

Publishing while female

Are women held to higher standards? Evidence from peer review.

Erin Hengel, University of Liverpool, Chatham Street, Liverpool L69 7ZH. erin.hengel@liverpool.ac.uk

According to raw numerical counts, women produce less than men. For example, 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).

Yet there is another side to female productivity that is often ignored: when evaluated by narrowly defined quality measures, women often outperform. For example, 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., 2016).

In a recent study, I show that female economists surpass men on another dimension: writing clarity. Using five readability measures, I find that female-authored articles published in top economics journals are better written than equivalent papers by men.

Why? Because they have to be. In a model of an author's decision making process, I show that tougher editorial standards and/or biased referee assignment are uniquely consistent with women's observed pattern of choices. I then document evidence that higher standards affect behaviour and lower productivity.

Higher standards impose a quantity/quality tradeoff that likely contributes to academia's "Publishing Paradox" and "Leaky Pipeline".¹ 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. Moreover, evidence of this tradeoff is present in a variety of occupations—*e.g.*, doctors, lawyers and real estate agents—suggesting higher standards distort women's productivity, more generally.

Is it really discrimination?

To determine readability, I rely on a well-known relationship: simple vocabulary and short sentences are easier to understand and straightforward to quantify. Using the five most widely used, studied and reliable formulas to exploit this, I analyse 9,123 article abstracts published in the *American Economic Review*, *Econometrica*, *Journal of Political Economy* and *Quarterly Journal of Economics*.²

First, female-authored abstracts are 1–6 percent better written than similar papers by men. The difference cannot be explained by year, journal, editor, topic, institution, English language ability or with various proxies for article quality and author productivity. This means the readability gap probably wasn't (i) a response to specific policies in earlier eras; (ii) caused by women writing on topics that are easier to explain; (iii) generated by factors correlated with gender but really related to knowledge, intelligence and creativity; nor (iv) due to a lopsided concentration of female native English speakers.³

Second, the gap widens precisely while papers are being reviewed. To show this, I analyse readability before and after review by comparing published articles to earlier drafts released by the National Bureau of Economic Research (NBER) Technical and Working Paper Series.⁴

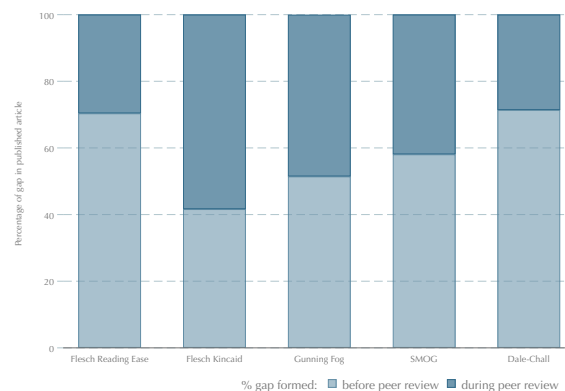


Fig. 1. Figure compares the gap formed before peer review (light blue) to the gap formed in peer review (dark blue) as a percentage of the gender gap in the published article.

Fig. 1 compares the gap formed before peer review (light blue) to the gap formed in peer review (dark blue) as a percentage of the gender gap in the published article. It suggests peer review is directly responsible for almost half of the gender readability gap.

Why does peer review cause women to write more clearly? There are two possible explanations. Either women voluntarily write better papers—*e.g.*, because they're more sensitive to referee criticism or overestimate the importance of writing well—or better written papers are women's response to higher standards imposed by referees and/or editors.

Both explanations imply women spend too much time rewriting old papers and not enough time writing new papers. However, my evidence suggests the latter is primarily to blame. To show this, I model an author's decision-making process over time. The model establishes three sufficient conditions to test for higher standards in peer review.

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.

The intuition behind these conditions is simple. Assuming preferences do not change over time, authors improve readability today relative to yesterday only if they believe better writing leads to higher acceptance rates. Of course, oversensitivity and/or poor information may distort their beliefs—and affect readability—but the impact declines with experience. Holding acceptance rates constant, this implies that a widening readability gap between equivalent authors is caused by discrimination—*i.e.*, asymmetric editorial standards and/or biased referee assignment beyond women's control.

¹ "Publishing Paradox" and "Leaky Pipeline" refer to phenomena in academia whereby women publish fewer papers and disproportionately leave the profession, respectively.

² Readability scores are highly correlated across an article's abstract, introduction and discussion sections (Hartley et al., 2003).

³ It is not clear how—or even if—native English speakers write more clearly than non-native speakers. In fact, Hayden (2008) found that peer reviewed articles by the latter are more readable, on average.

⁴ NBER persistently releases its working papers two to three years before publication (mean 2.1 years)—precisely the length of time papers spend in peer review (Ellison, 2002; Goldberg, 2015).

On average, conditions 1 and 2 hold. Experienced female economists write better than equivalent male economists and women improve their writing over time (but men don't) (Fig. 2). Between authors' first and third published articles, the readability gap increases by up to 12 percent. Although my data do not identify probability of acceptance, conclusions from extensive study elsewhere suggest no gender difference (see, *e.g.*, Ceci et al., 2014).

Technically, however, each condition must hold for the same author in two different situations—before and after gaining experience and when compared to an equivalent, experienced author of the opposite gender. To account for this, I match prolific female authors to similarly productive male authors on characteristics that predict the topic, novelty and quality of research.

I found evidence of discrimination in 60–70 percent of matched pairs. I then subtracted experienced male scores from experienced female scores within each of these matched pairs. Fig. 3 displays their distribution.

In the absence of systemic discrimination against women (or men), differences in Fig. 3 should symmetrically distribute around zero. They obviously don't. Not only is discrimination usually against women, but instances of obvious discrimination predominately are too: differences in Fig. 3 are, on average, 8.5 times more likely to be one standard deviation above zero (indicating discrimination against women) than below it (indicating discrimination against men).

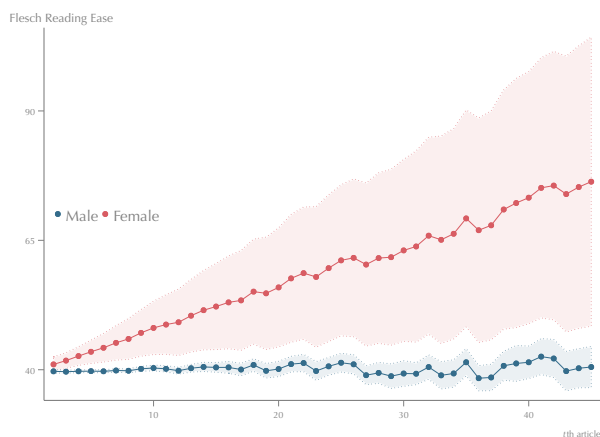


Fig. 2. Flesch Reading Ease marginal mean scores for author's first, second, ..., *t*th publication in the data. Pink represents women co-authoring only with other women; blue are men co-authoring only with other men.

Within pair differences from Fig. 3 can also be used to generate unconditional (conservative) estimates of the effect of higher standards on authors' readability (for details, see Hengel, 2017). On average, they suggest that discrimination causes senior female economists to write (at least) nine percent more clearly than they otherwise would.⁵

Prolonged peer review

Writing well takes time, so higher standards probably delay peer review. To evaluate this hypothesis, I investigate submit-accept times at *Econometrica*.

Fig. 4 displays review time distribution by author sex. Women's times (pink) are disproportionately clustered in Fig. 4's right tail: articles by female authors are six times more likely to experience delays above the 75th percentile than they are to enjoy speedy revisions below the 25th.

Using a more precise estimation strategy, I find that male-authored papers take (on average) 18.5 months to complete

all revisions; equivalent papers by women need half a year longer. These estimates are based on a model by Ellison (2002, table 6, p. 963). In addition to the statistically significant variables he incorporates—author productivity, article length, number of co-authors, order in an issue, citation count and field dummies—I also control for motherhood and childbirth.⁶

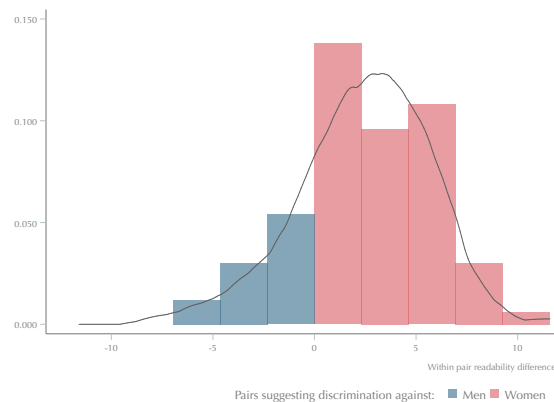


Fig. 3. Distribution of within pair differences in readability for pairs in which one member satisfies conditions 1 and 2 according to the SMOG score. Blue bars represent matched pairs in which the man satisfies 1 and 2 (indicative of discrimination); pink bars are pairs in which the woman does. Because male scores are subtracted from female scores, differences are positive in pairs suggesting discrimination against women and negative in pairs suggesting discrimination against men. Estimated density function drawn in grey.

How do women react to higher standards?

As a final exercise, I investigate how women react to higher standards as they update beliefs about referees' expectations. Fig. 5 compares papers pre- and post-review at increasing publication counts. Solid circles denote NBER draft readability; arrow tips reflect readability in the final, published versions of those same papers; dashed lines trace changes made as papers undergo peer review.

All things equal, economists who anticipate referees' demands are rejected less often; economists who don't enjoy more free time. Fig. 5 implies little—if any—gender difference in this tradeoff: senior economists of both sexes sacrifice time upfront to increase acceptance rates.

Moreover, Fig. 5 emphasises that only *inexperienced* women make changes during peer review. Assuming choices by senior economists express optimal tradeoffs with full information, this implies that women initially underestimate referees' expectations.

Men, however, do not. Draft and final readability choices remain relatively stable over the course of their careers.

Are men just better informed about referees' expectations? Yes and no. Male and female draft readability scores for first-time publications are exactly the same. This suggests that men and women start out with identical beliefs. But those beliefs reflect standards that apply only to men. Women are then mistaken by thinking they apply to them, too.

⁵ This estimate averages results over all five scores. It assumes women are accepted in a subset of states in which men are accepted and within pair differences are zero for the 30–40 percent of matched pairs that fail to satisfy conditions 1 and 2. See Hengel (2017) for alternative estimates based on weaker assumptions. (Conclusions drawn from those estimates mirror the conclusions discussed here.)

⁶ Ellison (2002)'s work evaluates how non-gender author compositional effects contribute to higher mean-accept times at the *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics* and *Review of Economic Studies*. (Although his analysis controls for female authorship, it did not investigate gender differences specifically.)

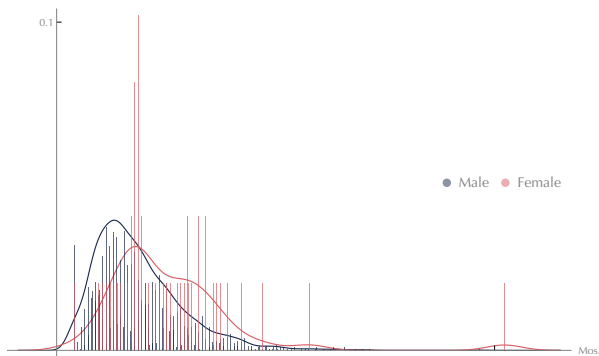


Fig. 4. Distribution of review times by author sex. Blue bars represent papers written only by men; pink bars are papers written only by women.

Policy implications

Fig. 5 suggests that women respond to biased treatment in ways that not only obscure the line between personal preferences and external constraints but can paint a rosier picture than even preferences justify. This raises a couple of concerns about identifying discrimination from narrow viewpoints. For example, if we only concentrate attention on a cross section of papers written by senior economists, we might conclude that women simply prefer writing more clearly. Alternatively, if we limit our focus to the gap formed inside peer review, we might decide it declines with experience.

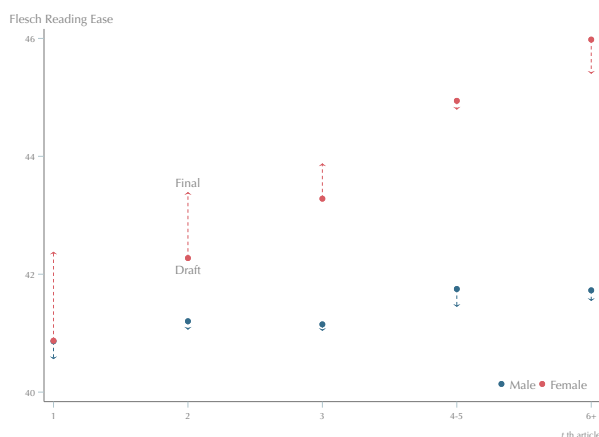


Fig. 5. Flesch Reading Ease marginal mean scores for authors' first, second, third, 4th–5th and sixth and up publications in the data. Solid circles denote estimated readability of NBER working papers; arrow tips show the estimated readability in the published versions of the same papers. Pink represents women co-authoring only with other women; blue are men co-authoring only with other men.

But neither conclusion is supported when the data are analysed from a broader perspective. A smaller gap in peer review is completely offset by a wider gap before peer review. Senior female economists did not enjoy writing so well when they were junior economists.

My evidence also emphasises that discrimination impacts more than just obvious outcomes. It corrupts productivity, too. 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 reduces women's wages—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—and distorts measurement of female productivity—billable hours and client revenue decline; female lawyers appear less productive than they truly are.

Unfortunately, there is no easy way to eliminate implicit bias. But least intrusive—and arguably most effective—is simple awareness and constant supervision. Monitoring referee reports is difficult, but it isn't impossible—especially if peer review were open. Several science and medical journals not only reveal referees' identities, they also post reports online. Quality does not decline (it may actually increase); referees still referee (even those who initially refuse) (van Rooyen et al., 1999; Walsh et al., 2000). And given what's at stake, is spending an extra 25–50 minutes reviewing a paper really all that bad (van Rooyen et al., 2010)?

References

- Azmat, G. and R. Ferrer (2017). “Gender Gaps in Performance: Evidence from Young Lawyers”. *Journal of Political Economy* 125 (5), pp. 1306–1355.
- 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.
- Ceci, S. J. et al. (2014). “Women in Academic Science: A Changing Landscape”. *Psychological Science in the Public Interest* 15 (3), pp. 75–141.
- Ellison, G. (2002). “The Slowdown of the Economics Publishing Process”. *Journal of Political Economy* 110 (5), pp. 947–993.
- Goldberg, P. K. (2015). “Report of the Editor: American Economic Review”. *American Economic Review* 105 (5), pp. 698–710.
- Hartley, J., J. W. Pennebaker, and C. Fox (2003). “Abstracts, Introductions and Discussions: How Far Do They Differ in Style?” *Scientometrics* 57 (3), pp. 389–398.
- 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.
- Hayden, J. D. (2008). “Readability in the British Journal of Surgery”. *British Journal of Surgery* 95 (1), pp. 119–124.
- Hengel, E. (2017). “Publishing while Female: Are women held to higher standards? Evidence from peer review.” Mimeo.
- 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 (February), pp. 811–825.
- 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.
- Tsugawa, Y. et al. (2016). “Comparison of Hospital Mortality and Readmission Rates for Medicare Patients Treated by Male vs. Female Physicians”. *JAMA Internal Medicine* 02138, pp. 1–8.
- Van Rooyen, S., T. Delamothe, and S. J. W. Evans (2010). “Effect on Peer Review of Telling Reviewers that Their Signed Reviews Might Be Posted on the Web: Randomised Controlled Trial”. *British Medical Journal* 341 (c5729).
- Van Rooyen, S. et al. (1999). “Effect of Open Peer Review on Quality of Reviews and on Reviewers' Recommendations: A Randomised Trial”. *British Medical Journal* 318 (7175), pp. 23–27.
- Walsh, E. et al. (2000). “Open Peer Review: A Randomised Controlled Trial”. *British Journal of Psychiatry* 176 (1), pp. 47–51.