Publishing while female

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

Erin Hengel
University of Liverpool

NBER Summer Institute
17 July 2018

Background

Women are underrepresented in economics

□ Roughly 25–30 percent of PhDs, assistant professors and associate professors.

□ Almost 15 percent of full professors (Lundberg, 2017).

Women are *really* underrepresented at top journals

□ In 2015, the average ratio of female authors was 15 percent. Only 7.5 percent of papers were majority female-authored. Just 4 percent were written entirely by women.

□ *QJE* did not publish a single exclusively female-authored paper in 2015…or 2016…or 2017…

□ ...in four of the fifteen years between 2001–2015, *Econometrica* and *JPE* didn't either.

Is peer review affirmative action for men?

Background

Women are held to higher standards

- □ Men are rated more competent when compared to otherwise equally competent women (Foschi, 1996).
- □ Male undergraduate students underestimate female classmates' ability (Grunspan et al., 2016).
- □ Female graduate students are rated less qualified for laboratory management positions (Moss-Racusin et al., 2012).
- □ When collaborating with men, women are given less credit for mutual work (Heilman and Haynes, 2005; Sarsons, 2017).
- □ Manuscripts by female authors are rated lower quality (Goldberg, 1968; Paludi and Bauer, 1983; Krawczyk and Smyk, 2016).

> ***"Women must do twice as well to be thought half as good."***
>
> –Charlotte Whitton

Gender discrimination in peer review

Are women held to higher standards in peer review?

- □ Little evidence gender impacts acceptance rates (see Blank, 1991; Gilbert et al., 1994; Ceci et al., 2014).
- □ Most papers undergo major referee-requested revisions (Abrevaya and Hamermesh, 2012).
- □ Are referees, *e.g.*, more likely to double-check technical details, demand robustness checks or require clearer exposition in a female-authored paper?
  - □ If so, then female-authored papers should be better quality on the dimension in which they are held to higher standards.

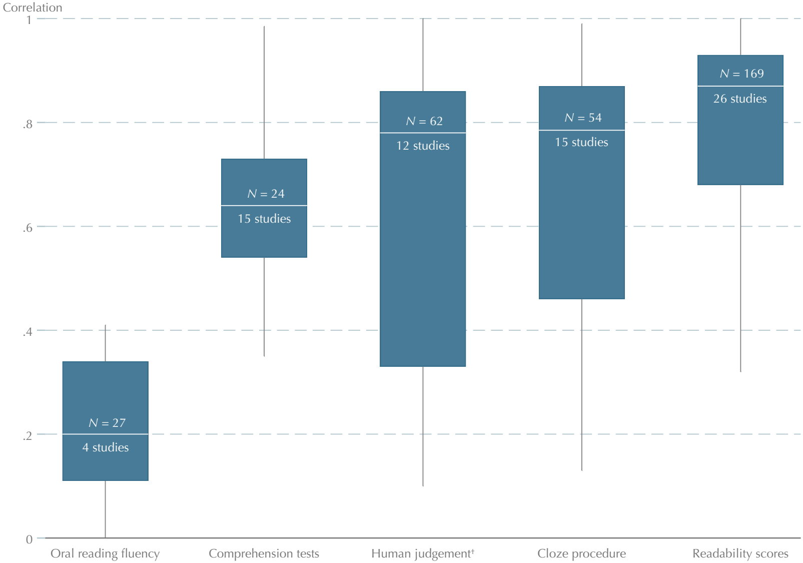
"I have no doubt that one of [discrimination's] results has been that those women who do manage to make their mark are much abler than their male colleagues."

–Milton Friedman

Writing clarity

1. Clear writing is valued by journals.
   - □ Stated explicitly in submission guidelines.
   - □ "Evaluate adequacy of the language" is one of the most frequent tasks editors make of referees (Chauvin et al., 2015).
2. Good writing is highly correlated with simple vocabulary and short sentences.
   - □ Flesch Reading Ease, Flesch-Kincaid, Gunning Fog, SMOG and Dale-Chall.
   - □ Developed primarily for adults and tested on technical documents (see DuBay, 2004).
   - □ Used in research, particularly in finance and political science (see Loughran and Mcdonald, 2016; Benoit et al., 2017).
   - □ Validated against surrogate masures of reading comprehension, including readership (Swanson, 1948; Richardson, 1977), reading persistence, efficiency and retention (Klare et al., 1957; Klare and Smart, 1973).
   - □ Readable academic articles win more awards (Sawyer et al., 2008), are downloaded more often (Guerini et al., 2012) and cited more frequently.

Correlation with alternative measures



Text used in the analysis

- Every article abstract published in the *AER*, *Econometrica*, *JPE* and *QJE* since 1950.
  - Largely exist as machine readable text.
  - Contain few citations and equations which distort readability scores.
  - Most read portion of a paper (King et al., 2006).
  - Standardised layout—readability less influenced by non-textual cues.
  - Readability scores highly correlated across abstract, introduction and discussion sections of a paper (Hartley et al., 2003; Plavén-Sigray et al., 2017).

Strategy

Identification

1. Causally link the gender gap to the peer review process.
2. Establish sufficient conditions to verify discrimination is present in academic publishing.
   - □ Conditions are satisfied on average for two different measures of research quality: readability *and* citation counts.
   - □ Use matching to make the causal link between women's better writing and higher standards by referees and/or editors.

Consequences

- □ **Time tax**. Female-authored papers take longer in peer review.
- □ **Behaviourial change**. As women update beliefs about referees' standards, they increasingly meet those standards before peer review.

Causal impact of peer review

	FGLS			OLS
	Working paper	Published article	Difference	Change in score
Flesch Reading Ease	2.26**	3.21***	0.95*	0.94
	(1.00)	(1.21)	(0.57)	(0.60)
Flesch-Kincaid	0.31	0.75***	0.44**	0.44**
	(0.23)	(0.28)	(0.18)	(0.19)
Gunning Fog	0.44*	0.86***	0.42**	0.42**
	(0.24)	(0.29)	(0.19)	(0.20)
SMOG	0.33**	0.56***	0.24**	0.24*
	(0.15)	(0.19)	(0.12)	(0.12)
Dale-Chall	0.32***	0.45***	0.13**	0.13**
	(0.10)	(0.11)	(0.05)	(0.05)
Editor effects	✓	✓		✓
Journal effects	✓	✓		✓
Year effects	✓	✓		
Journal×Year effects	✓	✓		✓
Quality controls	✓^{2}	✓^{2}		✓^{3}
Native speaker	✓	✓		✓

Peer review causes a large increase in the readability gap

- Readability gap is 2–3 times as large in the published article.
- Suggests peer review causes female-authored abstracts to become about 2–5 percent more readable.

Robustness

- Using the change in score as the dependent variable implicitly controls for field.
- Adding field controls to FGLS estimates does not change results. table
- No significant gap under double-blind review. table figure
  - *Caution*: small samples, particularly of female-authored papers.
- Abstract word limits do not seem to drive results. table
- Timing independence: female-authored manuscripts are submitted to journals *first*; released as NBER Working Papers *second*. figure

Causal impact of discrimination: theory

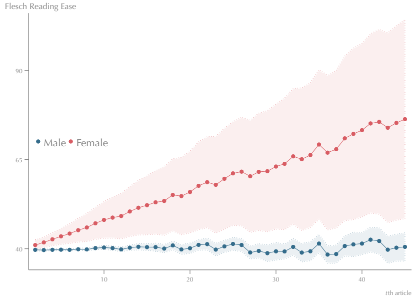
Why does peer review cause women to write more clearly?

Possibility 1 Women voluntarily write better papers—*e.g.*, they're more sensitive to referee criticism.

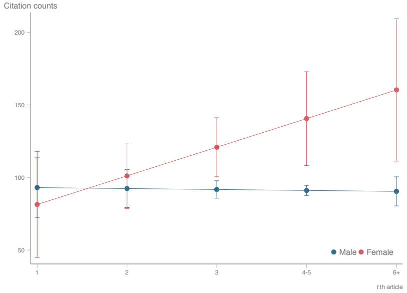
Possibility 2 Better written papers are women's response to higher standards imposed by referees and/or editors.

- □ Model an author's decision making process within a subjective expected utility framework.
- □ Establish 3 sufficient conditions that distinguish Possibility 1 from Possibility 2.
  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.

Causal impact of discrimination: evidence (I)



1. Experienced female economists write better than equivalent male economists
2. Women improve their writing over time.



1. Experienced female economists are cited more than equivalent male economists.
2. Women increase citation counts over time.

No female advantage in acceptance rates (Ceci et al., 2014).

Causal impact of discrimination: evidence (II)

- Use a matching estimator to account for the fact that each condition must hold for the same author in two different situations:
  - Before and after gaining experience.
  - When compared to an equivalent, experienced author of the opposite gender.
- Matches based on observable characteristics: primary *JEL* category, citation counts, decade, institution, *etc*.

Results figure table

- Evidence of discrimination in 60–70 percent of matched pairs; almost always against women.
- Suggests discrimination causes women to write 9 percent more clearly than they otherwise would.

Prolonged peer review

	(1)	(2)	(3)	(4)	(5)	(6)
Female ratio	5.29**	6.63***	6.64***	5.54***	6.65***	8.80***
	(2.01)	(2.16)	(2.14)	(2.05)	(2.15)	(2.72)
Max. t_j	-0.16**	-0.17**	-0.17**	-0.16**	-0.16**	-0.17*
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.09)
No. pages	0.18***	0.18***	0.18***	0.18***	0.18***	0.21***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
N	1.02**	0.97**	0.96**	1.01**	0.97**	1.149
	(0.44)	(0.44)	(0.44)	(0.44)	(0.44)	(0.70)
Order	0.22**	0.22**	0.22**	0.22**	0.22**	0.50**
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.22)
No. citations	0.00	0.00	0.00	0.00	0.00	-0.00***
	(0.000)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mother			-6.66**		-10.93***	-17.67***
			(2.68)		(3.21)	(3.29)
Birth				-2.25	7.58*	12.34**
				(3.36)	(4.17)	(5.59)
Constant	37.71***	37.60***	37.79***	37.69***	37.89***	14.85***
	(2.04)	(2.08)	(2.05)	(2.05)	(2.06)	(2.79)
Editor effects	✓	✓	✓	✓	✓	✓
Year effects	✓	✓	✓	✓	✓	✓
Institution effects	✓	✓	✓	✓	✓	✓
JEL effects						✓
No. observations	2,626	2,610	2,626	2,626	2,626	1,281

Econometrica

□ 5–9 months longer in peer review

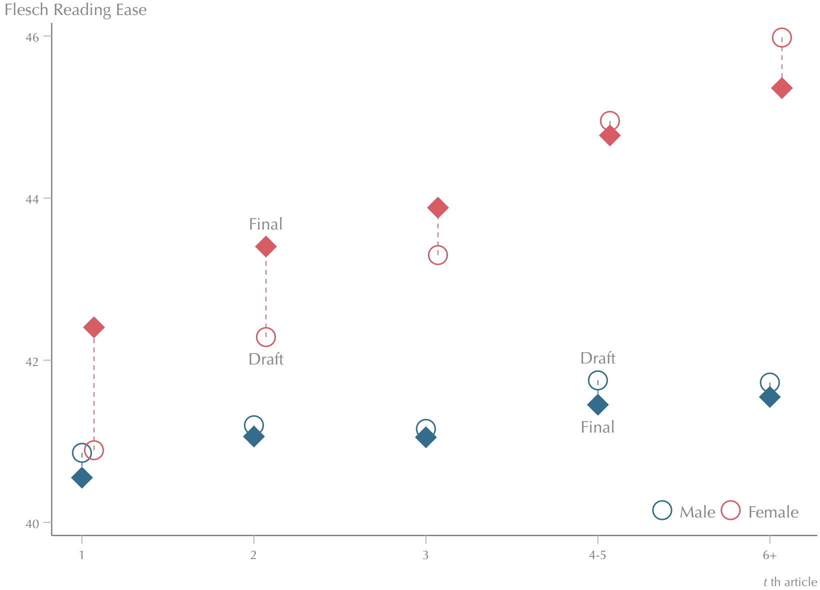
Energy Economics

□ 27–29 days longer in peer review

□ More revision rounds & referee reports

□ Desk rejected at higher rates

Behaviourial changes



Conclusions for academia

Implications for measuring productivity

□ Women may produce better quality output…

□ But quality costs time, so women produce less.

□ Women appear less productive than they actually are.

"Publishing Paradox" may not be so paradoxical…

References I

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.

Benoit, K., K. Munger, and A. Spirling (2017). “Measuring and Explaining Political Sophistication through Textual Complexity”. Mimeo.

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.

Ceci, S. J. et al. (2014). “Women in Academic Science: A Changing Landscape”. *Psychological Science in the Public Interest* 15(3), pp. 75–141.

References II

Chauvin, A. et al. (2015). “The most important tasks for peer reviewers evaluating a randomized controlled trial are not congruent with the tasks most often requested by journal editors”. *BMC Medicine* 13(1), pp. 1–10.

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

Foschi, M. (1996). “Double Standards in the Evaluation of Men and Women”. *Social Psychology Quarterly* 59(3), pp. 237–254.

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.

Goldberg, P. (1968). “Are Women Prejudiced against Women?”. *Trans-action* 5(5), pp. 28–30.

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.

References III

Guerini, M., A. Pepe, and B. Lepri (2012). “Do Linguistic Style and Readability of Scientific Abstracts affect their Virality?”. *Proceedings of the Sixth Interntaional AAAI Conference of Weblogs and Social Media*, pp. 475–478. arXiv: 1203.4238.

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.

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.

King, D. W., C. Tenopir, and M. Clarke (2006). “Measuring Total Reading of Journal Articles”. *D-Lib Magazine* 12(10), pp. 1082–9873.

References IV

Klare, G. R., W. H. Nichols, and E. H. Shuford (1957). “The relationship of typographic arrangement to the learning of technical training material”. *Journal of Applied Psychology* 41(1), pp. 41–45.

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.

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.

Loughran, T. and B. Mcdonald (2016). “Textual Analysis in Accounting and Finance: A Survey”. *Journal of Accounting Research* 54(4), pp. 1187–1230.

Lundberg, S. (2017). “Committee on the Status of Women in the Economics Profession (CSWEP)”. *American Economic Review* 107(5), pp. 759–776.

References V

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.

Paludi, M. A. and W. D. Bauer (1983). “Goldberg Revisited: What’s in an Author’s Name”. *Sex Roles* 9(3), pp. 387–390.

Plavén-Sigray, P. et al. (2017). “The Readability Of Scientific Texts Is Decreasing Over Time”. *bioRxiv*, p. 119370.

Richardson, J. V. J. (1977). “Readability and Readership of Journals in Library Science”. *Journal of Academic Librianship* 3(1), pp. 20–22.

Sarsons, H. (2017). “Recognition for Group Work: Gender Differences in Academia”. *American Economic Review* 107(5), pp. 141–145.

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.

References VI

Swanson, C. E. (1948). “Readability and Readership: A Controlled Experiment”. *Journal ism Bulletin* 25(4), pp. 339–343.

APPENDIX

Double-blind review

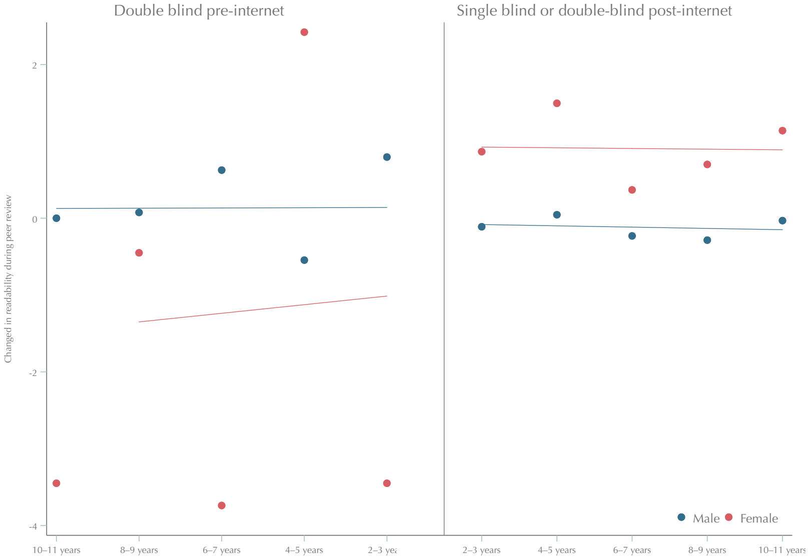
	Flesch Reading Ease	Flesch-Kincaid	Gunning Fog	SMOG	Dale-Chall
Non-blind	0.93	0.43**	0.41**	0.23*	0.12**
	(0.60)	(0.19)	(0.20)	(0.12)	(0.05)
Blind	-1.51	-0.56	-0.54	-0.36	-0.13
	(3.05)	(0.70)	(0.82)	(0.59)	(0.18)
Difference	2.44	1.00	0.95	0.59	0.25
	(3.14)	(0.75)	(0.87)	(0.61)	(0.18)
Editor effects	✓	✓	✓	✓	✓
Journal effects	✓	✓	✓	✓	✓
Journal × Year effects	✓	✓	✓	✓	✓
Quality controls	✓^{3}	✓^{3}	✓^{3}	✓^{3}	✓^{3}
Native speaker	✓	✓	✓	✓	✓

Notes. Sample 1,988 NBER working papers; 1,986 published articles. Standard errors clustered by year in parentheses. Quality controls denoted by ✓^{3} includes max. t_j, only. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

No significant gap under double-blind review

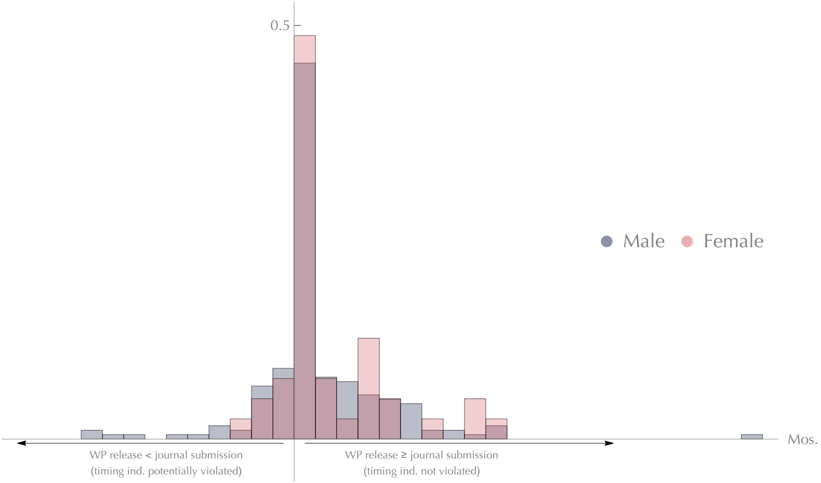
causal impact of peer review

Double-blind review



causal impact of peer review

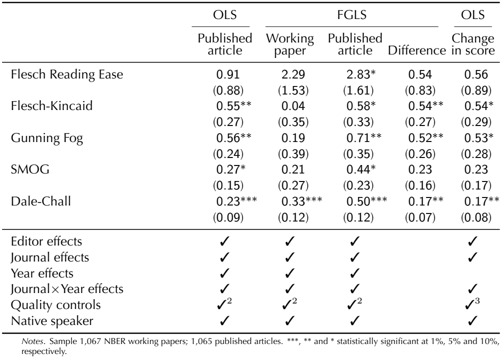
Timing independence



Female-authored manuscripts are submitted to journals *first*; released as NBER Working Papers *second*.

causal impact of peer review

Are abstract word limits driving results?



	OLS	FGLS			OLS
	Published article	Working paper	Published article	Difference	Change in score
Flesch Reading Ease	0.91	2.29	2.83*	0.54	0.56
	(0.88)	(1.53)	(1.61)	(0.83)	(0.89)
Flesch-Kincaid	0.55**	0.04	0.58*	0.54**	0.54*
	(0.27)	(0.35)	(0.33)	(0.27)	(0.29)
Gunning Fog	0.56**	0.19	0.71**	0.52**	0.53*
	(0.24)	(0.39)	(0.35)	(0.26)	(0.28)
SMOG	0.27*	0.21	0.44*	0.23	0.23
	(0.15)	(0.27)	(0.23)	(0.16)	(0.17)
Dale-Chall	0.23***	0.33***	0.50***	0.17**	0.17**
	(0.09)	(0.12)	(0.12)	(0.07)	(0.08)
Editor effects	✓	✓	✓		✓
Journal effects	✓	✓	✓		✓
Year effects	✓	✓	✓		
Journal × Year effects	✓	✓	✓		✓
Quality controls	✓[2]	✓[2]	✓[2]		✓[3]
Native speaker	✓	✓	✓		✓

Notes. Sample 1,067 NBER working papers; 1,065 published articles. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

causal impact of peer review

Sample restricted to abstracts below journals' official word limits

No meaningful impact

- □ Sample size is smaller.
- □ Coefficients and standard errors are generally larger.

Accounting for field

	OLS	FGLS		
	Published article	Working paper	Published article	Difference
Flesch Reading Ease	1.32**	2.80***	3.68***	0.88
	(0.58)	(1.04)	(1.17)	(0.59)
Flesch-Kincaid	0.55***	0.46*	0.90***	0.44**
	(0.18)	(0.24)	(0.30)	(0.20)
Gunning Fog	0.51***	0.53**	0.92***	0.39*
	(0.18)	(0.24)	(0.32)	(0.21)
SMOG	0.29**	0.39***	0.60***	0.21
	(0.12)	(0.15)	(0.19)	(0.13)
Dale-Chall	0.14***	0.32***	0.42***	0.10*
	(0.05)	(0.10)	(0.10)	(0.05)
Editor effects	✓	✓	✓	
Journal effects	✓	✓	✓	
Year effects	✓	✓	✓	
Journal×Year effects	✓	✓	✓	
Quality controls	✓^{2}	✓^{2}	✓^{2}	
Native speaker	✓	✓	✓	
JEL (primary) effects	✓	✓	✓	

Notes. Sample 1,505 NBER working papers; 1,503 published articles. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Adding field controls does not change results.

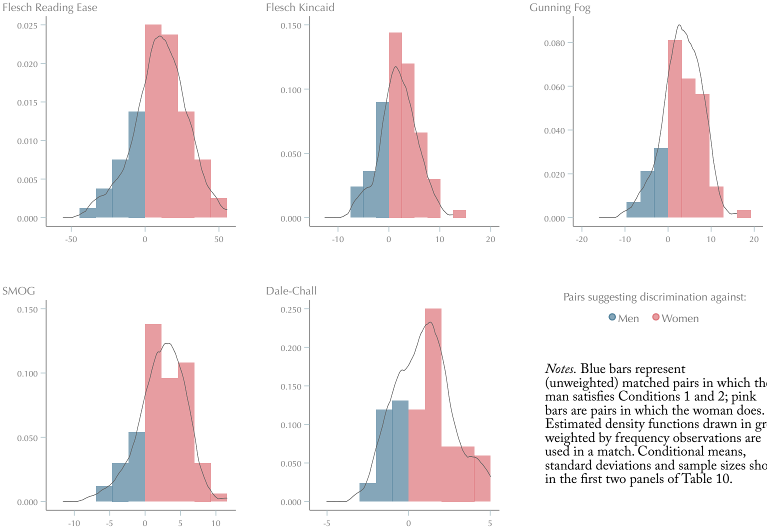
Causal impact of discrimination: evidence (II)

- □ Determine whether conditions 1 and 2 hold for one member in each matched pair.
- □ If so, then discrimination is present within that matched pair.
- □ If not, then my test for discrimination is inconclusive.

	Discrimination against women ($\underline{D}_{ik} > 0$)			Discrimination against men ($\underline{D}_{ik} < 0$)			Mean, all observations	
	Mean	S.D.	*N*	Mean	S.D.	*N*	(1)	(2)
Flesch Reading Ease	18.32	12.94	58	-12.42	10.58	21	6.69***	6.02***
							(1.62)	(1.68)
Flesch Kincaid	3.70	2.68	61	-2.05	2.11	25	1.40***	1.22***
							(0.34)	(0.35)
Gunning Fog	5.11	3.31	62	-3.12	2.57	17	2.23***	2.03***
							(0.42)	(0.44)
SMOG	3.64	2.35	63	-2.44	1.95	16	1.58***	1.44***
							(0.30)	(0.32)
Dale-Chall	1.94	1.30	48	-0.96	0.65	23	0.57***	0.51***
							(0.15)	(0.16)

Notes. Sample 121 matched pairs (104 and 121 distinct men and women, respectively). First and second panels display conditional means, standard deviations and observation counts of $\underline{D}_{ik}$ from subpopulations of matched pairs in which the woman or man, respectively, satisfies Conditions 1 and 2. Third panel displays mean $\underline{D}_{ik}$ over all observations. To account for the 30–40 percent of pairs for which Theorem 1 is inconclusive, (1) sets $\underline{D}_{ik} = 0$, while (2) sets $\underline{D}_{ik} = \widehat{R}_{i3} - \widehat{R}_{k3}$ if $\widehat{R}_{i3} < \widehat{R}_{k3}$ (*i* female, *k* male) and zero, otherwise. Male scores are subtracted from female scores; $\underline{D}_{ik}$ is positive in panel one and negative in panel two. $\underline{D}_{ik}$ weighted by frequency observations are used in a match; degrees-of-freedom corrected standard errors in parentheses (panel three, only). ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Causal impact of discrimination: evidence (II)


Flesch Reading Ease

Flesch Kincaid

Gunning Fog

SMOG

Dale-Chall

Notes. Blue bars represent (unweighted) matched pairs in which the man satisfies Conditions 1 and 2; pink bars are pairs in which the woman does. Estimated density functions drawn in grey, weighted by frequency observations are used in a match. Conditional means, standard deviations and sample sizes shown in the first two panels of Table 10.