INSTITUTE OF ECONOMICS
Centre for Economic and Regional Studies

KRTK KTI

Közgazdaság- és Regionális Tudományi Kutatóközpont
KÖZGAZDASÁG-TUDOMÁNYI INTÉZET

CERS-IE WORKING PAPERS | KRTK-KTI MŰHELYTANULMÁNYOK

The Role of the Gender Composition of Performance Feedback on Peers in Shaping Persistence and Performance

SÁNDOR KATONA - ANNA LOVÁSZ

CERS-IE WP – 2021/5

January 2021

https://www.mtakti.hu/wp-content/uploads/2021/01/CERSIEWP202105.pdf

ABSTRACT

We study the impact of the gender composition of a scoreboard on the persistence and performance of players in an online game. Players were randomly selected into eight groups, defined along two dimensions: they saw high or average scores on a scoreboard (score level), and within each of these, they saw either 3 male, 2 male and one female, 1 male and 2 female, or 3 female names associated with the scores (gender composition). Based on 1140 participants, we find that males are generally less responsive to performance information on other participants. Compared to the baseline of all male names on the scoreboard, females play fewer games when they see only female names, but more games when they see mixed gender names with high scores. Their performance (best score) increases significantly when they see at least one female name and high scores. This result is in line with the importance of female-specific reference points – or role models - in encouraging females' participation and higher performance in competitive settings. It supports the use of policies aimed at providing these, such as the introduction of female role models and the public acknowledgement of high performing females.

JEL codes: I20, J16, J24, M54

Keywords: Gender Gaps, Competition, Performance Feedback, Gender Composition, Reference Point

Sandor Katona, Centre for Economic and Regional Studies, Toth Kalman u. 4. Budapest, 1097 Hungary and Eotvos Lorand University, Pazmany Peter setany 1/a, Budapest, 1117 Hungary

Anna Lovasz, Centre for Economic and Regional Studies, Toth Kalman u. 4. Budapest, 1097 Hungary and University of Washington Tacoma, 1900 Commerce Street, Tacoma, WA 98402-3100, USA

A másokról adott teljesítmény információ nemek szerinti összetételének hatása a kitartásra és teljesítményre

KATONA SÁNDOR - LOVÁSZ ANNA

ÖSSZEFOGLALÓ

Egy online játék során vizsgáljuk az eredménytábla nemek szerinti összetételének hatását a játékosok kitartására és teljesítményére. A játékosok véletlenszerűen nyolc kezelési csoportba kerültek, melyeket két dimenzió szerint definiáltunk: átlagos vagy magas pontszámokat mutattunk (teljesítményszint), és ezeken belül 3 férfi, 3 női, 2 férfi és egy női, vagy 1 férfi és 2 női név szerepelt az eredménytáblán (nemek szerinti összetétel). 1140 játékos adatai alapján azt találtuk, hogy a férfiak általában kevésbé érzékenyek a más játékosok teljesítményéről közölt eredményekre. A csak férfi nevet tartalmazó eredménytáblás csoporthoz viszonyítva a nők kevesebb játékot játszottak amikor csak női neveket láttak, viszont többet játszottak amikor kevert nemű táblát láttak. A nők teljesítménye (pontszáma) szignifikánsan (17-25 százalékkal) magasabb volt, amikor legalább egy magas pontszámú női nevet láttak az eredménytáblán. Ez az eredmény arra utal, hogy a nők inkább a saját nemükre vonatkozó teljesítményinformáció alapján ítélik meg a saját várható teljesítményüket. Szakpolitikai szempontból a visszajelzések tervezésének ezt fontos figyelembe venni, mert a magas teljesítményű nőkről látott információk növelhetik a nők részvételét és teljesítményét a versenyhelyzetekben.

JEL: I20, J16, J24, M54

Kulcsszavak: Nemek közötti eltérések, verseny, teljesitményvisszajelzések, nemek szerinti összetétel, referenciapont

The Role of the Gender Composition of Performance Feedback on Peers in Shaping Persistence and Performance

Sandor Katona[(a)(b)] and Anna Lovasz[*,(a)(c)]

[(a)] *Centre for Economic and Regional Studies, Toth Kalman u. 4. Budapest, 1097 Hungary*
[(b)] *Eotvos Lorand University, Pazmany Peter setany 1/a, 1117 Hungary*
[(c)] *University of Washington Tacoma, 1900 Commerce Street, Tacoma, WA 98402-3100, USA*

January 24, 2021

Abstract

We study the impact of the gender composition of a scoreboard on the persistence and performance of players in an online game. Players were randomly selected into eight groups, defined along two dimensions: they saw high or average scores on a scoreboard (*score level*), and within each of these, they saw either 3 male, 2 male and one female, 1 male and 2 female, or 3 female names associated with the scores (*gender composition*). Based on 1140 participants, we find that males are generally less responsive to performance information on other participants. Compared to the baseline of all male names on the scoreboard, females play fewer games when they see only female names, but more games when they see mixed gender names with high scores. Their performance (best score) increases significantly (17-25 percent) when they see at least one female name and high scores. This result is in line with the importance of female-specific reference points – or role models - in encouraging females' participation and higher performance in competitive settings. It supports the use of policies aimed at providing these, such as the introduction of female role models and the public acknowledgement of high performing females.

Keywords: Gender Gaps, Competition, Performance Feedback, Gender Composition, Reference Point

JEL codes: I20, J16, J24, M54

Acknowledgements: The research was supported by grant FK 124658 of the National Research, Development and Innovation Office of Hungary. The experiment is registered under AEA RCT registry ID #AEARCTR-0003966. The project does not have IRB approval as our institution did not have an IRB. We would like to thank Andor Zöldesi for developing the game website. We also thank Daniel Horn, Janos Hubert Kiss, and commenters at Eotvos Lorand University for their valuable feedback.

**Corresponding author: University of Washington Tacoma, 1900 Commerce Street, Tacoma, WA, 98402 USA, email: plovi@uw.edu or lovasz.anna@krtk.hu*

1. Introduction

Competition-related attitudes and traits have been noted as key factors contributing to gender gaps in educational (*Buser et al 2014, Ors et al 2013*) and labor market outcomes (*Azmat and Petrongolo 2014, Bertrand 2011, Joensen and Nielsen 2009*). Studies have shown that women tend to choose competitive situations less often (*Gneezy et al 2009, Niederle and Vesterlund 2011, Healy and Pate 2011, Booth and Nolen 2012, Wozniak et al 2014*) and to perform worse in competitive settings (*Gneezy et al 2003, Cai et al. 2019, Cotton et al 2013*). Niederle and Vesterlund (2007) suggests that an individual's expectation of their own performance, self-confidence, and attitude towards competition are important determinants of the decision to compete. While these expectations and attitudes have been shown to differ by gender, they are also not necessarily constant: they depend on culture, age, biological traits, task, as well as the informational environment in which they are made. From a policy point of view, the important question is what can be done to decrease gender gaps in competition-related attitudes and performance expectations.

Studies have shown that the introduction of female role models can positively impact female expectations (*Del Carpio and Guadalupe 2018*), as can gender quotas through this role model effect (*Balafoutas and Sutter 2010, Niederle et al 2013*). Another line of research has focused on the provision of relative performance information (*Niederle and Vesterlund 2007, Ertac and Szentes 2010*), finding that such feedback can improve females' choices and outcomes. The content, source, and timing of feedback all appear to be key determinants of its effectiveness. In order to provide policy recommendations regarding best practices for feedback provision in educational and workplace settings, it is important to study how these factors impact performance expectations in more detail. In this study, we test one particular aspect of performance feedback content: the gender composition of information provided on other participants' performance. Policies related to role models and quotas rely on the idea that females' performance expectations depend on seeing examples of high-performing females. We examine whether the same gender-dependence exists in the response to performance feedback. We test how providing performance information on mixed gender or all female participants - versus only males - impacts females' and males' persistence and performance.

Previous empirical evidence suggests that such gender-dependence in the evaluation of performance feedback on peers may indeed exist. Baier et al (2018) found that when participants received information about the performance of other members of their group, males' confidence increased. When the performance information was linked to gender, they found a positive effect on the confidence of females as well, leading to a lower gender gap in confidence and in the choice to compete. This supports the idea that females take gender into consideration when weighting new performance information on others in their own performance expectations. They weight information that pertains to other females more highly compared to information that does not contain gender. For males, on the other hand, the gender of those whose performance information is seen is less important. In this study, we compare the impact of information that pertains to all males, mixed gender participants, or all females. This allows us to see whether females consider performance information on males as relevant as that on females. We further compare the impact to that of information on both genders, which provides gender-level relative performance information. We assess the impacts on males as well, to see whether they are impacted by information on females' performance differently than mixed or all-male information.

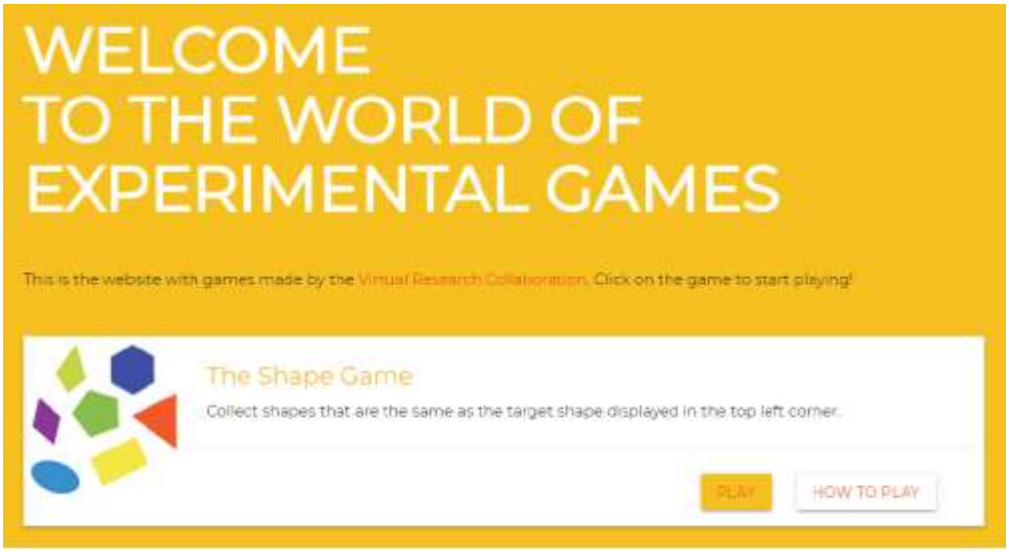
To do so, we use a simple online game with randomized treatment in the form of the content of a scoreboard that is shown before and during the game. Players are selected into eight treatment groups, in which we vary (1) the score levels shown, and (2) the three names attached to the scores shown. Players either see "Recent scores" representing average scores, or "Recent High Scores" representing scores in the top ten percentile. Within both of these categories, we vary the names between all male, mixed gender, and all female compositions. The names and scores shown are derived from real performance data from previous players, by finding players with recognizably male or female names, who achieved the same scores. We test the impacts on the number of games played (persistence), and the best score of each player (performance). Given this specification of treatment groups, we are able to compare the impact of the gender composition of the scoreboard when it shows average performers, and when it shows high performers. This allows us to gain insight into the motivation behind any observed behavioral responses, and see whether players are impacted differently by information on participation (by gender), or information on the highest performance levels (by gender).

The results, based on 1140 participants, are in line with the previous evidence, in that females appear to be more sensitive to the gender composition of the performance information provided. Males do not show a significant impact of any treatment on persistence or performance. Females, on the other hand, see significant performance gains (17-25 percent) and an increase in persistence when the gender composition of the scoreboard is mixed and high scores are shown. This supports that females do take the gender aspect of performance information on other participants into consideration, and are impacted positively when they see successful females. In a sense, this can also be seen as an example of the role model effect, which arises not through meeting such females personally, but simply by seeing public feedback on their performance. The content of information that reveals the gender of public performance feedback, and the gender composition of such feedback may therefore be key aspects of feedback design for educators and managers to consider in order to encourage gender equality in competitive settings.

2. Methodology

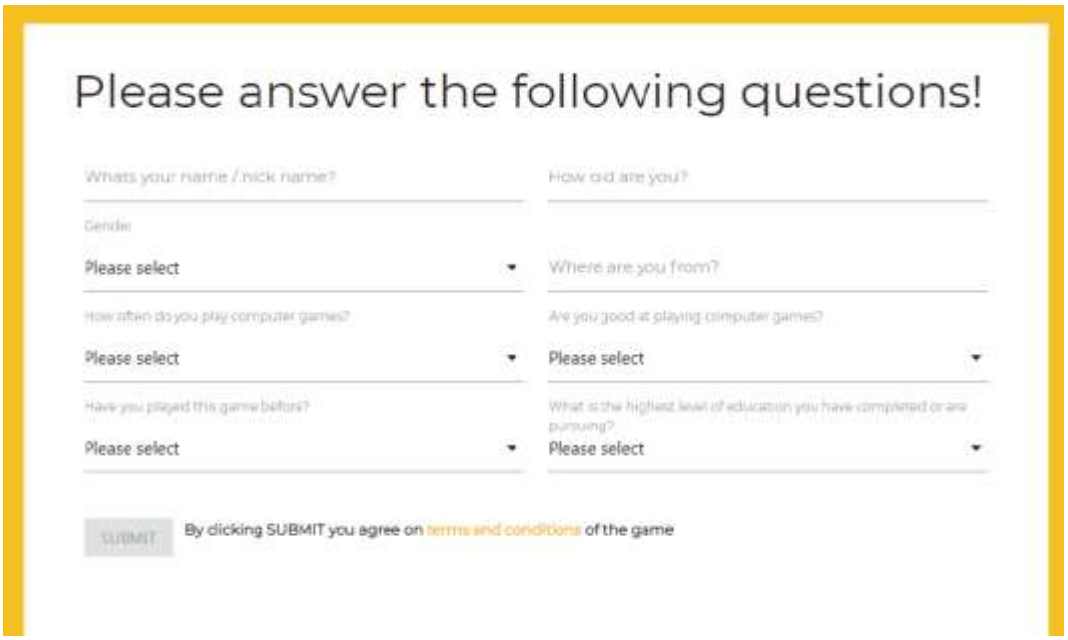
We used a simple online computer game, available on a website, to test the impact of scoreboard gender composition and score levels. We advertised the game on social media sites, with advertisements targeted towards the age group of 18-45-year-olds and four countries (Czech Republic, Hungary, Poland, Slovakia). When individuals clicked on the link in the advertisement, they went to a webpage on which they were given a short description of the game, including a demo video, as well as a description of the research purposes and terms and conditions (Figure 1).

Figure 1: Experimental website



A simple survey precedes the game (Figure 2), which asks for basic demographic information: gender, age, country, and level of education. The survey was designed to be quick and easy to fill out, asking for anonymous information similar to those requested on many typical game sites. Players are informed of the experimental purpose and the details of data collection and storage, but otherwise, the goal was to focus the player's attention on the game itself in order to observe real-life behavior in a natural game setting. The survey includes two further questions related to the individual's own experience with games (plays often, sometimes, never), and to their task-related confidence in playing online games (excellent, pretty good, ok, pretty bad, very bad). Additional data was automatically collected to account for whether the device the game is played on is a touchscreen or not and on device screen size, as these can impact performance.

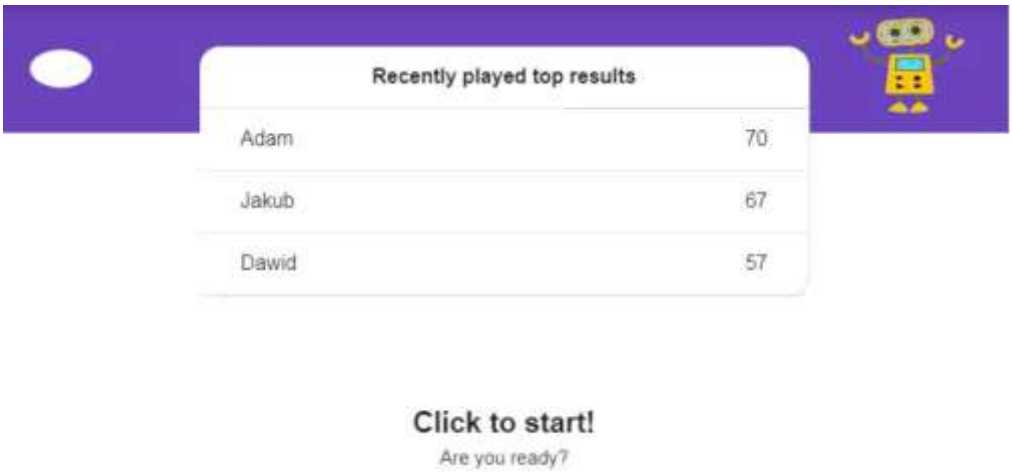
Figure 2: Pre-game survey



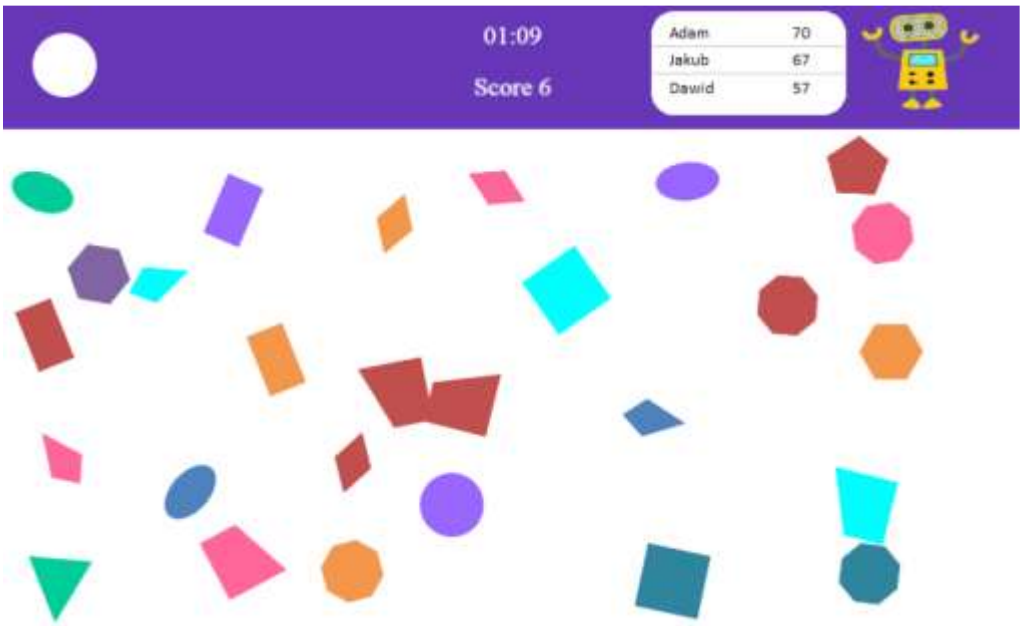
The game involved a visual perception task that requires both concentration and effort (Figure 3.b.). There were many geometric shapes moving around the screen. The task was to click on those matching the target shape that was displayed in the top left corner of the screen. Players had to find and click on all of the shapes that matched the target, then the target shape changed. Players received a point for every shape they clicked on that matched the target. The game took two minutes, and the goal was to score as many points as possible. Players could choose to play again as many times as they wished to.

Figure 3: The Shape Game

a.



b.



When players submitted the survey to start playing the game, they were randomly selected to be in one of the eight treatment groups described in Table 1. There were eight total groups

among which randomization took place, which differed in terms of the scoreboard they saw before (Figure 3.a) and during (Figure 3.b) the game, in addition to the remaining game time and their cumulative score. The scoreboard shown differed along two dimensions. Players saw either a "Recent Scores" scoreboard with scores ranging from 45-55 points, or they saw a "Recent High Scores" scoreboard with higher scores ranging from 73-87 points. All scores and names shown were from real life outcomes collected during a previous experiment using the game. The score levels were chosen to reflect average scores (Recent Scores specification) and scores in the top 10 percentile (Recent High Scores specification) based on the previous data.

Within the lower and higher scoreboard specifications, we distinguished four groups that differed in terms of the names shown, and specifically, in terms of the gender composition of the names shown. Each scoreboard included three names. The first group consisted of all male names, the second included 2 male names and 1 female name, the third included 1 male name and 2 female names, and the fourth group included 3 female names. Male and female names shown were chosen from the existing dataset, matching male and female players' outcomes that had achieved the same scores. We chose names that are internationally recognizable as male or female in gender. This specification allowed us to test the impact of seeing higher vs. lower scores and the gender composition of the names on the scoreboard separately, as well as their joint impacts. We considered the group with lower scores and all male names shown our baseline group.

Table 1: Scoreboard specifications

	3 males	**2 males 1 female**	**1 male 2 females**	**3 females**
Recent Scores	Jozef 55 Adam 48 Paul 45	Jozef 55 Victoria 48 Paul 45	Anna 55 Adam 48 Petra 45	Anna 55 Victoria 48 Petra 45
Recent High Scores	Jozef 87 Adam 78 Paul 73	Jozef 87 Victoria 78 Paul 73	Anna 87 Adam 78 Petra 73	Anna 87 Victoria 78 Petra 73

Once players played the game, we collected detailed performance data, recording every event that took place: players' clicks, score, pauses, and target shape changes. Our initial dataset was at the level of events, which we aggregated to the player-session level. We defined a session as a

single browser session during which the individual played one or more games. Players received the same treatment during every game in a gaming session. This allows us to study longer-run impacts of the scoreboard treatments. We only analyze the first session of each player, so there is no within-player variation in treatment. Identification is therefore based on between-player variation. It is important to note that different individuals played a different number of games. The number of games played could itself be impacted by treatment, and is therefore one of the outcome variables we study. We calculated session-level outcomes: the number of games played (persistence), and the best score in the session (performance), and linked these to the individual level variables collected in the pre-game survey. The analysis of session level performance (best score) means that the estimated impacts of the treatments include any effect that is realized through persistence (number of games played) and learning.

We assess the impact of the scoreboard treatments on session-level outcomes based on OLS equations that include three-way interaction terms of gender, score level, and scoreboard gender composition. They additionally control for observable characteristics (the age, country, and education level of the individual, and whether they are playing on a touchscreen device). The estimated regressions are of the following form:

$$outcome_i = \alpha_0 + \alpha_1 \cdot higher\ scores_i + \alpha_2 \cdot female_i + \alpha_3 \cdot higher\ scores_i \cdot female_i + \alpha_4 \cdot group\_3\ females_i + \alpha_5 \cdot group\_2females\ 1\ male_i + \alpha_6 \cdot group\_1female\ 2\ males_i + \alpha_7 \cdot group\_3\ females_i \cdot higher\ scores_i + \alpha_8 \cdot group\_2females\ 1\ male_i \cdot higher\ scores_i + \alpha_9 \cdot group\_1female\ 2\ males_i \cdot higher\ scores_i + \alpha_{10} \cdot group\_3\ females_i \cdot female_i + \alpha_{11} \cdot group\_2females\ 1\ male_i \cdot female_i + \alpha_{12} \cdot group\_1female\ 2\ males_i \cdot female_i + \alpha_{13} \cdot group\_3\ females_i \cdot higher\ scores_i \cdot female_i + \alpha_{14} \cdot group\_2females\ 1\ male_i \cdot higher\ scores_i \cdot female_i + \alpha_{15} \cdot group\_1female\ 2\ males_i \cdot higher\ scores_i \cdot female_i + \alpha_{16}{}'X_i + \vartheta_i \quad (1)$$

where *outcome*$_i$ represents the player-session level outcome variables for individual i, and X_i represents control variables (age group, education level, region, touchscreen, screen size). Based on these OLS results, we calculate treatment effects and the significance of the impacts of the gender composition treatments, given the lower or higher score levels shown, and the impacts of seeing higher scores, given the different gender compositions. We calculate the treatment effects separately by gender.

3. Results

3.1. Descriptive statistics

Table 2 summarizes the descriptive statistics of the sample used in the analysis. 1140 individuals participated in the game, 64% of whom are female, playing a total of 2335 games. The sampling method (online advertising on social media) resulted in a sample that is composed of individuals aged between 18 and 45, with a mean age of around 29. Approximately 44% of the individuals reported having some college or university education, 39% report secondary education, and 16% report lower education levels. The sample is dominated by the four countries targeted in the ads (not shown in the table): Hungary, Poland, Slovakia, and the Czech Republic.

Table 2: Descriptive statistics of the sample

		Recent Scores				Recent High Scores			
	Total	3 females	3 males	2 females, 1 male	1 female, 2 males	3 females	3 males	2 females, 1 male	1 female, 2 males
N (individuals)	1140	138	157	142	143	126	134	149	151
N (games)	2335	215	333	293	293	237	275	350	339
Games/player	3.915	3.587	3.968	3.915	3.986	3.667	4.157	4.168	3.834
Female	0.64	0.581	0.630	0.689	0.683	0.614	0.600	0.644	0.681
Age	29.527	28.757	32.471	29.688	31.055	28.917	31.819	27.435	24.938
Education:									
Elementary	0.165	0.188	0.159	0.169	0.189	0.175	0.127	0.154	0.159
Secondary	0.392	0.413	0.389	0.380	0.392	0.357	0.418	0.403	0.384
College	0.443	0.399	0.452	0.451	0.420	0.468	0.455	0.443	0.457
Plays games									
Never	0.156	0.188	0.140	0.127	0.154	0.111	0.149	0.168	0.205
Sometimes	0.522	0.442	0.529	0.556	0.566	0.563	0.582	0.483	0.464
Often	0.322	0.370	0.331	0.317	0.280	0.325	0.269	0.349	0.331
Confidence									
Low	0.273	0.261	0.248	0.282	0.301	0.262	0.261	0.309	0.258
Middle	0.539	0.536	0.567	0.535	0.580	0.548	0.507	0.483	0.550
High	0.188	0.203	0.185	0.183	0.119	0.190	0.231	0.208	0.192
Touchscreen	0.521	0.536	0.561	0.528	0.531	0.587	0.537	0.510	0.391

3.2. Main results

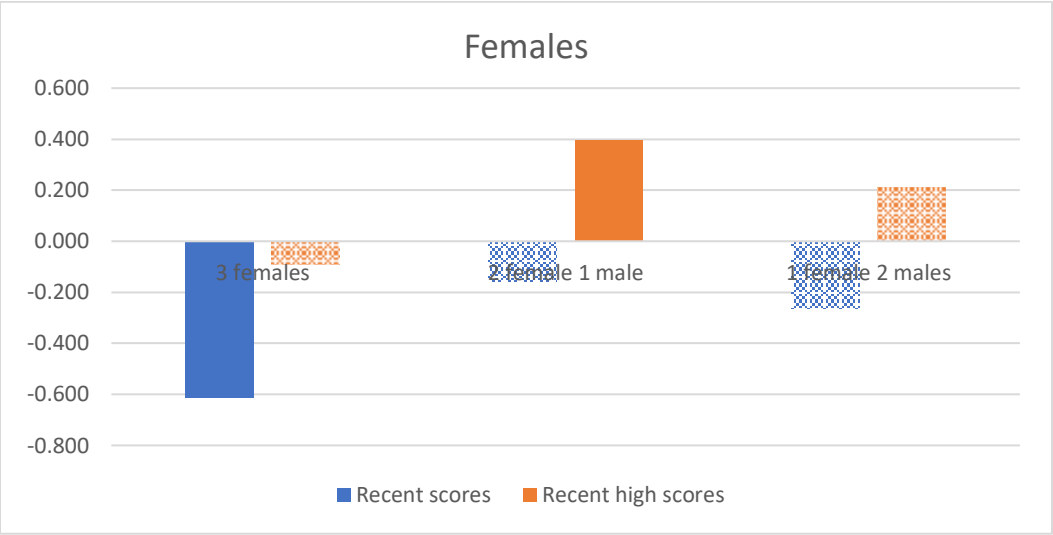
We now turn to our main results, highlighting the estimated treatment effects. The full OLS results are shown in Table A1. Table 3 and Figure 4 summarize the estimated impacts of various scoreboard gender composition treatments on the number of games played. For females, seeing a scoreboard with three female names has a significant negative impact when average scores (Recent Scores) are shown. When high scores are shown, seeing a board with both male and female names has a positive impact, which is significant for the specification with two female names and one male name. For males, we do not see any significant impacts on persistence. In terms of sign, the impact of seeing three female names is negative – regardless of score level – and the impact of seeing mixed gender names tends to be more positive. The gender gaps in the impacts, defined as the female effect minus the male effect, are generally not significant.

Table 3: Treatment effects of gender composition on the number of games played, by score level and gender

		Females		**Males**		**Difference**	
Score level	**Gender composition**	**Estimate**	**p-value**	**Estimate**	**p-value**	**Estimate**	**p-value**
Recent Scores	3 females	-0.614	0.007	-0.249	0.519	-0.365	0.415
	2 female 1 male	-0.158	0.485	0.189	0.617	-0.348	0.431
	1 female 2 males	-0.265	0.236	0.513	0.147	-0.778	0.084
Recent High Scores	3 females	-0.091	0.705	-0.175	0.674	0.083	0.862
	2 female 1 male	0.396	0.089	0.057	0.884	0.339	0.454
	1 female 2 males	0.212	0.189	0.118	0.760	0.094	0.835

Notes: Treatment effects calculated based on OLS estimates of equation (1). Full OLS results are shown in Table A1. Dependent variable is the number of games played in the session. Independent variables include the interaction terms of gender, score level, and gender composition, and controls for age, education, region, and device type.

Figure 4: Treatment effects of gender composition on the number of games played, by score level and gender



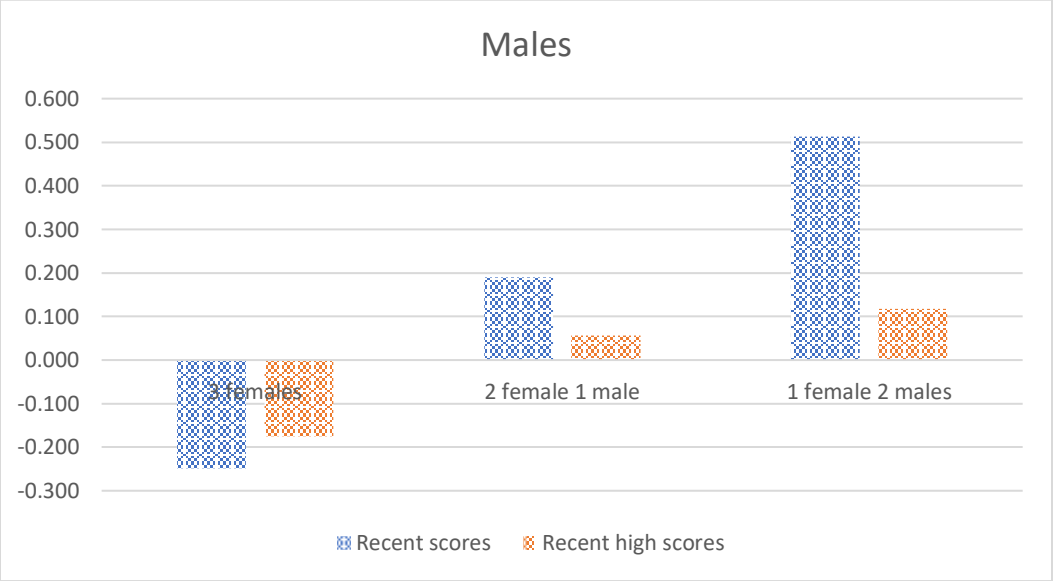


Table 4 and Figure 5 summarize the treatment effects on performance, i.e. the best score the player achieved in the gaming session. For female players, we do not see any significant impacts when average scores are shown. When high scores are shown, however, we see that all treatments that include at least one female name on the scoreboard have a significant positive impact compared to the baseline of all male names. For males, we again do not see any significant impacts. The trend of the sign of the estimates indicates that males are not motivated to perform better when they see only female names on a scoreboard. The impacts of mixed gender scoreboards when the scores

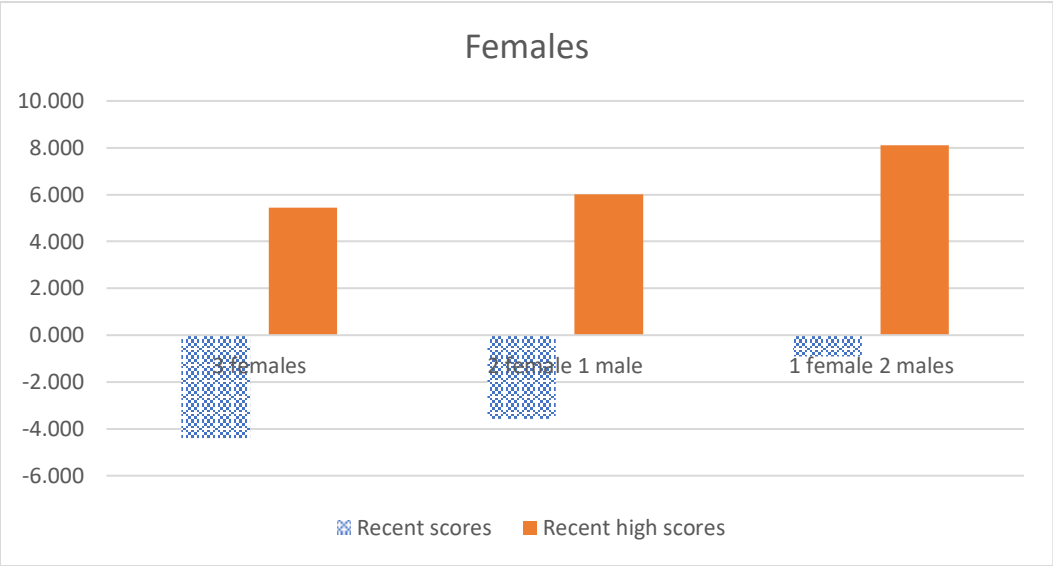
shown are high are positive – though not significant – compared to the only male names baseline. Interestingly, the gender gaps in the impacts suggest that females respond more negatively to seeing female participants' names with average scores than males do. The gender gaps in the average score specifications are generally negative, while those in the high score specifications are positive. Females appear to respond more positively to seeing high scoreboards with female names than men do, while seeing female names with average scores has a more negative impact.

Table 4: Treatment effects of gender composition on best score, by score level and gender

		Females		Males		Difference	
Score level	**Gender composition**	**Estimate**	**p-value**	**Estimate**	**p-value**	**Estimate**	**p-value**
Recent Scores	3 females	-4.393	0.133	-1.508	0.761	-2.885	0.617
	2 female 1 male	-3.577	0.221	7.345	0.132	-10.923	0.055
	1 female 2 males	-0.916	0.750	7.298	0.147	-8.214	0.157
Recent High Scores	3 females	5.434	0.080	-2.934	0.584	8.369	0.177
	2 female 1 male	6.020	0.044	0.693	0.890	5.327	0.361
	1 female 2 males	8.107	0.007	-0.197	0.968	8.304	0.152

Notes: Treatment effects calculated based on OLS estimates of equation (1). Full OLS results are shown in Table A1. Dependent variable is the best score in the session. Independent variables include the interaction terms of gender, score level, and gender composition, and controls for age, education, region, and device type.

Figure 5: Treatment effects of gender composition on best score, by score level and gender



We next look at the impact of seeing high scores compared to average scores on the number of games played, conditional on the gender composition of the names shown (Table 5 and Figure 6). Overall, we do not see significant impacts for either males or females. The pattern of the signs of the estimates suggests that the impact of seeing high scores is more positive for females when they see at least one female name on the scoreboard. The impact is significant at the 10% level when two female names and one male name is shown. Males, on the other hand, seem to respond more negatively when a mixed gender scoreboard is shown.

Table 5: Treatment effect of seeing higher scores on the number of games played, by gender composition of scoreboard seen and gender

	Females		Males	
Gender composition of names shown	**Estimate**	**P-value**	**Estimate**	**P-value**
3 female names	0.371	0.122	0.125	0.759
3 male names	-0.151	0.505	0.051	0.897
2 female 1 male name	0.403	0.082	-0.081	0.827
1 female 2 male names	0.325	0.158	-0.345	0.362

Notes: Treatment effects calculated based on OLS estimates of equation (1). Full OLS results are shown in Table A1. Dependent variable is the number of games played in the session. Independent variables include the interaction terms of gender, score level, and gender composition, and controls for age, education, region, and device type.

Figure 6: Treatment effect of seeing higher scores on the number of games played, by gender composition of scoreboard seen and gender

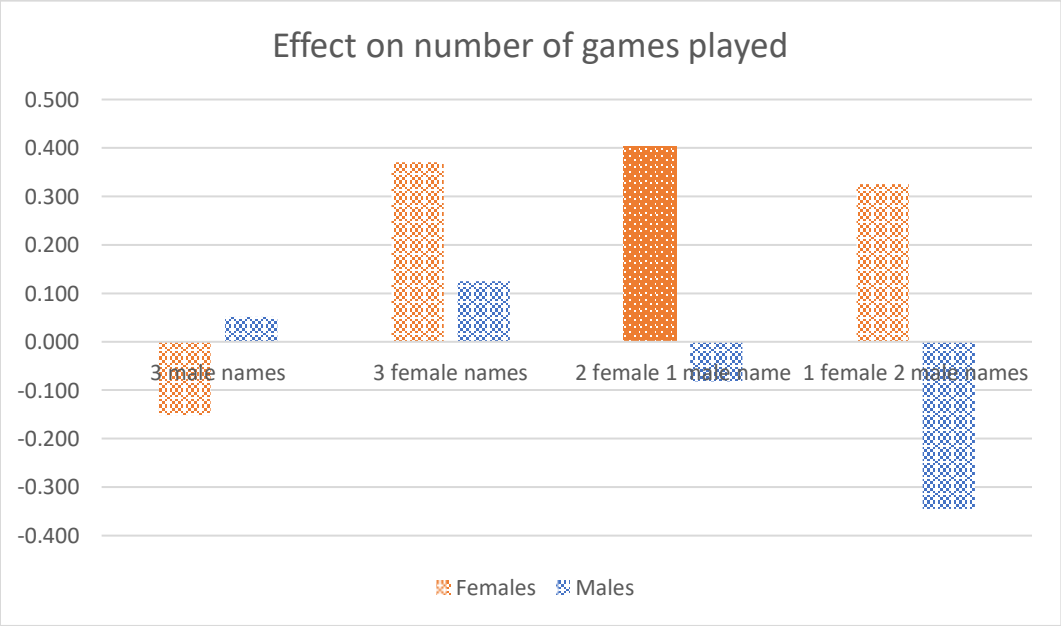


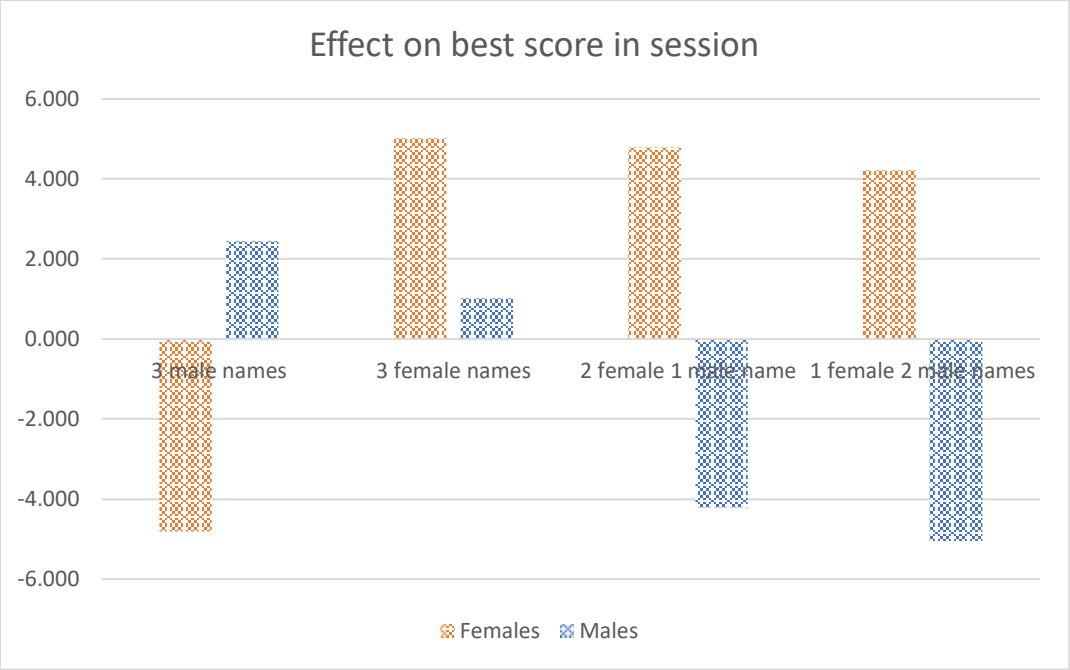
Table 6 and Figure 7 show the impact of high scores on the best score achieved in the session. Again, we see know significant treatment effect estimates. The signs of the estimates suggest that females respond more positively to seeing high scores when there is at least one female name on the scoreboard. Again, the response of males is more negative when a mixed gender scoreboard is shown.

Table 6: Treatment effect of seeing higher scores on the player's best score, by gender composition of scoreboard seen and gender

	Females		Males	
Gender composition of names shown	**Estimate**	**P-value**	**Estimate**	**P-value**
3 female names	5.021	0.104	1.019	0.846
3 male names	-4.806	0.101	2.445	0.631
2 female 1 male name	4.791	0.108	-4.207	0.380
1 female 2 male names	4.217	0.155	-5.050	0.299

Notes: Treatment effects calculated based on OLS estimates of equation (1). Full OLS results are shown in Table A1. Dependent variable is the best score in the session. Independent variables include the interaction terms of gender, score level, and gender composition, and controls for age, education, region, and device type.

Figure 7: Treatment effect of seeing higher scores on the player's best score, by gender composition of scoreboard seen and gender



Overall, the results shown in this section suggest that, in line with previous evidence, females' performance expectations, and thereby, their persistence and performance is more dependent on the gender composition of performance feedback on peers than those of males. We find no evidence of a significant impact on males. Female players, on the other hand, respond positively to seeing female names on a scoreboard showing previous high achievers, in terms of their

persistence and performance. The impact on performance may be due to the impact on persistence and learning, and/or better performance within games. The magnitude of these impacts is non-negligible. Seeing high-performing female names on the scoreboard increases the best score females achieve in the session by about 5.5-8 points, which translates to an increase of 17-25 percent compared to the mean baseline score of 32 points. The results suggest that public acknowledgement of the high performance of females – along with males – in a way that reveals the gender of high performers may be a tool that can be used to decrease gender differences in competitive settings.

4. Conclusion

In this study, we analyze data on the persistence and performance of players of a simple online game, in which randomized treatment was given in the form of a scoreboard, which show either average or high scores, with names indicating different gender compositions. Our findings indicate that females respond more to performance information on other participants when that information is linked to gender, and contains female participants. In particular, females' persistence and performance increases when the names and scores of high performing female participants are shown. Males, on the other hand, show no significant response to the gender composition of the scoreboard shown.

These impacts are estimated based on a sample of players who chose to click on an ad for the game in social media advertisements. This self-selection into participation resulted in a sample skewed towards females, suggesting that males are less likely to participate in such games, and that our sample is not representative of the population. It is also important to note that we observe these impacts based on the behavior of participants in their natural setting, which is likely to capture real-life behaviors since an anonymous online game does not incent individuals to alter their behavior. Overall, our results still provide evidence of the impact of the gender composition of performance information on peers on females' outcomes.

Females appear to consider performance information on other females to be relevant as reference points, while males are less sensitive to the gender of high performing participants. Similarly to the introduction of female role models, seeing scores of previous high performing

females can encourage females to be more persistent and to perform better in competitive settings. The policy implication of this finding is that feedback design in educational and workplace settings needs to take the gender composition of performance information on others into account. Public acknowledgement of high-performing females may be a tool for decreasing the gender gaps in performance expectations, competitive attitudes, and outcomes.

References

Azmat, G., & Petrongolo, B. (2014). Gender and the labor market: What have we learned from field and lab experiments?. Labour Economics, 30, 32-40.

Baier, Alexandra; Davis, Brent J.; Jaber-Lopez, Tarek; Seidl, Michael (2018) : Gender, competition and the effect of feedback and task: An experiment, Working Paper Forschungsförderung, No. 062, Hans-Böckler-Stiftung, Düsseldorf, http://nbn-resolving.de/urn:nbn:de:101:1-2018040415835

Balafoutas, L. and Sutter, M. (2012). Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory. Science (New York, N.Y.). 335. 579-82.

Bertrand, M. (2011): New Perspectives on Gender. In David E. Card, Orley Ashenfelter (Eds.): Handbook of labor economics. Volume 4B, vol. 4. 1st ed. Amsterdam, New York, New York, N.Y., U.S.A: North-Holland (Handbooks in economics, 5), pp. 1543–1590.

Booth, A. and Nolen, P. (2012). "Choosing to compete: How different are girls and boys?" Journal of Economic Behavior & Organization, Volume 81, Issue 2, pp. 542-555.

Buser, T; Niederle, M; Oosterbeek, H. (2014): Gender, Competitiveness, and Career Choices. The Quarterly Journal of Economics 129 (3), pp. 1409–1447.

Cai, X; Lu, Y; Pan, J; Zhong, S. (2019): Gender Gap under Pressure: Evidence from China's National College Entrance Examination. The Review of Economics and Statistics 101 (2), pp. 249–263.

Cotton, C., McIntyre, F. and Price, J. (2013). "Gender differences in repeated competition: Evidence from school math contests," Journal of Economic Behavior & Organization, Volume 86, February, pp. 52-66.

Del Carpio, Lucia and Maria Guadalupe, "More Women in Tech? Evidence from a Field Experiment Addressing Social Identity," 2018. CEPR Discussion Paper DP13234.

Ertac, Seda, and Balazs Szentes. 2010. "The Effect of Performance Feedback on Gender Differences in Competitiveness: Experimental Evidence." Working Paper, Koc University, Turkey.

Gneezy, U.; Niederle, M.; Rustichini, A. (2003): Performance in Competitive Environments: Gender Differences. The Quarterly Journal of Economics 118 (3), pp. 1049–1074.

Gneezy, Uri; Leonard, Kenneth L. and List, John A. (2009). "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society," Econometrica, Vol. 77, No. 5 September 2009, 1637–1664.

Healy, Andrew and Jennifer Pate (2011). Can Teams Help to Close the Gender Competition Gap? The Economic Journal, 121: 1192–1204.

Joensen, Juanna Schroter and Helena Skyt Nielsen. 2009. "Is There a Causal Effect of High School Math on Labor Market Outcomes?" Journal of Human Resources 44(1): 171-198

Niederle, Muriel, and Lise Vesterlund, (2007). "Do Women Shy away from Competition? Do Men Compete too Much?" Quarterly Journal of Economics, August 2007, Vol. 122, No. 3: 1067-1101.

Niederle, Muriel; Vesterlund, Lise (2011): Gender and Competition. In Annu. Rev. Econ. 3(1), pp. 601–630.

Niederle Muriel, Segal Carmit, Vesterlund Lise. How costly is diversity? Affirmative action in light of gender differences in competitiveness // Management Science. 2013. 59, 1. 1–16.

Ors, Evren, Frédéric Palomino, and Eloïc Peyrache. 2013. "Performance Gender Gap: Does Competition Matter?" Journal of Labor Economics 31 (3): 443–99. https://doi.org/10.1086/669331.

Wozniak, D., Harbaugh, W., & Mayr, U. (2014). The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices. Journal of Labor Economics, 32(1), 161-198.

Appendix

Table A1: Full OLS results

	(1)	(2)
Dependent variable	**Number of games played**	**Best score in session**
fe	0.625**	8.039**
	(0.316)	(4.068)
highB	0.0509	2.445
	(0.395)	(5.086)
3F	-0.249	-1.508
	(0.386)	(4.965)
2F1M	0.189	7.345
	(0.379)	(4.873)
1F2M	0.513	7.298
	(0.391)	(5.030)
highB_3F	0.0739	-1.426
	(0.568)	(7.308)
highB_2F1M	-0.132	-6.652
	(0.542)	(6.974)
highB_1F2M	-0.396	-7.495
	(0.546)	(7.031)
fe_highB	-0.202	-7.251
	(0.456)	(5.867)
fe_3F	-0.365	-2.885
	(0.448)	(5.762)
fe_2F1M	-0.348	-10.92*
	(0.442)	(5.685)
fe_1F2M	-0.778*	-8.214
	(0.450)	(5.799)
fe_highB_3F	0.448	11.25
	(0.658)	(8.466)
fe_highB_2F1M	0.687	16.25**
	(0.632)	(8.142)
fe_highB_1F2M	0.872	16.52**
	(0.635)	(8.173)
dagegroup2	-0.103	3.594
	(0.244)	(3.141)
dagegroup3	0.247	8.406***
	(0.244)	(3.137)
dagegroup4	0.191	3.871
	(0.220)	(2.826)
dagegroup5	0.111	-0.0807

	(0.192)	(2.475)
dagegroup6	0.0444	-11.54***
	(0.223)	(2.870)
deduc1	-0.0432	1.813
	(0.152)	(1.952)
deduc3	0.240**	5.293***
	(0.112)	(1.445)
dregion1	0.256	-1.092
	(0.451)	(5.810)
dregion3	-0.157	-6.021
	(0.452)	(5.821)
dregion4	-0.128	0.354
	(0.489)	(6.300)
dregion5	-0.0393	-4.537
	(0.519)	(6.681)
dtouch1	0.233	12.66***
	(0.184)	(2.374)
dpixel2	-0.207	-4.479*
	(0.183)	(2.359)
dpixel3	-0.0926	-1.519
	(0.214)	(2.753)
dpixel4	0.168	-9.779
	(0.592)	(7.625)
Constant	1.354**	20.58***
	(0.546)	(7.032)
Observations	1,140	1,140
R-squared	0.081	0.189

Notes: Specifications include the interactions of gender, score level shown, and gender composition of scores shown. The baseline group is males, shown lower scores ("Recent scores"), and 3 male names on the scoreboard. The variable fe refers to a female dummy, highB refers to being shown higher scores ("Recent high scores"), 3F indicates player was shown 3 female names, 2F1M refers to 2 female and 1 male name shown, 1F2M refers to 1 female 2 male names shown. Standard errors are shown in parentheses.