Author gender differences in psychology citation impact 1996-2018[1]

Academic psychology in the USA is a gender success story in terms of overturning its early male dominance but there are still relatively few senior female psychology researchers. To assess whether there are gender differences in citation impact that might help to explain either of these trends, this study investigates psychology articles since 1996. Seven out of eight Scopus psychology categories had a majority of female first-authored journal articles by 2018. From regression analyses of first and last author gender and team size, female first authors associate with a slightly higher average citation impact, but extra authors have a ten times stronger association with higher average citation impact. Last author gender has little association with citation impact. Female first authors are more likely to be in larger teams and if team size is attributed to the first author's work, then their apparent influence of female first authors on citation impact doubles. Whilst gender differences in average citation impact are too small to account for gender-related trends in academic psychology, they warn that male dominated citation-based ranking lists of psychologists do not reflect the state of psychology research today.
Keywords: Gender; citation impact; bibliometrics; academic publishing; science of science

Introduction

Over the past century, academia has progressed from being almost universally male to the current more balanced situation, where male academics still predominate in most subjects in the USA, but with a diminishing overall gender gap. There are huge variations between subjects, however, with a factor of 100 difference between the most male (mathematical logic) and most female (nursing review and exam preparation) specialities in the USA (Thelwall, Bailey, Tobin, & Bradshaw, 2019). Although psychology in the USA has switched to majority female, females lag in terms of "power, status, and salary" (American Psychological Association, Committee on Women in Psychology, 2017). For the most recent data available, in 2013-14, 46% of males and 28% of females in U.S. graduate psychology departments were full professors (American Psychological Association, Committee on Women in Psychology, 2017). The scarcity of senior female psychologists is partly demographic (today's seniors may have joined the field when it had a male majority) but may also be due to career choices, cultural issues or hiring and promotion biases (e.g., Ceci & Williams, 2011).

This study assesses whether there are gender differences in research success in psychology that might explain either the female majority or the female underrepresentation within senior positions. It exploits one narrow but conveniently accessible data source: the citation impact of individual (standard) journal articles. Whilst citation counts are weak indicators of the impact of research and poorer indicators of its quality, when aggregated on a sufficiently large scale with careful methods they can reveal statistically significant evidence of differences, even when these differences are small. Although citations are sometimes valued as evidence that work has proven useful to future researchers (Merton, 1973), articles can be cited even when they are only weakly relevant (e.g., as introductory background) and biasing factors (e.g., geography) may affect which articles an author chooses to cite (Borgman & Furner, 2002). Articles can also be useful without generating citations if they inform practitioners, help students or discredit prior findings.

[1] Thelwall, M. (in press). Author gender differences in psychology citation impact 1996-2018. *International Journal of Psychology*.

Several previous studies have analysed gender differences in citation rates within psychology. An analysis of all Web of Science (WoS) Psychology articles from 2009 found research with male first or last authors to be more cited, as were larger teams (González-Álvarez & Cervera-Crespo, 2019). Several studies have also examined individual psychology journals, tending to find no overall gender differences in citation rates (Farrell, Corcoran, Sandoz, & McHugh, 2019; González-Alvarez, 2017). Male authors associated with higher citation impact across thirty neuroscience journals 2009-10, however (González-Álvarez & Cervera-Crespo, 2017). Thus, previous research has found either an (international) male citation advantage or statistically insignificant gender differences in citation rates.

None of the above citation analysis studies focus on the USA and all contain methodological weaknesses that undermine their findings. The investigation of all psychology (González-Álvarez & Cervera-Crespo, 2019) did not consider speciality differences in citation impact. This is a limitation because male specialisation in high citation areas within psychology would result in an overall male citation advantage. International studies are also limited because gender demographics and citation impact differences between nations can translate to an apparent male citation advantage. As a hypothetical example, if there is a greater proportion of male researchers in countries that tend to publish more highly cited papers then this would translate into an international apparent male citation advantage. In addition, the above studies have all also used the arithmetic mean, which can be misleading for highly skewed data (Olivier, Johnson, & Marshall, 2008). It is sensitive to individual highly cited articles and may therefore reflect these rather than underlying trends or differences between typical papers (Zitt, 2012).

This article assesses whether female authorship is associated with higher citation impact research for US psychology, considering narrow field specialities, focusing on a single country and using the largest sample size yet. It is based on the simplifying assumption that the gender of the first author is most important because they usually are the main contributor, with the last author sometimes also making a substantial contribution (Larivière, Desrochers, Macaluso, Mongeon, Paul-Hus, & Sugimoto, 2016). Research with more authors tends to be more cited (Larivière, Gingras, Sugimoto, & Tsou, 2015) and so team size is a possible contributory factor, especially if it is associated with gender. The article uses the terminology female and male (with the social sciences meaning of the social construct of gender role rather than biological sex: American Psychological Association, 2015) rather than man and woman because its focus is on the gender role portrayed and enacted by each academic, for example as represented by their first name, decisions and interactions, rather than their biological sex.

Methods

Scopus standard journal articles (of type "Journal Article", excluding editorials, reviews, brief communications etc.) from the Psychology category 1996-2018 were downloaded from Scopus during November 2018 to January 2019 (for the 2018 data). All articles with all authors from the USA were then extracted for the USA sample. Scopus was used in preference to the Web of Science for its greater coverage (Mongeon & Paul-Hus, 2016). Additional articles could have been found by adding other citation indexes, but a single source was required so that (a) its field classifications could be used and (b) the citation counts would be consistent. The starting year of 1996 corresponds to an expansion in Scopus. The final year for the citation analysis component, 2014, gives a minimum of three years for all articles to attract citations, which is adequate for comparison purposes (Abramo, Cicero, & D'Angelo, 2011). After three years, articles may well continue to attract citations, but their citation counts are mature enough for comparisons between articles published in the same year to be reasonable. Although book chapters and monographs are valuable in areas of social

psychology, the focus on standard journal articles is preferable because it gives a standard base for comparisons. Book chapters take longer to be cited, are less comprehensively indexed in major citation databases and are classified with a broader level of granularity by subject (Leydesdorff & Felt, 2012), making them problematic in citation analysis.

Raw citation counts are not useful for comparison purposes because older articles tend to be more cited and there are natural differences in citation rates between psychology subdisciplines. These are related to research type (pure vs. applied) and technical factors including typical number of references per article and the types of document cited. The solution is to field normalise for each citation count by dividing it by the average across all articles from the same narrow field and year, so that articles have higher values only when they are more cited than average for their field and publication year. This is known as the Mean Normalised Citation Score (MNCS) (Waltman, van Eck, van Leeuwen, Visser, & van Raan, 2011). This article uses the MNCS variant that log transforms all citation counts before performing any calculations, the Mean Normalised Log-transformed Citation Score (MNLCS) (Thelwall, 2017). The natural log transformation is necessary because citation counts are highly skewed and the normalisation calculation would otherwise have to take the arithmetic mean of a skewed set of citation counts for the MNCS denominator, which is not robust. For each article, its citation count c was therefore transformed to $\ln(1+c) / \overline{\ln(1+f)}$, where the average is taken over all articles f in the same narrow field and year. For articles in multiple Scopus categories, the denominator $\overline{\ln(1+f)}$ was replaced with the average of all the averages, $\overline{\overline{\ln(1+f)}}$, for every narrow Scopus field containing the article. The result is the Normalised Log-transformed Citation Score (NLCS). Averaging these for a set of articles (e.g., female-authored psychology articles) gives a MNLCS that can be fairly compared to another MNLCS even if the latter is based on a different balance of fields and years. This is because, by design, NLCS is always a ratio to the world average: above 1 only when the original citation count is above average for its narrow field(s) and year of publication. The skewness of the resulting eight data sets varied between -0.286 (*Clinical Psychology* 1996-2014) and 0.304 (*Psychology (misc)* 1996-2014) and the excess kurtosis varied between -0.031 (*Neuropsychology & Physiological Psychology* 1996-2014) and 0.505 (*Psychology (misc)* 1996-2014).

First and last author gender were dependant variables along with team size. Team size was recorded as 1,2,3,4 or at least 5. Larger team sizes were not coded separately (e.g., 5, 6, 7) because there were fewer articles with larger team sizes and so their regression parameters would be less accurate. These five sizes were coded as separate binary variables rather than through a function (e.g., log transformation) because there is no agreement on the best function to use.

Gender was guessed from the first name of each author, when strongly gendered. Papers where the first or last author gender could not be determined were removed. To illustrate, if for a three-author paper the first author was female but the last was unknown then the article would be rejected. The first name heuristic matched with a list of names that are 90% male or female in the US 2010 census or as reported by gender-api.com from social web profiles. The algorithm was cross-checked by applying it to a set of 1010 US first authors from 2017 that had been manually checked online (taken from: Thelwall et al., 2019). The algorithm gave nine errors out of 695 classifications (1.3%), almost equal in direction of gender swap (5 and 4). Overall, the algorithm estimated 38.4% female and the human checkers estimated 37.4% female for the test set, so the algorithm seems to overestimate the proportion of females by 1%. This compensation factor should be halved for the current data set, because only 16% of the authors did not have a gender assigned by the algorithm in comparison to 31% from the test set, so there is half as much scope for female overestimation based on the authors without assigned genders.

The results are reported and field normalised using the Scopus categories of the articles. These are classified by journal, using Elsevier's All Science Journal Classification (ASJC) codes. Unfortunately, these are not defined, except by name, and the scope of the category *Psychology (all)* is not clear since it does not contain all psychology journals. It contains general journals, with the two largest being *Psychological Reports* and *Psychological Science*, as well as specialist journals, such as *Appetite* and *Computers in Human Behaviour*. *Psychology (misc)* seems to contain journals that do not match the other psychology categories. The results are less reliable for the broader categories *Psychology (all)* and *Psychology (misc)* because heterogeneity creates the possibility that the category contains high and low citation specialities with different gender balances that would bias the field normalisation process.

Ordinary least squares regression was first tested to assess whether gender and team size factors associate with higher citation counts. This is a plausible method because NLCS are typically not highly skewed and have moderate kurtosis values, even if the data do not closely follow the normal distribution shape. The NLCS variance is inflated by differences between years in the average for the USA, however, so publication year was added as a set of binary independent variables (one per year after the first). Variations by year could also be addressed by multilevel modelling (a random intercepts model) but there is sufficient data to avoid this, given that variations in gender differences over time may have systematic causes and there is sufficient data to accommodate an extra parameter for each year.

As shown in the supplementary material (see residual plots and Bartlett tests), there is heteroscedasticity in the regression model because of the different ranges of citation counts for each year. Heteroscedasticity would make the regression coefficient statistical tests unreliable. Since the ratio of male to female authors also changes over time, this heteroscedasticity cannot be fully eliminated by running a two-step regression, first for publication year and second for gender. Thus, because the heteroscedasticity could not be fully eliminated, a maximum likelihood method was used to fit the data that does not require homoscedasticity (Yohai, 1987), using the *lmrob* function in the R package robustbase (Rousseeuw, Croux, Todorov, Ruckstuhl, Salibian-Barrera, Verbeke, & Maechler, 2019), incorporating publication years to reduce heteroscedasticity and therefore improve the accuracy of the model fit.

This study used only publicly available data and was exempt from ethical approval.

Results

Both first and last author genders were assigned for at least three quarters of US psychology articles 1996-2018 in all eight fields (Table 1). Most first authors were male in 1996 and most were female in 2018, except for *Psychology (miscellaneous)* (49% female). The female percentage increase was highest in *Clinical Psychology* and lowest in *Neuropsychology and Physiological Psychology*. Female last authors achieved a majority in only three fields by 2018, with more moderate increases, except in *Clinical Psychology*. Presumably the last authors are more senior, reflecting the lower female share of higher ranked posts.

Table 1. Gender identification and share statistics for the eight Scopus narrow fields 1996-2018.

Narrow field	Articles with US only authors	Gendered first & last authors %	Female first in 1996 %	Female first in 2018 %	Change 1996-2018 %	Female last in 1996 %	Female last in 2018 %	Change 1996-2018 %
Psychology (all)	52492	76.9	35.2	52.9	17.8	34.4	47.8	13.3
Psychology (miscellaneous)	7058	74.4	30.7	49.2	18.5	33.1	41.0	7.9
Applied Psychology	42646	76.0	36.7	55.2	18.5	35.0	47.0	12.0
Clinical Psychology	66882	78.6	38.0	58.5	20.5	37.0	47.2	10.2
Developmental & Edu. Psych.	78854	76.6	48.8	66.6	17.7	46.8	58.9	12.1
Experimental & Cognitive Psych.	40222	76.1	32.8	50.6	17.8	29.0	43.9	14.8
Neuropsych. & Physiological Psych.	15801	75.2	39.4	51.1	11.7	30.7	39.9	9.2
Social Psychology	54810	77.2	43.3	57.9	14.5	44.2	53.7	9.5

There is an increasing proportion of female first-authored research in all eight narrow psychology fields (Figures 1-8), steadily achieving a female majority in almost all cases. Except for *Psychology (miscellaneous)*, female first-authored research had higher citation impact than male first-authored research for most years, but there is not a common temporal pattern for either.

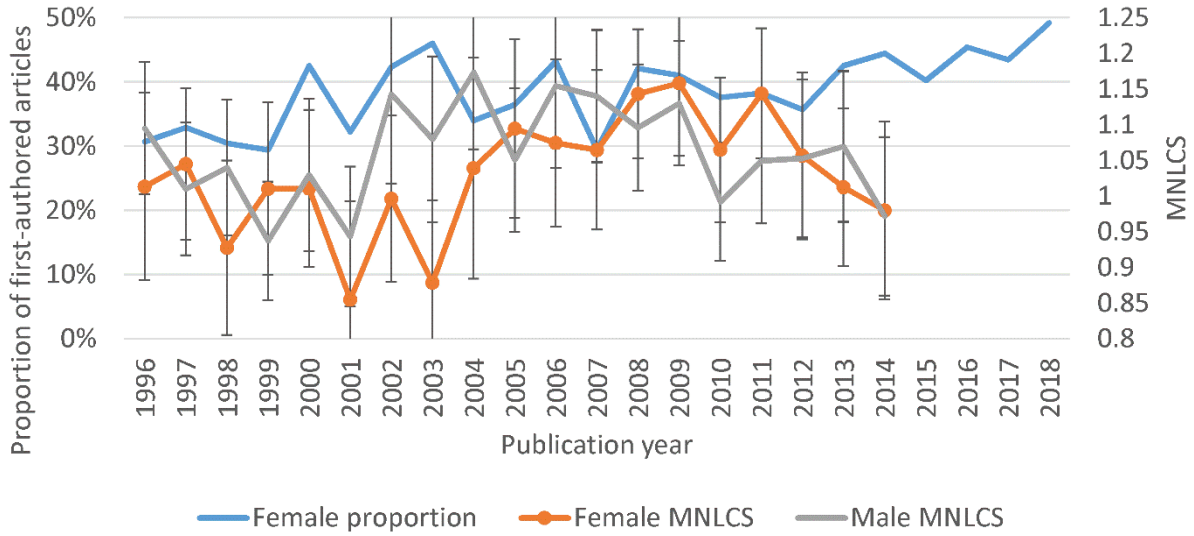

Figure 1. Psychology (miscellaneous): The proportion of female first-authored articles and the average citation impact of female and male first-authored articles. Data: Scopus-indexed journal articles with US authors and first and last author genders detected.

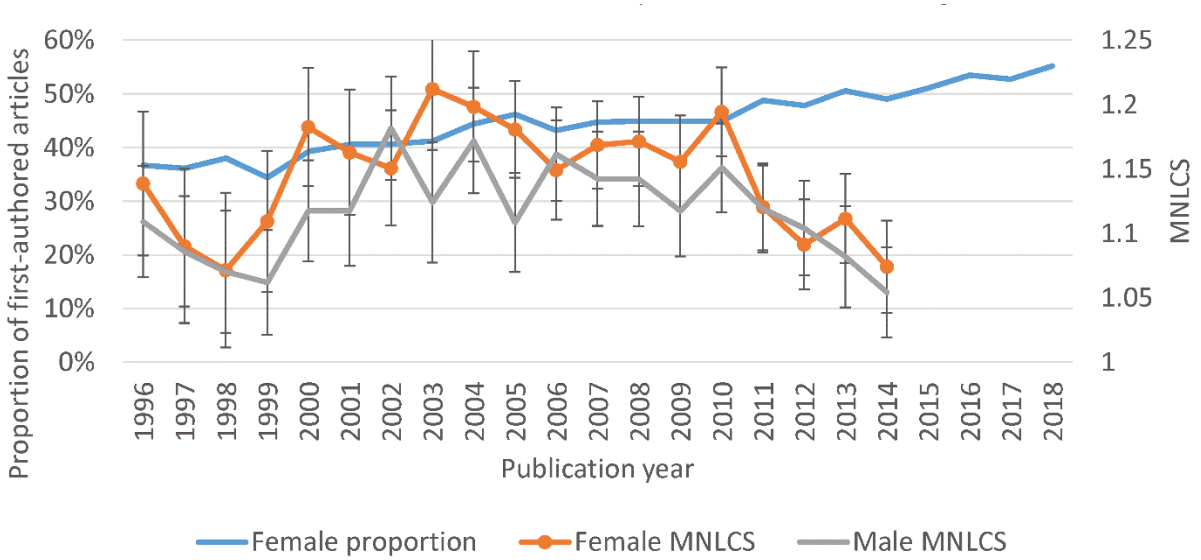

Figure 2. Applied Psychology: The proportion of female first-authored articles and the average citation impact of female and male first-authored articles. Data: Scopus-indexed journal articles with US authors and first and last author genders detected.

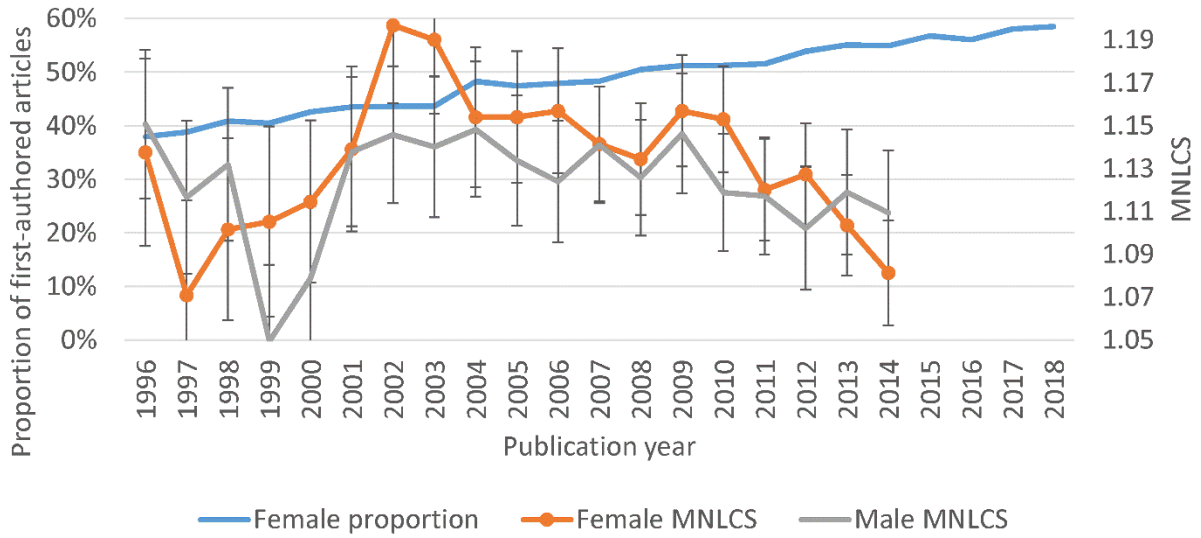

Figure 3. Clinical Psychology: The proportion of female first-authored articles and the average citation impact of female and male first-authored articles. Data: Scopus-indexed journal articles with US authors and first and last author genders detected.

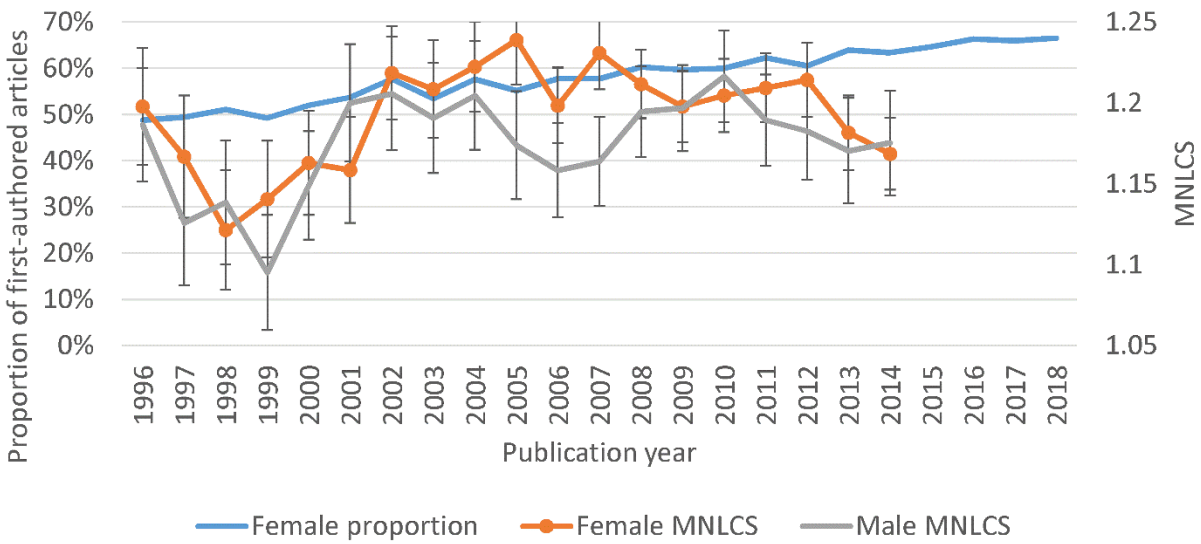

Figure 4. Developmental and Educational Psychology: The proportion of female first-authored articles and the average citation impact of female and male first-authored articles. Data: Scopus-indexed journal articles with US authors and first and last author genders detected.

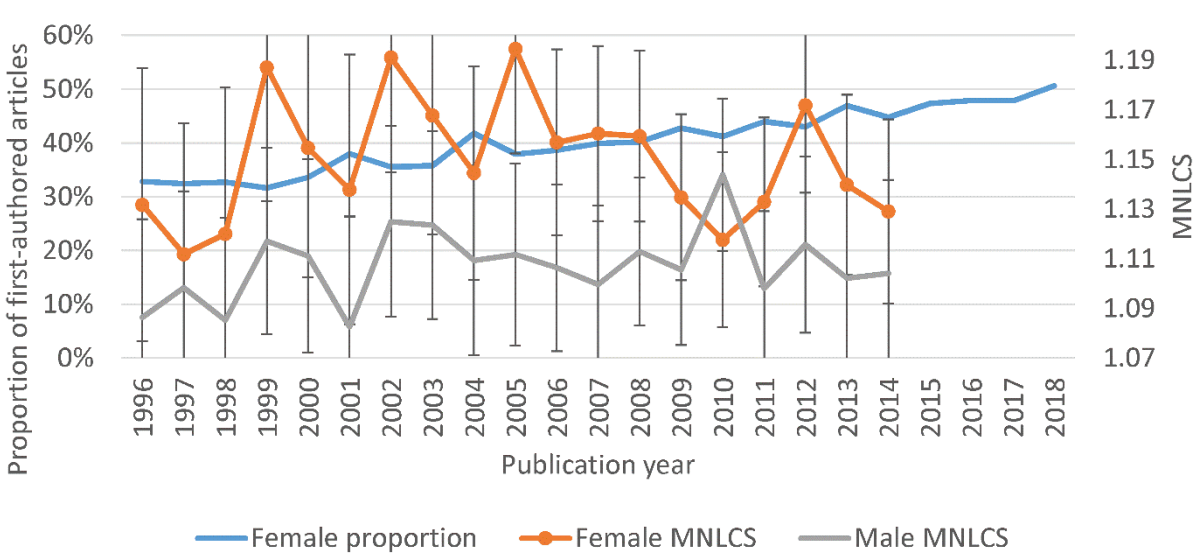

Figure 5. Experimental and Cognitive Psychology: The proportion of female first-authored articles and the average citation impact of female and male first-authored articles. Data: Scopus-indexed journal articles with US authors and first and last author genders detected.

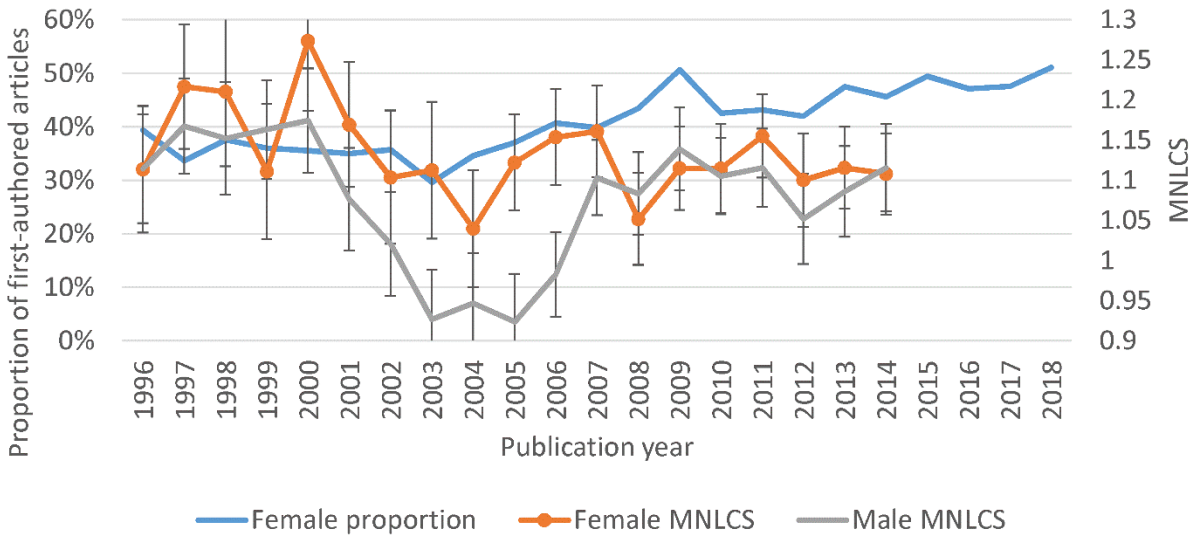

Figure 6. Neuropsychology and Physiological Psychology: The proportion of female first-authored articles and the average citation impact of female and male first-authored articles. Data: Scopus-indexed journal articles with US authors and first and last author genders detected.

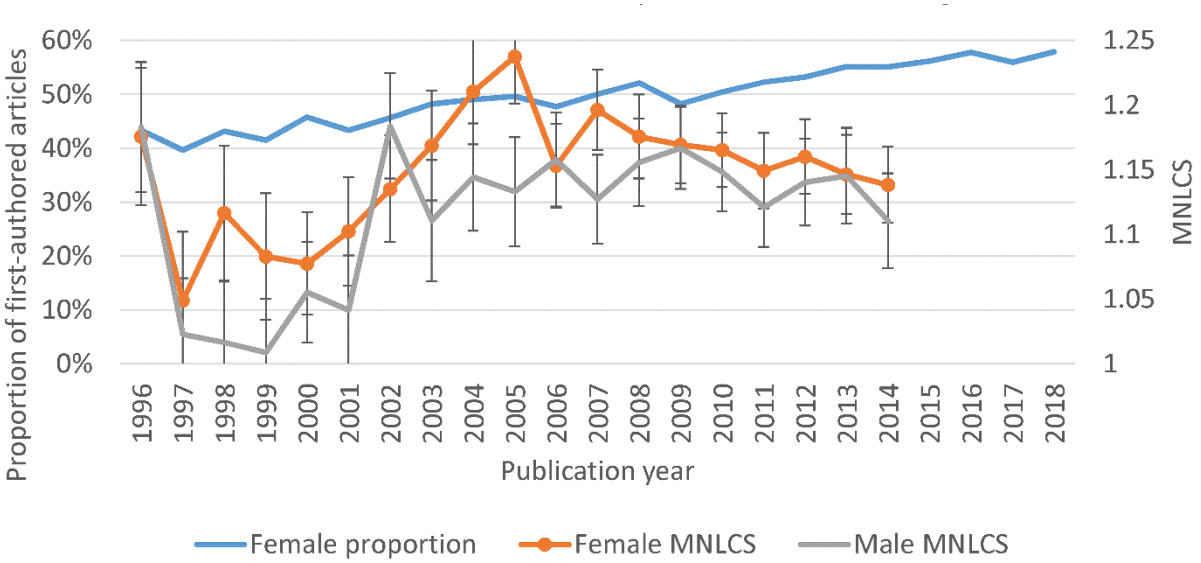

Figure 7. Social Psychology: The proportion of female first-authored articles and the average citation impact of female and male first-authored articles. Data: Scopus-indexed journal articles with US authors and first and last author genders detected.

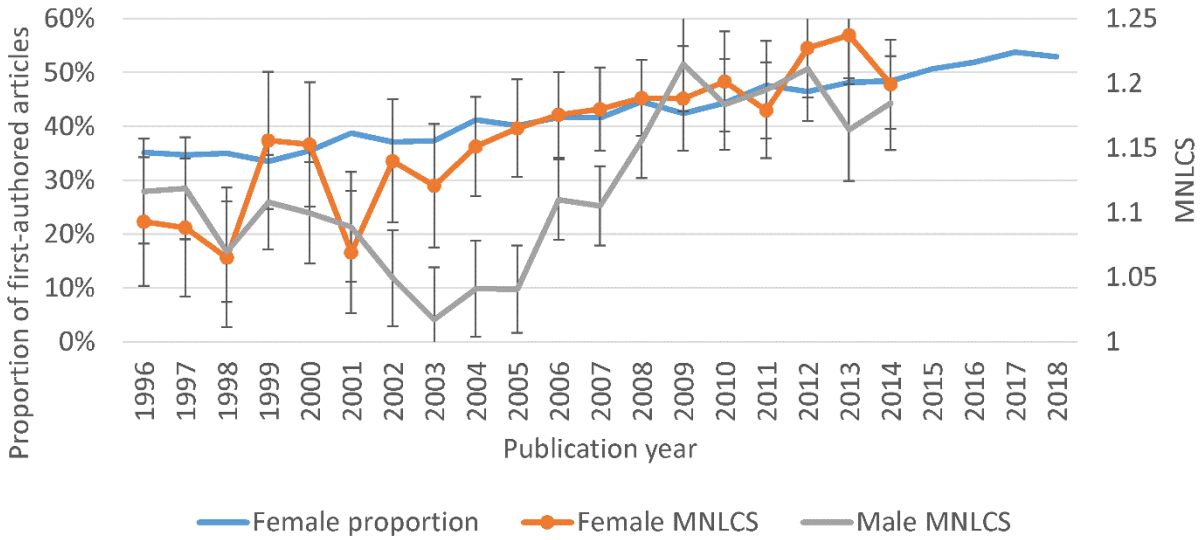

Figure 8. Psychology (all): The proportion of female first-authored articles and the average citation impact of female and male first-authored articles. Data: Scopus-indexed journal articles with US authors and first and last author genders detected.

Female first authors led larger teams than male first authors in all cases (Table 2). The same is true for last authors, except that the difference is not always statistically significant and is smaller.

Table 2. Average team sizes 1996-2014 for Scopus narrow Psychology fields (US only) by first and last author gender. Gender differences are statistically significant with $p<0.001$ except where underlined (Wilcoxon, $p>0.05$).

Narrow field	Articles	Female first team size	Male first team size	Female last team size	Male last team size
Psychology (all)	32320	2.12	1.84	2.05	1.89
Psychology (miscellaneous)	4207	1.92	1.62	1.91	1.63
Applied Psychology	24697	2.24	2.00	2.18	2.05
Clinical Psychology	39811	2.62	2.40	2.49	2.52
Developmental and Educational Psych.	45665	2.30	2.12	2.23	2.22
Experimental & Cognitive Psych.	23778	2.28	2.00	2.19	2.06
Neuropsychology & Physiological Psych.	9369	2.70	2.32	2.50	2.45
Social Psychology	31893	2.03	1.85	2.00	1.88

In the full regression model, considering team size, first and last author gender (and publication year), a female first author gave a statistically significant advantage in five out of eight fields (Table 3). A female last author was a statistically significant advantage in one field and a statistically significant disadvantage in another. The magnitude of the gender advantage is approximately a tenth of the magnitude of the extra author advantage.

Table 3. Regression coefficients for the full regression model for Scopus narrow Psychology fields (US only, first and last author gendered). Publication year regression coefficients are not shown.

	Model 1					
Narrow field	**Female first**	**Female last**	**2 auth**	**3 auth**	**4 auth**	**5+ auth**
Psychology (all)	0.023***	-0.013	0.235***	0.282***	0.314***	0.356***
Psychology (miscellaneous)	-0.010	-0.026	0.103***	0.196***	0.212***	0.195***
Applied Psychology	0.004	0.006	0.237***	0.289***	0.316***	0.348***
Clinical Psychology	-0.004	-0.021***	0.290***	0.348***	0.378***	0.452***
Developmental and Educational Psych.	0.015**	-0.009	0.131***	0.213***	0.221***	0.287***
Experimental & Cognitive Psych.	0.023***	0.016*	0.158***	0.139***	0.158***	0.203***
Neuropsychology & Physiological Psych.	0.030**	0.008	0.244***	0.262***	0.296***	0.334***
Social Psychology	0.018**	0.000	0.190***	0.272***	0.285***	0.294***

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Three reduced regression models were also run, ignoring the team size variable (Table 4). These treat the number of co-authors as under the control of the first (Model 3: excluding the last author and team size) or last (Model 4: excluding the first author and team size) author or both (Model 2: excluding team size), attributing any advantage to these author positions rather than the team itself. With this model, a female first author is a statistically significant citation advantage for all narrow fields except Psychology (miscellaneous) in both Model 2 and Model 3. A female last author is a statistically significant disadvantage in two fields and an advantage for one in Model 2. In model 4, a female last author is a statistically significant advantage in three cases and a statistically significant disadvantage in one. The magnitude of the gender advantage in the reduced model is larger than in the full model and is approximately a fifth of the magnitude of the extra author advantage.

A fifth reduced model was run to take into account interactions between first and last author gender, using male first and last authors as the default (Table 4). Articles with at least one female (either first or last author) are associated with higher citations in all fields except *Psychology (miscellaneous)* and *Clinical Psychology*. In the latter case, a female first and last author is associated with fewer citations than a male first and last author. In general, the optimal combination is a female first author with a male last author since this gives the highest coefficient for all fields.

Table 4. Regression coefficients for four reduced regression models for Scopus narrow Psychology fields (US only, first and last author gendered), ignoring team size. Publication year regression coefficients are not shown.

	Model 2			Model 3		Model 4	Model 5		
Narrow field	**Female first**	**Female last**		**Female first**		**Female last**	**Female 1st & last**	**Female 1st male last**	**Male 1st female last**
Psychology (all)	0.049***	-0.006		0.047***		0.011	0.026**	0.137***	0.094***
Psychology (miscellaneous)	0.007	-0.011		0.002		-0.009	-0.016	0.061*	0.045
Applied Psychology	0.023**	0.010		0.026***		0.017*	0.020*	0.116***	0.115***
Clinical Psychology	0.019***	-0.027***		0.013*		-0.022***	-0.018**	0.115***	0.095***
Developmental & Educational	0.026***	-0.013**		0.022***		-0.006	0.022***	0.106***	0.086***
Experimental & Cognitive.	0.037***	0.020**		0.042***		0.030***	0.048***	0.061***	0.049***
Neuropsych. & Physiological	0.056***	0.004		0.057***		0.016	0.037*	0.106***	0.072***
Social Psychology	0.033***	0.004		0.035***		0.016**	0.036***	0.131***	0.115***

* p<0.05; ** p<0.01; *** p<0.001

A sixth model was run to investigate the interaction of team size with first author gender (Table 5; the last four columns are the interaction terms). For this model, the default was a solo male article and the regression coefficients show increases or decreases in citation impact association relative to this. The interaction coefficients are mostly statistically significant, negative and much smaller than the main team size coefficients. Thus, larger teams tend to associate with greater increases in citation counts for male first authors than for female first authors.

Table 5. Regression coefficients for a full regression model for Scopus narrow Psychology fields (US only, first and last author gendered), with first author gender interactions with team size. Publication year regression coefficients are not shown.

	Model 6									
Field	**Female first**	**Female last**	**2 auth**	**3 auth**	**4 auth**	**5+ auth**	**2 auth fem 1st**	**3 auth fem 1st**	**4 auth fem 1st**	**5+ auth fem 1st**
All	0.058***	-0.021**	0.253***	0.301***	0.335***	0.375***	-0.046**	-0.052*	-0.053*	-0.048*
Mis	0.024	-0.033	0.146***	0.161***	0.184***	0.217***	-0.106*	0.067	0.047	-0.055
App	0.076***	-0.009	0.270***	0.317***	0.363***	0.396***	-0.083***	-0.072**	-0.111***	-0.111***
Clin	0.023	-0.025***	0.301***	0.350***	0.395***	0.472***	-0.026	-0.008	-0.038*	-0.045**
Dev	0.03**	-0.012*	0.145***	0.200***	0.245***	0.294***	-0.024	0.021	-0.042*	-0.014
Exp	0.08***	0.008	0.190***	0.150***	0.177***	0.221***	-0.089***	-0.040	-0.061**	-0.058**
Neu	0.141***	-0.006	0.290***	0.292***	0.340***	0.375***	-0.141***	-0.106**	-0.138***	-0.130***
Soc	0.094***	-0.020**	0.227***	0.321***	0.351***	0.367***	-0.081***	-0.105***	-0.135***	-0.148***

* p<0.05; ** p<0.01; *** p<0.001

A seventh model was run to investigate the interaction of team size with *last* author gender (Table 6). For this model, the default was a solo male article and the regression coefficients show increases or decreases in citation impact association relative to this. The interaction coefficients are again mostly statistically significant, negative and much smaller than the main team size coefficients. This gender difference tends to be smaller than for first authors. Thus, larger teams tend to associate with slightly greater increases in citation counts for male last authors than for female last authors.

Table 6. Regression coefficients for a full regression model for Scopus narrow Psychology fields (US only, first and last author gendered), with last author gender interactions with team size. Publication year regression coefficients are not shown.

	Model 7									
Field	**Female first**	**Female last**	**2 auth**	**3 auth**	**4 auth**	**5+ auth**	**2 auth fem 1st**	**3 auth fem 1st**	**4 auth fem 1st**	**5+ auth fem 1st**
All	0.018**	0.006	0.242***	0.280***	0.33***	0.375***	-0.020	0.002	-0.039	-0.047*
Mis	-0.017	0.008	0.153***	0.165***	0.185***	0.189***	-0.123**	0.065	0.051	0.004
App	-0.007	0.059***	0.261***	0.311***	0.343***	0.387***	-0.062***	-0.056*	-0.066**	-0.094***
Clin	-0.005	-0.013	0.293***	0.343***	0.383***	0.460***	-0.006	0.013	-0.012	-0.021
Dev	0.016**	-0.014	0.132***	0.183***	0.224***	0.283***	-0.001	0.06**	-0.005	0.008
Exp	0.018**	0.049**	0.172***	0.160***	0.165***	0.218***	-0.042*	-0.061*	-0.023	-0.045*
Neu	0.021*	0.091**	0.275***	0.283***	0.333***	0.371***	-0.098**	-0.064	-0.114**	-0.116***
Soc	0.001	0.066***	0.229***	0.301***	0.328***	0.346***	-0.084***	-0.064*	-0.092***	-0.112***

* p<0.05; ** p<0.01; *** p<0.001

Combining all the evidence, a female first author is associated with higher citation counts in seven out of eight psychology fields (Model 2, 3), including all six specialist areas. This is partly because the first author tends to gather a larger team because when team size is treated as not attributable to the first author then there is a female first author citation advantage in only five fields. The magnitude of the gender advantage is small, however, and much smaller than the advantage associated with an extra author. The gender combination tending to be associated with the highest citation impact is a female first author with a male last author, which accounts for 19% of articles overall (Table 7). Male first or last authors tend to be associated with slightly greater increases in citations for larger team sizes than do female first or last authors, respectively.

Table 7. Percentages of first and last author gender combinations (US only, first and last author gendered).

Narrow field	**Female 1st & last**	**Female 1st male last**	**Male 1st female last**	**Male 1st & last**
Psychology (all)	26%	17%	15%	42%
Psychology (miscellaneous)	23%	15%	14%	47%
Applied Psychology	27%	18%	15%	39%
Clinical Psychology	28%	22%	15%	35%
Developmental & Educational	38%	21%	15%	26%
Experimental & Cognitive.	21%	20%	15%	44%
Neuropsych. & Physiological	19%	23%	14%	44%
Social Psychology	33%	18%	15%	35%
Average	**27%**	**19%**	**15%**	**39%**

Limitations

The results are limited by the use of citations, which only partially reflect the impact of academic research. They are also limited by the Scopus classification scheme because anomalies in this can affect the field normalisation results. A quarter of the papers had to be rejected for unknown first or last author gender (Table 1). The gender identification heuristic therefore biases the results towards majority ethnic groups in the USA, which are more likely to have first names that are common enough to be included on the list. Cultures tending to use

gender neutral names (e.g., Sikh) or sometimes using names that become gender neutral when written in the Latin alphabet (e.g., Chinese) will also be underrepresented.

The models used do not consider demographic shifts and so female citation advantages may be second order effects of a greater share of females amongst younger authors (i.e., if younger authors are more cited and there are more females amongst the youngest authors, then this would increase the female citation average). The regression assumes that gender differences in citation impact have not systematically changed over time. Whilst this is plausible in the context of the graphs, it is unproven. The proportion of female authors may have been overestimated by 0.5% according to the manual check reported in the methods.

Discussion

Females first-authored at least as many papers as males by 2018 in seven of the eight categories, which is unusual within the context of male domination of academia. Females may be attracted to psychology because it is a "people" career, with some evidence of a relative female preference for interacting with people (Su, Rounds, & Armstrong, 2009). Alternatively, psychology may be regarded as a directly helping discipline (e.g., clinical psychologists), and therefore attractive to females, who tend to value communal career goals more (Diekman, Steinberg, Brown, Belanger, & Clark, 2017). Psychology may also be perceived as more welcoming to females than many other scientific subjects (Tellhed, Bäckström, & Björklund, 2017). From a skills perspective, more females may believe themselves to be empathetic and therefore suitable for clinical or counselling psychology careers (Harton & Lyons, 2003). Possible more prosaic reasons include amenability to an appropriate work-life balance or the promise of understanding everyday social interactions better.

Female last authors have achieved parity or are close to parity in most of the psychology fields studied. Presumably the smaller proportions of female last authors are due to a lower proportion of senior females occupying this position as team leaders, research supervisors or recruited collaborators (Clay, 2017). For psychology graduate students in the USA, there have been three times as many females as males for more than ten years (Fowler, Cope, Michalski, Christidis, Lin, & Conroy, 2019) but this has not yet translated into a substantial female majority of research publications, perhaps because of the wide variety of careers available to higher graduates and PhDs (e.g., Christidis, Stamm, Lin, & Conroy, 2019).

In terms of gender differences in citation rates, the results point to a small female first author advantage (Model 3), in contrast to the previous studies reviewed above that found either a male advantage or no significant difference. The current paper uses the finest-grained field normalisation yet for this issue, a method that does not allow individual highly cited articles to dominate, and has the most data. In combination, these account for the statistically significant differences. It is nevertheless possible that there are some areas of psychology in which male first authors have more cited research. Visual inspection of the graphs suggests that there isn't a general trend for increasing or decreasing relative female citation advantages in any field, so the gender difference seem likely to persist.

There is a clear relationship between gender and team size, with female first authors leading larger teams in all fields and female last authors contributing to larger teams in most fields. When team size is combined with first and last author genders in one model, the female first author citation advantage is smaller and not always statistically significant. Thus, the above evidence suggests that part of the female first author citation advantage is due either to her ability or willingness to build larger teams or more frequently finding herself leading larger teams for other reasons (e.g., female PhD students more frequently co-

authoring with supervisors; female researchers being more likely to work on topics requiring multidisciplinary inputs). The gender combination associating with the highest citation counts in most fields is a female first author with a male last author. This seems likely to often be a female PhD student (because female PhDs predominate) with a senior supervisor (because senior males predominate), so this might be due to senior academics supervising the most capable students.

Conclusions

The results point to a small female first author citation advantage in all six specialist areas of psychology, as well as one general psychology category but not the miscellaneous category. This female citation advantage is partly due to leading larger teams of co-authors and is not consistent over time. The advantage is small, however, and never greater than 0.057, on average (Model 4). Thus, female first authorship is associated with a 6% higher field normalised logged citation count. This seems unlikely to be large enough to account for the relatively high female participation rate in psychology.

The results point to US females in psychology at least equalling males in terms of research citation impact. This is a more statistically robust result than from previous research which has found either a male citation advantage or no statistically significant differences. Thus, from the new findings, citation-related factors cannot explain the male dominance of senior US positions and awards. It is possible that misinterpretations of citation counts, such as use of the h-index, career citations, or un-normalised citation counts compared between fields, have incorrectly led to males gaining greater recognition. In this context, the current article should help to dispel any myths that female research in psychology is less impactful. Moreover, if citations are used to make award, tenure and promotion decisions in psychology and the results are male dominated, then committees should look very carefully at their methods because the current results show that it should not usually happen.

Ethical Compliance Statement

Funding: None.

Compliance with Ethical Standards: No human subjects involvement.

Conflicts of Interest: The authors declare that they have no conflict of interest.

Informed Consent (when applicable): Not applicable.

References

Abramo, G., Cicero, T., & D'Angelo, C. A. (2011). Assessing the varying level of impact measurement accuracy as a function of the citation window length. Journal of Informetrics, 5(4), 659-667.

American Psychological Association. (2015). Guidelines for psychological practice with transgender and gender nonconforming people. American Psychologist, 70(9), 832-864.

American Psychological Association, Committee on Women in Psychology (2017). The changing gender composition of psychology: Update and expansion of the 1995 task force report. https://www.apa.org/pi/women/programs/gender-composition

Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. Annual Review of Information Science and Technology, 36, 2-72.

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. Proceedings of the National Academy of Sciences, 108(8), 3157-3162.

Christidis, P., Stamm, K., Lin, L. & Conroy, J. (2019). Educating the next generation. Monitor on Psychology, 50(7), 21.

Clay, R.A. (2017). Women outnumber men in psychology, but not in the field's top echelons. Monitor on Psychology, 48(7), 18.

Diekman, A. B., Steinberg, M., Brown, E. R., Belanger, A. L., & Clark, E. K. (2017). A goal congruity model of role entry, engagement, and exit: Understanding communal goal processes in STEM gender gaps. Personality and Social Psychology Review, 21(2), 142- 175.

Farrell, L., Corcoran, F., Sandoz, E., & McHugh, L. (2019). Examination of sex-specific publication trends within the Journal of Contextual Behavioral Science between 2012 and 2017. Journal of Contextual Behavioral Science, 11(1), 27-33.

Fowler, G., Cope, C., Michalski, D., Christidis, P., Lin, L., & Conroy, J. (2019). Women outnumber men in psychology graduate programs. Monitor on Psychology, 49(11), 21.

González-Álvarez, J., & Cervera-Crespo, T. (2017). Research production in high-impact journals of contemporary neuroscience: a gender analysis. Journal of Informetrics, 11(1), 232-243.

González-Álvarez, J., & Cervera-Crespo, T. (2019). Contemporary psychology and women: A gender analysis of the scientific production. International Journal of Psychology, 54(1), 135-143.

González-Alvarez, J. (2017). Gender of Personality and Individual Differences (PAID) contributors: An analysis of recent years (2008–2016). Personality and Individual Differences, 117(1), 106-110.

Harton, H. C., & Lyons, P. C. (2003). Gender, empathy, and the choice of the psychology major. Teaching of Psychology, 30(1), 19-24.

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science*, 46(3), 417-435.

Larivière, V., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7), 1323-1332.

Leydesdorff, L., & Felt, U. (2012). Edited volumes, monographs and book chapters in the Book Citation Index (BKCI) and Science Citation Index (SCI, SoSCI, A&HCI). Journal of Scientometric Research, 1(1), 28-34.

Merton, R. K. (1973). The sociology of science: Theoretical and empirical investigations. Chicago, IL: University of Chicago Press.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics, 106(1), 213-228.

Olivier, J., Johnson, W. D., & Marshall, G. D. (2008). The logarithmic transformation and the geometric mean in reporting experimental IgE results: what are they and when and why to use them? Annals of Allergy, Asthma & Immunology, 100(4), 333-337.

Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., & Maechler, M. (2019). robustbase: Basic Robust Statistics. R package version 0.92-8.

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: a meta-analysis of sex differences in interests. Psychological bulletin, 135(6), 859.

Tellhed, U., Bäckström, M., & Björklund, F. (2017). Will I fit in and do well? The importance of social belongingness and self-efficacy for explaining gender differences in interest in STEM and HEED majors. Sex Roles, 77(1-2), 86-96.

Thelwall, M., Bailey, C., Tobin, C. & Bradshaw, N. (2019). Gender differences in research areas, methods and topics: Can people and thing orientations explain the results? Journal of Informetrics, 13(1), 149-169.

Thelwall, M. (2017). Three practical field normalised alternative indicator formulae for research evaluation. *Journal of Informetrics*, 11(1), 128–151.

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. (2011). Towards a new crown indicator: An empirical analysis. Scientometrics, 87(3), 467-481.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15(2), 642-656.

Zitt, M. (2012). The journal impact factor: Angel, devil, or scapegoat? A comment on JK Vanclay's article 2011. Scientometrics, 92(2), 485-503.