DO GENDERED LANGUAGES FAIL WOMEN IN MATH?

Tamar Kricheli Katz* and Tali Regev**

Abstract:

Four studies examined whether addressing women and men in the masculine – a prominent grammatical practice in gendered languages – affects their performance. All studies used large, random, representative samples of the native Hebrew-speaking adult population in Israel. When addressed in the masculine, women meaningfully decreased their performance in math. These effects were stronger among participants who acquired the Hebrew language early rather than later in life, suggesting that it is the extent of language proficiency that generates one's sensitivity to being addressed in the masculine or in the feminine. Similar but smaller effects were found when men were addressed in the feminine compared to the masculine. In addition, when women were addressed in the masculine, their effort exertion levels declined and they reported feeling that 'science is for men' more than when addressed in the feminine. Three additional studies explored the role of gender stereotypes in creating the effects of the masculine and feminine address on performance. Whereas addressing women in the masculine in stereotypically male-type tasks (Study 1) and gender-neutral-type tasks (Studies 3 and 4) decreased their performance, it improved their performance in female-type tasks (Study 2). The results of our study demonstrate the powerful role of language in activating stereotypes and cultural beliefs and, in turn, the powerful role of stereotypes and cultural beliefs in affecting actual performance.

Key Words: Gender, Languages, Stereotypes

Introduction

Languages play a powerful role in creating and maintaining unequal realities for women and men. They reflect and perpetuate the prominent social distinctions and the stereotypes and cultural schemas associated with them, such as the distinctions, schemas, and stereotypes associated with sex and gender. Languages distinguish between sex and gender categories, signal them, and mirror the common cultural beliefs about who women and men are and should be (Ibrahim 1973, Parks and Roberton 1998, Stahlberg et al. 2007, Wolff and Holmes 2011). One of the ways in which gender is embodied in languages is via the use of grammatical gender.

Languages vary by their usage of grammatical gender. In gendered languages like French, Spanish, German, and Hebrew, parts of speech – pronouns, nouns, adjectives, and verbs – all have feminine and masculine forms. In such languages, generic use of the masculine form for both females and males is more prominent than in gender-neutral ones (Ibrahim 1973, Stahlberg et al. 2007). Moreover, gendered languages that do not have a third gender-neutral form tend to use the masculine form as an unmarked form to refer to an unknown or mixed-gender audience.

Research suggests that gendered languages are associated with gender inequality. Studies have shown that the countries in which gendered languages are spoken tend to be associated with greater sex inequality in labor, credit, and education than are countries whose languages feature gender-neutral grammatical systems (Davis and Reynolds 2018, Gay et al. 2013, Mavisakalyan 2015, Prewitt-Freilino et al. 2012). Other studies demonstrate experimentally that addressing people in languages with grammatical gender affects their attitudes, perceptions, and motivations. In one study, answering a survey about sexist attitudes in a language with grammatical gender (French or Spanish) was found to increase the reported sexist attitudes, compared to answering the same survey in English (Wasserman and Weseley 2009). In a related experimental study, addressing women in the masculine in an academic-motivation questionnaire generated lower reports of task value and intrinsic goal orientation compared with addressing women in the gender-neutral form of the language (Vainapel et al. 2015). While the literature has established a cross-language correlation between gendered languages and gender inequality, and has shown the effects of gendered language on *attitudes*, it has not yet provided evidence for the causal effects of gendered language on women and men's actual *performance*.

Does the grammatical practice of addressing both women and men in the masculine affect their performance in math and other tasks? This question is important because the practice is prominent in gendered languages and might have significant implications for women's underperformance in math and for gender inequality more broadly.

Addressing women in the masculine (or men in the feminine) forms may affect their performance because their sense of alienation is activated. Women addressed in the masculine form may find it harder to view themselves as the prototypical test taker than they would when addressed in the feminine. These perceptions may lead women to believe less in their ability to succeed and therefore to decrease the levels of effort, concentration, and performance they invest in the task. In a different context, it was shown that when job advertisements were constructed to use wordings that were semantically more masculine (such as "leader," "competitive," and "dominant") rather than feminine (such as "supportive," "understanding," and "interpersonal"), women participants found these jobs less appealing and reported feeling less likely to be fitting for them (Gaucher et al. 2011).

Addressing women in the masculine (and men in the feminine) may also make gender and gender stereotypes more salient. Indeed, stereotypes and cultural beliefs about gender are easily activated; once activated, they may affect the quality of women's performance (Spencer et al. 1999, Steele and Aronson 1995). In general, people hold highly defined and consensual stereotypes about who women and men are. People tend to associate communion (expressive traits) with women and agency and competence (instrumental traits) with men (Cuddy et al. 2007, Diekman and Eagly 2000, Fiske et al. 2002, Glick et al. 2004). In addition to the general stereotypes and cultural beliefs about gender, people tend to hold task-specific stereotypes and cultural beliefs about the ways in which women and men perform in specific tasks (Berger et al. 1977, Ridgeway 2011, Wagner and Berger 2002). Men, for example, are perceived to be better than women in math and science, although the evidence that supports these perceptions is relatively weak (Correll 2001, Hyde 2005, Nosek et al. 2002).

Stereotypes and cultural beliefs about women's lower ability in math and science generate a 'stereotype threat' and negatively affect girls' and women's actual performance, as well as their willingness to attribute their success to their abilities rather than to their efforts (Correll 2001, Spencer et al. 1999, Steele and Aronson 1995). The tendency to preform worse when negative

stereotypes are salient has been shown to be related to the anxiety, distraction, and decreased efforts (Stone 2002, Stone et al. 1999)[1] caused by the perceived lower expectations by others (Bosson and Pinel 2004, Spencer et al. 1999) and lower expectations of oneself (Cadinu et al. 2003). Indeed, stereotypes about women and math are so easily activated that merely asking women to indicate their gender before taking a math test negatively affected their performance (Danaher and Crandall 2008).

It follows from the literature that addressing women in the masculine would decrease their efforts and negatively impact their performance in math, whereas addressing men in the feminine would result in only a minor (if any) effect on men's efforts and performance. This is because in addition to alienating and confusing men, addressing men in the feminine might activate the general gender stereotypes about men being more capable and competent than women (tested in the present research in Study 1). It also follows that addressing women in the masculine would improve their performance on stereotypically female-type tasks (Study 2). This is because unlike with male-type tasks (like math), female-type tasks (like reading comprehension tests about emotions) would activate *positive* stereotypes about women's performance. Finally, it follows that addressing women in the masculine in relatively gender-neutral tasks would alienate women and activate the general stereotypes about men's superior competence and thus negatively affect women's performance (Studies 3 and 4).

Current Research

The current research examined whether addressing women and men in the feminine and in the masculine affects their performance, using a large representative sample of Hebrew-speaking participants living in Israel. Notably, Hebrew grammar is highly gendered. It assigns a gender to most nouns, verbs, and pronouns. The grammatical rule and the prominent practice is to address an anonymous person in the masculine singular form and all audiences in the masculine plural form, unless the audience is all female. Thus, exam instructions are often written in the masculine singular form.[2]

---

[1] Studies have also shown that when negative stereotypes become salient, people are motivated to disconfirm them (the 'mere effort account'). This motivation may sometimes compensate for the negative effects of threat on performance (Harkins 2006, Jamieson and Harkins 2007).

[2] In a 2002 directive confronting the issue of gender equality in education, the Israeli Ministry of Education instructed educators to use address boys and girls either in the masculine plural form, or using a slash (2nd person feminine / 2nd person masculine). http://cms.education.gov.il/educationcms/applications/mankal/arc/sc4dk9_4_4.htm

In all studies, we used the masculine singular and the feminine singular forms of the language to address women and men participants taking a test (by varying the form of the verb "answer" in the instructions). To eliminate peer group effects that may interact with the effects of being addressed in the masculine or in the feminine, participants took the tests online and not in a lab together with other participants. In particular, we did not want women participants to wonder whether men are being addressed in the feminine or in the masculine.

STUDY 1

In Study 1, we tested whether addressing women and men in the masculine compared to the feminine affects their performance in math. We also tested whether the form of gender address affects participants' efforts (as measured by the time they spend on the test) and on their implicit and explicit biases and beliefs. Finally, we tested whether the effect of being addressed in the masculine compared to the feminine was stronger for native speakers compared to immigrants.

Method

Participants. A survey company that specializes in Internet-based surveys and uses a representative panel of the Israeli population recruited 963 participants (51% female). The company randomly selected a sample of adult (18+) Hebrew speakers from its panel. Of the sample, 763 participants (82%) were born in Israel and the other 204 had immigrated to Israel. We focused our analyses on native Hebrew speakers and included people who immigrated to Israel and acquired the Hebrew language later in life as a particular sub-set, only when indicated.

Stimuli and task. Participants were asked to complete a math test addressed either in the masculine or in the feminine. They were asked to answer six math questions that were taken from the qualitative reasoning section of a previous Psychometric Entrance Exam (an SAT-type exam). The questions were designed to assess the ability of test takers to use numbers and mathematical concepts to solve quantitative problems (see Appendix 2). We did not limit the time participants could spend on answering the questions.

Procedure and Measures. The experiment was carried out online. Of the initial sample, 490 participants were assigned to the masculine condition and 473 participants were assigned to

the feminine condition[3] and were asked to complete the math test. Upon completion, participants were asked to fill out an implicit association test (Nosek et al. 2005),[4] followed by an explicit bias questionnaire addressing their attitudes and beliefs about the associations of women and men with the natural sciences and the liberal arts. Finally, participants were asked to report their own attitudes toward science. We reordered the time participants spent on the math test. The survey company provided information about the demographic characteristics of participants (including their age at the time of immigration).

Results

Descriptive Statistics. Table S1 in the Appendix presents the descriptive statistics for the variables we used in the analysis for native Hebrew speakers and immigrants. The average score of participants in the math test was 63/100. The average time they spent was 1.1 minutes per question. Whereas all participants completed the math test, only 77% of them completed the full questionnaire, which included the Implicit Association Test and the attitudes and explicit biases survey.

Performance. Figure 1 graphically presents native Hebrew speakers' mean scores in math by gender and experimental condition.

(Figure 1 about here)

[3] Participants were recruited in two waves, each nationally representative of the adult population in Israel**.** In the first wave, all men were assigned to the "masculine" condition and all women to the "feminine" condition. In the second wave, all women were assigned to the "masculine" condition and all men to the "feminine" condition. Participants were assigned to one of the two experimental conditions by waves because of a technical difficulty. However, each of the waves was nationally representative, and the two samples were balanced.

[4] The Gender-Science Implicit Association Test is part of Harvard's Implicit Project (https://implicit.harvard.edu/); it is a computer-based test that is designed to assess participants' implicit biases about gender and science. It measures the time it takes participants to associate words that represent men and women (e.g., "man," "son," "boy," vs. "woman," "aunt," "wife") to words that represent the sciences and the liberal arts (e.g., "Astronomy," "Math," "Chemistry," vs. "History," "Art," "Music"). The implicit bias score measures the difference between the average time it takes a participant to correctly associate women with the sciences and men with liberal arts (the "hard association") and the average time it takes her/him to correctly associate women with liberal arts and men with the sciences (the "easy association").

On average, women scored lower than men. Whereas the average score of women was 58, the average score of men was 69 ($p<0.001$, N=759). Addressing participants in the feminine increased women's scores. Whereas the average score of women who were addressed in the masculine was 55, the average score of women who were addressed in the feminine was 60 ($p=0.059$, N=383). While the average score of men who were addressed in the masculine was 70, the average score of men who were addressed in the feminine was 64 ($p<0.05$, N=376).

To better understand the results, we conducted a multivariate analysis. We ran OLS regression models predicting the effect of addressing participants in the feminine form on their test performance. Results are presented in Table S2 in the Appendix. When addressed in the masculine, women received scores that were lower by 15 percentage points than the scores men received (Model 1; $p<0.01$, N=759). When addressed in the feminine, women scored higher by five percentage points compared to when addressed in the masculine (or a 0.17 standard deviation increase; Model 1; $p<0.05$, N=759). The effect of addressing men in the masculine compared to the feminine was smaller: when addressed in the feminine, men scored lower by seven percentage points than when addressed in the masculine (Model 1; $p <0.1$, N=759). In fact, when both women and men were addressed in the feminine, the gender gap in math became statistically insignificant. When women were addressed in the feminine and men in the masculine, women scored only ten percentage points lower on average than men did (Model 1; $p<0.001$, N=759). When the demographic characteristics of participants were accounted for, the effects remained similar (Model 2).

Language Proficiency. We predicted that the effects would be weaker for people who acquired the Hebrew language later in life and are therefore less proficient in it. To test this prediction, we ran an OLS regression model on the full sample of participants (native and immigrants). We used the age at immigration to Israel as a proxy for language proficiency (Stevens 1999) and included it in the model together with an interaction term between it and the form of address (masculine or feminine). Being a year older when immigrating to Israel decreased the effect of being addressed in the feminine for women by one percentage point (Table 1, Model 3; $p<0.1$, N=926).

Efforts. The time participants spent on the math test reflects their efforts and motivation to succeed. Indeed, the time participants spent on the math test correlated with the score they

attained (a correlation of 0.24). We therefore tested whether the time participants spent on their math test was affected by whether they were addressed in the feminine or in the masculine. In Figure 2, we graphically present participants' mean time invested in the math test, by gender and experimental condition (for native Hebrew speakers only). Addressing participants in the feminine form increased the time women spent on their math test and decreased the time men did. In fact, when comparing women and men who were both addressed in the masculine, we find that women spent significantly less time on their math test relative to men ($p<0.05$). However, when comparing women and men who were both addressed in the feminine, we find that women spent significantly more time on their math test relative to men ($p<0.10$). There was no statistical difference in the time women and men spent on their math test when women were addressed in the feminine and men were addressed in the masculine.[5]

(Figure 2 about here)

Similar results were obtained in OLS regression models predicting the time participants spent on their math test (Appendix Table S3). When addressed in the masculine, women spent on average 1.87 minutes less on the test than men did (Model 1, $p<0.05$, N=688). When addressed in the feminine, women spent 1.18 minutes more on the math test compared to women addressed in the masculine (an increase of 0.3 standard deviations; Model 1, $p<0.01$, N=688). When men were addressed in the feminine, they spend 1.44 minutes less on the math test compared to men who were addressed in the masculine ($p<0.1$, N=688). Accounting for participants' demographic characteristics left the effects practically unchanged (Model 3).

[5] We do not report the results of a mediation model predicting the score in math by the gendered form of address and by the time participants spent on the questions, because the time spent is an outcome in the study.

Prototypical Test Takers and Sense of Alienation. After completing the math test, participants were asked to mark the extent to which they agreed (on a scale of 1 to 7) with the statement that 'science is for men.' As predicted, when addressed in the feminine, participants reported agreeing less with the statement that 'science is for men'; when addressed in the masculine, the average response of participants was 4.98, whereas when addressed in the feminine, the mean was 4.78 ($p<0.05$). No statistical significant differences were found between the effects on female and male participants. Participants were also asked to note the extent to which they agreed with the statement that 'arts and humanities are for women.' No statistically significant differences were found between the responses of participants who were addressed in the masculine and the responses of those who were addressed in the feminine. Similar results were found in the regression models (Table S4 in the Appendix). These finding suggest that gendered languages affect perceptions of who the prototypical test taker is and the sense of alienation. Specifically, when women and men are addressed in the feminine in an SAT-type math test, they view women as more representative of the prototypical scientist, compared to when addressed in the masculine.

STUDY 2

In study 2 we tested whether addressing women and men in the masculine compared to the feminine affected people's scores in a reading comprehension test about empathy. Here, unlike in Study 1, we chose an assignment that was stereotypically associated with women's strengths and men's weaknesses. If indeed addressing women in the masculine makes gender stereotypes and cultural beliefs more salient, then we would expect women to perform *better* in such an assignment when addressed in the masculine compared to when addressed in the feminine. We therefore predicted that men would perform worse in this task when addressed in the feminine compared to the masculine, but that the effect would be weaker than it is for women. Because the stereotypes about men's general competency are strong and ubiquitous, we predicted that there would only be a weak effect on men of being reminded of their weakness in performing empathy-related tasks.

Method

Participants. The survey company recruited 680 participants (49% female). The sample was randomly selected from the company's panel to be representative of the adult (18+) native Hebrew speaking population (excluding immigrants).

Stimuli and task. Participants were tasked with completing a reading comprehension test about empathy. They were addressed in the masculine or in the feminine and asked to answer six reading comprehension questions (see Appendix 2). There was no time limit for completing the test.

Procedure and Measures. The experiment was carried out online. Participants were randomly assigned to being addressed in the masculine or in the feminine (altogether, 333 participants were addressed in the masculine and 347 in the feminine) and completed the reading comprehension test. The survey company recorded and provided the time participants spent on the test, as well as information about the demographic characteristics of participants.

Results

Positive Stereotypes. In the empathy reading comprehension test – unlike the math test in Study 1 – women performed better than men. Whereas the average score of women was 53.7, the average score of men was 50.2 ($p=0.058$; $N=680$).

(Figure 3 about here)

In Figure 3, we report the scores of women and men in the empathy reading comprehension test, broken down into the test conditions of being addressed in the masculine or in the feminine.

As predicted, being addressed in the masculine improved women's performance. When addressed in the masculine, women's average score was 56.6; whereas when addressed in the feminine, women's average score was only 51 ($p<0.05$). The average score of men when addressed in the masculine was 50.9, and it declined to only 49.5 when addressed in the feminine. This gap was not statistically significant.

We have shown here that when the gender stereotypes that are relevant to the task are positive, the performance of women improved when addressed in the masculine compared to when addressed in the feminine. This finding suggests that being addressed in the masculine evokes gender stereotypes and cultural beliefs and thus affects performance. Note that if the effects of being addressed in the masculine were merely the results of being indirectly addressed or of not

viewing oneself as the prototypical test taker, then we would expect women to perform *worse* when addressed in the masculine, regardless of the content of the gender stereotypes evoked.

STUDY 3

In Study 3, we tested whether addressing women and men in the masculine compared to the feminine affects people's scores in a reading comprehension test about playwrights and philosophy. Here, unlike in Study 1 or 2, we chose an assignment that is associated with neither women's strengths nor weaknesses. We sought to investigate the effects of being addressed in the masculine on women and men when the assignment was relatively gender neutral, so that gender-specific stereotypes are irrelevant. Addressing participants in the masculine compared to the feminine in gender-neutral tasks may evoke perceptions of prototypical test takers and evoke general gender stereotypes. Therefore, we predicted that women would perform worse when addressed in the masculine, compared to when addressed in the feminine, even when tasks are relatively gender neutral. For men, who are commonly viewed as prototypical test takers and as more competent overall, we predicted the effects of being addressed in the masculine to be positive but marginal (if at all present).

Method

Participants. The survey company recruited 780 participants (53% female). The sample was randomly selected from the company's panel to be representative of the adult (18+) population of native Hebrew speakers.

Stimuli and task. Participants were asked to complete a reading comprehension test about playwrights and philosophy. They were asked (in the imperative 2nd person feminine or masculine) to read and answer six reading comprehension questions about a short text. The text and questions were taken from the qualitative reasoning sections of a previous Psychometric Entrance Exam, published on the official website of the Israeli National Center for Testing and Evaluation (see Appendix 2). Again, no time limit was given for completing the test.

Procedure and Measures. The experiment was carried out online. Participants were randomly assigned to being addressed in the masculine or in the feminine (altogether, 395 participants were addressed in the masculine and 385 in the feminine) and were asked to complete

the reading comprehension test. The time participants spent on the test was recorded. The survey company provided information about the demographic characteristics of participants.)

Results

***Gender-Neutral* Reading Comprehension.** On average, women preformed worse than men on the reading comprehension test about playwrights and philosophy. Whereas men scored on average 54, women scored only 48 ($p<0.01$). As predicted, being addressed in the masculine resulted in women's lower performance compered to being addressed in the feminine. When addressed in the masculine, women's average score was 44.2, whereas when addressed in the feminine, it was 49.4 ($p<0.1$). The average score of men when addressed in the masculine was 55, it declined to 51.5 when addressed in the feminine. The gap in the men's performance was not statistically significant.

STUDY 4

In Study 4, we tested whether addressing women and men in the masculine compared to the feminine affects people's scores in a word completion task. This study was designed to provide additional evidence on the effects of being addressed in the masculine compared to the feminine when the assignment was relatively gender neutral. Our predictions were identical to those in Study 3: we predicted that women would perform worse when addressed in the masculine, compared to when addressed in the feminine, and that the effects of being addressed in the masculine on men would be positive but marginal (if at all present).

Method

Participants. The survey company recruited 674 participants (50% female). The sample was randomly selected from the company's panel to be representative of the adult (18+) native Hebrew speaking population.

Stimuli and task. Participants were told that they had one minute to write as many words as they could. The words had to begin with consecutive letters of the alphabet (a word that begins with the letter "a," followed by a word that begins with the letter "b," etc.). They were told that they would be scored on accuracy of alphabetical order and the total length of all the words they provide (see Appendix 2).

Procedure and Measures. The experiment was carried out online. Participants were randomly assigned to being addressed in the masculine or in the feminine (altogether, 343 participants were addressed in the masculine and 331 in the feminine) and asked to complete the word completion task. The survey company provided information about participants' demographic characteristics.

Results

***Gender-Neutral Word Completion*.** On average, *women performed better than men* in this assignment. *Whereas the average total number of letters women provided* was 42.8, *the average number of letters that men provided* was *only* 39.6 ($p<0.05$). As predicted, being addressed in the masculine improved women's performance in this relatively gender-neutral task. When addressed in the masculine, the average number of letters women provided was 40.8, whereas when addressed in the feminine it was 45 ($p<0.05$). The average number of letters men provided was 38.2 when addressed in the masculine and 41.51 when addressed in the feminine. This gap was not statistically significant ($p>0.10$).

General Discussion

Addressing women in the masculine – a prominent practice in gendered languages – negatively affects women's performance. In four studies, we explored the effects of addressing women and men in the masculine compared to the feminine on their performance in math and other tasks. Because the samples for the experiments were both representative and uniquely large (N=~400 per condition), this study offers both the internal validity that characterizes experiments conducted in controlled settings and the external validity that characterizes studies of large, random samples of the population.

When addressed in the masculine, women substantially decreased their performance in math. In fact, when women were addressed in the feminine and men in the masculine, the

performance gap between women and men was reduced by a third, compared to when both women and men were addressed in the masculine. These effects were stronger among participants who acquired the Hebrew language early in childhood rather than later in life. This finding suggests that it is the extent of language proficiency that generates one's sensitivity to being addressed in the masculine or in the feminine. The effect of addressing men in the feminine on their performance in math was negative, but only marginally significant. When women were addressed in the masculine, their efforts (in terms of time spent on the math test) decreased and they reported feeling that 'science is for men' more than when addressed in the feminine (Study 1). Addressing women in the masculine actually improved women's performance on stereotypically female-type tasks (Study 2). Finally, with stereotypically gender-neutral tasks, addressing women in the masculine decreased women's performance, suggesting that the masculine address alienates women from the task and activates general stereotypes about men's greater competence (Studies 3 and 4).

The results demonstrate the powerful role of language in activating stereotypes and cultural beliefs, as well as the strong impact of stereotypes and cultural beliefs in affecting actual performance. If indeed people are sensitive to the salience of gender in languages, then all the other ways by which gender becomes present may have similar effects on performance. The results further highlight the potent effects of everyday social interactions in maintaining and reinforcing systems of inequality: women and men use gendered languages constantly. When they do so, stereotypes and cultural beliefs about gender frame their performance and thus further legitimize and sustain the unequal gender system.

Our findings have some policy implications regarding the ways in which women and men ought to be addressed in exams and in other learning environments. Recall (Study 1) that women's math scores were on average higher by 0.17 standard deviations when they were addressed in the masculine compared to the feminine. These results suggest that the form of address can have significant and meaningful implications for the gender gap in achievement.[6] However, further research needs to be conducted in order to carefully craft accurate policy recommendation. The effects of activating stereotypes and a sense of alienation may vary by context and environment. Studies have shown that in mixed-sex groups where gender is more salient, sex stereotypes are

[6] Currently, the Israeli psychometric exam addresses participants in the masculine-plural form; see, e.g., the latest exam form: https://nite.org.il/files/psych/psychometric_sept_2018.pdf

activated more easily (Cota and Dion 1986, Ridgeway 2011). As a result, women perform worse in math in mixed-sex settings (Inzlicht and Ben-Zeev 2000). Women may therefore be more vulnerable to the negative effects of being addressed in the masculine when taking the tests in mixed-sex settings compared to when taking the test surrounded by women only or when alone by themselves. On the other hand, in a mixed-gender classroom, women may feel discomfort with being addressed in the feminine, if they assume or wonder whether men in the classroom are also addressed in the feminine, which is a highly uncommon practice. Thus, it is hard to predict whether women would perform better in mixed-sex settings when addressed in the feminine or in the masculine.

Note also that the effects on men of being addressed in the feminine are negative in some contexts, although smaller than the effects on women and marginally significant. It follows, therefore, that in order to improve women's performance while not harming men's, women and men should possibly be granted the right to choose whether they wish to be addressed in the feminine or in the masculine in exams. Because the stereotypes and cultural beliefs embodied in the language may affect women's learning experiences and not only their performance in exams, the language spoken in classrooms should also be modified to include feminine generics and neutral forms.

Naturally, modifying the languages of exams, and even the language spoken in classrooms, would not altogether eliminate gender gaps in math and reading comprehension performance. Gender inequality is persistent and over-determined: it is consistently and simultaneously generated and maintained in multiple spheres of life and spanning different levels of analysis (Ridgeway 2011, Ridgeway and Correll 2004, Ridgeway and Smith-Lovin 1999, Risman 1998). Yet, tackling such inequality within each realm or level of analysis is important in generating the possibility for change. The ability to minimize gender gaps in standardized testing results merely by changing the gender of the language used is indicative of the power of language in perpetuating inequalities.

Acknowledgements

Rabia Agrabia and Michal Danielli provided excellent research assistance. We are grateful to Catherine Albiston, Ronen Avraham, Ian Ayers, Michal Barak, Lauren Edelman, Nadav Levy,

Hadas Mandel, Ariel Porat, Haggai Porat, Cecilia Ridgeway, the participants of the Berkeley Institute for Jewish Law and Israel Studies Seminar, the Tel Aviv University Law and Economics Workshop, and the Tel Aviv University Gender Studies Seminar, and two anonymous referees for their insightful comments. The Israel Science Foundation provided financial support for this research (grant 483/15).

References

Berger, J., Fisek, H., Norman, R., & Zelditch, M. (1977). *Status Characteristics and Social Interaction*. New York, NY: Elsevier.

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology, 40,* 247–255.

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology, 33,* 267–285.

Correll, S. J., (2001). Gender and the Career Choice Process: The Role of Biased Self Assessment. *American Journal of Sociology, 106,* 1691–1730.

Cota, A. A. & Dion, K. L. (1986). Salience of gender and sex composition of ad hoc groups: An experimental test of distinctiveness theory. *Journal of Personality and Social Psychology, 50(4),* 770-776.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2007). The BIAS Map: Behaviors From Intergroup Affect and Stereotypes. *Journal of Personality and Social Psychology, 92,* 631-48.

Danaher, K. & Crandall, C. S. (2008). Stereotype Threat in Applied Settings Re-Examined. *Journal of Applied Social Psychology, 38,* 1639–1655.

Davis, L. & Reynolds, M. (2018). Gendered language and the educational gender gap. *Economics Letters, 168,* 46-48.

Diekman, A. B. & Eagly. A. H. (2000). Stereotypes as Dynamic Constructs: Women and Men of the Past, Present, and Future. *Personality and Social Psychology Bulletin, 26,* 1171-88.

Hyde, J. S. (2005). The Gender Similarities Hypothesis. *American Psychologist, 60,* 581- 92.

Fiske, S. T., Cuddy A. J., Glick P., and Xu J. (2002). A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow From Perceived Status and Competition. *Journal of Personality and Social Psychology, 82*, 878-902.

Gaucher, D., Friesen, J. & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology, 101(1),* 109-128.

Gay, V., Santacreu-Vasut, E. & E., Shoham, A. (2013). The grammatical origins of gender roles. In: *Working papers Berkeley Economic History Laboratory Paper Series*.

Glick, P., Lameiras M., Fiske S. T., Eckes T., Masser B., Volpato C, Manganelli A. M., Jolynn C. X., Pek, J. C. X., Huang L., Sakalli-Ugurlu N., Rodriguez Castro, Y., D'Avila Pereira M. L., Willemsen T. M., Brunner A., Six-Materna I. & Wells, R. (2004). Bad But Bold: Ambivalent Attitudes Toward Men Predict Gender Inequality in 16 Nations. *Journal of Personality and Social Psychology, 86,* 713–28.

Harkins, S. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology, 91,* 436 – 455.

Ibrahim, M. H. (1973). *Grammatical gender: Its origin and development*. The Hague, Netherlands: Mouton.

Inzlicht, M. & Ben-Zeev, T. (2000). A Threatening Intellectual Environment: Why Females Are Susceptible to Experiencing Problem-Solving Deficits in the Presence of Males. *Psychological Science, 11(5),* 365 – 371.

Jamieson J. P. & Harkins, G. S. (2007). Mere Effort and Stereotype Threat Performance Effects. *Journal of personality and social psychology. 93*. 544-64.

Mavisakalyan, A (2015). Gender in language and gender in employment. *Oxford Development Studies, 43(4),* 403–424

Nosek B. A., Greenwald A. G. & Banaji M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*, 166–180.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math= male, me= female, therefore math≠ me. *Journal of personality and social psychology, 83(1),* 44.

Parks, J. B., & Roberton, M. A. (1998). Influence of age, gender, and context on attitudes toward sexist/nonsexist language: Is sport a special case? *Sex Roles, 38,* 477– 494.

Prewitt-Freilino, J., Caswell, T. A., & Laakso, E. K. (2012). The gendering of language: A comparison of gender equality in countries with gendered, natural gender, and genderless languages. *Sex Roles, 66,* 268 –281.

Ridgeway, C.L. (2011). *Framed by Gender: How Gender Inequality Persists in the Modern World*. Oxford University Press.

Ridgeway, C. L.& Correll, S. J. (2004). Unpacking the Gender System: A Theoretical Perspective on Cultural Beliefs and Social Relations. *Gender and Society*, *18*:510-31.

Ridgeway, C. L. and L. Smith-Lovin (1999) The Gender System and Interaction. Annual Review of Sociology, 25:191-216.

Risman, B. J. (1998) Gender Vertigo: American Families in Transition. New Haven, CT: Yale University Press.

Stahlberg, D., Braun, F., Irmen, L., & Sczesny, S. (2007). Representation of the sexes in language. In K. Fiedler (Ed.), Social communication: Frontiers of social psychology (pp. 163– 187). New York, NY: Psychology Press.

Steele, C. M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797-811.

Stevens, G. (1999). Age at immigration and second language proficiency among foreign-born adults. Language in Society, 28(4), 555–578.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. Journal of Experimental Social Psychology, 35, 4 –28.

Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self handicapping in White athletes. Personality and Social Psychology Bulletin, 28, 1667–1678.

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. Journal of Personality and Social Psychology, 77, 1213– 1227.
The Academy of the Hebrew Language. (2010). Q&A: Relating both to man and woman. Retrieved from http://hebrew-academy.huji.ac.il [in Hebrew].

Vainapel, S., Shamir, O.Y., Tenenbaum, Y., Gilam, G. (2015). The Dark Side of Gendered Language: The Masculine-Generic Form as Cause of Self-Report Bias. Psychological Assessment, 27, 1513–1519.

Wagner, David G. and Joseph Berger. 2002. "The Evolution of Expectation States Theories." Pp. 41-78 in Contemporary Sociological Theories, edited by M. Zelditch Jr. and J. Berger. New York, NY: Rowman & Littlefield.

Wasserman, B. D., & Weseley, A. J. (2009). ¿Qué? quoi? Do languageswith grammatical gender promote sexist attitudes? Sex Roles, 61, 634 –643.

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. Wiley Interdisciplinary Reviews: Cognitive Science, 2, 253–265.