roles* 21 (11-12), pp. 725–747.

Roodman, D. (2009). "Practitioners' corner: a note on the theme of too many instruments". *Oxford Bulletin of Economics and Statistics* 71 (1), pp. 135–158.

Sawyer, A. G., J. Laran, and J. Xu (2008). "The readability of marketing journals: are award-winning articles better written?" *Journal of Marketing* 72 (1), pp. 108–117.

Schmidt, B. (2015). "Gender bias exists in professor evaluations". *New York Times*. `http://www.nytimes.com/roomfordebate/2015/12/16/is-it-fair-to-rate-professors-online/gender-bias-exists-in-professor-evaluations` (last accessed: 4 October 2016).

Sirico, L. J. (2007). "Readability studies: how technocentrism can compromise research and legal determinations". *Quinnipiac Law Review* 26 (1), pp. 147–172.

Spirling, A. (2016). "Democratization and linguistic complexity: the effect of franchise extension on parliamentary discourse, 1832–1915". *Journal of Politics* 78 (1), pp. 120–136.

Swanson, C. E. (1948). "Readability and readership: a controlled experiment". *Journalism Bulletin* 25 (4), pp. 339–343.

Thörnqvist, T. (2015). "Sophistication, news and individual investor trading". Mimeo.

Torgler, B. and M. Piatti (2013). *A Century of American Economic Review*. New York, New York: Palgrave Macmillan.

Tregenza, T. (2002). "Gender bias in the refereeing process?" *Trends in Ecology & Evolution* 17 (8), pp. 349–350.

Wallace, L. S. et al. (2008). "Suitability and readability of consumer medical information accompanying prescription medication samples". *Patient Education and Counseling* 70 (3), pp. 420–425.

Weisberg, Y. J., C. G. De Young, and J. B. Hirsh (2011). "Gender differences in personality across the ten aspects of the big five". *Frontiers in Psychology* 2 (178).