

Pre-Analysis Plan: Gender 'n Computational Social Science Publications

01 May, 2021 - 11:53:21 (CEST)

Contents

1	Expectations	2
2	Research Desing and Protocol	2
2.1	Sample	2
3	Measures	3
3.1	Dependent Variables	3
3.2	Independent Variable	4
3.3	Control Variables	4
3.4	Exclusion Criteria & Missing Values	4
4	Analysis	4
4.1	Hypothesis 1 & 2	4
4.2	Statistical Significance	5
	References	6

1 Expectations

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 political 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 X pathways that could distort the submission pool: a) playing it safe due to the gender perception gap, and b) the Mathilda effect. Papers using 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.

2 Research Desing and Protocol

2.1 Sample

We have conducted an expert survey, inviting all scholars published a scientific article between January 2016 and September 2020 ($N = 45,437$). Using a keyword search on the Web of Science, we then identified a total of 7,296 *potentially* relevant articles whose abstracts referred to some kind of textual contents or text analytic procedures. We then accessed the full text of these articles to determine whether the presented research included any form of quantitative textual analysis. 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. Analyses that relied solely on metadata or pre-existing classifications 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 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 was processed. 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). This screening yielded a total of $N = 854$ articles, for which the authors were looked up. This yielded us with 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 421 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*.

3 Measures

3.1 Dependent Variables

I rely on four measures reflecting different aspects of “playing it safe” and “embracing the Mathilda effect”.

1. *Number of Validation Strategies (H1a)* In this measure, respondents are asked to tick as many options as used to validate their CTAM, choosing from:
 - I check the documentation of the software developer
 - I check whether findings are plausible and interpretable
 - I compare machine-made classifications (i.e. coded outputs) against manually coded/given standards
 - I check whether the classification rules/criteria are valid (e.g., check for meaningful indicators in a dictionary)
 - I evaluate how well the algorithmic procedures match my concept of interest (e.g., checking adequacy of bag-of-words assumption)
 - Other: (Open ended)
2. *Reported Challenges (H1b)* In this measure, respondents are asked to indicate the challenges encountered when using/ or reasons not to use CTAM, choosing from:
 - Time/effort required (e.g. technical requirements, experience)
 - Funding required (e.g., for training, fees)
 - Availability of required training in computational methods
 - Limited methodological guidance/documentation of tools
 - Level of instruction and materials higher than needed
 - Level of instruction/materials lower than needed
 - Availability of suitable computational tools for certain languages
 - Comparability of computational tools in different languages
 - Issues concerning measurement validity/limited nuance
 - Loss of manual contact with the material
 - Reviewers’/editors’ skepticism toward computational methods among
 - Peers’ skepticism toward computational methods
 - I am skeptical toward computational methods myself
 - Other challenges: (Open ended)
3. *More training needs – general (H2a)* measured with the question “*Do you believe that, in the context of your training activities, there are important training needs in relation to computational text analysis in any of the following areas?*” Respondents had the following answer options:
 - Data and open access tools, please specify: (Open ended)
 - Programming and software skills, please specify: (Open ended)
 - Theory and concepts, please specify: (Open ended)
 - Research integrity, ethics, please specify: (Open ended)
 - Other: (Open ended)
4. *More training needs – individual (H2b)* was measured by asking whether respondents would be interested themselves to receive (further) training. Respondents had the following answer options:
 - Yes, I would like to learn more about: (Open ended)
 - No, my skills are sufficient
 - No, I do not plan to use computational text analysis anytime soon
 - No, for other reasons: (Open ended)

3.2 Independent Variable

We will use the gender reported by the scholar as the independent variable in the analyses.

3.3 Control Variables

As control variables, the following will be included in the analyses: *Academic discipline*, *time in academia*, *software knowledge*, *types of content analysis used*.

- *Academic discipline* is measured by asking in which field(s) the respondent is active as a researcher: Communications, Economics, Political Science, Psychology, Sociology, and Other.
- *time in academia* is measured by asking whether the respondent is a PhD student, an early-career researcher (<5 years since PhD), a mid-career researcher (5-15 years since PhD), or a senior researcher (>15 years since PhD).
- *software knowledge* is measured by asking much you use the following kinds of software in your own research, ranging from not at all till primary method of my research on a 5 point scale. The options are:
 - Statistical Software (e.g., SPSS, Stata, SAS)
 - Mathematical Software (e.g., Matlab, Octave)
 - Network Analysis Software (e.g., Gephi, Pajek, UCInet)
 - Qualitative Data Analysis Software (e.g., MaxQDA, Atlas.ti, Nvivo)
 - Specialized Text Mining Software (e.g., LIWC)
 - Generalized Text Analysis Platforms (e.g., AmCAT)
 - Open Source Coding Platforms (e.g., Python, R, Julia)
 - Other: (Open ended)
- *types of content analysis used* is measured by asking much you use the following kinds of text analysis methods in your own research, ranging from not at all till primary method of my research on a 5 point scale. The options are:
 - Qualitative text analysis (e.g., discourse analysis, conversation analysis)
 - Quantitative manual text analysis (e.g., manual content analysis)
 - Computational text analysis (e.g., automated content analysis, machine learning, text mining)
 - Other methods: (Open ended)

3.4 Exclusion Criteria & Missing Values

Participants are requested to respond to each question. We employ the following criteria: If 10% or less of the values on the dimension are missing, then we re-code the missing values to the overall mean. If 11% or more of the values on the dimension are missing, then we re-code the missing values to a constant (for instance 0) and include a dummy variable indicating whether the response on the covariate was missing or not.

4 Analysis

We test the hypotheses formulated in [Section 1](#) by fitting linear regression models for each dependent variable. In each model, will estimate the coefficient for gender of participant.

4.1 Hypothesis 1 & 2

I test the *playing it safe hypothesis* and the *Mathilda effect hypothesis* using Equation 1 below. The coefficient of β_1 denotes the difference in number of validation strategies reported (H1a), the difference in number of challenges reported (H1b), the difference in general training needs reported (H1a), and the difference in individual training needs reported (H1b), by women and men scholars. If the effect of β_1 is positive and statistically significant, the hypotheses will be confirmed.

$$\hat{Y} = \beta_0 + \beta_1 Female + \beta_{2-5} Controls + \varepsilon \quad (1)$$

4.2 Statistical Significance

All the hypotheses are directional, and therefore all of the tests will be one-tailed. I will use an α -value of 0.05 as the value for statistical significance in all models above.

References

- 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). Cambridge University Press: 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 & Politics* 53 (1). Cambridge University Press: 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). Cambridge University Press: 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). Cambridge University Press: 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 & Politics* 51 (4). Cambridge University Press: 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). Cambridge University Press: 889–922. <https://doi.org/10.1017/S0020818313000209>.
- Rossiter, Margaret W. 1993. "The Matthew Matilda Effect in Science." *Social Studies of Science* 23 (2). Sage Publications London: 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 & Politics* 50 (2). Cambridge University Press: 433–47. <https://doi.org/10.1017/S1049096516002985>.