Gender Differences in the Effect of Subjective Feedback in an Online Game

Anna Lovász , Ewa Cukrowska-Torzewska , Mariann Rigó , Ágnes Szabó-Morvai

PII: S2214-8043(22)00029-5
DOI: https://doi.org/10.1016/j.socec.2022.101854
Reference: JBEE 101854

To appear in: *Journal of Behavioral and Experimental Economics*

Received date: 1 February 2021
Revised date: 8 December 2021
Accepted date: 25 February 2022

Please cite this article as: Anna Lovász , Ewa Cukrowska-Torzewska , Mariann Rigó , Ágnes Szabó-Morvai , Gender Differences in the Effect of Subjective Feedback in an Online Game, *Journal of Behavioral and Experimental Economics* (2022), doi: https://doi.org/10.1016/j.socec.2022.101854

Gender Differences in the Effect of Subjective Feedback in an Online Game

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

[(a)] *Corresponding author: University of Washington Tacoma, 1900 Commerce St, Tacoma, WA 98402, plovi@uw.edu*

[(b)] *University of Warsaw, Faculty of Economic Sciences, Długa 44/50, 00-241 Warsaw, Poland*

[(c)] *Institute of Medical Sociology, University of Düsseldorf, Medical Faculty, Moorenstr. 5, 40225 Düsseldorf, Germany*

[(d)] *Centre for Economic and Regional Studies, Toth Kalman u. 4. Budapest, 1097 Hungary*

[(e)] *University of Debrecen, Debrecen, Egyetem ter 1, 4032 Hungary*

November 21, 2021

Highlights

- We use an online game to test for gender differences in the impact of two positive subjective feedback types given in the form of simple text and graphics
- Encouragement increases the persistence of female players and their score, while it impacts male players negatively
- Praise has no significant impact
- Players with lower task-related confidence are more responsive to subjective feedback
- Individualized feedback provision could be a potential tool for decreasing gender gaps in educational and workplace outcomes

Abstract

We test for gender differences in the impact of subjective feedback, using an online game with randomized treatment in the form of pop-up texts and graphics. We estimate the impact of giving encouragement or praise on performance (score, accuracy) and persistence (number of games played). We find some evidence of gender differences: while praise has little impact, encouragement has a positive impact on the persistence and performance of females and a negative impact on the performance of males. These are driven by those with low task-related confidence. The findings suggest that giving more individualized feedback – such as encouragement to females with low task-related confidence – can potentially improve gender equality, especially in traditionally male fields.

Keywords: Gender differences, supervisory feedback, non-cognitive traits

JEL codes: I24, J16, J24, M50

1. Introduction

In this study, we use a randomized online experiment to test whether the subjective content of supervisory feedback could be an environmental element that contributes to gender gaps in persistence and performance. Recently, gender differences in various preferences and traits - competitiveness, risk aversion, cooperation, altruism - have been documented in a significant strand of laboratory and field experiments (Eckel and Grossman 2008; Croson and Gneezy 2009; Niederle 2016), which have been shown to contribute to gender gaps in real-life educational (Buser, Niederle, and Oosterbeek 2014; Ors, Palomino, and Peyrache 2013), and labor market outcomes (Azmat and Petrongolo 2014; Bertrand 2011). As Niederle (2016) describes, one approach to how these gaps should be addressed focuses on elements of the institutional environment that favor traits unequally distributed by gender, which could be adjusted to achieve outcomes that more closely reflect underlying abilities *("fixing institutions")*. For instance, previous research has shown that the provision of objective feedback on relative performance can eliminate the gender gap in the choice to compete (Ertac and Szentes 2010; Wozniak, Harbaugh, and Mayr 2014, 2016). We show that the subjective content of feedback can also impact gender gaps in persistence and performance. Individualized feedback may therefore be a potential low-cost intervention for improving gender equality in educational and workplace settings.

Teaching and HR management strategies have long considered supervisory feedback to be a key element of motivational strategy. A growing body of literature in psychology, education, management, and economics analyzes its impact on performance (Wong 2015; Posner and Kouzes 1999; Khan et al. 2014). Economics has so far focused on the effect of objective feedback on choices, performance, and the updating of beliefs (Azmat and Iriberri 2010; Kuhnen and Tymula 2012; Bandiera, Larcinese, and Rasul 2015). Several studies document gender differences in the impact. Berlin and Dargnies (2016) show that women react more strongly to relative performance feedback in their competition entry decisions. In a laboratory experiment, Wozniak et al (2014) show that the provision of relative performance feedback increases high-ability females' participation in competitive compensation schemes and eliminates gender differences in choices. In a recent study, Alnamlah and Gravert (2020) study gender differences in the impact of feedback given after a failure, and find that women respond negatively in terms of

persistence in a competitive setting when failure is attributed to a lack of ability, but positively when it is attributed to luck. This suggests that the content of feedback beyond objective performance information plays a role in shaping choices.

Subjective feedback has not yet recieved much attention in the economics literature. Praise – a positive valuation of performance or effort - has been studied in its role as a verbal performance incentive, with the finding that its impact can be greater than that of a financial reward (Ariely 2016). The psychology literature provides evidence on a broad range of feedback types, regarding their impact on individual motivation and performance and the underlying mechanisms (e.g. Deci and Ryan 1985; Locke 1996). These highlight differences in the impact by individual characteristics such as self-confidence (Chang et al. 2012), as well as gender (Healy and Pate 2011; Wozniak, Harbaugh, and Mayr 2016). Our study adds to the existing literature by measuring the effect of subjective feedback using an online randomized experiment, and assessing whether subjective feedback is an element of the institutional design that can impact gender gaps. We focus on two common types of subjective feedback given by supervisors: encouragement and praise. Encouragement refers to an expression of support or a positive expectation regarding an individual's future performance, such as "You can do it!". Praise refers to a positive valuation of an individual's performance, such as "Well done!", which is given after some successful action.

We collect data using a 2-minute online game of visual perception advertised on social media sites. Players are randomly selected to receive either no subjective feedback (control group) or positive subjective feedback (treatment group). Feedback is given in the form of simple phrases and emoticons that pop up during the game in addition to the player's score (see Figure A1 and Table A1). We test the impacts of encouragement and praise separately, which may differ due to both their information content and their perception. We record the actions and performance outcomes of the players, and link the resulting dataset to pre-game survey information on basic demographics, technical data, and measures of task-related experience and confidence. We evaluate gender differences in the impacts on performance (score) and persistence (number of games played), then examine the role of task-related confidence in shaping these impacts. Our setup does not provide monetary incentives for

participation or performance, nor any publicity of player outcomes. This is similar to the setup of many recent online games that aim to test and improve skills and relieve stress.[1] Individuals who play are motivated by intrinsic incentives such as improving their personal high score.

Based on a sample of 869 individuals and 1486 games, the results provide some evidence of gender differences in the effects of subjective feedback. For women, encouragement has a strong significant positive impact on persistence and some positive impact on performance. For men, the impact on performance is negative, stemming from a drop in accuracy. On the other hand, praise does not have a significant impact on the outcomes of men or women, with the exception of a significant negative effect on the performance of women with lower task-related confidence. The effects are driven by those with low task-related confidence, who are generally more responsive to subjective feedback. Overall, the results suggest that the impact of subjective feedback varies by gender, feedback type, and task-related confidence. The main implication of the results is to point out the important heterogeneity that exists in the response to subjective feedback. This suggest that uniform feedback can lead to significant performance losses and contribute to gender inequality.

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. However, it is in a gaming context with low stakes and short interactions, so our results may not be generalizable to paid job settings where external monetary and further social incentives are also present. Some previous studies have confirmed the validity of hypothetical choices in real life settings,[2] so the behavior we observe during a game may similarly be indicative of behavior in a school or workplace context. Our findings have direct relevance for some intrinsically-motivated settings that are important for labor market success, such as everyday human capital investment or self-improvement activities. Our experiment captures elements of basic personality traits, which have been shown to impact long run success. For example, attitudes towards new

[1] For example, popular puzzle games and brain teasers such as I Love Hue and Polyforge.

[2] In particular, studies have compared preferences related to risk taking. Taylor (2013) found that mean risk preferences do not differ significantly when measured in hypothetical and real settings. Holt and Laury (2002) found no significant differences under low incentive conditions, but found significant differences in in the case of high stakes settings, where individuals tended to underestimate their actual risk aversion.

tasks and challenges are known to be critical determinants of educational achievement (Henderson and Dweck 1991; Hong et al. 1999), and impact gender differences in career choices and outcomes (Lloyd, Walsh, and Yailagh 2005; Dweck 2006).

Recent studies based on online experiments have shown that certain low-cost interventions can significantly impact individuals' perseverance and performance, for example, growth mindset interventions that highlight the malleability of skills can positively impact math students (Bettinger et al. 2018). The provision of subjective feedback is a similarly low-cost intervention, yet one that is constantly present in real-life supervisory communication. We show that it is a factor that can affect gender gaps. Our analysis highlights the economic importance of better targeted, or more individualized supervisory feedback. The results do not suggest that solely gender-based targeting of subjective feedback should be adopted, but rather that more individualized feedback – which takes the task and various psychological factors into consideration – can help decrease existing gender gaps. This may be especially crucial for closing gender gaps in traditionally male-dominated fields and occupations, such as STEM fields or leadership positions, where women have less experience, fewer role models, less societal support, and lower task-related confidence (Sterling et al 2020).

2. The effect of subjective feedback

Studies in psychology (e.g. Henderlong and Lepper 2002) emphasize that it is vital to view the process of receiving and interpreting feedback as a complex interaction, considering the role of the feedback content, the environment in which it is given, as well as the personality of the recipient. Positive subjective feedback may affect individuals' persistence and performance through three main channels. First, individuals may directly gain utility (or disutility) from the subjective content of feedback, which we term the **utility channel**. Some individuals might enjoy a task more if it is carried out in a friendly, pleasant environment, or if a supervisor acknowledges their effort or performance. In this sense, positive feedback can serve as a form of reward. On the other hand, it could also impose a cost, as some individuals – e.g., those who prefer tough, competitive environments – may dislike it and view it annoying and therefore decrease their effort to avoid it. The sign and magnitude of

the effect of subjective feedback through the utility channel depend on individual preferences towards the specific subjective content, which may differ by personality or gender (e.g. Deci and Ryan 1985).[3]

The second channel impacting individuals' effort choices is the **belief-updating channel**. Receiving feedback can affect an individual's expected performance by providing information on their past performance or environmental expectations regarding their future performance. The subjective content of feedback will only affect an individual's expected performance if it is perceived as relevant information. For example, receiving encouragement may improve task-related confidence if encouragement is interpreted as a sign of good faith from the supervisor. The sign and size of the effect depends on how the given individual perceives[4] and values such information when updating their beliefs, which may vary by personality and gender. Previous studies found that low-confidence individuals are more averse to obtaining objective feedback (Mobius et al. 2011), but also place relatively more weight on subjective valuations from the environment when forming their performance expectations (Chang et al 2012). The performance-relevant information content of praise and encouragement differs, so they may be valued differently when individuals update their beliefs. Praise is a positive valuation of past performance, while encouragement expresses positive expectations of future performance. Individuals with lower confidence may be more affected by encouragement compared to confident individuals if they indeed place a higher weight on environmental expectations.

Finally, subjective feedback may affect performance through a third channel as well, the **productivity channel**. Emotions elicited by the manipulation of the task environment have been previously shown to affect productivity in various tasks (Baker, Frith, and Dolan 1997; Brand, Verspui, and Oving 1997; Ellis et al. 1997; McKenna and Lewis 1994). Stressful situations such as competition, high stakes, or the presence of an audience have been shown to decrease productivity (Ariely et al. 2009; Baumeister 1984; Azmat Ghazala, Calsamiglia Caterina, and Iriberri

[3] For example, cognitive evaluation theory (Deci and Ryan 1985) emphasizes the basic need for autonomy and competence. Accordingly, incentives (or external interventions) that are interpreted by recipients as decreasing their feelings of self-control and competence, termed as having a "controlling" aspect, will create a detrimental effect on motivation. Differences in personal characteristics can result in the same positive feedback being interpreted as either informational or controlling by different individuals.

[4] Whether participants perceive feedback as sincere is an aspect that has received considerable attention in the psychological literature (Henderlong and Lepper 2002).

Nagore 2016). Positive subjective feedback may set individuals at ease by signaling a friendly environment, and thereby improve their performance given their effort level. On the other hand, if the feedback is perceived annoying, this may detract from the individual's concentration and performance. Since praise may raise the performance-dependent stakes of playing, it may increase the stressfulness of the environment and lead to lower performance. On the other hand, since encouragement rewards effort rather than performance, it may lower performance-related stress, and lead to higher productivity.

In our experiment, we measure the effect of the subjective content of feedback when it is provided in addition to the objective content, to see how adding subjective content to objective performance feedback affects different types of individuals. The magnitudes and signs of the effects depend on the relative roles of the channels described above. Gender differences in the effects may arise due to differences in preferences towards the specific subjective content, differences in the perception and the weighting of subjective feedback in belief updating, and/or differences in the environmental impact on productivity.

3. Methodology

3.1. Experimental design

To obtain the data for testing the impacts of the two subjective feedback types, we utilize a simple online computer game developed for this purpose. The link to the game was advertised on social media sites, targeted towards the age group of 18-45-year-olds, and four countries (Czech Republic, Hungary, Poland, Slovakia). Prior to choosing to play, individuals are given a short description of the game, which is a simple game of visual perception that requires both concentration and effort. There are many different geometric shapes moving around the screen. The task is to click on those matching the target shape that is displayed in the top left corner of the screen (see Figure A1 for a screenshot). Players must find and click on all of the shapes that match the target, then the target shape changes. Players receive a point for every shape they click on that

matches the target. The game takes two minutes, and the goal is to score as many points as possible during that time. Players see the remaining game time and their cumulative score at the top of the screen during the entire game. When the game ends, players see their game end score, and have the option to play again.

A simple survey precedes the game (Appendix Table A2). It asks for basic demographic information: gender, age, country, and level of education. Data is also automatically collected to account for whether the device the game is played on is a touchscreen or not, as this can impact performance. The survey was designed to be quick and easy to fill out, asking only for anonymous information similar to those requested on other game sites. Players are informed of the experimental purpose of the game and the details of data collection, 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.[5] Since the question is asked after players see the game description, their responses likely reflect their beliefs regarding how well they will play this type of game, i.e. their task-related confidence. Women have lower task-related confidence despite having similar performance (Table 2), which may be due to less task-related experience or the different weighting of their prior experience in their performance expectations.

When players click to start playing the game, they are randomly selected to receive one of the four treatment types described in Table 1 (further details in Table A1). We study the impact of the two subjective feedback types separately, each with their own control group. All groups receive messages during the game. The treatment involves variation in the content of those messages. The control groups see only objective feedback, the treatment groups see additional positive subjective feedback. The praise group and its control are shown messages after a certain number of successfully completed targets, when the target shape change occurs. The number of targets completed is indicated as the objective performance information for both. Although the number of targets completed is not the performance goal of the game, the shape change

[5] Specifically, players were asked "How often do you play computer games?" and "Are you good at playing computer games?" Based on the five confidence categories, we define three categories used in the analysis: high confidence (excellent or pretty good), medium confidence (ok), and low confidence (pretty bad or very bad). Alternatively, we carry out the analysis with 5 categories and different definitions of 2 category confidence measures. The significance of the estimates depends on how detailed the categories are, but the main findings we present are unchanged by these specifications.

represents a natural break in the flow of the game when feedback can be given following the successful completion of a sub-task (a target). Players in the praise treatment group are shown phrases and graphics expressing positive valuations in addition to the objective feedback. Encouragement and its control are shown messages independently of any successful action, at specific points in time. These messages can interrupt a player's series of clicks, and is a key difference between the two control groups, which we discuss later on. The score is shown in the message as the objective feedback for both groups. Players in the encouragement treatment group are shown phrases and graphics expressing support in addition to the objective feedback. These timing and feedback specifications ensure that (1) praise is clearly performance-dependent as it is given after a success (target completed), while encouragement is clearly independent of performance and simply given at specific times, and (2) the treatment and control groups for each feedback type do not differ in any other dimension besides the subjective feedback (text and graphics).

Table 1: Treatment specifications

Treatment type	**Specification**	**Feedback content and timing**
Praise Treatment	Objective feedback + Praise	Number of targets completed + praise graphics and text, after a given numbers of completed targets
Praise Control	Objective feedback	Number of targets completed, after a given numbers of completed targets
Encouragement Treatment	Objective feedback + Encouragement	Score + encouragement graphics and text, every 30 seconds
Encouragement Control	Objective feedback	Score, every 30 seconds

Note: see Table A1 in the Appendix for exact details.

The feedback messages are communicated via pop-up textboxes, which have to be clicked for the game to continue. This ensures that they are noticed by players. Since our goal was to collect data internationally, we used commonly known English phrases and simple, culturally

neutral emoticons in the treatments. Given their widespread use in online media, these are likely to convey the same meaning for players of different nationalities and cultural backgrounds. The phrases in the praise treatment include positive valuations of past performance (e.g. "Good job!"). The phrases in the encouragement treatment refer to expectations of future performance and expressions of support (e.g. "You can do it!"). The emoticons used for each treatment were selected to convey the same general meaning as the phrases. Our treatment includes the impact of both the phrases and emoticons, so we measure the combined impact of these exact feedback text and graphic specifications. Appendix Table A1 gives the details of the timing, phrases, and graphics.

3.2. Data and empirical analysis

The effort performance data from the online data collection is highly detailed, recording every event that takes place during the game to the thousandth of a second. This includes the timing of every click a player makes, its success, messages shown, and target changes. Our initial dataset is therefore at the level of events. We aggregate this to the game level, calculating game level outcomes such as game end score, accuracy, and total clicks in the game. We then aggregate to the player-session level. A session is defined as a single browser session during which the individual played one or more games. Players receive the same treatment during every game in a gaming session, as our goal is to allow for longer term treatments and impacts. Different players play a different number of games, and this itself can be affected by the treatment. We calculate session level outcomes: the number of games played, mean score (the average of game level end scores), best score (the highest game level end score), accuracy, and total clicks. We analyze the first game score along with the session-level outcomes in order to present a dynamic view of the impact of feedback during an entire task: a players' initial performance (first game score), their persistence (number of games played, total clicks), and their overall performance in the session (mean and best score, accuracy). We link game and session-level outcomes to the individual level variables collected in the pre-game survey.

We analyze the complete first session of each player, during which they receive the same feedback treatment.[6] The best score represents the typical goal set by players in a game where there is no publicity (leaderboard) or relative performance feedback. They are motivated to keep clicking and to play the game again in order to beat their personal best score, and this motivation may be impacted by our treatments. Since players play a different number of games, the estimated impact of the feedback treatments on players' best scores includes any impact realized through increased persistence and learning. We evaluate the role of this learning channel in our estimates. It is also important to note that some players who start playing quit before the 2-minute time limit of their first game. These players are included in our estimation sample. Their game end score (and best score in the session) is the cumulative score they achieved in the game, regardless of the timing of their last click. This means that our analysis does not suffer from bias due to dropouts: our estimated treatment effects also include any impact on performance that is realized through players quitting the game. If a treatment causes players to stop clicking earlier, then, similarly to a lower number of games played, this represents lower persistence, which translates to lower scores and leaves less opportunity for improvement. Appendix Table A4 summarizes the number of players who did not click after 30 seconds (6.5%) or 90 seconds (13.5%). We find no significant differences by treatment group or gender, so this does not appear to be a key channel for the impact of feedback.[7]

We assess the impact of subjective feedback based on OLS equations. The dependent variables are the session level outcomes described above. The main independent variables are the gender and treatment dummies and their interactions. We also control for a variety of observable characteristics (the age[8], country, and education level of the individual, and whether they are playing on a touchscreen device[9]).[10] We estimate

[6] In order to mimic other games in the online gaming market and encourage players to see our game similarly, we do not limit the number of games they can play, or the number of sessions (whether they return to the website later on and play more games). However, since no players returned for a second session in our final sample, this had no impact on our analysis, and did not lead to longer-term outcome measures to be studied.

[7] We checked the robustness of our results and the role quitting may play by running our main regressions when quitters were dropped from the sample based on various definitions (did not click after 5, 15, 30, 60 seconds, total clicks less than 3,5, or 10). The main patterns presented here were unchanged.

[8] We include age categories (under 23, 23-39, 40 or above) in our main specification. Alternatively, we also defined different age categories and a quadratic formulation with age and age squared. These did not impact our findings, and show mostly insignificant impacts with some evidence of a linear relationship where younger people tend to score higher.

[9] The device type is detected automatically using internet cookies.

treatment effects separately for praise and encouragement, using separate data samples that include only the relevant control groups for each feedback type. We assess impacts based on equations of the following form:

$$outcome_i = \alpha_0 + \alpha_1 \cdot treatment_i + \alpha_2 \cdot female_i + \alpha_3 \cdot treatment_i \cdot female_i + \alpha_4{}' X_i + \vartheta_i \qquad (1)$$

where i refers to individual i, and X_i represents the control variables.

We use further specification to assess the mechanisms behind any observed impacts. We run regressions with best score as the dependent variable while controlling for the number of games played, in order to see what part of the impacts on performance are due to increased persistence and learning. We assess how changes in effort (clicks) and accuracy translated into performance impacts. Next, we evaluate whether the effects of the subjective feedback differ by task-related confidence, based on the self-reported confidence measure.[11] This allows us to assess whether mean gender differences in the treatment effects (a) arise due to differences in task-related confidence and differential feedback effects by confidence level, or (b) are gender-based and exist within confidence levels as well. The latter case would suggest that feedback targeted by gender may be the most effective, while the first case points to the importance of subjective feedback targeted by task-related confidence. This is important since previous literature shows that the relative confidence of males and females vary in different tasks. A female disadvantage is more likely to arise in tasks that are stereotypically or traditionally male. In contrast, in traditionally female tasks, males may have lower task-related confidence, and require appropriate feedback themselves to overcome the disadvantage.

[10] Controlling for these characteristics is important, as the sample size may not be large enough to guarantee randomness among groups in terms of individual characteristics, especially within gender and treatment cells.
[11] See Question 6 in the pre-game survey in Appendix Figure A2.

4. Results

4.1. Descriptive statistics

Table 2 summarizes the characteristics of the sample used in the analysis. It consists of 866 individuals, 39% of whom are female, who played a total of 1486 games.[12] The lower ratio of female players in our sample may be due to two reasons. First, it is possible that women were less likely to click on our ad and play the game, which suggests different selection mechanisms by gender. Second, we do not observe the gender composition of those who saw our social media ads. It is possible that fewer women saw our ads, especially if ad targeting algorithms take gender into account, and more male users tend to play online games. This could further strengthen any existing gender differences. Our estimation results may therefore only be representative of the population of online game players, and any gender differences we find are based on these differently selected groups of men and women.

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 26. Approximately 71% of the individuals report having some college or university education, 28% report secondary education, and barely any have lower education levels. The age and education distributions are comparable across men and women. The sample is dominated by the four countries targeted in the ads (not shown in the table): Hungary, Poland, Slovakia, and the Czech Republic, however, there are some observations from a total of about 20 countries. There are noticeable gender differences in terms of game-play related characteristics: females play on a higher ratio of touchscreen devices[13] compared to males, and generally play computer games less often. The

[12] The table shows some unevenness in the distribution of players by feedback group. However, we checked the likelihood of such a distribution based on our sample size, and found no evidence of an issue with the randomization that takes place when players click to play the game. The numbers are well within the possible distribution along treatment group. A more even distribution can only be expected at much larger sample sizes.

[13] Our estimates suggest that being on a touchscreen device decreases a player's score and accuracy (Appendix Table A2). This is reflective of players using mobile phones with smaller screens, which make it harder to click on small shapes. To check whether there are any gender differences in the impact of device type, we ran a regression including interaction terms of the female and touchscreen dummies. We found no significant gender gaps in the impact on scores.

table shows that there are some minor differences in observed characteristics across treatment and control groups, which supports the use of OLS regressions that control for observable characteristics when estimating the treatment effect.

Table 2: Summary statistics of the sample

	Total	Encouragement		Praise	
		Control	Treatment	Control	Treatment
Full sample					
N (individuals)	866	190	207	247	222
N (games)	1486	303	358	447	378
Mean number of games/player	1.72	1.59	1.73	1.81	1.70
Maximum number of games/player	16	8	16	10	16
SD number of games/player	1.35	0.90	1.47	1.35	1.55
Female	0.39	0.38	0.39	0.40	0.37
Age	26.25	26.71	25.92	26.24	26.2
Education:					
Elementary	0.01	0.01	0.01	0.01	0.01
Secondary	0.28	0.28	0.27	0.29	0.29
College or university	0.71	0.71	0.72	0.69	0.7
Plays games often	1.23	1.26	1.23	1.22	1.25
Confidence in games:					
Low	0.18	0.17	0.20	0.19	0.18
Middle	0.49	0.48	0.50	0.53	0.46
High	0.32	0.34	0.29	0.28	0.36
Touchscreen	0.63	0.66	0.65	0.62	0.60

Females					
N (individuals)	348	75	85	102	86
N (games)	675	123	188	204	160
Number of games/player	1.94	1.64	2.21	2.00	1.86
Maximum number of games/player	16	8	16	10	16
SD number of games/player	1.81	1.10	2.10	1.77	2.04
Age	25.5	25.2	25	26	25.7
Education:					
Elementary	0.00	0.00	0.01	0.00	0.00
Secondary	0.27	0.29	0.24	0.26	0.27
College or university	0.73	0.71	0.75	0.74	0.72
Plays games often	1.00	0.98	1.03	0.97	0.97
Confidence in games:					
Low	0.29	0.25	0.31	0.30	0,29
Middle	0.56	0.56	0.58	0.57	0.55
High	0.14	0.18	0.11	0.13	0.16
Touchscreen	0.73	0.8	0.73	0.72	0.68
Males					
N (individuals)	518	115	122	145	136
N (games)	811	180	170	243	218
Number of games/player	1.57	1.57	1.39	1.68	1.60
Maximum number of games/player	8	5	5	7	8
SD number of games/player	0.90	0.74	0.67	0.93	1.14
Age	26.7	27.6	26.5	26.4	26.5
Education:					

Elementary	0.01	0.01	0.01	0.01	0.02
Secondary	0.29	0.28	0.28	0.32	0.3
College or university	0.69	0.71	0.7	0.66	0.68
Plays games often	1.39	1.43	1.32	1.39	1.41
Confidence in games:					
Low	0.11	0.11	0.12	0.11	0.11
Middle	0.45	0.43	0.45	0.50	0.45
High	0.44	0.45	0.43	0.39	0.44
Touchscreen	0.57	0.56	0.59	0.54	0.56

4.2. Treatment effects by gender

The estimated treatment effects for each subjective feedback type are summarized in Table 3 (full results in Table A2). For encouragement, we see a significant gender difference in the impact on performance in the first game: women respond much more positively than men. The treatment effect is positive, but only significant at the 10 percent level for women, and negative, but not significant for men. Women play a significantly higher number of games when they receive treatment, encouragement increases their persistence (column 2). Men, however, do not play more games, and the resulting gender difference in the number of games played is significant. The other columns show performance impacts based on the best score (columns 3 and 7) and mean score (columns 4 and 8) achieved in the session. For encouragement, we find a negative impact on males' best score, as well as their mean score. For women, we see a positive impact on the best score and the mean score; however, these are not significant. The gender difference is significant for all outcome measures: encouragement increases female persistence relative to males, while it decreases male performance relative to females.

For praise, we find that treatment does not impact the persistence or performance of either gender significantly. Contrary to the impact of encouragement, the signs of the estimated effects of praise for women tend to be negative, while the effects of men are close to zero. As we

mentioned earlier, we used different timing specifications for praise (given after a success when the target shape changes) and for encouragement (given every 30 seconds) in order to clearly signal performance-dependence and independence. However, the timing itself may have impacted players' performance and effort, because the praise treatment was given at a natural break in game play, while encouragement may have interrupted players' clicking. We therefore included a control group for each that mimics the given treatment groups' timing and objective feedback content (targets completed vs score). To get an idea of how these feedback message differences impacted players, we compare the two control groups' outcomes using an OLS regressions similar to equation 1 (see Appendix Table A5[14]). These indeed show weak evidence of a negative impact of the encouragement timing on the number of games played, but no impact on scores in the first game or best score.

Table 3: Estimated treatment effects on the number of games played, best score, and mean score in the session by gender and feedback type

	Encouragement				**Praise**			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Dependent variable	**First game score**	**Number of games played**	**Best score in session**	**Mean score in session**	**First game score**	**Number of games played**	**Best score in session**	**Mean score in session**
Male treatment effect (α_1)	-3.886	-0.120	-5.288*	-4.629*	1.289	-0.0682	-0.597	0.647
	(0.107)	(0.652)	(0.050)	(0.058)	(0.553)	(0.707)	(0.800)	(0.762)
Female treatment effect $(\alpha_1 + \alpha_3)$	4.873	0.772**	3.577	3.443	-1.844.	-0.096	-2.604	-1.810
	(0.099)	(0.016)	(0.267)	(0.236)	(0.491)	(0.676)	(0.387)	(0.505)
Gender difference in treatment effect (α_3)	8.759**	0.892**	8.865**	8.072**	-3.133	-0.028	-2.007	-2.458
	(0.021)	(0.031)	(0.034)	(0.033)	(0.363)	(0.923)	(0.599)	(0.476)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	397	397	397	397	469	469	469	469
R-squared	0.118	0.067	0.104	0.101	0.171	0.064	0.179	0.196

[14] Table A5 presents the results from regressions run on the sample containing only the control groups. We also checked the impact of the encouragement control using the full sample, with treatment and control type interactions included, and got very similar results.

*Notes: Player-session level data. Controls include female dummy, age, country, education, and device type. P-values are indicated in parentheses. Significance is indicated by stars: *** p<0.01, ** p<0.05, * p<0.1. Full OLS regression results with standard errors are shown in Table A2.*

We next explore the mechanisms behind the performance impacts (Table 4). First, we evaluate the role of increased persistence and learning in achieving a higher performance. Columns 1 and 5 show the same estimated effects on best score as we saw in Table 3. In columns 2 and 6, we add a control for the number of games played in the session, to see how much of the estimated impacts on performance are realized through changes in the number of games played. We can see that, in line with learning during the task, the number of games played has a significant positive impact on the best score achieved in the session. The negative impact of encouragement on males' best scores decreases slightly in magnitude and significance when we control for the number of games played in column 2, as does the positive impact on females' best scores. The gender difference in the impact encouragement on performance becomes insignificant. This suggests that encouragement impacts the performance of men and women differently in part due to its differential impact on persistence and learning.

Table 4: OLS estimates of treatment effects by gender on best score, the number of clicks, and accuracy

	Encouragement				**Praise**			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Specification	**Best score**	**Best score**	**Mean clicks per game**	**Accuracy**	**Best score**	**Best score**	**Mean clicks per game**	**Accuracy**
Male treatment effect (α_1)	-5.288**	-4.974*	-3.823	-0.079**	-0.597	-0.289	2.038	0.003
	(0.050)	(0.057)	(0.175)	(0.042)	(0.800)	(0.896)	(0.433)	(0.917)
Female treatment effect ($\alpha_1 + \alpha_3$)	3.577	1.556	4.012	0.020	-2.604	-2.168	-1.900	0.000
	(0.267)	(0.620)	(0.233)	(0.659)	(0.387)	(0.443)	(0.565)	(0.991)
Gender difference in treatment effect (α_3)	8.865**	6.530	7.834*	0.099*	-2.007	-1.879	-3.938	-0.003
	(0.034)	(0.109)	(0.073)	(0.099)	(0.599)	(0.600)	(0.348)	(0.956)

Number of games		2.62***				4.52***		
		(0.000)				(0.000)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	397	397	397	397	469	469	469	469
R-squared	0.095	0.094	0.101	0.057	0.189	0.109	0.092	0.106

*Notes: Player-session level data. Specifications 1 & 5: dependent variable is best score. Specifications 2 & 6: dependent variable is best score, with the number of games played in the session as an additional control (learning). Specifications 3 & 7: dependent variable is the mean clicks per game. Specifications 4 & 8: dependent variable is the mean accuracy (score/clicks) in the session. Controls include female, age, country, education, and device type. Standard errors are clustered at the player level. P-values are indicated in the parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1. Full OLS results with standard errors shown in Table A2.*

Next, we evaluate the roles of changes in effort (number of clicks) and accuracy in shaping performance impacts. From columns 3 and 4, we can see that for encouragement, the negative effect on male scores is due to a significant negative impact on accuracy. For women, on the other hand, encouragement leads to more clicks per game (effort), but no significant impact on accuracy. Overall, males' performance appears to be impacted negatively by encouragement due to a decrease in accuracy, suggestive of an impact through the productivity channel described earlier. For females, encouragement increases persistence and effort, translating to an increase in performance, though not a significant one. Again, we do not see significant impacts or gender differences in the case of the praise treatment.

4.3. The role of task-related confidence

The summary statistics in Table 3 show significant pre-game differences in the task-related confidence of males and females. While 29% of women responded that they are "bad" or "very bad" at games, only 11% of men categorized themselves similarly. At the other end of the spectrum, 14% of women had high confidence ("pretty good" or excellent"), while 44% of men did. The differences observed in mean treatment effects by gender could be related to the lower confidence of women, if those with low confidence respond differently to subjective feedback than those with high confidence do. On the other hand, it is also possible that women and men respond differently to feedback, even within

confidence levels. To see which of these drives the observed mean gender effects, we estimate OLS equations in which the gender and treatment dummy variables are additionally interacted with confidence level. We use the number of games played and the best score in the session as dependent variables. The resulting estimates are shown in Table 5.

The results indicate that the treatment effects of encouragement are mainly driven by impacts on those with low task-related confidence. For males with low confidence, the number of games played is not affected , but the best score is impacted very negatively. For females with low confidence, the number of games played is impacted very positively, while the effects on the best score are not significant. The gender differences among those with low confidence are highly significant, pointing to differential impacts by gender. For praise, the only significant impact we see is on the best scores of women with low confidence, who respond negatively. Overall, we find evidence of significant gender differences in the effect of encouragement within the low confidence level. It follows that the mean gender differences found earlier are not due to the different confidence distributions by gender, but differences in the response to feedback by gender within confidence groups.

Table 5: Treatment effects on score by gender and task-related confidence level

		Encouragement		Praise	
		Estimate	**P-value**	**Estimate**	**P-value**
		Number of games played			
Low confidence	Treatment effect, males	-0.247	0.566	-0.107	0.864
	Treatment effect, females	0.927***	0.003	-0.032	0.940
	Gender difference	1.174**	0.027	0.076	0.920
Middle Confidence	Treatment effect, males	-0.232	0.270	-0.309	0.256
	Treatment effect, females	0.094	0.677	-0.173	0.585
	Gender difference	0.326	0.289	0.136	0.744
High confidence	Treatment effect, males	0.044	0.831	0.121	0.657
	Treatment effect, females	-0.146	0.733	-0.056	0.926
	Gender difference	-0.191	0.689	-0.177	0.788
		Best score in session			
Low confidence	Treatment effect, males	-23.997***	0.004	-1.965	0.808
	Treatment effect, females	6.868	0.150	-9.105*	0.093
	Gender difference	30.866***	0.003	-7.140	0.463
Middle confidence	Treatment effect, males	2.189	0.587	-2.030	0.562
	Treatment effect, females	1.995	0.643	2.092	0.607
	Gender difference	-0.193	0.974	4.122	0.442
High confidence	Treatment effect, males	-8.187**	0.040	0.133	0.970
	Treatment effect, females	0.565	0.945	-8.779	0.256
	Gender difference	8.752	0.337	-8.912	0.295

*Notes: Dependent variables are number of games played, and best score in session. Controls include female dummy, age, country, education, device type. P-values are indicated in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1. Full OLS results with standard errors shown in Table A3.*

4.5 Discussion

Several aspects of the task environment are vital to the interpretation and external relevance of the results. We observe changes in behavior in the short-run, based on relatively short-term interactions. The subjective feedback in the game is based on pre-programmed messages, and are not given by a real-life supervisor. Players play the game within their home setting, completely anonymously and without any public or social aspect. 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. In general, the stakes of effort (playing) and especially performance (playing well) are relatively low. The time and energy costs of playing are minor, just 2 minutes of clicking. There are no financial rewards or prizes, competition, or publicity in the game. The potential benefits of playing are therefore limited to the individual's sense of achievement, learning, and enjoyment of the game. The potential costs are also low, as there is no public shame or loss involved. The results are also likely to be task-specific. 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).

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. However, other methods of data collection, such as incentivized laboratory experiments, are also subject to the selection of individuals who are willing to participate. Arechar et al. (2017) discuss the benefits and problems of online experiments, and conclude that data collected using such methods is reliable and can provide the basis for valuable contributions to the empirical evidence. As noted earlier, our sample (around 39 percent female) shows some imbalance by gender, which may be due to differential targeting of our social media ads and/or to differential self-selection into playing the game. This means the estimated impacts and gender gaps may not be representative of the entire population. If individuals who did not choose to play are less frequent game players and have lower game-playing confidence, our results based on confidence levels would suggest even larger gender gaps. However, we have no evidence on the selection mechanism to verify that hypothesis. Finally, it is important to keep in mind that our results are particular to the subjective content given. The goal of this study is not to provide specific suggestions for subjective content, but to test for heterogeneity by gender in its impacts.

Keeping these caveats in mind, the results support the importance of more subtle elements of the institutional design that can play a role in shaping gender gaps in outcomes.

Although our results are not directly comparable to empirical evidence from previous papers – due to different sampling methods, feedback content, task content, and outcome measures – they point to similar implications regarding the interplay of individual characteristics and feedback. We provide some evidence that subjective supervisory feedback affects those with low confidence more, and it affects women differently than men. They are in line with previous findings of the positive effect of encouragement on women (Unkovic, Sen, and Quinn 2016), and with studies that suggest that performance-based feedback such as praise has a more negative effect on women than effort-dependent feedback such as encouragement (Zeldin and Pajares 2000; Roberts and Nolen-Hoeksema 1989). In order to increase gender equality, it is therefore especially important for supervisors to individualize their subjective feedback content in stereotypically male tasks, where women may have lower task-related confidence. The more individualized provision of subjective feedback can improve the persistence and the relative performance of females, potentially leading to increased task-related confidence in the longer run.

5. Conclusion

We estimate the effect of positive subjective feedback on the persistence and performance of individuals, and provide evidence of gender differences in these effects. Women exhibit, on average, a positive response to encouragement, increasing their persistence, effort, and performance. Men respond negatively to encouragement, as it lowers their accuracy and scores. The gender differences are even more apparent among men and women with low task-related confidence. Evidence of such differences suggests that one possible contributor to existing gender inequalities may be that current environments utilize communication that is, on average, better suited to the needs of males. More individualized feedback, such as encouragement given to less confident women could potentially decrease gender gaps, especially in traditionally male tasks.

The important implication of our experiment is that even within such a brief, low-stakes task, as the 2-minutes on-line game, and based on minimal changes to the subjective content received during the task, we see significant differences in the response by gender and task-related confidence. The results point to economic losses that may occur due to the provision of uniform supervisory feedback, for example, due to lower confidence women not receiving sufficient encouragement in STEM fields. Further research is needed to develop truly relevant practical recommendations for the targeting of subjective content.

Acknowledgments

The project was supported by funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 691676. The research was also supported by grants FK 124658 and 121267-PD of the National Research, Development and Innovation Office of Hungary. Anna Lovasz received support from the Bolyai grant of the Hungarian Academy of Sciences and the ÚNKP Bolyai plusz grant ÚNKP-18-4-ELTE-665. We would like to thank Andor Zöldesi for the development of the experimental game and website. We are grateful to Andrea Kiss, Corina Haita-Falah, Daniel Horn, Hubert Janos Kiss, Barbara Pertold-Gebicka, and participants at conference and seminar presentations for their comments. This research was carried out in a collaboration within the Virtual Research Collaboration on Gender and Family in the Labor Market.

Declarations of interest: none.

Appendix

Figure A1: Screenshot of experimental game

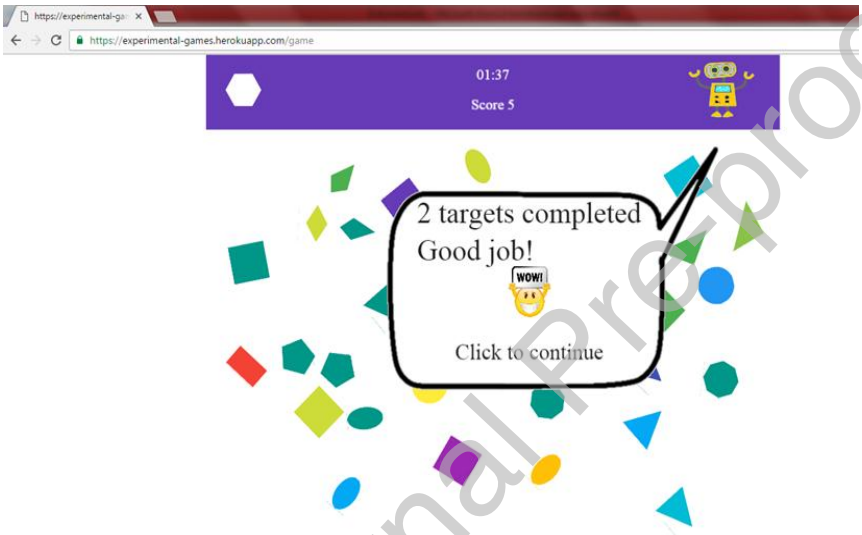

Table A1: Details of the treatment specifications

Praise					Encouragement				
Treatment			Control		Treatment			Control	
Timing of feedback	Picture	Text	Timing of feedback	Picture	Timing of feedback	Picture	Text	Timing of feedback	Picture
start screen	-	Are you ready? Click to start	start screen	Are you ready? Click to start	start screen		Are you ready? "Good luck!" "Click"	start screen	Are you ready? Click to start
after 2nd, 5th, 10th, 15th, 20th... shape change		"X targets completed" + 3 texts alternate: "Good job!" or "Well done!" or "You're great!"	after 2nd, 5th, 10th, 15th, 20th... shape change	"X targets completed"	30th second		Score: X + "You can do it!"	30th second	Score: X
END		Congratulations! Score: XX Play again!	END	Score: XX Play again!	60th second		Score: X + "Keep it up!"	60th second	Score: X
					90th second		Score: X + "Almost there!"	90th second	Score: X
					END		Score: XX + Play again!	END	Score: XX Play again!

Figure A2: 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)

Table A2: Full OLS results for various session outcomes, by gender and treatment type

	Encouragement			Praise		
	1	**2**	**3**	**4**	**5**	**6**
Dependent variable	**Number of games**	**Best score**	**Mean score**	**Number of games**	**Best score**	**Mean score**
treatment	-0.120	-5.288*	-4.629*	-0.0682	-0.597	0.647
	(0.266)	(2.690)	(2.429)	(0.181)	(2.362)	(2.133)
Female	0.0803	-3.140	-3.325	0.398**	3.694	3.947*
	(0.302)	(3.057)	(2.760)	(0.201)	(2.623)	(2.369)
Female* Treatment	0.892**	8.865**	8.072**	-0.0282	-2.007	-2.458
	(0.412)	(4.173)	(3.768)	(0.293)	(3.815)	(3.445)
Age 23-39	0.235	-2.915	-2.759	0.429**	-1.665	-2.872
	(0.231)	(2.335)	(2.108)	(0.166)	(2.165)	(1.955)
Age 40-45	0.355	-7.040*	-7.734**	0.120	-3.113	-2.687
	(0.409)	(4.140)	(3.738)	(0.337)	(4.388)	(3.963)
Hungary	-0.115	-5.154	-3.223	0.0266	-3.563	-4.151
	(0.380)	(3.843)	(3.470)	(0.327)	(4.264)	(3.851)
Poland	-0.400	-6.776*	-4.308	0.0635	-3.190	-4.576
	(0.392)	(3.966)	(3.581)	(0.328)	(4.276)	(3.861)
Slovakia	-0.550	-5.865	-3.441	0.00276	-4.266	-5.512
	(0.566)	(5.731)	(5.174)	(0.373)	(4.856)	(4.385)
Touchscreen	0.0354	-9.328***	-9.043***	-0.429***	-17.18***	-16.49***
	(0.215)	(2.176)	(1.964)	(0.157)	(2.041)	(1.843)
High School	-2.014*	-10.65	-7.994	0.295	13.51	10.60
	(1.129)	(11.42)	(10.31)	(0.751)	(9.781)	(8.832)
University	-1.647	-5.333	-4.480	0.272	16.94*	14.42
	(1.119)	(11.32)	(10.22)	(0.755)	(9.839)	(8.885)
Constant	3.314**	57.41***	50.28***	1.348*	34.14***	33.46***
	(1.189)	(12.03)	(10.86)	(0.758)	(9.868)	(8.911)
Observations	397	397	397	469	469	469

R-squared	0.067	0.104	0.101	0.064	0.179	0.196

*Notes: Player-session level dataset with various session outcomes. Controls include female dummy, age, country, education, and device type. Standard errors are shown in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1.*

Table A3: Full OLS results for various session outcomes, by gender, confidence level, feedback type

	Encouragement			Praise		
	1	**2**	**3**	**4**	**5**	**6**
Dependent variable	**Number of games**	**Best score**	**Mean score**	**Number of games**	**Best score**	**Mean score**
Treatment	-0.247	-24.00***	-21.85***	-0.107	-1.965	0.880
	(0.430)	(8.232)	(7.444)	(0.626)	(8.062)	(7.275)
Female	-0.127	-14.63*	-13.68**	0.488	1.503	3.371
	(0.392)	(7.500)	(6.781)	(0.483)	(6.230)	(5.622)
Female* Treatment	1.174**	30.87***	26.73***	0.0759	-7.140	-8.756
	(0.529)	(10.13)	(9.162)	(0.754)	(9.717)	(8.768)
Medium confidence	-0.110	-10.15	-9.476	0.102	-1.017	0.562
	(0.350)	(6.700)	(6.059)	(0.429)	(5.527)	(4.987)
High confidence	-0.204	-5.936	-6.051	0.139	2.302	3.091
	(0.347)	(6.653)	(6.016)	(0.445)	(5.729)	(5.169)
Female* Medium confidence	0.290	16.78*	14.65*	-0.137	3.725	1.475
	(0.452)	(8.661)	(7.831)	(0.553)	(7.130)	(6.434)
Female* High confidence	0.185	10.06	9.596	0.125	6.706	4.787
	(0.501)	(9.596)	(8.677)	(0.687)	(8.856)	(7.991)
Treatment* Medium confidence	0.0148	26.19***	24.12***	-0.201	-0.0651	-1.750
	(0.479)	(9.165)	(8.287)	(0.680)	(8.761)	(7.906)
Treatment* High confidence	0.291	15.81*	14.49*	0.229	2.098	0.431
	(0.478)	(9.152)	(8.276)	(0.683)	(8.799)	(7.940)
Treatment* Female* Medium confidence	-0.849	-31.06***	-26.05**	0.0600	11.26	12.95
	(0.614)	(11.75)	(10.63)	(0.860)	(11.08)	(10.00)
Treatment * Female * High confidence	-1.365*	-22.11	-17.69	-0.253	-1.772	-2.231
	(0.710)	(13.60)	(12.30)	(1.006)	(12.97)	(11.70)
Age 23-39	0.0681	-3.277	-3.058	0.451***	-1.622	-2.869
	(0.122)	(2.331)	(2.108)	(0.171)	(2.200)	(1.985)
Age 40-45	0.315	-7.028*	-7.703**	0.174	-2.457	-2.124
	(0.218)	(4.172)	(3.773)	(0.346)	(4.455)	(4.020)

Hungary	0.257	1.531	1.000	0.0392	-2.981	-3.780
	(0.250)	(4.782)	(4.324)	(0.331)	(4.269)	(3.852)
Poland	0.144	0.414	0.322	0.0237	-3.499	-4.926
	(0.255)	(4.885)	(4.417)	(0.332)	(4.283)	(3.865)
Slovakia	0.491	7.577	4.943	-0.0321	-3.912	-5.185
	(0.299)	(5.732)	(5.183)	(0.376)	(4.846)	(4.373)
High School	-1.956***	-10.27	-8.015	0.292	13.48	10.68
	(0.597)	(11.44)	(10.34)	(0.755)	(9.731)	(8.781)
University	-1.683***	-4.590	-4.129	0.268	16.96*	14.58*
	(0.592)	(11.33)	(10.25)	(0.760)	(9.797)	(8.840)
Constant	3.215***	57.52***	52.97***	1.233	33.46***	31.70***
	(0.674)	(12.90)	(11.66)	(0.857)	(11.04)	(9.961)
Observations	397	397	397	469	469	469
R-squared	0.151	0.136	0.134	0.073	0.204	0.222

*Notes: Player-session level dataset with various session outcomes. Controls include female dummy, age, country, education, and device type. Standard errors indicated in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1.*

Table A4: Number of quitters by timing threshold, feedback type, and gender

	Number of players	**No clicks after 30 seconds**				**No clicks after 90 seconds**			
		Overall	Male	Female	% female	Overall	Male	Female	% female
Control - encouragement	190	12	7	5	42	26	16	10	38
Control - praise	247	15	9	6	40	30	19	11	37

Encouragement	207	14	8	6	43	28	19	9	32
Praise	222	16	9	7	44	33	21	12	36
Total	866	57	33	24	42	117	75	42	36

Table A5: The impact of the difference in feedback timing and objective feedback content between the two control groups

	Number of games	**Best score**	**First score**	**Number of games**	**Best score**	**First score**
Control encouragement	-0.207	-0.029	1.515	-0.118	2.152	3.073
	(0.091)	(0.988)	(0.375)	(0.448)	(0.381)	(0.163)
Female				0.381	3.003	1.571
				(0.019)	(0.239)	(0.492)
Female* Control encouragement				-0.237	-5.485	-3.886
				(0.333)	(0.155)	(0.261)
Controls	yes	yes	yes	yes	yes	yes
Observations	385	385	385	385	385	385
R-squared	0.072	0.122	0.12	0.087	0.101	0.123

References

Alnamlah, Manar, Christina Gravert. 2020. "She Could Not Agree More: the Role of failure Attribution in Shaping the Gender Gap in Competition Persistence." CEBI Working Paper Series, Working Paper 25/20. ISSN 259644TX.

Arechar, Antonio A., Simon Gächter, and Lucas Molleman. 2017. "Conducting Interactive Experiments Online." *Experimental Economics*, May, 1–33. https://doi.org/10.1007/s10683-017-9527-2.

Ariely, Dan. 2016. *Payoff: The Hidden Logic That Shapes Our Motivations*. TED Books. https://www.amazon.com/Payoff-Hidden-Logic-Shapes-Motivations/dp/1501120042.

Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar. 2009. "Large Stakes and Big Mistakes." *The Review of Economic Studies* 76 (2): 451–69. https://doi.org/10.1111/j.1467-937X.2009.00534.x.

Azmat Ghazala, Calsamiglia Caterina, and Iriberri Nagore. 2016. "Gender Differences in Response to Big Stakes." *Journal of the European Economic Association* 14 (6): 1372–1400. https://doi.org/10.1111/jeea.12180.

Azmat, Ghazala, and Nagore Iriberri. 2010. "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment Using High School Students." *Journal of Public Economics* 94 (7): 435–52. https://doi.org/10.1016/j.jpubeco.2010.04.001.

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

Baker, S. C., C. D. Frith, and R. J. Dolan. 1997. "The Interaction between Mood and Cognitive Function Studied with PET." *Psychological Medicine* 27 (3): 565–78.

Bandiera, Oriana, Valentino Larcinese, and Imran Rasul. 2015. "Blissful Ignorance? A Natural Experiment on the Effect of Feedback on Students' Performance." *Labour Economics*, European Association of Labour Economists 26th Annual Conference, 34 (June): 13–25. https://doi.org/10.1016/j.labeco.2015.02.002.

Bandura, Albert. 1997. *Self-Efficacy: The Exercise of Control*. Worth Publishers.

Baumeister, R. F. 1984. “Choking under Pressure: Self-Consciousness and Paradoxical Effects of Incentives on Skillful Performance.” *Journal of Personality and Social Psychology* 46 (3): 610–20.

Berlin, Noémi and Dargnies, Marie-Pierre. 2016. "Gender differences in reactions to feedback and willingness to compete." Journal of Economic Behavior & Organization, 130©, p. 320-336.

Bertrand, Marianne. 2011. “New Perspectives on Gender.” In *Handbook of Labor Economics*, 4B:1543–90. Elsevier. https://ideas.repec.org/h/eee/labchp/5-17.html.

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.

Brand, N., L. Verspui, and A. Oving. 1997. “Induced Mood and Selective Attention.” *Perceptual and Motor Skills* 84 (2): 455–63. https://doi.org/10.2466/pms.1997.84.2.455.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2014. “Gender, Competitiveness, and Career Choices.” *The Quarterly Journal of Economics* 129 (3): 1409–47. https://doi.org/10.1093/qje/qju009.

Chang, Chu Hsiang, D. Lance Ferris, Russell E. Johnson, Christopher C. Rosen, and James A. Tan. 2012. “Core Self-Evaluations.” *Journal of Management* 38 (1): 81–128. https://doi.org/10.1177/0149206311419661.

Croson, Rachel, and Uri Gneezy. 2009. “Gender Differences in Preferences.” *Journal of Economic Literature* 47 (2): 448–74. https://doi.org/10.1257/jel.47.2.448.

Deci, Edward L., and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press.

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.

Eckel, Catherine C., and Philip J. Grossman. 2008. “Differences in the Economic Decisions of Men and Women: Experimental Evidence.” SSRN Scholarly Paper ID 1883696. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=1883696.

Ellis, H. C., S. A. Ottaway, L. J. Varner, A. S. Becker, and B. A. Moore. 1997. “Emotion, Motivation, and Text Comprehension: The Detection of Contradictions in Passages.” *Journal of Experimental Psychology. General* 126 (2): 131–46.

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., and A. Imas. 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. https://doi.org/10.1016/bs.hefe.2016.08.003.

Healy, Andrew, and Jennifer Pate. 2011. “Can Teams Help to Close the Gender Competition Gap?*.” *The Economic Journal* 121 (555): 1192–1204. https://doi.org/10.1111/j.1468-0297.2010.02409.x.

Henderlong, Jennifer, and Mark R. Lepper. 2002. “The Effects of Praise on Children’s Intrinsic Motivation: A Review and Synthesis.” *Psychological Bulletin* 128 (5): 774–95.

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. http://www.hup.harvard.edu/catalog.php?isbn=9780674050365.

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. http://dx.doi.org/10.1037/0022-3514.77.3.588.

Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects," American Economic Review, American Economic Association, 92(5): 1644-1655.

Khan, Aqeel, Roslee Ahmad, Abdul Rahim Hamdan, and Mohamed Sharif Mustaffa. 2014. “Educational Encouragement, Parenting Styles, Gender and Ethnicity as Predictors of Academic Achievement among Special Education Students.” *International Education Studies* 7 (2): 18. https://doi.org/10.5539/ies.v7n2p18.

Kuhnen, C. M., & Tymula, A. (2012). Feedback, self-esteem, and performance in organizations. Management Science, 58(1), 94–113. https://doi.org/10.1287/mnsc.1110.1379.

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. http://dx.doi.org/10.2307/4126476.

Locke, Edwin A. 1996. "Motivation through Conscious Goal Setting." *Applied and Preventive Psychology* 5 (2): 117–24. https://doi.org/10.1016/S0962-1849(96)80005-9.

McKenna, F. P., and C. Lewis. 1994. "A Speech Rate Measure of Laboratory Induced Affect: The Role of Demand Characteristics Revisited." *The British Journal of Clinical Psychology* 33 (Pt 3) (September): 345–51.

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. https://press.princeton.edu/titles/10874.html.

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.

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.

Posner, Barry Z., and James M. Kouzes. 1999. *Encouraging the Heart: A Leader's Guide to Rewarding and Recognizing Others*. Jossey-Bass: Wiley. http://www.wiley.com/WileyCDA/WileyTitle/productCd-0787953172.html.

Roberts, Tomi-Ann, and Susan Nolen-Hoeksema. 1989. "Sex Differences in Reactions to Evaluative Feedback." *Sex Roles* 21 (11–12): 725–47. https://doi.org/10.1007/BF00289805.

Sterling, Adina, Thompson, Marissa, Wang, Shiya, Kusimo, Abisola, Gilmartin, Shannon, Sheppard, Sheri. 2020. "The confidence gap predicts the gender pay gap among STEM graduates." Proceedings of the National Academy of Sciences Dec 2020, 117 (48) 30303-30308; https://doi.org/10.1073/pnas.2010269117.

Taylor, Matthew P. 2013. "Bias and Brains: Risk Aversion and Cognitive Ability across Real and Hypothetical Settings." *Journal of Risk and Uncertainty*, 46(3): 299–320.

Unkovic, Cait, Maya Sen, and Kevin M. Quinn. 2016. “Does Encouragement Matter in Improving Gender Imbalances in Technical Fields? Evidence from a Randomized Controlled Trial.” *PLOS ONE* 11 (4): e0151714. https://doi.org/10.1371/journal.pone.0151714.

Wong, Y. Joel. 2015. “The Psychology of Encouragement: Theory, Research, and Applications Ψ.” *The Counseling Psychologist* 43 (2): 178–216. https://doi.org/10.1177/0011000014545091.

Wozniak, David, William T. Harbaugh, and Ulrich Mayr. 2014. "The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices." Journal of Labor Economics, 32, issue 1, p. 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. https://papers.ssrn.com/abstract=1976073.

Zeldin, Amy L., and Frank Pajares. 2000. “Against the Odds: Self-Efficacy Beliefs of Women in Mathematical, Scientific, and Technological Careers.” *American Educational Research Journal* 37 (1): 215–46. https://doi.org/10.2307/1163477.