ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Lovasz, Anna; Bat-Erdene, Boldmaa; Cukrowska-Torzewska, Ewa; Rigo, Mariann; Szabo-Morvai, Agnes

Working Paper

Competition , subjective feedback, and gender gaps in performance

CERS-IE Working Papers, No. CERS-IE WP - 2021/1

Provided in Cooperation with:
Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences

Suggested Citation: Lovasz, Anna; Bat-Erdene, Boldmaa; Cukrowska-Torzewska, Ewa; Rigo, Mariann; Szabo-Morvai, Agnes (2021) : Competition , subjective feedback, and gender gaps in performance, CERS-IE Working Papers, No. CERS-IE WP - 2021/1, Hungarian Academy of Sciences, Institute of Economics, Centre for Economic and Regional Studies, Budapest

This Version is available at:
https://hdl.handle.net/10419/237545

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

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

Competition , Subjective Feedback, and Gender Gaps in Performance

ANNA LOVÁSZ - BOLDMAA BAT-ERDENE

EWA CUKROWSKA-TORZEWSKA - MARIANN RIGÓ

ÁGNES SZABÓ-MORVAI

CERS-IE WP – 2021/1

January 2021

https://www.mtakti.hu/wp-content/uploads/2021/01/CERSIEWP202101.pdf

CERS-IE Working Papers are circulated to promote discussion and provoque comments, they have not been peer-reviewed.
Any references to discussion papers should clearly state that the paper is preliminary.
Materials published in this series may be subject to further publication.

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES, BUDAPEST, 2020

ABSTRACT

We study gender differences in the impacts of competition and subjective feedback, using an online game with pop-up texts and graphics as treatments. We define 8 groups: players see a Top 10 leaderboard or not (competitiveness), and within these, they receive no feedback, supportive feedback, rewarding feedback, or "trash talk" (feedback type). Based on 5191 participants, we find that competition only increases the performance of males. However, when it is combined with supportive feedback, the performance of females also increases. This points to individualized feedback as a potential tool for decreasing gender gaps in competitive settings such as STEM fields.

JEL codes: I20, J16, J24, M54

Keywords: Gender Gaps, Competition, Supervisory Feedback

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

Boldmaa Bat-Erdene, Eotvos Lorand University, Pazmany Peter setany 1/a, Budapest, 1117 Hungary

Ewa Cukrowska-Torzewska, University of Warsaw, Faculty of Economic Sciences, Długa 44/50, 00-241 Warsaw, Poland

Mariann Rigo, University of Düsseldorf, Institute of Medical Sociology, Moorenstr. 5, 40225 Düsseldorf, Germany

Agnes Szabo-Morvai, Centre for Economic and Regional Studies, Toth Kalman u. 4. Budapest, 1097 Hungary and University of Debrecen, Economics Department, Böszörményi út 132, Debrecen, 4032 Hungary

Verseny, szubjektív visszajelzések, és nemek közötti eltérések a teljesítményben

LOVÁSZ ANNA - BOLDMAA BAT-ERDENE

EWA CUKROWSKA-TORZEWSKA - RIGÓ MARIANN

SZABÓ-MORVAI ÁGNES

ÖSSZEFOGLALÓ

Egy online játék alapján vizsgáljuk a nemek közötti eltéréseket a felettesi kommunikáció elemeinek hatásában. A randomizált kezelések egyszerű szövegek és grafikák formájában jelennek meg a játék során. Nyolc csoportot különböztetünk meg: a játékosok látnak Top 10 táblát vagy nem (*verseny*), és, ezeken belül, nem kapnak visszajelzést, illetve bíztatást, dícséretet, vagy cukkoló visszajelzést kapnak (*visszajelzés*). 5191 résztvevő adatai alapján azt látjuk, hogy a verseny önmagában cask a férfiak teljesítményét javítja, azonban ha a verseny bíztatással párosul, akkor a nők teljesítménye is hasonlóan növekszik. Az eredmények a személyre szabott visszajelzések fontosságát támasztják alá mint olyan tényező, ami csökkentheti a nemek közötti eltéréseket a versenyhelyzetekben, például a STEM szakirányokban.

JEL: I20, J16, J24, M54

Kulcsszavak: nemek közötti eltérések, verseny, felettesi visszajelzések

Competition, Subjective Feedback, and Gender Gaps in Performance

Anna Lovász[*,(a)(b)], Boldmaa Bat-Erdene[(c)] Ewa Cukrowska-Torzewska[(d)], Mariann Rigó[(e)], Ágnes Szabó-Morvai[(a)(f)]

[(a)] *Centre for Economic and Regional Studies, Toth Kalman u. 4. Budapest, 1097 Hungary*
[(b)] *University of Washington Tacoma, 1900 Commerce Street, Tacoma, WA 98402-3100, USA*
[(c)] *Eotvos Lorand University, Pazmany Peter setany 1/a, Budapest, 1117 Hungary*
[(d)] *University of Warsaw, Faculty of Economic Sciences, Długa 44/50, 00-241 Warsaw, Poland*
[(e)] *University of Düsseldorf, Institute of Medical Sociology, Moorenstr. 5, 40225 Düsseldorf, Germany*
[(f)] *University of Debrecen, Economics Department, Böszörményi út 132, Debrecen, 4032 Hungary*

January 4, 2021

Abstract

We study gender differences in the impacts of competition and subjective feedback, using an online game with pop-up texts and graphics as treatments. We define 8 groups: players see a Top 10 leaderboard or not (*competitiveness*), and within these, they receive no feedback, supportive feedback, rewarding feedback, or "trash talk" (*feedback type*). Based on 5191 participants, we find that competition only increases the performance of males. However, when it is combined with supportive feedback, the performance of females also increases. This points to individualized feedback as a potential tool for decreasing gender gaps in competitive settings such as STEM fields.

Keywords: Gender Gaps, Competition, Supervisory Feedback

JEL codes: I20, J16, J24, M54

Acknowledgements: The project leading to this application received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 691676. The research was supported by grants FK 124658 and 121267-PD of the National Research, Development and Innovation Office of Hungary. Anna Lovasz also received support from the Bolyai grant of the Hungarian Academy of Sciences and the ÚNKP Bolyai plusz grant ÚNKP-18-4-ELTE-665. The experiment is registered under AEA RCT registry ID #AEARCTR-0003984. The project does not have IRB approval as our institutions did not have IRBs. We would like to thank Andor Zöldesi for developing the game website. We also thank Daniel Horn, Janos Hubert Kiss, Andrea Kiss, Barbara Pertold-Gebicka, and commenters at various conference and seminar presentations for their valuable feedback.

**Corresponding author: University of Washington Tacoma, 1900 Commerce Street, Tacoma, WA, 98402 USA, phone: (1)253-537-4782, email: plovi@uw.edu or lovasz.anna@krtk.mta.hu*

1. Introduction

The economics literature documents significant gender differences in psychological traits and preferences (Niederle 2016, Eckel and Grossman 2008; Croson and Gneezy 2009). In particular, women tend to compete less[1] and perform weaker in competitive situations.[2] These competition-related gender differences have been noted as possible 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). Differences in attitudes towards competition can impact outcomes through key decisions, such as field of study and occupation (Buser et al 2014, Osborne et al. 2003, Kirkeboen et al. 2016), which contribute significantly to the gender gaps in earnings that we still observe today (Bertrand 2020, Macis 2017). Competitive attitudes may also impact choices and outcomes within a field, occupation, or workplace. For example, women may choose to participate less often in challenging tasks and striving for promotions (Bertrand 2011, Kauhanen and Napari 2015), and perform worse in high-stakes, competitive settings (Jurajda and Münich 2011, Ors, Palomino and Peyrache 2013).

Given these documented differences, there is much debate about what can be done to improve the performance of females in relatively disadvantageous competitive situations, and thereby decrease gender gaps in outcomes. One approach, termed "fix institutions," emphasizes the idea that certain institutional elements can be altered to "achieve outcomes that better reflect underlying abilities" (Niederle 2016). It is possible that certain elements of the current institutional design – such as competitiveness, feedback culture, hiring procedures, or assessment methods - favor males due to gender differences in individual traits.[3] Previous evidence shows that altering particular elements of the institutional design can decrease or even eliminate gender differences in choices and outcomes in competitive settings.[4] One such element that has received attention is the provision of

[1] See Niederle and Vesterlund 2007, Gneezy et al 2009, Niederle and Vesterlund 2011, Healy and Pate 2011, Booth and Nolen 2012, and Wozniak et al 2014 for some well-known studies on the topic.

[2] Examples include Gneezy et al 2003, Gneezy and Rustichini 2004, Cai et al. 2019, and Cotton et al 2013.

[3] For example, confidence is frequently studied as a key underlying trait of gender gaps in outcomes. Competition may motivate high confidence individuals more, or hurt the performance of lower confidence individuals due to increased stress (Azmat et al 2015). Females tend to have lower confidence, even conditional on their ability, especially in traditionally male tasks (McCarty 1986, Lloyd et al 2005), and may therefore suffer a relative disadvantage in competitive settings. Subjective feedback has been shown to impact those with lower confidence more (Chang et al 2012), so it could potentially benefit women and other groups with lower confidence levels.

[4] Several institutional elements, such as single-sex tournaments (Gneezy et al 2003; Datta Gupta et al , 2005), quota-style affirmative policies (Niederle and Vesterlund 2008; Balafoutas and Sutter 2012), feedback about relative performance (Wozniak et al 2011; Wozniak et al. 2014; Ertac, Szentes, 2010; Wozniak et al. 2016), and performance

objective feedback (Ertac and Szentes 2010, Azmat and Iriberri 2010, Bandiera, Larcinese, and Rasul 2015, Wozniak et al 2016). Nevertheless, in real-life settings, feedback often contains an objective as well as a subjective element; thus, we extend the literature by testing the impact of subjective feedback. In particular, we study its interaction with competition, to see whether the provision of certain types of positive subjective feedback can counteract the disadvantage of females in competitive settings.

We test the impact of subjective content that is given in addition to objective performance feedback and contains a positive or negative qualification of past or expected future performance, such as praise or encouragement. Subjective feedback has received significant attention in the pedagogy, psychology, and human resource management literature (e.g., Deci and Ryan 1985; Locke 1996, Posner and Kouzes 1999, Dweck 2007, Johnson 2013, Khan et al. 2014, Wong 2015). Differences have been highlighted in the perception and impact by confidence and gender (Chang et al. 2012, Healy and Pate 2011; Wozniak et al 2016). However, subjective feedback has not been studied as a factor that contributes to gender differences in the education or labor economics literature, or in terms of its interaction with competition. We study the impacts of three common types of subjective feedback from the previous literature: supportive feedback (encouragement), rewarding feedback (praise), and "trash talk." We evaluate whether any of these feedback types can mitigate the relative disadvantage of females due to competition.

To measure the impacts, we analyze a sample of 5191 individuals who played a simple online game of visual perception[5] that was advertised on social media. During the game, players received randomized treatment in the form of simple text and graphics, which appeared as pop-ups on the screen before, during, and after the game. Treatment was randomized among a total of eight groups, along two dimensions: competitiveness and subjective feedback type. Players either saw a leaderboard or did not (*competitiveness*),[6] and, within each of these categories, they received either no subjective feedback, supportive feedback, rewarding feedback, or trash talk (*feedback type*). This experimental design allows us to test the impacts of competition and the three feedback types when they are given separately, as well as their joint impacts, compared to the control group

feedback followed by sequential choice in entering a tournament (Niederle, Muriel & Yestrumskas, Alexandra 2008) have been shown to improve women's participation in competitive settings.

[5] https://experimental-games.herokuapp.com/.

[6] Competition - in our case, a leaderboard - therefore consists of both relative performance information and public acknowledgement of the highest performers.

of no leaderboard and no feedback. We estimate treatment effects on various outcome measures capturing persistence (number of games played) and performance (accuracy, mean score, best score), and assess gender differences in these.

As a first step, we replicate previous results regarding the unfavorable impact of competition on females compared to males,[7] and estimate the impacts of the three subjective feedback types separately when there is no competition. Next, we analyze how these elements interact. We compare how males and females are impacted by competition when there is no subjective feedback, versus when competition is coupled with subjective feedback. To our knowledge, there is no previous empirical evidence on these interactions, although in real-life settings, these elements are present simultaneously. The results point to significant heterogeneity by gender in the impacts. While competition increases the persistence of both genders, it only improves the performance of males significantly. However, when competition is coupled with supportive feedback, the performance of females increases as well, similarly to males. Competition combined with supportive feedback increases the performance of both genders and shows no gender gap in the overall impact. We acknowledge that these estimates need to be evaluated keeping in mind that they may be specific to the task, context, sample, and specific feedback content. Despite this caveat, the results do provide evidence of significant heterogeneities in the impact of competition and subjective feedback, which can contribute to gender differences in outcomes. Better targeted feedback can potentially decrease gender gaps in performance in competitive settings.

Individualized feedback could be a key factor for decreasing gender gaps in educational and labor market outcomes in the future. Subjective feedback is an unavoidable part of everyday supervisory communication, and providing more targeted feedback is a relatively low-cost intervention.[8] One aim of our study is to draw the attention of educators and managers to the importance of individualized feedback, as blanket policies such as "tough love" can lead to quantifiable performance losses and contribute to inequalities. Furthermore, recent advances - and the COVID-19 pandemic - have made the use of educational and HR software increasingly widespread.

[7] As Niederle (2016) notes, due to the limited external validity of both laboratory and field experiments, it is important to confirm the empirical evidence of gender differences in competitiveness in various settings and on various populations.

[8] Our study is thus related to recent studies highlighting the impacts of low-cost behavioral interventions in educational settings (e.g., Aronson et al 2002, Bettinger et al 2018). For a summary of behavioral nudges in education, see Damgaard and Nielsen 2018.

Software developments such as Personalized Tutoring Systems and the use of AI in education (Luckin et al 2016, Perotta and Selwyn 2019) allow for the provision of much more frequent individualized feedback compared to in-person supervision. Our study does not answer the question of what the targeting of feedback should be based on, nor do we suggest that it should be based on gender. Research is being carried out to improve targeting algorithms based on highly detailed observable data on individual characteristics, behavior, and performance (e.g., Narciss et al 2014). As these data-driven feedback provision mechanisms improve, we may see a decrease in gender inequalities.

2. Methodology

2.1. Experimental Design

We utilize a simple online game developed for this research, during which players receive randomized treatment. The Shape Game (Figure 1) is a simple game of visual perception that requires both concentration and effort. The task is to click on a given geometric shape that is displayed in the top left corner of the screen (target shape), from the set of shapes that appear on the screen. The shapes move around the screen, and players must find and click on all shapes that match the target shape shown. The target shape then changes to a new shape. The game takes two minutes, and the goal is to score as many points as possible. All players see the remaining game time and their cumulative score in the upper corners of the screen during the entire game.

Figure 1: Screenshot of the Shape Game

Prior to choosing to play, individuals are given a description of the game, and can view a demo video as well. The game is preceded by a simple survey (see Appendix Figure A1). This asks for basic demographic information: gender, age, country, and level of education. 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.[9] The survey was designed to be quick and easy to fill out, asking only for anonymous information similar to what is often requested on gaming sites. The survey also asks players to give a nickname, which is used if the player achieves a high score in the Top 10 leaderboard. Players are informed of the experimental purpose of the game and the details of data collection, but otherwise, the goal was to focus player's attention on the game itself, in order to observe real-life behavior in a natural game setting. Additionally, data is also collected automatically to account for whether the device the game is played on is a touchscreen or not, as well as screen size, both of which may also affect performance.

[9] Players are asked how good they consider themselves to be at computer games: excellent, pretty good, ok, pretty bad, or very bad. Since the question is asked after the game description, the players' responses likely reflect their beliefs regarding how well they will play this particular game.

When players click to start playing the game, they are randomly selected to be in one of the treatment groups, as summarized in Table 1. Treatment is varied along two dimensions: whether players see a Top 10 leaderboard or not (*competitiveness*), and in terms of the subjective feedback they receive (*feedback type*). This gives us a total of 8 groups. Seeing a leaderboard or no leaderboard is interacted with four types of subjective feedback (including a control with no feedback, supportive feedback, praise of performance, and trash talk). This setup allows us to estimate the effects of the leaderboard and each of the feedback types individually, as well as their joint effects.

Table 1: Summary of Treatment Groups

Group	Leaderboard	Feedback
1 (Baseline)	No	None
2	No	Supportive feedback (encouragement)
3	No	Rewarding feedback (praise)
4	No	Trash talk
5	Yes	None
6	Yes	Supportive feedback (encouragement)
7	Yes	Rewarding feedback (praise)
8	Yes	Trash talk

Table 2 summarizes the categories and gives details regarding the specifications and the exact timing and feedback content. Subjective feedback was given in the form of text and graphics. As our goal was to collect data internationally, we used commonly known English phrases and simple, culturally neutral emoticons and pictures in the treatments. Our choice of subjective feedback types was motivated by previous evidence on their impact. Supportive feedback, or encouragement, has been shown to positively impact females' participation in competitive settings (Unkivoc et al 2016, Kahn et al 2014). The phrases in this treatment referred to expressions of support regarding future performance ("You can do it!") and acknowledgment of effort ("Great effort!").

Rewarding feedback, or praise of performance, has been shown to have a less beneficial impact on females: studies suggest that performance-based feedback (praise) may have a negative effect on women, and is less beneficial compared to effort-based feedback (Zeldin and Pajares 2000; Roberts and Nolen-Hoeksema 1989). On the other hand, such feedback can motivate individuals,

and has shown to impact performance as a verbal reward (Ariely 2016). The phrases in the rewarding feedback treatment included generally used positive valuations of past performance ("Good job!").

Finally, we decided to include a "trash talk" treatment. This was motivated by previous evidence on its impact as a motivator (Yip et al 2017), its widespread use in gaming, as well as the feedback we received from participants in our pilot experiment (Lovasz et al 2017). Several male testers noted that they are motivated by this type of verbal feedback, termed "competitive incivility." This treatment included phrases such as "Is that the best you can do?" and "Are you asleep?!", meant to capture the spirit of such feedback. However, as our results show, the phrases were likely not uncivil and/or humorous enough to motivate those with a preference for this type of communication in gaming.

Table 2: Treatment specifications and timing

Group number	Leaderboard	Subjective Feedback type	Graphic 1	Graphic 2	Graphic 3	Graphic 4	Graphic 5
Timing	Beginning, end		Beginning	2nd shape change	5th shape change	8th shape change	10th shape change
1	No	None					
2	No	Supportive feedback					
3	No	Rewarding feedback					
4	No	Trash talk					
5	Yes	None					
6	Yes	Supportive feedback					
7	Yes	Rewarding feedback					
8	Yes	Trash talk					

2.2. Data

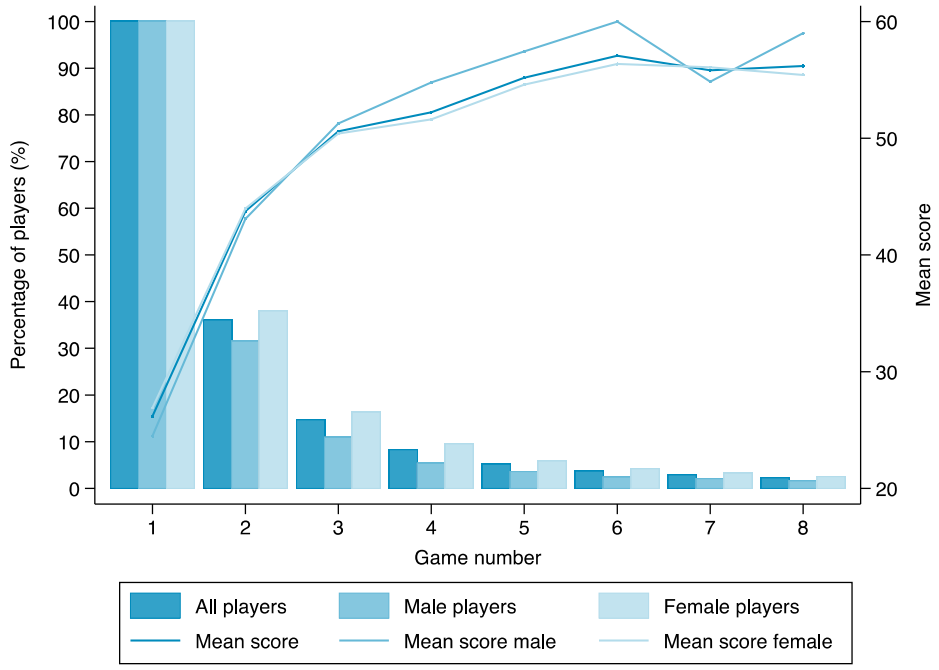
The Shape Game is freely available on a website.[10] Participants were recruited using paid social media advertisements, targeted at the age group of 18-45-yearolds, from European countries, where we found the response rate to be highest in the pilot experiment. The resulting data sample is comprised of 5191 individuals, who played a total of 9557 games. It is important to note that different players played a different number of games. During a single gaming session – defined as all the games played in a single web browser session – players received the same treatment. A small portion of players returned for a second gaming session, but we only included the first

[10] https://experimental-games.herokuapp.com

session of each player in our sample. This means that randomization took place at the player level, and players received the same treatment throughout their observed session. This allows us to study longer-run impacts on persistence and performance.

The data was collected at the event level, meaning we observed every click made by players as well as target shape changes and feedback shown. This information on player behavior (clicks) and performance (score) is linked to the demographic and other individual information given in the survey, and the automatically collected technical data on device type (as well as screen size). The event-level data was aggregated to the game level (game total clicks, end score), and to the individual player level (number of games played, total score, mean game score, and best game score, total clicks, accuracy). We focus our analysis on player level outcomes: number of games played as a measure of persistence, and mean game score, as well as best game score and player accuracy (score/clicks) as measures of performance in the session. Figure 2 shows the ratio of players by game number, as well as the mean score by game number. It shows that only about a third of players played a second game, and the ratio of players decreases sharply with game number. We can also see a sharp increase in mean score for those who play further after the first game, which is in line with learning.

Figure 2: Number of games played and mean game score



The descriptive statistics of the estimation sample are presented in Table 3. These show that the sample is generally skewed towards younger individuals with higher education. About 55 percent of players are aged 24 or below, and about 78-80 percent have secondary or higher level of education. The sample consists of only around 27 percent frequent game players, and the majority of players consider themselves to be "okay" at playing games (55 percent), only 7 percent consider themselves to be excellent. It is likely that serious gamers are less likely to click on an ad for a game in their social media feed. The randomization among treatment groups is supported by the balanced distribution of characteristics.

Table 3 also shows that our sample is heavily skewed towards female players, who comprise about two thirds of the observations. This suggests a selection bias in our sample. Assuming that the social media ads were shown equally to males and females, females were twice as likely to click on the ad, fill out the survey, and play the game compared to males. This may be due to the above-mentioned reason: more males are frequent game players, who could be less likely to try out such a simple game, or to try a game based on a social media ad. Another reason for this selection could be that there is a gender difference in the willingness to support research activities. As the game gives no financial rewards, participation is based on intrinsic motivators, as is usual in the market for games. While we cannot say with certainty what the underlying selection mechanisms are, it is likely that the sample tends to be under-representative of highly competitive, frequent gamers, which impacts the male subsample more than the female subsample. The unbalanced gender composition of our sample means that our results are not representative of the entire population, which may impact our estimates of the treatment effects. If the most competitive, frequent gamers tend to be excluded, we are likely underestimating the impact of competition as a motivator, especially among males.

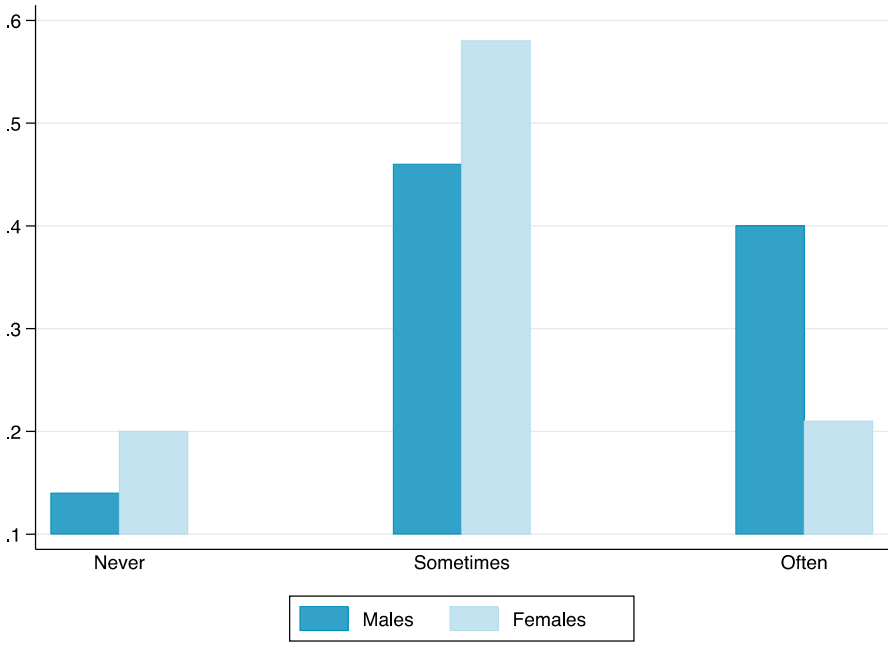
Table 3: Descriptive statistics of the sample

	Total	1. No LB + no FB (control)	2. No LB + supportive FB	3. No LB + rewarding FB	4. No LB + trash talk FB	5. LB + no FB	6. LB + supportive FB	7. LB + rewarding FB	8. LB + trash talk FB
N (individuals)	5191	644	688	655	620	652	644	659	629
N (games)	9557	1110	1167	1114	1037	1310	1312	1372	1135
Number of games played	1.84	1.72	1.70	1.70	1.67	2.01	2.04	2.08	1.80
Female	0.70	0.69	0.70	0.69	0.71	0.69	0.71	0.68	0.72
Age									
>17	0.25	0.24	0.26	0.26	0.26	0.27	0.22	0.26	0.25
18-24	0.30	0.30	0.29	0.30	0.30	0.34	0.30	0.27	0.30
25-34	0.18	0.18	0.17	0.18	0.18	0.15	0.21	0.19	0.18
35-44	0.13	0.15	0.14	0.13	0.12	0.12	0.14	0.14	0.14
45-64	0.11	0.10	0.12	0.11	0.10	0.11	0.11	0.11	0.11
<65	0.02	0.02	0.02	0.03	0.04	0.02	0.02	0.03	0.02
Education									
Elementary	0.18	0.17	0.18	0.17	0.21	0.17	0.18	0.17	0.20
Secondary	0.36	0.32	0.36	0.36	0.33	0.41	0.34	0.36	0.37
College or university	0.46	0.51	0.46	0.46	0.46	0.42	0.48	0.47	0.43
Plays games often									
Never	0.18	0.18	0.19	0.18	0.19	0.17	0.21	0.16	0.21
Sometimes	0.55	0.55	0.56	0.53	0.55	0.55	0.52	0.58	0.53
Often	0.27	0.27	0.26	0.29	0.25	0.28	0.27	0.26	0.26
Confidence in game playing									
Very bad	0.07	0.06	0.08	0.07	0.08	0.06	0.08	0.05	0.06
Pretty bad	0.14	0.15	0.15	0.12	0.16	0.16	0.14	0.14	0.15
Ok	0.54	0.55	0.51	0.56	0.53	0.52	0.55	0.54	0.56
Pretty good	0.18	0.17	0.21	0.17	0.17	0.19	0.17	0.19	0.17
Excellent	0.07	0.07	0.06	0.08	0.06	0.07	0.06	0.08	0.07
Region									
Hungary	0.46	0.46	0.44	0.46	0.49	0.47	0.45	0.47	0.49
North America	0.06	0.06	0.06	0.06	0.07	0.06	0.05	0.05	0.05
Other	0.24	0.24	0.26	0.25	0.23	0.21	0.24	0.24	0.23
Poland, Czech Republic, Slovakia	0.18	0.18	0.17	0.17	0.17	0.18	0.19	0.18	0.17
Western Europe	0.07	0.07	0.07	0.06	0.05	0.08	0.07	0.07	0.06
Touchscreen	0.69	0.68	0.67	0.71	0.70	0.71	0.68	0.71	0.68

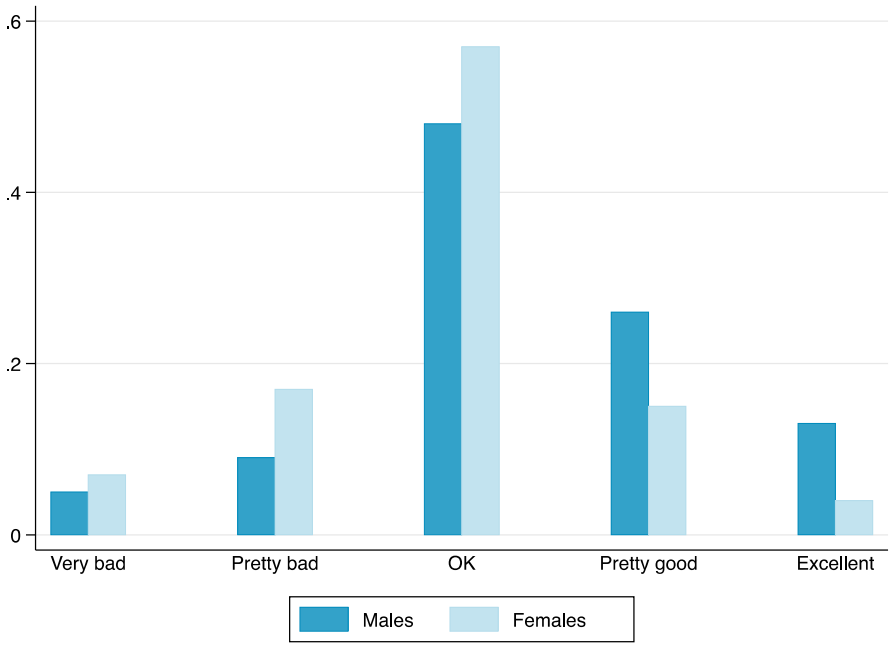
Appendix Table A1 shows the descriptive statistics of the sample by gender. These support that randomization was well balanced within gender as well. Figure 3 depicts the distributions of game playing frequency and self-reported game playing confidence by gender. The distributions suggest that even in our selected sample, males tend to play computer games more often than females, and they tend to have higher confidence in their game-playing ability than females. As we will see in the next section, the lower confidence of females is not in line with their actual performance, suggesting that they tend to undervalue their ability.

Figure 3: Frequency of game playing and self-reported confidence by gender

a. How often do you play computer games?



b. How good are you at playing computer games?



2.3. Empirical Analysis

The empirical analysis relies on OLS regressions for the estimation of the various treatment effects. We estimate the impact of each treatment compared to the control group, who saw no leaderboard and did not receive any subjective feedback. The estimated equations also control for the observable characteristics in our data seen in Table 3: the age, country, and education level of the individual, whether they are playing on a touchscreen device, and their screen size. Controlling for these characteristics should only impact estimates if the sample size is not large enough to guarantee the randomness among groups in terms of individual characteristics, or if there is some problem with the randomization. The results shown do not differ significantly from treatment effects estimated without the control variables, or from simple comparisons of means by group.

We estimate the effects of the various treatments on player-session level outcomes: score in the first game, the number of games the player played in the session, the mean game score, the best score they achieved in the session, and accuracy in the session. We use the pooled sample of all treatment groups in our estimation. We include dummy variables indicating whether a player saw a leaderboard, whether they received one of the subjective feedback types, the player's gender, and

the interactions of these as explanatory variables. Including the interaction terms makes it possible to estimate the combined effects of competition and feedback and the differential impacts by gender. The estimated regressions are of the following form:

$$outcome_i = \alpha_0 + \alpha_1 \cdot leaderboard_i + \alpha_2 \cdot female_i + \alpha_3 \cdot leaderboard_i \cdot female_i + \alpha_4 \cdot supportive\ FB_i + \alpha_5 \cdot rewarding\ FB_i + \alpha_6 \cdot trashtalk\ FB_i + \alpha_7 \cdot supportive\ FB_i \cdot leaderboard_i + \alpha_8 \cdot rewarding\ FB_i \cdot leaderboard_i + \alpha_9 \cdot trashtalk\ FB_i \cdot leaderboard_i + \alpha_{10} \cdot supportive\ FB_i \cdot female_i + \alpha_{11} \cdot rewarding\ FB_i \cdot female_i + \alpha_{12} \cdot trashtalk\ FB_i \cdot female_i + \alpha_{13} \cdot supportive\ FB_i \cdot leaderboard_i \cdot female_i + \alpha_{14} \cdot rewarding\ FB_i \cdot leaderboard_i \cdot female_i + \alpha_{15} \cdot trashtalk\ FB_i \cdot leaderboard_i \cdot female_i + \alpha_{16}' X_i + \vartheta_i \quad (1)$$

where $outcome_i$ represents the various player-session level outcome variables for individual i, and X_i represents control variables (age group, education level, region, touchscreen, screen size).

The coefficient estimates are then used to calculate the different treatment effects and their significance: for the leaderboard when no subjective feedback is given, the effects of the three subjective feedback types, and the combined effects of the leaderboard and each subjective feedback type. They can also be used to calculate gender differences in the treatment effects and in mean outcomes. Table 4 summarizes the calculations based on the linear combinations of the coefficient estimates.

Table 4: Calculation of treatment effects and gender gaps based on the OLS coefficient estimates

Effect	Linear combination of coefficients
Treatment effect of leaderboard with no feedback, males	leaderboard
Treatment effect of leaderboard with no feedback, females	leaderboard+female*leaderboard
Treatment effect of leaderboard and feedback, males	leaderboard+feedback+leaderboard*feedback
Treatment effect of leaderboard and feedback, females	leaderboard+female*leaderboard+feedback+female*feedback +leaderboard*feedback+female*leaderboard*feedback
Gender difference in the treatment effect of leaderboard (with no feedback)	female*leaderboard
Gender difference in the treatment effect of leaderboard and feedback	female*leaderboard+female*feedback+female*leaderboard*feedback
Baseline gender gap in outcomes	female
Gender gap in outcomes with leaderboard and no feedback	female+female*leaderboard
Gender gap in outcomes with leaderboard and feedback	female+female*leaderboard+female*feedback+female*leaderboard*feedback

Note: Treatment effects calculated separately for each subjective feedback type, "feedback" refers to the given type of feedback.

2.4 Robustness and Relevance

We performed several robustness checks to assess the estimation results. We calculated mean outcomes for each treatment group by gender, and tested the significance of the differences. We estimated the OLS equations with no controls. We estimated the impacts of competition and the feedback types on subsamples, rather than pooling all groups.[11] Additionally, we estimated the effects on subsamples by gender, using the pooled groups specification of equation 1. Estimates based on subsamples had lower significance levels due to the smaller sample sizes, but were comparable to the main results. We also checked the sensitivity of the results to certain sample restrictions: dropping players who clicked less than 3 (or 5) times in the game, and dropping observations with extremely high scores (two players achieved scores above 150). Finally, we estimated the impact of treatments using the game level dataset, on game level performance (game end score). The game level analysis (Appendix Table A4) yielded higher significance due to the sample size, but is complicated by the fact that different players play different number of games, which is also a potential impact of treatment. In these specifications, we calculate clustered standard errors. All checks supported the robustness of the conclusions described here, and are available upon request.

It is important to discuss the relevance of our estimates. The use of an online game represents a sort of lab in the field method – as discussed in Gneezy and Imas (2017) - in the sense that it allows us to maintain experimental control while observing real life behavior in a natural setting. Although this is not a labor market setting, the results reflect the impact on individual behavior when facing a new task. Gaining a deeper understanding of such behavior is important, because attitudes towards new challenges have been shown to be key lifelong determinants of educational achievement (Henderson and Dweck 1991; Hong et al. 1999), and to impact gender differences in career choices and outcomes (Lloyd, Walsh, and Yailagh 2005; Dweck 2006).

Several aspects of the task environment and experimental design are key to the interpretation and external relevance of the results. First, we observe changes in behavior in the short-run, based on short-term interactions. These short-term estimation results cannot be directly extended to longer-

[11] For example, subsamples containing two groups for the impact of the leaderboard (Group 1 and Group 5), the impact of supportive feedback (Group 1 and Group 2), or the joint impact of leaderboard and supportive feedback (Group 1 and Group 6). We also tested subsamples with four groups, for example, the impact of a leaderboard and supportive feedback based on Groups 1, 2, 5, and 6.

term interactions. For example, it is possible that in the longer run, the supportive feedback given becomes routine, loses its credibility, and thus becomes less effective. On the other hand, it could be that in the longer-run, such feedback may help build a stronger relationship with the supervisor, and thus become more important and effective.

Second, the source of the subjective feedback in the game is clearly pre-programmed, not a real-life supervisor. It is therefore not given by someone whose opinion may be important to individuals and their success. Players play the game within their homes, anonymously. In general, the stakes of effort (playing) and performance are relatively low. The time and energy costs of playing are minor, just 2 minutes of clicking. There are no financial rewards or prizes. These factors may influence both the perception of the feedback and its effect. Individual reactions would likely differ in the case of more tailored, personal feedback received from a real-life supervisor, as the stakes would be higher: it is more important to impress a supervisor than to perform well in an anonymous online environment.

Third, the results are also likely to be task-specific. The game is a relatively fun task, rather than a tedious one. However, it is also somewhat difficult and it requires concentration. We chose such a task because our goal was to measure differences in reactions to feedback when individuals are faced with a somewhat challenging task, where confidence and performance expectations are likely to play a role. The effects are likely to be very different in a setting where the tedious or unpleasant nature of the task leads to lower effort, rather than a lack of confidence. Online gaming in general, and visual perception in particular, are often considered to be stereotypically male tasks. Gender differences in effort and performance are generally smaller in tasks perceived as stereotypically more female (Niederle 2016).

Fourth, as mentioned earlier, the sampling method used – online advertising - is also likely to impact the results. Out of the potential pool of those who saw the ad for the game, the sample of players is comprised of those who chose to play. As we saw, our sample has a heavily skewed gender distribution, suggesting different selection mechanisms by gender. This is important to keep in mind throughout the analysis. Arechar et al. (2017) discuss the benefits and problems of online experiments,[12] and conclude that data collected from these can be reliable and provide the

[12] The authors discuss the available evidence and carry out a well-known experiment using both a laboratory setting and the online labor market platform Amazon Mechanical Turk. They find that basic behavioral patterns are replicable

basis for valuable contributions to the empirical evidence. They also highlight the potential selection bias due to participant dropouts as the most important issue. This is a potential limitation on the external relevance of our results: the behavior of individuals who are obligated to take part in a task, and their reaction to subjective feedback, may be different. However, other methods of data collection, such as incentivized laboratory experiments, are also subject to the selection of individuals who are willing to participate.

Finally, our results may be specific to the given subjective content (phrases, graphics, etc.) in the experiment. The goal of this study is not to provide specific suggestions for subjective content. It is rather to highlight the heterogeneity in its impact, and the importance of this heterogeneity as a potential contributor to gender inequalities in outcomes. We suggest that supervisory communication should be more individualized, rather than argue for the uniform or gender-targeted provision of any particular subjective content. These implications are less dependent on the external relevance of our specific estimates, but are rather supported by the evidence of heterogeneity in the impacts.

3. Results

In this section, we first look at persistence and performance outcomes by gender in the baseline case. We then look at the estimated treatment effects of seeing a leaderboard or receiving one of the three subjective feedback types. Next, we look at the combined treatment effects of seeing a leaderboard and receiving one of the feedback types. Finally, we discuss the overall implications regarding gender differences in outcomes under various treatment schemes. We present simple figures depicting mean outcomes and/or treatment effect estimates, where the significance of the estimates is indicated by shading. The full OLS results - corresponding to the estimations of equation 1 with the various outcome measures as the dependent variables - can be seen in Appendix Table A3.

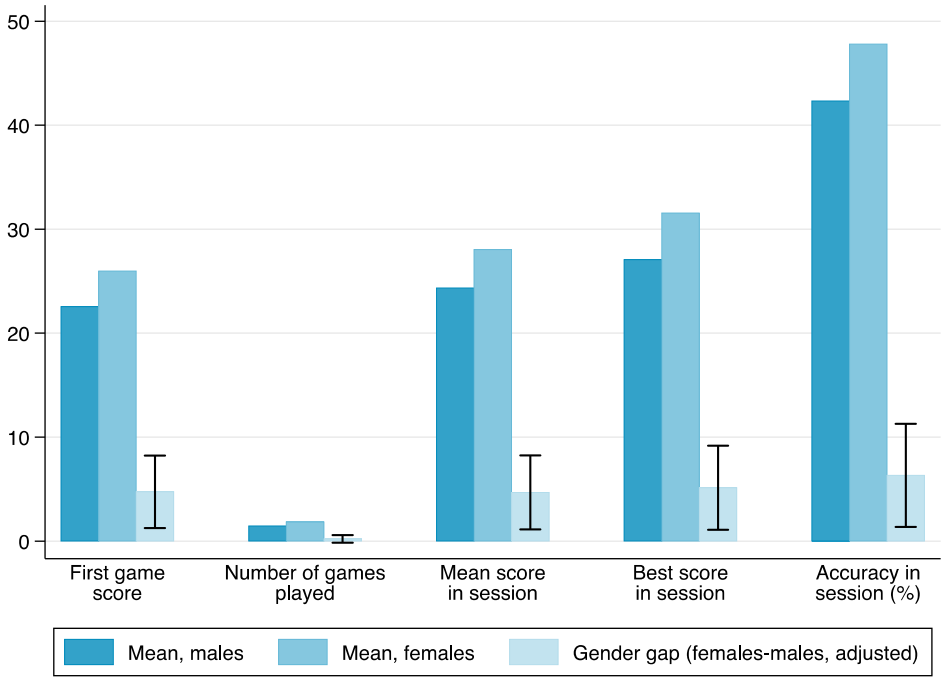
3.1. Baseline Outcomes and Gender Gaps

online, and conclude that online experiments can provide data of adequate quality and are a valuable complement to laboratory studies.

We first assess the various outcomes by gender in the baseline case, where players see no leaderboard, and do not receive any subjective feedback (Group 1). Figure 4 depicts the mean outcomes by gender and the gender difference in each outcome, for the players' first game score, number of games played in the session, mean score in the session, best score in the session, and accuracy (score/clicks) in the session. We can see that in the baseline group, female players have similar persistence (number of games played), but better performance, indicated by higher scores in the first game and session, and better accuracy.

The performance advantage of females is likely related to selection into our sample, as discussed earlier. If a higher ratio of males in the population are serious (high ability) gamers compared to females, and serious gamers are less likely to play such a game, this could lead to the higher mean female scores we observe in our sample. It is also possible that the males in our sample are simply not motivated to play as well when leaderboard is not shown, while females are more motivated. As there is no previous evidence suggesting that males have lower ability in this type of task (Shaquiri et al 2016), the gender gap observed in favor of females is likely reflective of one of these two mechanisms.

Figure 4: Mean outcomes by gender, baseline group



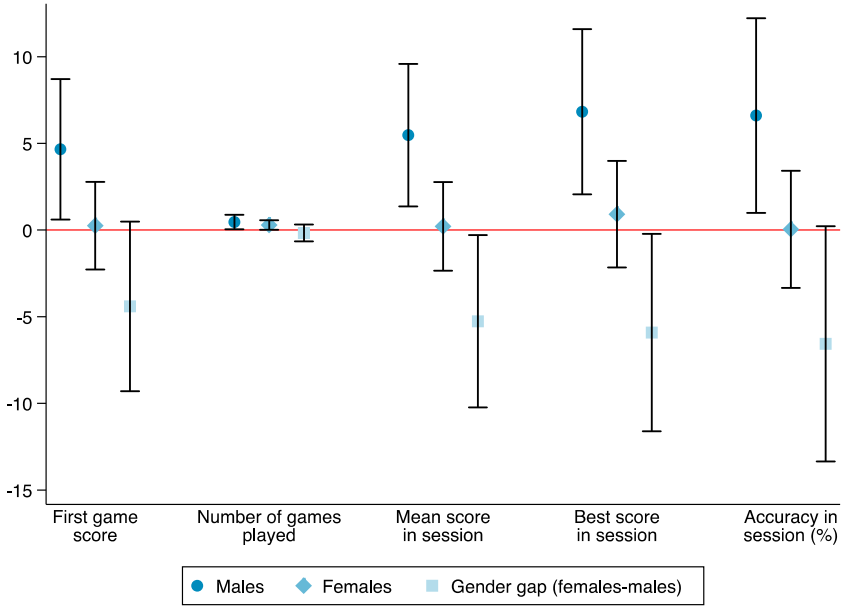
Notes: 95% confidence intervals are shown for the gender gaps. Detailed information on the coefficients can be found in Tables A2 and A3.

3.2. The Impact of Competition

Given the baseline performance advantage of female players, we next estimate the impact of competition when there is no subjective feedback given. The estimated treatment effects are shown in Figure 5 and are calculated as the mean performance difference between the treatment (leaderboard, no feedback) and no treatment (no leaderboard, no feedback) groups. The results suggest that both genders increase the number of games they play - their persistence - when they see a leaderboard at the beginning and end of each game, by about 0.3-0.45 games (significant at the 5%). Though the increase is higher for males, there is no significant difference in the impact by gender. The performance measures, on the other hand, indicate clear gender differences in the response to competition. Males' performance increases in the first game in terms of score and accuracy: their mean scores and best scores improve by about 5-7 points. This is a significant magnitude: compared to the baseline mean score of around 26, it represents a positive impact of around 19-27 percent. Females' performance, on the other hand, does not increase as a result of

being shown a leaderboard. Even though they play more games, their accuracy and scores do not benefit from competition. The results are in line with previous evidence pointing to the performance disadvantage of females in competitive settings, though the difference in the impact on persistence is not significant in our experiment.

Figure 5: Treatment Effects of Competition (Leaderboard) with no Subjective Feedback



Notes: 95% confidence intervals are shown for the treatment effect estimates and the gender gaps in the treatment effects. Detailed information on the coefficients can be found in Tables A2 and A3.

3.3. The Effects of Subjective Feedback Types and Competition

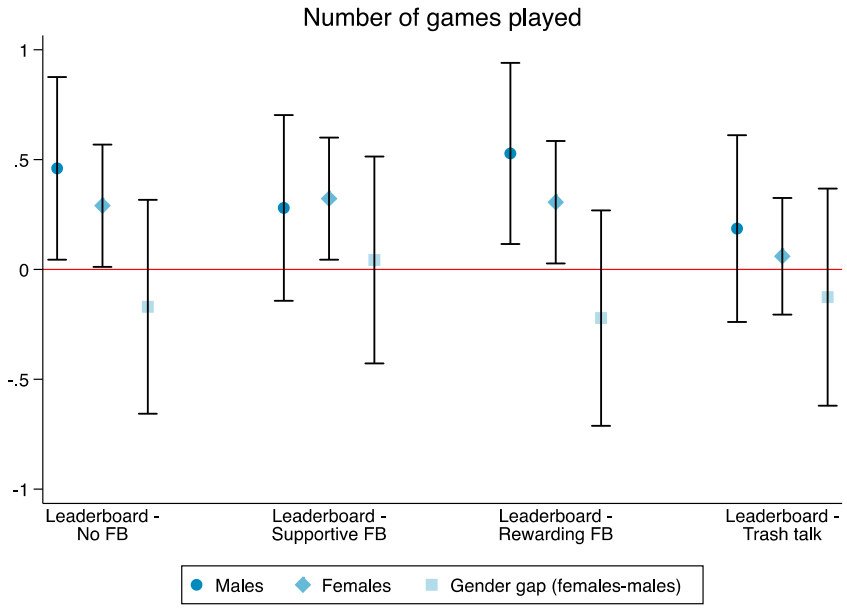
Appendix Figure A2 shows the treatment effects of the three subjective feedback types when there is no competition. Overall, the three feedback types have small or zero impacts on persistence and performance when given in a setting without competition. None of the feedback types impact the number of games played. The estimates of the impact of supportive feedback on performance of both males and females are positive, but insignificant. Rewarding feedback (praise) seems to have the strongest effect among the three feedback types. It has a positive impact on the performance of both genders, though it is stronger for males. However, the effects are only significant for the first game score. Trash talk has the least impact of all. The purpose of this feedback treatment was to mimic the competitive incivility often seen in online gaming, which can engender competitive

motivation (Yip 2017). Our finding of no impact does not provide evidence of the lack of this mechanism, it is likely due to our particular specifications, which were not believable or motivational enough. The inclusion of trash talk does provide a robustness check, showing that the specific content of the subjective feedback treatments matter.

Next, we turn our attention to our main question. Given the disadvantage of females in the competitive setting, can better outcomes be achieved if competition is paired with subjective feedback? Appendix Figure A3 shows the combined impacts of seeing a leaderboard and receiving subjective feedback on the full set of outcomes, separately for the three subjective feedback types. Figures 6 and 7 focus on two outcomes representing persistence and performance, and compare the effects of the various combined treatments to the treatment with only a leaderboard.

We can see that seeing a leaderboard has a positive impact on persistence in general. For males, the positive impact is the highest and most significant when rewarding feedback is given in addition to the leaderboard. The effect is smaller and less significant when the leaderboard is combined with supportive feedback. For females, the impact is somewhat smaller in magnitude, and more stable across feedback types. Adding trash talk to the leaderboard appears to decrease the beneficial impact on persistence for both genders. Overall, the treatment scheme that achieves the most beneficial impact on persistence is the combination of a leaderboard and rewarding feedback for male players, while for females, any treatment scheme with a leaderboard increases persistence, with the exception of the scheme where it is paired with trash talk. The gender gaps in the treatment effects are not significant for any treatment group.

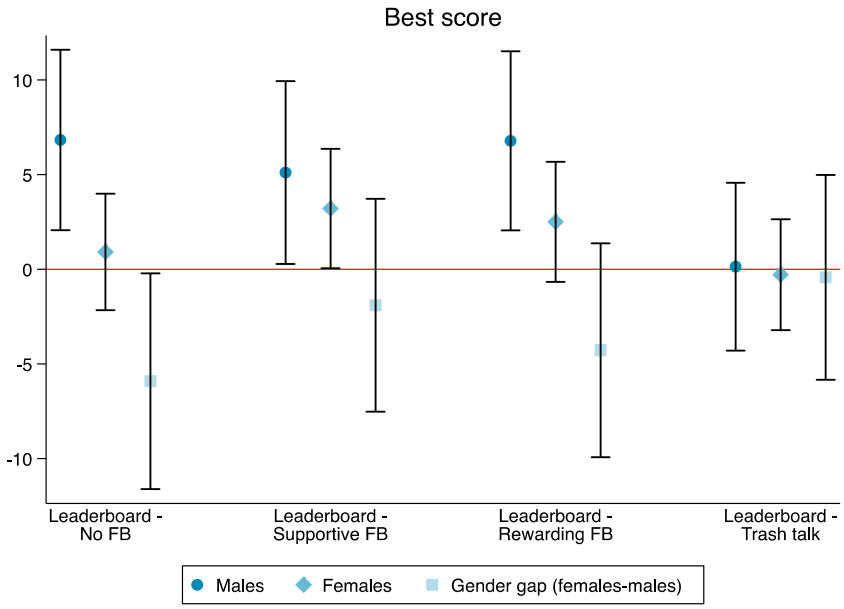
Figure 6: Treatment effects on persistence by gender



Notes: Persistence is measured as the number of games played by the player in the session. 95% confidence intervals are shown for the treatment effect estimates and the gender gaps in the treatment effects. "FB" refers to feedback. Detailed information on the coefficients can be found in Tables A2 and A3.

Figure 7 illustrates the treatment effects on performance (best score) separately for males and females. The results indicate significant gender differences in treatment effects by gender. Male players generally respond positively to seeing a leaderboard. They achieve higher scores in all treatments, with the exception of the treatment combining the leaderboard with trash talk. Adding trash talk appears to counteract the beneficial impact of competition. For female players, as we saw earlier, seeing a leaderboard does not in itself increase performance. However, the combined impact of a leaderboard and supportive feedback is positive and significant at the 5 percent level. The effect when the leaderboard is paired with rewarding feedback is positive but insignificant, while the combined treatment with trash talk has no impact. The figure also shows that the treatment with only a leaderboard has a significantly different impact by gender, favoring males, while the gender gaps in the other treatment effects are insignificant.

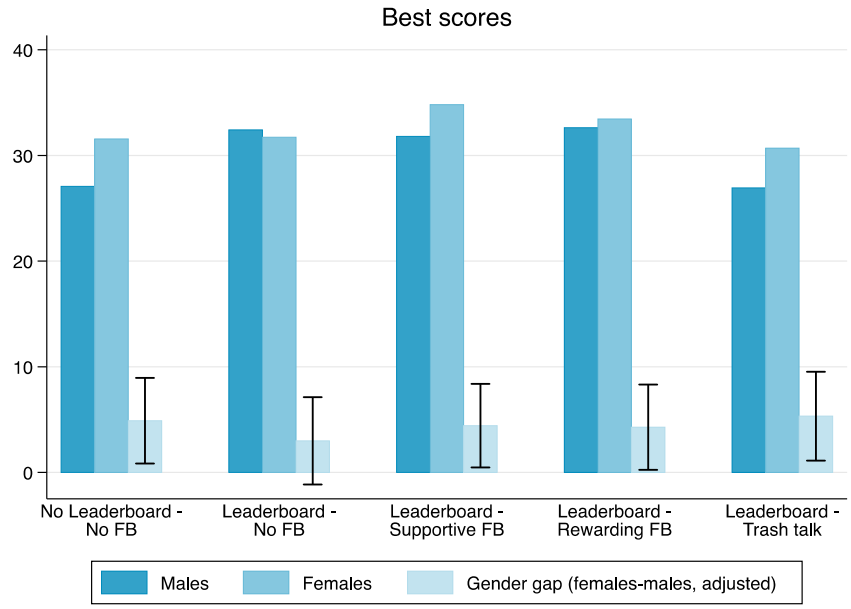
Figure 7: Treatment effects on performance (best score) by gender



Notes: Performance is measured by the best score achieved by the player in the session. 95% confidence intervals are shown for the treatment effect estimates and the gender gaps in the treatment effects. "FB" refers to feedback. Detailed information on the coefficients can be found in Tables A2 and A3.

Finally, we assess the levels of performance by gender for each treatment group in Figure 8, by looking at the means of the best session scores and the gender gap in these. We can see the previously documented performance advantage of female players in the baseline group. This advantage disappears when competition is added, due to the fact that the mean score of males increases, while that of females does not. The combined treatment of a leaderboard and supportive feedback, on the other hand, increases the mean scores of both males and females compared to the baseline group. We again see a gender performance gap in favor of females as a result. When competition is combined with rewarding feedback, females' performance does not increase much compared to the baseline, therefore, their performance advantage decreases to close to zero. The combined treatment with trash talk decreases mean scores (insignificantly), and preserves the baseline performance advantage of females.

Figure 8: Performance levels (best score) by gender and treatment group



Notes: 95% confidence intervals are shown for the treatment effect estimates and the gender gaps in the treatment effects. "FB" refers to feedback. Detailed information on the coefficients can be found in Tables A2 and A3.

The results imply that there are important heterogeneities in the impacts of competition, subjective feedback types, and their combinations by gender. The magnitudes of the treatment effects and their gender differences are far from negligible. For example, the treatments with the most beneficial impacts by gender can improve scores by 6.5 for males (leaderboard alone or leaderboard combined with rewarding feedback) and 3.5 for females (leaderboard combined with supportive feedback), representing increases of about 24 percent and 11 percent over the baseline scores, respectively. In terms of gender gaps in outcomes, the baseline advantage of females is around 18 percent, which decreases to zero when a leaderboard is shown, but no subjective feedback is given. Whatever the reasons behind the baseline performance advantage of females, the best overall performance outcomes are achieved with different treatment schemes for males and for females. Male players can be motivated by competition alone, and this motivation translates to higher persistence and performance. Female players appear to be motivated by competition as well, increasing their persistence. However, they only increase their performance when supportive feedback is provided at the same time. This means that the parallel provision of supportive feedback can counteract some of the disadvantage of females in competitive settings.

4. Conclusion

In this experiment, we show that (1) the subjective content of supervisory feedback is an important factor for performance, (2) the impacts of competition and subjective feedback elements are interdependent, and (3) there are significant heterogeneities in the impacts by gender. Males generally respond positively to competition in the form of a public leaderboard, increasing their persistence and performance. Females, on the other hand, do not experience similar gains in performance, unless the leaderboard is combined with supportive feedback. The best outcomes are achieved under different feedback schemes for males and females. The scheme that has the most beneficial impact overall is that of competition combined with supportive feedback. However, the main implication of our findings is not that this scheme should be used uniformly, or that feedback should be targeted by gender. We argue that the results highlight the importance of individualized feedback as a potential tool for decreasing existing gender gaps in competitive settings.

The effects of competition and feedback are likely dependent on individual characteristics, such as confidence, as well as the task and its context. The gender differences in our results can be due to underlying differences in such characteristics. Further research is needed to better determine what feedback is best suited for different individuals, i.e., how feedback should be targeted. However, supervisors – teachers and managers – should be made aware of the performance gains they can achieve if they implement more individualized feedback policies. The future developments of personalized learning software and HR software may enable better targeting mechanisms and the more frequent provision of such feedback. These changes could potentially decrease gender gaps in competitive settings, and lead to lower inequality in high-paying STEM fields and occupations.

References

Arechar, A., Gächter, S. and Molleman, L. 2017. "Conducting Interactive Experiments Online." Experimental Economics, May 2017, 1–33.

Ariely, Dan. 2016. Payoff: The Hidden Logic That Shapes Our Motivations. TED Books.

Azmat, G. & Iriberri, N. (2012). The Provision of Relative Performance Feedback Information: An Experimental Analysis of Performance and Happiness.

Azmat, G., C. Calsamiglia and N. Iriberri (2015), Gender Differences in Response to Big Stakes, CEP Discussion Paper No. 1314.

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.

Bandiera, O. & Larcinese, V. & Rasul, Imran, 2015. "Blissful ignorance? A natural experiment on the effect of feedback on students' performance," Labour Economics, Elsevier, vol. 34(C), pp. 13-25.

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.

Bertrand, M; Goldin, C; Katz, L. F. (2010): Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. In American Economic Journal: Applied Economics 2 (3), pp. 228–255.

Bertrand, M. (2020). Gender in the 21st Century. Nashville, TN: American Economic Association [publisher], 2020. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-10-21.

Bettinger, E., Ludvigsen, S., Rege, M., Solli, I. F., Yeager, D. 2018. "Increasing perseverance in math: Evidence from a field experiment in Norway." Journal of Economic Behavior & Organization 146: 115.

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.

Chang, C, Lance F.D., Johnson, R.E., Rosen, C. and Tan, J.A. (2012). "Core Self-Evaluations." Journal of Management 38 (1): 81–128.

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.

Croson, Rachel; Gneezy, Uri (2009). Gender Differences in Preferences. In Journal of Economic Literature 47 (2), pp. 448–474.

Damgaard, M. T. and Nielsen, H. S. (2018). Nudging in education, Economics of Education Review, Volume 64, pp. 313-342.

Datta Gupta, N., Poulsen, A. and Villeval, M. C. (2005). Male and Female Competitive Behavior: Experimental Evidence. IZA Discussion Paper No. 1833.

Dweck, C. S. 2006. "Is Math a Gift? Beliefs That Put Females at Risk." In Why Aren't More Women in Science? Top Researchers Debate the Evidence, edited by S. J. Ceci and W. Williams, 47–55. Washington, DC: American Psychological Association.

Dweck, C.S. (2007). The secret to raising smart kids. Scientific American: Mind. December/January, 36-43.

Eckel, Catherine and Grossman, Philip, (2008), Differences in the Economic Decisions of Men and Women: Experimental Evidence, ch. 57, p. 509-519 in Plott, Charles R. and Smith, Vernon L. eds., Elsevier.

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, and Aldo Rustichini, (2004). "Gender and competition at a young age," American Economic Review Papers and Proceedings, May 2004, pp. 377-381.

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.

Gneezy, U., and Imas, A. (2017). "Chapter 10 - Lab in the Field: Measuring Preferences in the Wild." In Handbook of Economic Field Experiments, edited by Abhijit Vinayak Banerjee and Esther Duflo, 1:439–64. Handbook of Field Experiments. North-Holland.

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

Henderson, V., and C. S. Dweck. 1991. "Adolescence and Achievement." In: At the Threshold: Adolescent Development, edited by S. Shirley Feldman and Glen R. Elliott, 197–216. Harvard University Press.

Hong, Y.-y., C.-y. Chiu, C. S. Dweck, D. M.-S. Lin, and W. Wan. 1999. "Implicit Theories, Attributions, and Coping: A Meaning System Approach." Journal of Personality and Social Psychology 77 (3): 588– 99.

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.

Johnson, D. A. (2013). A Component Analysis of the Impact of Evaluative and Objective Feedback on Performance. Journal of Organizational Behavior Management, 33(2), 89–103.

Jurajda, Štěpán, and Daniel Münich. 2011. "Gender Gap in Performance under Competitive Pressure: Admissions to Czech Universities." American Economic Review, 101 (3): 514-18.

Kauhanen, A.; Napari (2015): Gender Differences in Careers. In Annals of Economics and Statistics (117/118), p. 61.

Kirkeboen, Lars J.; Leuven, Edwin; Mogstad, Magne (2016): Field of Study, Earnings, and Self-Selection. The Quarterly Journal of Economics 131 (3), pp. 1057–1111.

Khan, A. Ahmad, R. Hamdan, A.R. and Mustaffa, M.S. (2014). "Educational Encouragement, Parenting Styles, Gender and Ethnicity as Predictors of Academic Achievement among Special Education Students." International Education Studies 7 (2): 18.

Lloyd, J. E. V., J. Walsh, and M. S. Yailagh. (2005). "Sex Differences in Performance Attributions, Self-Efficacy, and Achievement in Mathematics: If I'm So Smart, Why Don't I Know It?" Canadian Journal of Education 28 (3): 384–408.

Lovász A., Cukrowska-Torzewska E., Rigó M., Szabó-Morvai Á. (2017) Gender Differences in the Effect of Subjective Feedback. Budapest Working Papers on the Labour Market BWP1705.

Luckin, R., Holmes, W., Griffiths, M., and Forcier, L. B. (2016). Intelligence Unleashed: An argument for AI in Education. Pearson Education, London.

Macis, Mario (2017): Gender differences in wages and leadership. In IZA World of Labor.

McCarty P.A., (1986). Effects of feedback on the self-confidence of men and women. Academy of Management Journal 29(4): 840-847.

Mobius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. 2011. Managing Self-Confidence: Theory and Experimental Evidence. 17014. National Bureau of Economic Research.

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, and Lise Vesterlund. (2010). "Explaining the Gender Gap in Math Test Scores: The Role of Competition." Journal of Economic Perspectives, 24(2): 129–44.

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

Niederle, Muriel (2016). "Gender." In Handbook of Experimental Economics, edited by Alvin E. Roth and John H. Kagel, 2nd ed., 481–553. Princeton University Press.

Narciss, S. Sosnovsky, S. Schnaubert, L. Andrès, E. Eichelmann, A. Goguadze, G. Melis, E. (2014) Exploring feedback and student characteristics relevant for personalizing feedback strategies, Computers & Education, Vol. 71: 56-76.

Ors, E, Palomino, F. and Peyrache, E. (2013). Performance Gender Gap: Does Competition Matter? Journal of Labor Economics 31 (3): 443–99.

Osborne, Jonathan; Simon, Shirley; Collins, Sue (2003): Attitudes towards science: A review of the literature and its implications. In International Journal of Science Education 25 (9), pp. 1049–1079.

Perrotta, C. and Selwyn, N. (2019). Deep learning goes to school: toward a relational understanding of AI in education. Learning, Media and Technology, 45(3), pp. 251-269.

Posner, B.Z., and Kouzes, J.M. (1999). Encouraging the Heart: A Leader's Guide to Rewarding and Recognizing Others. Jossey-Bass: Wiley.

Roberts, Tomi-Ann, and Susan Nolen-Hoeksema. 1989. "Sex Differences in Reactions to Evaluative Feedback." Sex Roles 21 (11–12): 725–47.

Shaqiri, A., Andreas Brand; Maya Roinishvili; Marina Kunchulia; Guillaume Sierro; Julie Willemin; Eka Chkonia; Luisa Iannantuoni; Karin Pilz; Christine Mohr; Michael Herzog (2016). Gender differences in visual perception. Journal of Vision September 2016, Vol.16, 207.

Unkovic C, Sen M, Quinn K.M. (2016) Does Encouragement Matter in Improving Gender Imbalances in Technical Fields? Evidence from a Randomized Controlled Trial. PLoS ONE 11(4): e0151714.

Wong, Y. Joel. (2015). "The Psychology of Encouragement: Theory, Research, and Applications Ψ." The Counseling Psychologist 43 (2): 178–216.

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.

Wozniak, David, William T. Harbaugh, and Ulrich Mayr. (2016). "The Effect of Feedback on Gender Differences in Competitive Choices." SSRN Scholarly Paper ID 1976073. Rochester, NY: Social Science Research Network.

Yip, J. A., Schweitzer, M. E., & Nurmohamed, S. (2017). Trash-Talking: Competitive Incivility Motivates Rivalry, Performance, and Unethical Behavior. Organizational Behavior and Human Decision Processes, Vol. 144, pp. 125-144.

Zeldin, A.L., and Pajares, F. (2000). "Against the Odds: Self-Efficacy Beliefs of Women in Mathematical, Scientific, and Technological Careers." American Educational Research Journal 37 (1): 215–46.

Appendix

For Online Publication

Figure A1

The pre-game survey

1. What's your name / nick name?
2. How old are you?
3. Gender

□ Male □ Female

4. Where are you from?
5. How often do you play computer games?

□ Often □ Sometimes □ Never

6. Are you good at playing computer games?

□ I am excellent □ I am pretty good □ I am OK

□ I am pretty bad □ I am very bad

7. Have you played this game before?

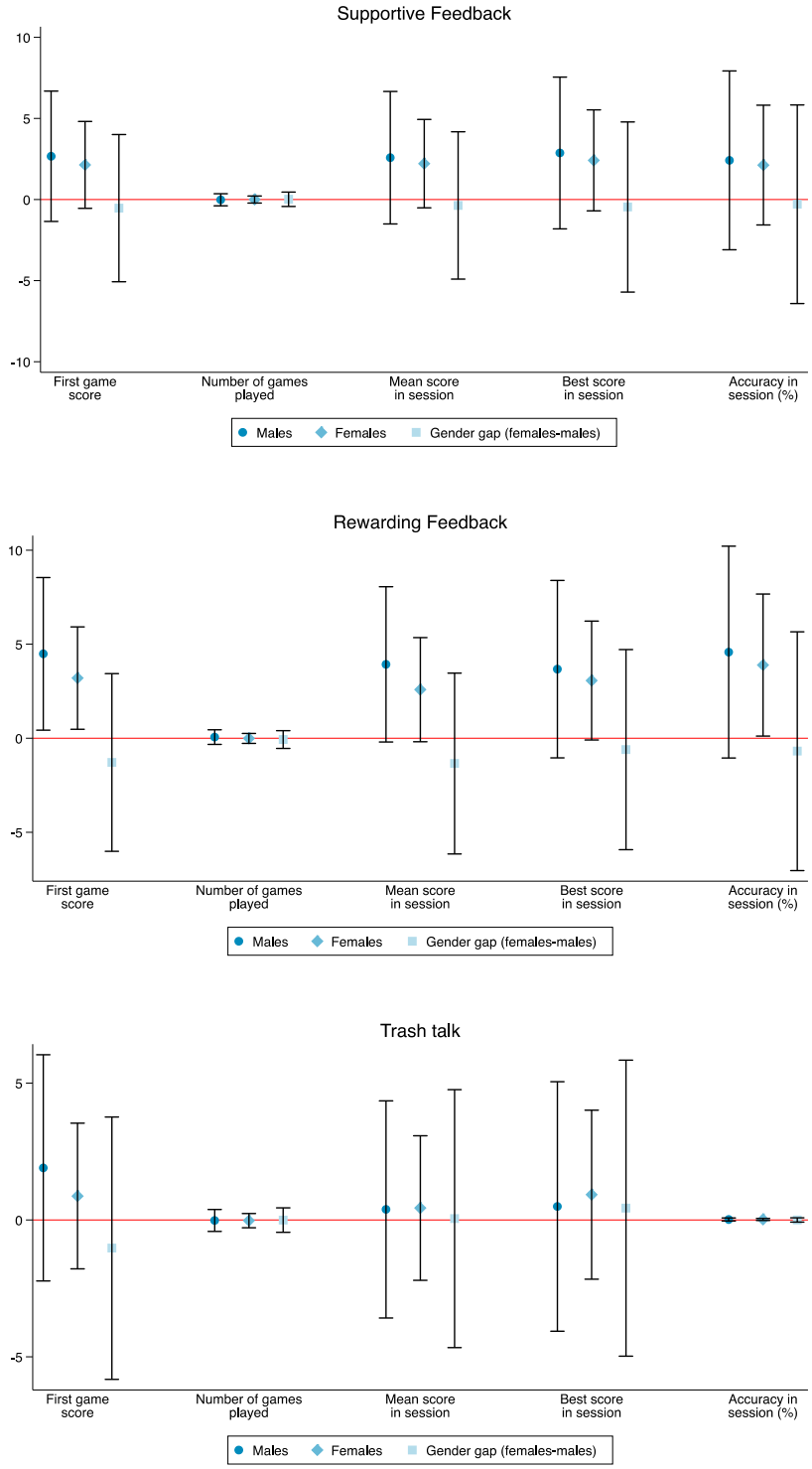
□ Yes □ No

8. What is the highest level of education you have completed or are pursuing?

□ Elementary □ Secondary (Vocational or High School)

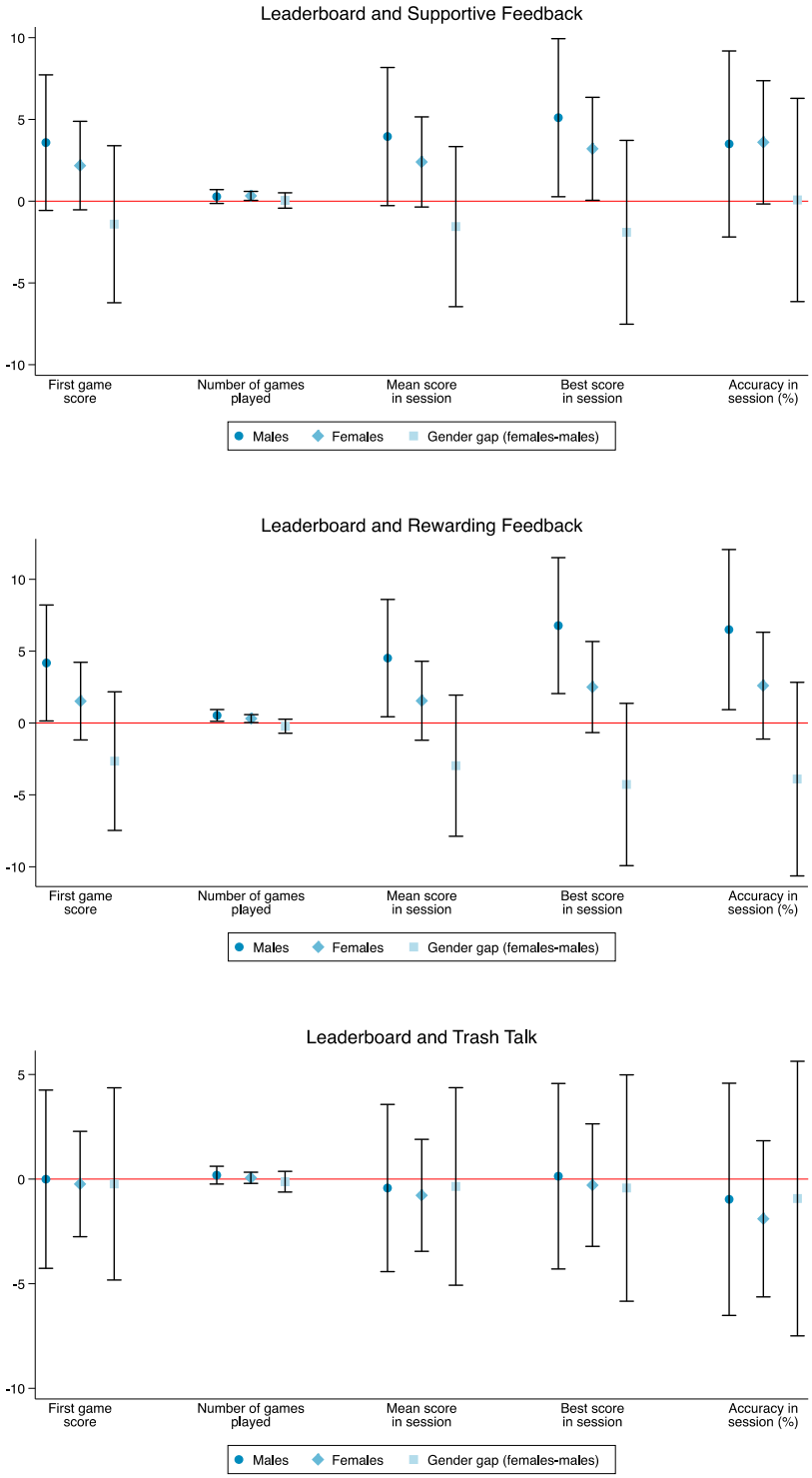
□ Higher education (College or University)

Figure A2: Treatment Effects of the Three Subjective Feedback Types



Notes: 95% confidence intervals are shown for the treatment effect estimates and the gender gaps in the treatment effects. Detailed information on the coefficients can be found in Tables A2 and A3.

Figure A3: Combined treatment effects of a leaderboard and subjective feedback



Notes: 95% confidence intervals are shown for the treatment effect estimates and the gender gaps in the treatment effects. "FB" refers to feedback. Detailed information on the coefficients can be found in Tables A2 and A3.

Table A1: Descriptive statistics of the sample by gender

		Treatment Group							
	Total	**1. No LB, no FB**	**2. No LB, supp. FB**	**3. No LB, rew. FB**	**4. No LB, trash FB**	**5. LB, no FB**	**6. LB, supp. FB**	**7. LB, rew. FB**	**8. LB, trash FB**
N (individuals)	5191	644	688	655	620	652	644	659	629
N (games)	9557	1110	1167	1114	1037	1310	1312	1372	1135
Female	1.84	1.72	1.70	1.70	1.67	2.01	2.04	2.08	1.80
Male									
N (individuals)	1556	201	206	201	178	202	185	210	173
N (games)	2544	293	296	300	251	380	322	415	287
Number of games/player	1.635	1.46	1.44	1.49	1.41	1.88	1.74	1.98	1.66
Age									
>17	0.31	0.29	0.30	0.33	0.31	0.35	0.28	0.32	0.31
18-24	0.30	0.35	0.31	0.30	0.28	0.31	0.34	0.27	0.27
25-34	0.16	0.15	0.17	0.12	0.16	0.13	0.16	0.17	0.17
35-44	0.10	0.08	0.11	0.12	0.09	0.08	0.11	0.09	0.10
45-64	0.09	0.07	0.09	0.08	0.08	0.10	0.08	0.10	0.12
<65	0.04	0.05	0.02	0.03	0.08	0.02	0.04	0.06	0.02
Education									
Elementary	0.20	0.18	0.23	0.18	0.26	0.19	0.20	0.17	0.19
Secondary	0.36	0.32	0.32	0.38	0.31	0.43	0.36	0.39	0.41
College or university	0.43	0.50	0.45	0.44	0.43	0.38	0.44	0.44	0.40
Plays games									
Never	0.14	0.10	0.17	0.17	0.12	0.12	0.16	0.12	0.14
Sometimes	0.46	0.49	0.41	0.37	0.54	0.45	0.44	0.51	0.50
Often	0.40	0.41	0.41	0.46	0.33	0.43	0.39	0.36	0.36
Confidence in game playing									
Very bad	0.05	0.04	0.07	0.06	0.06	0.04	0.06	0.05	0.02
Pretty bad	0.09	0.08	0.10	0.07	0.11	0.09	0.10	0.05	0.13
Ok	0.47	0.54	0.40	0.44	0.48	0.48	0.51	0.48	0.46
Pretty good	0.25	0.21	0.29	0.26	0.24	0.27	0.22	0.28	0.26
Excellent	0.13	0.12	0.14	0.17	0.11	0.11	0.11	0.14	0.12
Region									
Hungary	0.37	0.39	0.32	0.34	0.39	0.42	0.36	0.38	0.40
North America	0.06	0.06	0.08	0.07	0.08	0.05	0.05	0.06	0.06
Other	0.28	0.25	0.33	0.33	0.29	0.22	0.30	0.29	0.26

Poland, Czech Republic, Slovakia	0.21	0.22	0.19	0.21	0.19	0.24	0.23	0.22	0.21
Western Europe	0.06	0.08	0.07	0.05	0.04	0.07	0.06	0.05	0.07
Touchscreen	0.62	0.61	0.61	0.64	0.66	0.64	0.62	0.62	0.57
Female									
N (individuals)	3635	443	482	454	442	450	459	449	456
N (games)	7013	817	871	814	786	930	990	957	848
Number of games/player	1.929	1.84	1.81	1.79	1.78	2.07	2.16	2.13	1.86
Age									
>17	0.23	0.22	0.24	0.23	0.24	0.23	0.20	0.22	0.22
18-24	0.30	0.28	0.28	0.29	0.31	0.35	0.28	0.27	0.31
25-34	0.19	0.20	0.17	0.21	0.19	0.16	0.23	0.20	0.19
35-44	0.15	0.18	0.16	0.13	0.13	0.13	0.15	0.17	0.15
45-64	0.12	0.12	0.13	0.12	0.11	0.12	0.12	0.12	0.11
<65	0.02	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Education									
Elementary	0.17	0.16	0.16	0.17	0.19	0.16	0.17	0.16	0.21
Secondary	0.35	0.33	0.37	0.36	0.34	0.40	0.34	0.35	0.36
College or university	0.47	0.52	0.46	0.47	0.47	0.44	0.49	0.49	0.44
Plays games									
Never	0.20	0.22	0.19	0.18	0.22	0.19	0.22	0.18	0.23
Sometimes	0.58	0.58	0.62	0.60	0.56	0.59	0.56	0.61	0.54
Often	0.21	0.21	0.19	0.22	0.22	0.21	0.22	0.21	0.23
Confidence in game playing									
Very bad	0.07	0.07	0.08	0.07	0.08	0.07	0.09	0.06	0.07
Pretty bad	0.17	0.17	0.17	0.14	0.18	0.19	0.15	0.18	0.15
Ok	0.56	0.55	0.56	0.61	0.55	0.53	0.57	0.56	0.59
Pretty good	0.15	0.16	0.17	0.13	0.14	0.16	0.15	0.15	0.14
Excellent	0.04	0.05	0.02	0.04	0.05	0.05	0.04	0.05	0.05
Region									
Hungary	0.50	0.49	0.49	0.50	0.52	0.50	0.48	0.51	0.52
North America	0.05	0.06	0.05	0.06	0.06	0.06	0.05	0.05	0.04
Other	0.22	0.23	0.23	0.22	0.21	0.21	0.22	0.21	0.22
Poland, Czech Republic, Slovakia	0.16	0.16	0.16	0.15	0.16	0.15	0.17	0.16	0.15
Western Europe	0.07	0.06	0.07	0.06	0.05	0.08	0.07	0.07	0.06
Touchscreen	0.72	0.71	0.70	0.74	0.72	0.75	0.71	0.74	0.73

Table A2: Full OLS results, pooled groups

VARIABLES	(1) First game score	(2) Number of games played	(3) Mean score in session	(4) Best score in session	(5) Accuracy in session
female	4.225	0.284	4.308	4.89	0.0667
	(1.772)	(0.181)	(1.800)	(2.062)	(0.0246)
leaderboard	4.659	0.46	5.475	6.826	0.0661
	(2.068)	(0.212)	(2.102)	(2.407)	(0.0287)
supportive FB	2.666	-0.0160	2.579	2.872	0.0241
	(2.059)	(0.211)	(2.092)	(2.396)	(0.0286)
rewarding FB	4.488	0.0603	3.928*	3.674	0.0458
	(2.071)	(0.212)	(2.104)	(2.410)	(0.0287)
trash talk	1.903	-0.0188	0.386	0.493	0.0173
	(2.137)	(0.219)	(2.172)	(2.487)	(0.0297)
leaderboard x supportive FB	-3.742	-0.165	-4.100	-4.592	-0.0550
	(2.948)	(0.302)	(2.996)	(3.431)	(0.0409)
leaderboard x rewarding FB	-4.972	0.00693	-4.888	-3.720	-0.0467
	(2.912)	(0.298)	(2.959)	(3.389)	(0.0404)
leaderboard x trash talk FB	-6.568	-0.255	-6.288	-7.182	-0.0931
	(3.031)	(0.310)	(3.080)	(3.528)	(0.0421)
female x leaderboard	-4.406	-0.170	-5.265	-5.916	-0.0656
	(2.491)	(0.255)	(2.531)	(2.899)	(0.0346)
female x supportive FB	-0.531	0.0183	-0.365	-0.461	-0.00291
	(2.471)	(0.253)	(2.511)	(2.876)	(0.0343)
female x rewarding FB	-1.289	-0.0688	-1.344	-0.608	-0.00687
	(2.493)	(0.255)	(2.533)	(2.901)	(0.0346)
female x trash talk FB	-1.029	-0.00613	0.0496	0.431	0.00159
	(2.552)	(0.261)	(2.593)	(2.969)	(0.0354)
female x leaderboard x supportive FB	3.529	0.195	4.074	4.474	0.0693
	(3.529)	(0.361)	(3.586)	(4.107)	(0.0490)
female x leaderboard x rewarding FB	3.043	0.0172	3.640	2.245	0.0337
	(3.510)	(0.359)	(3.566)	(4.085)	(0.0487)
female x leaderboard x trash talk FB	5.207	0.0503	4.865	5.058	0.0547
	(3.610)	(0.369)	(3.668)	(4.201)	(0.0501)
age group 1	0.475	-0.467	-0.363	-1.140	-0.0741
	(1.117)	(0.114)	(1.135)	(1.300)	(0.0155)
age group 2	7.947	-0.27	7.386	7.795	0.0165
	(1.046)	(0.107)	(1.062)	(1.217)	(0.0145)
age group 3	8.077	-0.00320	8.247	8.935	0.05
	(1.113)	(0.114)	(1.131)	(1.296)	(0.0155)
age group 4	3.791	0.3	4.335	5.246	0.0352
	(1.181)	(0.121)	(1.200)	(1.374)	(0.0164)
age group 6	-1.933	-0.309	-2.668	-3.260	-0.0817

	(2.075)	(0.212)	(2.109)	(2.415)	(0.0288)
high school	0.943	0.113	0.946	1.279	0.0236
	(0.867)	(0.0887)	(0.881)	(1.009)	(0.0120)
university	4.523	0.33	4.499	5.457	0.0851
	(0.900)	(0.0920)	(0.914)	(1.047)	(0.0125)
region 1	-3.428	0.200	-3.08	-3.189	-0.106
	(1.300)	(0.133)	(1.320)	(1.512)	(0.0180)
region 3	-6.991	-0.0755	-7.137	-8.105	-0.158
	(1.356)	(0.139)	(1.378)	(1.579)	(0.0188)
region 4	-4.293	0.0952	-4.207	-4.362	-0.158
	(1.425)	(0.146)	(1.448)	(1.658)	(0.0198)
region 5	-3.731	-0.210	-4.305	-5.82	-0.11
	(1.658)	(0.170)	(1.684)	(1.929)	(0.0230)
touchscreen	11.98	0.423	12.08	13.48	0.241
	(1.459)	(0.149)	(1.482)	(1.697)	(0.0202)
pixel ratio 1	7.11	0.0307	7.602	7.692	0.221
	(2.666)	(0.273)	(2.708)	(3.102)	(0.0370)
pixel ratio 2	5.442	0.128	5.611	6.181	0.204
	(2.262)	(0.231)	(2.298)	(2.633)	(0.0314)
pixel ratio 3	8.377	0.281	8.517	10.12	0.162
	(2.265)	(0.232)	(2.301)	(2.635)	(0.0314)
Constant	8.468	1.119	10.29	11.58	0.216
	(3.138)	(0.321)	(3.189)	(3.652)	(0.0436)
Observations	5,190	5,190	5,190	5,190	5,190
R-squared	0.143	0.044	0.150	0.142	0.273

Table A3: Full linear combination estimates, pooled groups

Linear combination of coefficients	First game score		Number of games in		Best score in session		Mean score in session		Accuracy in session	
label	estimate	p	estimate	p	estimate	p	estimate	p	estimate	p
female	4.225	0.017	0.284	0.117	4.890	0.018	4.308	0.017	0.067	0.007
leaderboard+fe_leaderboard	0.253	0.855	0.290	0.041	0.911	0.573	0.211	0.881	0.000	0.982
leaderboard	4.659	0.024	0.460	0.030	6.826	0.005	5.475	0.009	0.066	0.021
suppFB+fe_suppFB	2.135	0.118	0.002	0.987	2.412	0.129	2.214	0.111	0.021	0.263
suppFB	2.666	0.195	-0.016	0.939	2.872	0.231	2.579	0.218	0.024	0.398
leaderboard+fe_leaderboard+suppFB+fe_suppFB+leaderboard_suppFB+fe_leaderboard_suppFB	2.176	0.115	0.322	0.023	3.204	0.046	2.398	0.088	0.036	0.061
leaderboard+suppFB+leaderboard_suppFB	3.584	0.090	0.280	0.196	5.106	0.038	3.954	0.066	0.035	0.230
suppFB+leaderboard_suppFB+fe_suppFB+fe_leaderboard_suppFB	1.923	0.163	0.032	0.819	2.294	0.153	2.188	0.118	0.036	0.063
suppFB+leaderboard_suppFB	-1.076	0.611	-0.181	0.403	-1.720	0.484	-1.521	0.479	-0.031	0.293
leaderboard+leaderboard_suppFB+fe_leaderboard+fe_leaderboard_suppFB	0.041	0.976	0.320	0.021	0.792	0.615	0.185	0.893	0.015	0.434
leaderboard+leaderboard_suppFB	0.918	0.662	0.296	0.169	2.234	0.361	1.375	0.520	0.011	0.704
fe_leaderboard	-4.406	0.077	-0.170	0.504	-5.916	0.041	-5.265	0.038	-0.066	0.058
fe_suppFB	-0.531	0.830	0.018	0.942	-0.461	0.873	-0.365	0.884	-0.003	0.932
fe_leaderboard+fe_suppFB+fe_leaderboard_suppFB	-1.408	0.577	0.043	0.868	-1.902	0.518	-1.556	0.545	0.001	0.983
fe_leaderboard+fe_leaderboard_suppFB	-0.877	0.726	0.025	0.924	-1.441	0.620	-1.190	0.639	0.004	0.917
fe_suppFB+fe_leaderboard_suppFB	2.998	0.234	0.213	0.409	4.014	0.171	3.709	0.148	0.066	0.058
female+fe_leaderboard+fe_suppFB+fe_leaderboard_suppFB	2.817	0.120	0.327	0.078	2.988	0.157	2.753	0.135	0.067	0.007
female+fe_leaderboard	-0.181	0.918	0.114	0.528	-1.026	0.617	-0.956	0.593	0.001	0.964
female+fe_suppFB	3.694	0.033	0.302	0.088	4.429	0.028	3.943	0.025	0.064	0.008
rewFB+fe_rewFB	3.199	0.021	-0.009	0.952	3.066	0.057	2.584	0.067	0.039	0.043
rewFB	4.488	0.030	0.060	0.776	3.674	0.127	3.928	0.062	0.046	0.111
leaderboard+fe_leaderboard+rewFB+fe_rewFB+leaderboard_rewFB+fe_leaderboard_rewFB	1.524	0.273	0.306	0.031	2.501	0.122	1.547	0.273	0.026	0.171
leaderboard+rewFB+leaderboard_rewFB	4.176	0.042	0.528	0.012	6.780	0.005	4.516	0.030	0.065	0.022
rewFB+leaderboard_rewFB+fe_rewFB+fe_leaderboard_rewFB	1.270	0.359	0.016	0.912	1.591	0.324	1.336	0.342	0.026	0.177
rewFB+leaderboard_rewFB	-0.484	0.813	0.067	0.748	-0.046	0.985	-0.959	0.645	-0.001	0.976
leaderboard+leaderboard_rewFB+fe_leaderboard+fe_leaderboard_rewFB	-1.676	0.225	0.314	0.026	-0.564	0.725	-1.037	0.460	-0.013	0.514
leaderboard+leaderboard_rewFB	-0.312	0.879	0.467	0.026	3.106	0.193	0.588	0.778	0.019	0.495
fe_rewFB	-1.289	0.605	-0.069	0.787	-0.608	0.834	-1.344	0.596	-0.007	0.843
fe_leaderboard+fe_rewFB+fe_leaderboard_rewFB	-2.652	0.285	-0.222	0.382	-4.279	0.138	-2.969	0.238	-0.039	0.259
fe_leaderboard+fe_leaderboard_rewFB	-1.363	0.581	-0.153	0.545	-3.671	0.202	-1.625	0.518	-0.032	0.352
fe_rewFB+fe_leaderboard_rewFB	1.754	0.478	-0.052	0.838	1.637	0.569	2.296	0.361	0.027	0.434
female+fe_leaderboard+fe_rewFB+fe_leaderboard_rewFB	1.573	0.368	0.062	0.728	0.611	0.764	1.340	0.450	0.028	0.249
female+fe_rewFB	2.936	0.096	0.215	0.234	4.282	0.037	2.964	0.098	0.060	0.015
trFB+fe_trFB	0.873	0.531	-0.025	0.861	0.924	0.569	0.436	0.759	0.019	0.328
trFB	1.903	0.373	-0.019	0.932	0.493	0.843	0.386	0.859	0.017	0.559
leaderboard+fe_leaderboard+trFB+fe_trFB+leaderboard_trFB+fe_leaderboard_trFB	-0.235	0.865	0.060	0.671	-0.289	0.857	-0.777	0.581	-0.019	0.324
leaderboard+trFB+leaderboard_trFB	-0.006	0.998	0.186	0.398	0.137	0.956	-0.427	0.845	-0.010	0.746
trFB+leaderboard_trFB+fe_trFB+fe_leaderboard_trFB	-0.488	0.723	-0.230	0.103	-1.200	0.455	-0.987	0.481	-0.019	0.311
trFB+leaderboard_trFB	-4.666	0.030	-0.274	0.213	-6.689	0.008	-5.902	0.007	-0.076	0.011
leaderboard+leaderboard_trFB+fe_leaderboard+fe_leaderboard_trFB	-1.108	0.424	0.085	0.548	-1.214	0.451	-1.212	0.389	-0.038	0.049
leaderboard+leaderboard_trFB	-1.909	0.390	0.205	0.366	-0.356	0.890	-0.813	0.718	-0.027	0.381
fe_trFB	-1.029	0.687	-0.006	0.981	0.431	0.885	0.050	0.985	0.002	0.964
fe_leaderboard+fe_trFB+fe_leaderboard_trFB	-0.228	0.929	-0.126	0.630	-0.427	0.886	-0.350	0.893	-0.009	0.793
fe_leaderboard+fe_leaderboard_trFB	0.801	0.759	-0.120	0.654	-0.858	0.778	-0.399	0.881	-0.011	0.764
fe_trFB+fe_leaderboard_trFB	4.178	0.102	0.044	0.866	5.489	0.065	4.915	0.058	0.056	0.112
female+fe_leaderboard+fe_trFB+fe_leaderboard_trFB	3.996	0.032	0.158	0.407	4.463	0.039	3.959	0.036	0.057	0.026
female+fe_trFB	3.195	0.084	0.278	0.142	5.321	0.013	4.358	0.020	0.068	0.008

Table A4: Game level OLS results: summary of treatment effects

	First game score						Game end score, pooled games					
	Supportive FB		**Rewarding FB**		**Trash talk**		**Supportive FB**		**Rewarding FB**		**Trash talk**	
	estimate	p	estimate	p	estimate	p	estimate	p	estimate	p	estimate	p
LB with no FB females	0.401	0.773	1.852	0.089	1.959	0.076	0.337	0.758	0.190	0.861	0.268	0.807
LB with no FB males	4.900	0.018	7.289	0.000	7.350	0.000	5.742	0.000	5.471	0.001	5.558	0.001
FB with no LB females	2.197	0.107	0.577	0.609	-0.334	0.772	2.275	0.034	2.606	0.016	0.499	0.650
FB with no LB males	2.789	0.175	4.022	0.030	-0.430	0.828	2.677	0.099	4.006	0.013	0.535	0.751
LB+FB females	2.199	0.111	1.601	0.138	0.473	0.677	2.364	0.030	1.486	0.171	-0.716	0.512
LB+FB males	3.652	0.084	7.627	0.000	3.062	0.108	4.045	0.015	4.586	0.004	-0.351	0.836
gender diff LB with noFB	-4.499	0.071	-5.437	0.009	-5.391	0.010	-5.405	0.006	-5.281	0.007	-5.290	0.007
gender diff FB with noLB	-0.592	0.811	-3.445	0.112	0.096	0.967	-0.402	0.836	-1.400	0.473	-0.036	0.986
gender diff LB+FB	-1.453	0.565	-6.026	0.003	-2.589	0.243	-1.682	0.398	-3.099	0.110	-0.366	0.856

Notes: Estimated treatment effects are presented, based on OLS regressions of the same form as equation (1), calculated as shown in Table 4. along with p-values. First game score results estimated using game level data from only the first game played by each individual, with the end of game score as the outcome measure. Pooled games results estimated on game level data consisting of all the games played by each individual, with the end of game score as the outcome measure, with estimated standard errors clustered at the player level.