Is Mathilda Playing it Safe? Gender in Computational Text Analysis Methods *

Mariken A.C.G. van der Velden *Vrije Universiteit Amsterdam*
Alona Dolinsky *University College Dublin*

The editors of top-journals in political science have investigated the role of the review process finding no indication for gender bias in the editorial process. They suggest that the submission pools are distorted by gender. We test some pathways that could distort the submission pool: a) playing it safe due to the gender perception gap, and b) the Mathilda effect. Papers using Computational Text Analysis Methods are more likely to be published in journals with a 'masculinized' perception gap. When women are aiming for these journals, they might 'play it safe' by conducting more validation checks than their male colleagues. Moreover, embracing the Mathilda effect – i.e. the systematic under-recognition of female scientists – women scholars are more likely to indicate that a) there are important training needs in more areas; and b) they themselves need (further) training in computational methods and use these reasons not to publish employing these methods. We test these claims using a) an unique content analysis of research articles published in the top 20 journals in communication science, political science, sociology and psychology between 2016 and 2020, identifying all 854 articles that involved some form of quantitative textual analysis; and b) an expert survey with all authors of quantitative text analytic research identified via said content analysis, which inquired about researchers' considerations and concerns in the application of computational text analytic strategies.

Keywords: Computational Text Analyses Methods, Gender, Expert Survey

Introduction

[Short motivation copied from Pre-Analysis Plan]

Gatto et al. (2020) demonstrates that while women cover significantly fewer methods courses in their doctoral training, when they do participate in methods training, they, however, show similar levels of method employment to their male colleagues. Yet, Teele and Thelen (2017) demonstrate that women are less likely to publish studies using quantitative and computational methods (see Maliniak, Powers, and Walter (2013) for a similar argument in International Relations). Brown et al. (2020) demonstrate that the *gender submission gap* is accompanied by a *gender perception gap*: Women report that they are more likely to submit to and get published in some journals, whereas men report as such with regard to other journals. Importantly, these gaps are observed even among scholars with the same methodological (i.e., quantitative or qualitative) approach. The editors of top-journals in politicial science have investigated the role of the review process finding no indication for gender bias in the editorial process (Breuning et al. 2018; König and Ropers 2018; Brown and Samuels 2018). The authors all suggest that the submission pools are distorted by gender.

We test two pathways that could distort the submission pool: a) playing it safe due to the gender perception gap, and b) the Mathilda effect. Papers using computational text analysis

*Corresponding author: Mariken A.C.G. van der Velden. Replication files are available on the corresponding author's Github account (https://anonymous.4open.science/r/gender-ctam-CBA3/README.md). **Current version**: May 02, 2022; **Author contributions**: data collection: MACGvdV; FL, CB & MS; data analysis: MACGvdV; writing of the paper: AD & MACGvdV

methods (CTAM) are more likely to be published in journals with a *"masculinized" perception gap*. When women are aiming for these journals, it could be that they "play it safe' ' by conducting more validation checks than their men. Moreover, embracing of the *Mathilda effect* (Rossiter 1993) – i.e. the systematic underrecognition of female scientists – women scholars are more likely to indicate that a) there are important training needs in more areas; and b) they themselves need (further) training in computational methods and use these reasons not to publish employing these methods.

Play it Safe hypothesis (*H1a*): Women scholars are more likely to play it safe and indicate more validation strategies than men scholars.

Play it Safe hypothesis (*H1b*): Women scholars are more likely to play it safe and indicate more challenges as reasons **not** to use CTAM than men scholars.

Mathilda effect hypothesis (*H2a*): Women scholars are more likely to indicate a higher number of important training needs than men scholars.

Mathilda effect hypothesis (*H2b*): Women scholars are more likely to indicate that they themselves require (further advanced) training than men scholars.

Is Mathilda Playing it Safe?

We have condected an expert survey, inviting all scholars who had published a scientific article using quantitative text analysis between January 2016 and September 2020 in one of the 20 highest ranked journals in Communications, Political Science, Sociology, and Psychology (for details, see Baden et al. 2021). Quantitative textual analysis was defined broadly to include any form of processing natural language that identified specific kinds of textual contents with the purpose of classification and quantitative analysis. Using a keyword search on the Web of Science, we identified a total of `7,296` *potentially* relevant articles out of the `45,437` published articles, whose abstract suggested any kind of textual content or test analytic procedures. Articles were considered relevant as soon as they used any form of quantitative textual analysis, even if it was used merely in an auxiliary capacity (e.g., a content analysis to identify frames to be used in an experiment; sentiment analyses of open-ended survey responses).[1] This screening yielded a total of `854` articles, for which the authors were looked up. This gave us `1,653` identifiable and working email-addresses. The experts have been invited to the questionnaire on March 4th of 2021, and received two reminders, each approximately a week after our last message (respectively on March 11th and March 16th of 2021). This yielded a responses of `433` responses (i.e. response rate of `25%`). The study has been approved by the Research Ethics Review Committee of the *Vrije Universiteit Amsterdam*, *Hebrew University Jerusalem Isreal*, and *University College Dublin*.

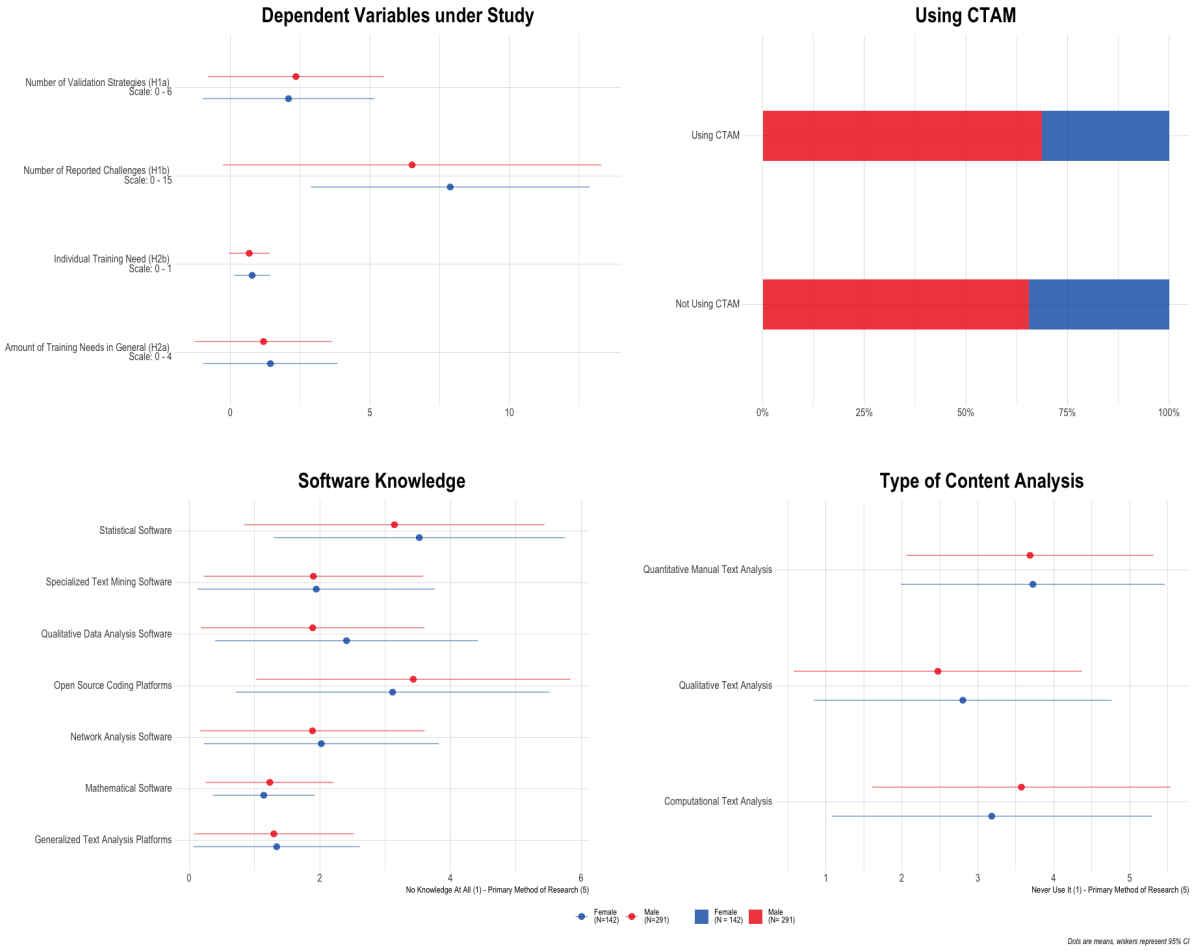
To survey the field on different aspects of potential 'playing it safe' strategies when it comes to CTAM, we have asked participants to list how many of five pre-defined validation strategies[2] they use, as well as if they use other strategies (open answer posibility). Additionally, we asked participants to what extent 14 pre-defined reasons are challenges for them *not* to use CTAM.[3]

[1] Analyses that relied solely on metadata or pre-existing classification were excluded, as were investigations accessing only formal properties of the sampled texts (e.g., length). We included analyses of multi-modal media (e.g., posters, television) as long as the textual contents were informative toward classification. Purely methodological contributions discussing specific potentials or limitations of available methods were excluded, unless they included applied demonstrations wherein actual textual data were processed.

[2] The pre-defined validation strategies are: 1) *I check the documentation of the software developer*; 2) *I check whether findings are plausible and interpretable*; 3) *I compare machine-made classifications (i.e. coded outputs) against manually coded/given standards*; 4) *I check whether the classification rules/criteria are valid (e.g., check for meaningful indicators in adictionary)*; and 5) *I evalute how well the algorithmic procedures match my concept of interest (e.g., checking adequacy of bag-of-words assumption)*.

[3] The pre-defined challenges are: 1) *Time/effort required (e.g. technical requirements, experience)*; 2) *Funding*

Figure 1: Gender Break-Down in Variables under Study



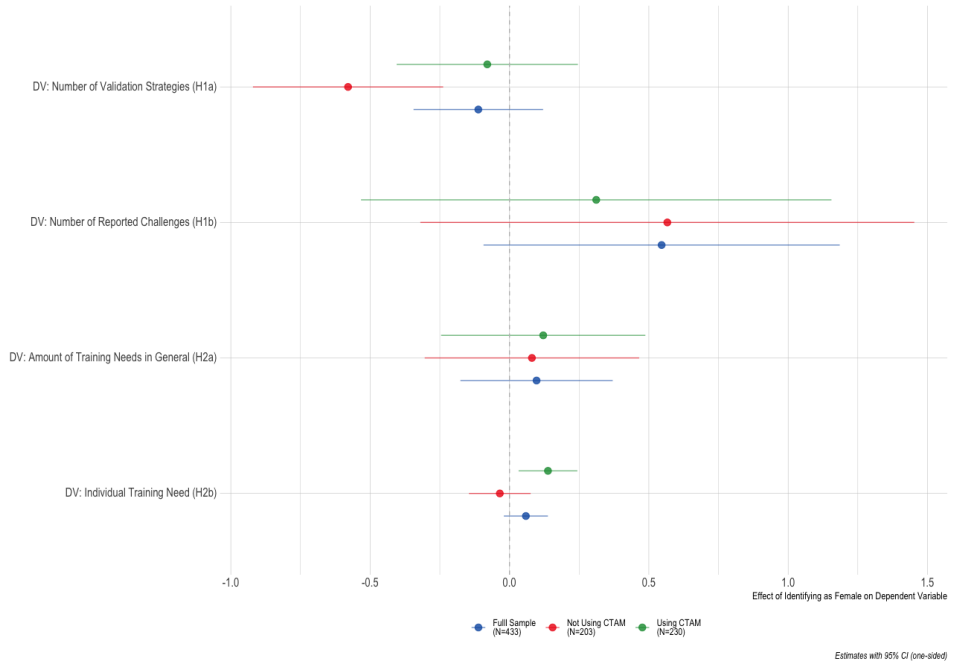
Answer options where `no challenge`, `minor challenge`, `major challenge`. In the analyses, we grouped minor and major challenges together and compared them to no challenges reported. To gauge a possible 'Mathilda effect', we have asked our experts to list if the field in general needs more training possibilities when it comes to data and open access tools, programming and software skills, theory and concepts, research integrity and ethics, or other training skills – for each option there was the possiblity to specify what was needed. Moreover, we have also asked the participants if they themselves need more training. In all our analyses, we control for whether or not the participant uses CTAM themselves, the types of content analyses they use in their work, and their pre-existing knowledge of statistical software. Figure 1 demonstrates the gender break-down in our sample for the dependent variables as well as the control variables. While the descriptive statistics (average

required (e.g., for training, fees); 3) *Availability of required training in computational methods*; 4) *Limited methodological guidance/documentation of tools*; 5) *Level of instruction and materials higher than needed*; 6) *Level of instruction/materials lower than needed*; 7) *Availability of suitable computational tools for certain languages*; 8) *Comparability of computational tools in different languages*; 9) *Issues concerning measurement validity/limited nuance*; 10) *Loss of manual contact with the material*; 11) *Reviewers'/editors'' skepticism toward computational methods among*; 12) *Peers' skepticism toward computational methods*; 13) *I am skeptical toward computational methods myself*; and 14) *Other challenges: (Open ended)*.

and standard deviation) do not show gender differences, we do see in the top-right panel that men in our sample are over-represented. Yet where we have roughly the same amount of women scholars using CTAM as not using CTAM (i.e. using quantitative or qualitative manual content analysis), our sample has more men scholars using CTAM than not using CTAM.

To test our pre-registered hypotheses, we ran a OLS regression and visualize the results in Figure 2 and 3 with an one-tailed α of `0.05`. We present the regression tables in Appendix **XX**. Looking at the variables measuring 'playing it safe' strategies, i.e. reporting number of validation strategies and number of challenges, Figure 2 demonstrates that women scholars do not report more validation strategies than men. While not statistically significant for the full sample as well as for the split sample using CTAM, we see the opposite: Women scholars report to use less validation strategies. In the top-left panel of Figure 3, we inspect the gender differences in the type of validation strategies reported. This shows that the only strategy women scholars are statistically significantly less likely to report compared to men scholars is checking the documentation of the software developer. Women scholars are more likely to report to use the validation strategies of checking algortihmic procedures and checking against the gold standard. This is only borderline non-significant, most likely due to the small number of women in our sample. For the other playing it safe strategy – i.e. number of challenges reported – Figure 2 demonstrates that women scholars are more likely to report challenges, but this effect is not statistically significant. Breaking down the challenges in the right-hand panel of Figure 3 do demonstrate interesting gender patterns. Women are more likely to report time/effort, funding, and training needs as well as limitied methodological guidance of the tools as a challenge to not using CTAM. Men scholars, however, are more taken back by peers' and editors' skepticism towards the method.

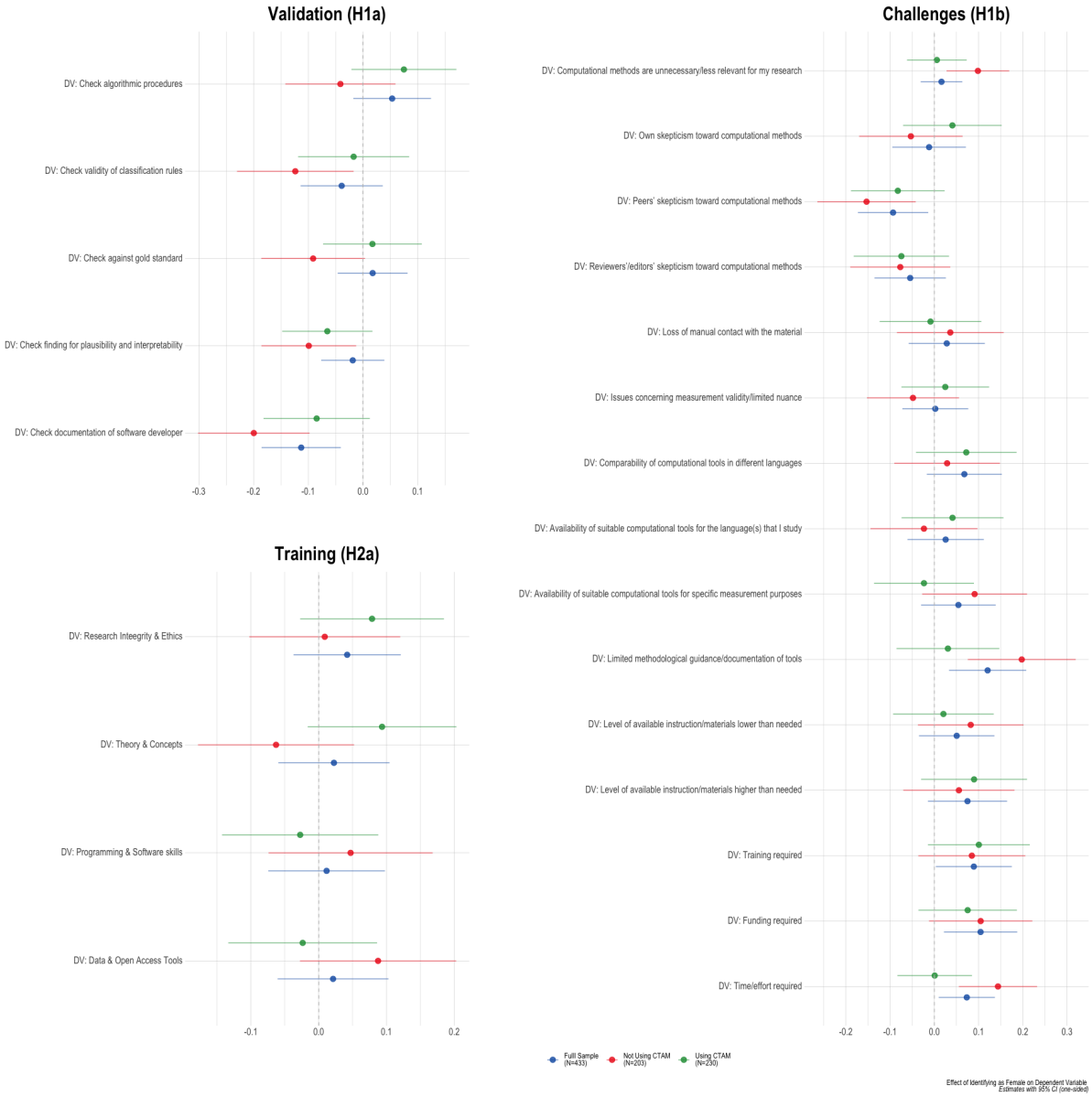
Figure 2: Effect of Being Female on Reported Validation Strategies, Challenges, and Training Needs



Investigating a possible Mathilda effect, Figure 2 demonstrates that women scholars are more likely to report more training needs, but this effect is not statistically significant. Looking at the type of training needs in the bottom-left panel of Figure 3, we see that the insignificance of

the aggregated training needs stems from women using CTAM and using manual content analysis reporting opposite needs: Women in CTAM, report general needs for training in research integrity and ethics as well as for training in concepts, whereas women not using CTAM report general training needs for programming and software skills as well as for data and open access tools. When it comes to individual training needs, women scholars using CTAM do report they need more training compared to men scholars using CTAM.

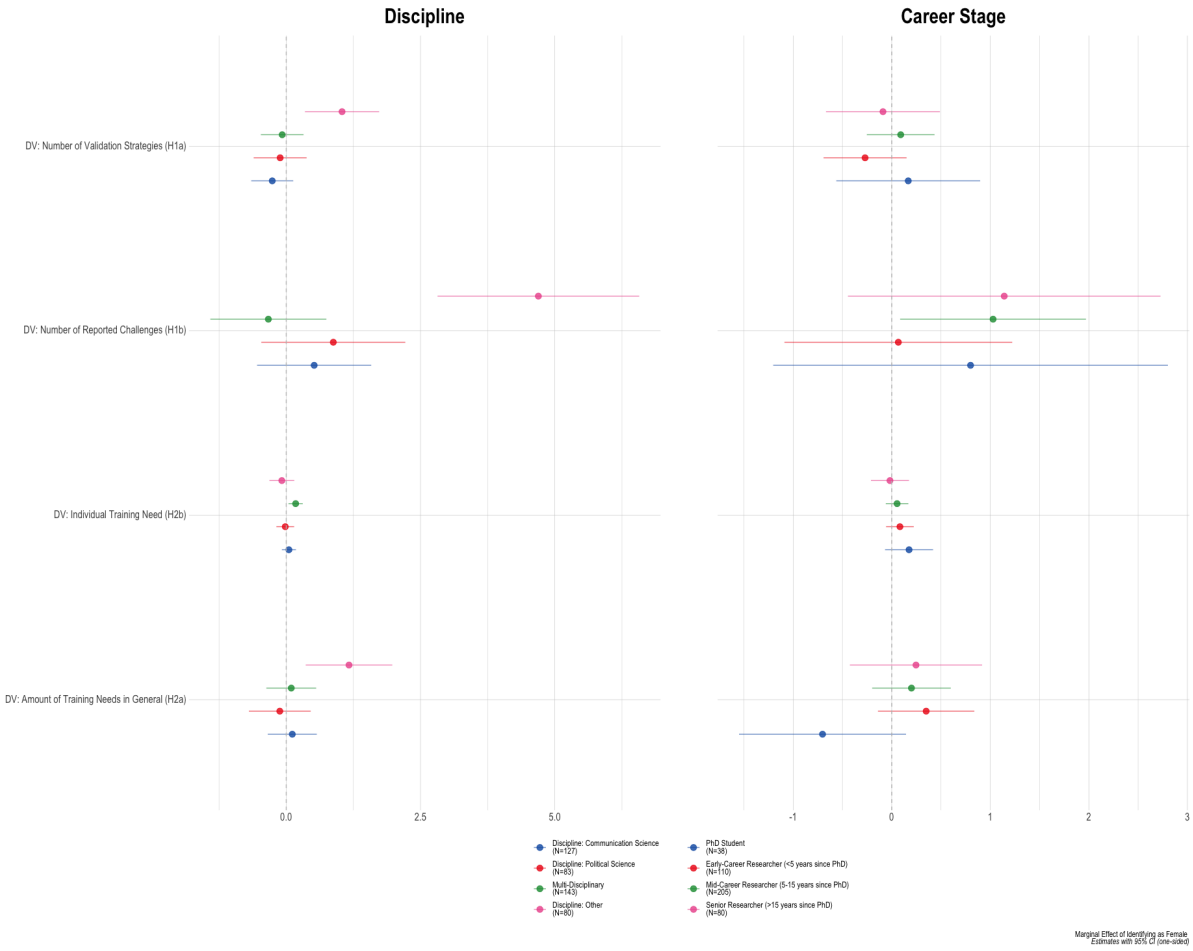
Figure 3: Effect of Being Female on Reported Type of Validation Strategies, Challenges, and Training Needs



In addition to the pre-registered effects, we have explored the interaction with discipline and stage of the career of scholars, as demonstrated in Figure 4. The left-hand panel of Figure 4 shows that compared to communication and political science, the other disciplines (grouping sociology,

psychology, and economy), where computational methods were less common, we see support for both our playing it safe and Mathilda effect hypotheses. This can be optimistically interpreted: When the CTAM is more prevalent in the discipline, women scholars catch up quickly. The right-hand panel of Figure 4 demonstrates that none of the reported results are driven by a particular career stage of scholars.

Figure 4: Marginal Effect of Being Female on Reported Validation Strategies, Challenges, and Training Needs



ADDITIONAL CHECK: ARE WOMEN LESS LIKELY TO BE FIRST AUTHOR?

Conclusion

[TBA]

References

Baden, Christian, Christian Pipal, Martijn Schoonvelde, and Mariken AC G van der Velden. 2021. "Three Gaps in Computational Text Analysis Methods for Social Sciences: A Research Agenda." *Communication Methods and Measures*, 1–18. https://doi.org/10.1080/19312458.2021.2015574.

Breuning, Marijke, Benjamin Isaak Gross, Ayal Feinberg, Melissa Martinez, Ramesh Sharma, and John Ishiyama. 2018. "Clearing the Pipeline? Gender and the Review Process at the American Political Science Review." *PS: Political Science& Politics* 51 (3): 629–34. https://doi.org/10.1017/S1049096518000069.

Brown, Nadia E., Yusaku Horiuchi, Mala Htun, and David Samuels. 2020. "Gender Gaps in Perceptions of Political Science Journals." *PS: Political Science &Amp; Politics* 53 (1): 114–21. https://doi.org/10.1017/S1049096519001227.

Brown, Nadia E., and David Samuels. 2018. "Beyond the Gender Citation Gap: Comments on Dion, Sumner, and Mitchell." *Political Analysis* 26 (3): 328–30. https://doi.org/10.1017/pan.2018.14.

Gatto, Malu AC, Anita R Gohdes, Denise Traber, and Mariken ACG Van Der Velden. 2020. "Selecting in or Selecting Out? Gender Gaps and Political Methodology in Europe." *PS: Political Science & Politics* 53 (1): 122–27. https://doi.org/10.1017/S1049096519001288.

König, Thomas, and Guido Ropers. 2018. "Gender and Editorial Outcomes at the American Political Science Review." *PS: Political Science &Amp; Politics* 51 (4): 849–53. https://doi.org/10.1017/S1049096518000604.

Maliniak, Daniel, Ryan Powers, and Barbara F. Walter. 2013. "The Gender Citation Gap in International Relations." *International Organization* 67 (4): 889–922. https://doi.org/10.1017/S0020818313000209.

Rossiter, Margaret W. 1993. "The Matthew Matilda Effect in Science." *Social Studies of Science* 23 (2): 325–41. https://doi.org/10.1177/030631293023002004.

Teele, Dawn Langan, and Kathleen Thelen. 2017. "Gender in the Journals: Publication Patterns in Political Science." *PS: Political Science &Amp; Politics* 50 (2): 433–47. https://doi.org/10.1017/S1049096516002985.