DISCUSSION PAPER SERIES

DP13552

GIRLS' AND BOYS' PERFORMANCE IN COMPETITIONS: WHAT WE CAN LEARN FROM A KOREAN QUIZ SHOW

Alison L Booth and Jungmin Lee

DEVELOPMENT ECONOMICS AND LABOUR ECONOMICS

ISSN 0265-8003

GIRLS' AND BOYS' PERFORMANCE IN COMPETITIONS: WHAT WE CAN LEARN FROM A KOREAN QUIZ SHOW

Alison L Booth and Jungmin Lee

Discussion Paper DP13552
Published 27 February 2019
Submitted 19 February 2019

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

GIRLS' AND BOYS' PERFORMANCE IN COMPETITIONS: WHAT WE CAN LEARN FROM A KOREAN QUIZ SHOW

Abstract

We compare the performance of high-ability adolescent girls and boys who participated in a a long-running Korean television quiz show. We find there is a gender gap in performance -- in favour of boys -- across episodes of the quiz show. To investigate underlying mechanisms that might explain this, we explore how male and female performance varies under different rules of the game. We find that there are no gender gaps when stress is kept to a minimum -- that is, in games without fastest-finger buzzer, knock-outs or penalties. However, in games with these features, there are significant gender gaps. In addition, we examine performance in Round 2 of the shows, where we find larger gender gaps. These are consistent with girls being increasingly hindered by psychological stress and risk aversion as competition is higher. Finally, we use panel data to estimate performance in the games in which all players stay in for 25 questions. Here we find that girls are less likely to respond faster especially when their winning probability is higher. Further, the gender gap is more salient at the end of the game. The results are also consistent with gendered behavioural responses to psychological pressure.

JEL Classification: J16, I21, D9, L83, M5

Keywords: gender and competition, tournaments, psychological pressure, risk

Alison L Booth - alison.booth@anu.edu.au
Australian National University and CEPR

Jungmin Lee - jmlee90@snu.ac.kr
Seoul National University, Korea

Girls' and Boys' Performance in Competitions: What We Can Learn from a Korean Quiz Show*

Alison Booth
Australian National University,
Australia

and

Jungmin Lee**
Seoul National University,
Korea

February 2019

Abstract

We compare the performance of high-ability adolescent girls and boys who participated in a a long-running Korean television quiz show. We find there is a gender gap in performance – in favour of boys – across episodes of the quiz show. To investigate underlying mechanisms that might explain this, we explore how male and female performance varies under different rules of the game. We find that there are no gender gaps when stress is kept to a minimum – that is, in games without fastest-finger buzzer, knock-outs or penalties. However, in games with these features, there are significant gender gaps. In addition, we examine performance in Round 2 of the shows, where we find larger gender gaps. These are consistent with girls being increasingly hindered by psychological stress and risk aversion as competition is higher. Finally, we use panel data to estimate performance in the games in which all players stay in for 25 questions. Here we find that girls are less likely to respond faster especially when their winning probability is higher. Further, the gender gap is more salient at the end of the game. The results are also consistent with gendered behavioural responses to psychological pressure.

Keywords: gender and competition; tournaments; psychological pressure; risk aversion

JEL Classification: J16, I21, D9, L83, M5

* Acknowledgments: Thanks to Tim Hatton for helpful comments.

** Corresponding author. Email: jmlee90@snu.ac.kr

1. Introduction

In this paper, we compare the performance of high-ability adolescent girls and boys who participated in a series of publicly-observed tournaments. These tournaments took place in the context of a weekly South Korean television quiz show, entitled *Janghak*.[1] Our aim is to gauge if there are gender differences in the behaviour of girls and boys of high-school age in this extremely competitive environment. We also wish to establish how contestants' behaviour alters as the game rules vary and the gender composition of the group varies. As we shall argue later in the paper, some of the rules are likely to be associated with greater psychological pressure that may affect the performance of girls and boys differently.

A growing experimental literature investigates whether or not gender gaps in economic outcomes might be due to inherent differences in male and female attitudes to competition or to risk, and a number of experimental studies do find that the competitive choices made by men and women differ.[2] In our present paper, we adopt a different but complementary approach to these experiments by analyzing unique performance data from a real-world activity that is competitive by its very nature – the television quiz show for young people, *Janghak*. While several related studies have utilized similar US quiz-show data to explore gender differences in competitive outcomes, none have done so using data from South Korea. Yet the gender pay gap in Korea is the highest in the OECD, according to the 2017 OECD Report *The Pursuit of Gender Equality.*[3] It is therefore interesting to investigate if there are gender gaps in

[1] *Janghak* means 'scholarship' in Korean. See: https://www.youtube.com/watch?v=2PAc74BPVJg for an example of an episode.

[2] Studies investigating gender differences in performance in competitive environments include Gneezy et al., 2003; Niederle and Vesterlund, 2007, 2011; Booth, 2009; Dreber et al., 2011; Cárdenas et al, 2012; and Niederle, 2014. Research exploring gender differences in preference to enter a competition include Gneezy, Leonard and List, 2009; Booth and Nolen, 2012; Apicella and Dreber, 2015; Buser et al, 2017; Booth et al, 2016, while analyses of the determinants of risk attitudes risk include Booth et al., 2014a; Dreber et al, 2014; and Khachatryan et al., 2015. Buser et al (2014) explores how preference for competition across genders affects academic task-choice.

[3] Korea has the highest gender pay gap of any OECD country and its women remain underrepresented in public life, in spite of the fact that Korean women and girls have above-average scores in PISA and in PIAAC, and younger women have higher levels of educational attainment than their male peers (OECD, 2017). See also https://www.oecd.org/korea/Gender2017-KOR-en.pdf .

competitive performance even amongst Korean children of high-school age, and to chart how these vary across different situations.

Two studies that are close to ours are by Jetter and Walker, (2016, 2017), who use data from the US television quiz show *Jeopardy!*. Utilising a sample of adult men and women, Jetter and Walker (2016) found that there are no gender differences in responding to questions and in accuracy in high-stakes situations. In a comparison of the behavior of children, teenagers and college students from *Jeopardy!*, Jetter and Walker (2017) found no noticeable gender differences throughout all three subsamples.[4]

Also relevant to our present paper are studies by Cai et al (2019) and Irriberi and Rey-Biel (2018), who look at gender differences in performance in two stage competitions. Irriberi and Rey-Biel (2018) use data from a two-stage math competition in Madrid in Spain, while Cai et al (2018) use data from the college entrance exam (*Gaokao)* in China*,* the first stage of which is a mock examination. In Section 4.3 of our present paper, we will be looking at the performance of girls and boys in the subset of our quiz-show episodes in which all players stayed in the game for 25 rounds, which may be thought of as a 25-stage competition.

In our present paper, our primary focus is on exploring how competitive performance differs with gender. Some of the quiz-show episodes in *Janghak* used different means of eliciting contestants' responses to the questions, which we call the 'response technologies'. These different 'response technologies' – details of which will be provided in Section 2 – can be thought of as varying the amount of psychological pressure faced by contestants. Our conjecture is that gender might be made more salient through the response technology and through the gender proportion of the competing group.[5]

---

[4] Save-Soderbergh and Lindquist (2017) used data from the Swedish children's version of *Jeopardy!* [4] to focus on risk, which we are unable to do with our data. Comparing the wagering behaviour of children aged 10-11 with adults, they find gender gaps in risk-taking for adults but none for the girls and boys. Their children are younger than our conestants, whose mean age is just over 17 years.

[5] Booth and Yamamura (2018), utilizing data from speedboat racing in Japan (a sport in which men and women racers are randomly assigned to single-sex or mixed sex races), found that the same woman performs relatively worse in terms of her race time in mixed-sex races as compared with single-sex races, while for the average male racer, the opposite is true.

To summarize, we find there is a gender gap in performance across the majority of episodes of the quiz show game. To explore underlying mechanisms that might explain this, we investigate how male and female performance varies under different rules of the game. We find that there are no gender gaps when stress is kept to a minimum – that is, in games without knock-outs or penalties. However, in games with these features, there are significant gender gaps. In addition, we examine performance in Round 2 of the shows, where we find larger gender gaps. These are consistent with girls being increasingly hindered by psychological stress and risk aversion as the competition proceeds.[6] Finally, we use panel data to estimate performance in the games in which all players stay in for 25 questions. Here we find that girls are less likely to respond faster even when their winning probability is higher. We also find that their probability of answering correctly is lower. These results are consistent with boys' over-confidence and girls' under-confidence, as well as with different behavioural responses to psychological pressure. It is interesting that we have found these gender gaps in game-show performance in Korea, whereas they have not been found in the US (see Jetter and Walker, 2016, 2017). This may be because of different cultural values between the two countries, as highlighted in OECD (2017), or because of other unobserved differences between the shows.

The remainder of our paper is set out as follows. In Section 2, we describe the institutional background of the quiz show, while in Section 3 we provide a data overview, including summary statistics. In Section 4, we present and discuss the estimates and we conclude in Section 5.

2. Game Show Data

Our data come from the Korean television game show *Janghak Quiz*, which is a weekly competition program where high-school students compete for scholarship (the show's title, *Janghak*, means scholarship in Korean). Each group of participants comprises five individuals who compete against one another. The prize (scholarship) is not substantial, amounting to about

[6] This result is consistent with Iriberri and Rey-Biel (2018), as will be discussed in Section 4.3 below.

1,000 or 2,000 USD for weekly winners, but as will be explained below, they are advanced to monthly or annual competitions or can continue over weeks, in which cases the prize can increase up to 30,000 or 40,000 USD. It is the oldest television game show in Korea, having been broadcasted since 1973. Each episode hosts five contestants from five different high schools. Contestants self-select into the game; any high-school students who are interested in participating may apply and take a brief preliminary qualifying test. Depending on their results in this test, they are then selected to participate.[7] We expect our participants to be more competitive than the bulk of the high school population, since they have volunteered to play and are then selected by the game show organisers after passing a preliminary test based on ability.

Exact studio audience size is not known, but seems to be small, mainly consisting of contestants' friends and family members. Contestants are seated in a row on a stage in front of the audience. The show host begins the show and explains the game rules briefly, and each contestant introduces his/herself. They usually say their name and school and make their resolution for the game, like "I can win" or "I will make my school be proud of me". During the show, the host talks to contestants in an informal way for the purpose of relaxing them. Questions are pre-recorded and read by a voice actor.

For our analysis, we collected individual-level data giving basic information about contestants, such as sex, high school, and grade, as well as their round-by-round score and rank. These data were collected by a research assistant who watched all the shows recorded from February 2008 to December 2011 (episodes 577 to 777).[8] We excluded some "special" episodes from the sample. For example, we excluded one special show where students and their teachers were paired and competed as a team. In the end, for our empirical analysis, we focus on 180 shows with 900 contestants (= 180*5).

The game rules changed a few times during the sample period, and we are able to exploit this variation. Table 1 provides the game formats for the sample period. From episodes

---

[7] We were unable to obtain this confidential information.

[8] Video files are publicly available at the show's website (http://home.ebs.co.kr/janghakquiz/main).

577 to 592, there were no rounds; instead, five contestants attempted to answer all of a fixed number of questions and the one with the highest score in the end became the winner.

From episodes 593 to 628, each weekly show consisted of three rounds, but there was no knocking out over rounds. The three rounds differed by question type and rules. Details are given in the top panel of Table 1. As will be noted again later, when there are multiple rounds, scores in the previous round are not carried over to the next round. This is an important game feature because this means that there is no dynamic effect in terms of score. In other words, different rounds are independent. If contestants behave differently over rounds, it is because of different stakes and emotional pressure over rounds as they advance to higher rounds.

Episodes 629 onwards are labelled as 'Survival' in the bottom panel of Table 1. Note that the game had three rounds. In the first round, the show began with five contestants and one or two with the lowest scores were knocked out. In the second round, only one contestant with the highest score survived and was advanced to the final round where he or she competed with the previous week's winner.

[INSERT TABLE 1]

We focus on three main game features; buzzer, knocking-out, and points reduction. Table 1 shows how episodes differ across rounds in terms of these three features. The game rules have changed a few times. In some shows, there was a buzzer with a buzzer press, that is, contestants were required to push a buzzer to obtain the chance to answer. Even among these shows, the type of buzzer is different. There was a standard buzzer that was operated by a finger press, but there were two unique types; in one type, the contestant was required to stand up to press the buzzer while in the other, the contestant needed to jump to the front of the stage fastest in order to reach the buzzer. There were also shows without any buzzer. In these shows, all contestants choose or write their answer on their personal screen. Also the shows differ by whether there is any knocking out from advancing to the next round and, if so, how many are knocked out. Lastly, the shows differ by whether is a point reduction or a penalty associated with providing a wrong answer.

The details of game rules differ across episodes and rounds. Typically, contestants are initially given some basic points. Points differ by the difficulty of each question. For some questions, there is also a bonus gift, such as a laptop computer or a digital camera. In some shows, contestants are given one "chance" to preempt the first opportunity to answer the question. Question types are also various, including true or false, multiple-choice, and short-answer open-ended questions. We will clarify how we use these various game and show types in our analysis below.

3. Descriptive Statistics

Tables 2 to 5 use data for the 900 individual observations who took part in the first round of *all* the episodes listed in Table 1.

[INSERT TABLE 2]

The numbers of observations for each group are reported at the bottom of Table 2. We have 900 observations with usable responses from Round 1. These 900 contestants comprise 325 girls and 575 boys. The last four columns of Table 2 provide details of the 454 contestants who advanced into Round 2 from Round 1, of whom 139 are girls and 315 boys. Thus, girls in Round 1 comprise 36% of observations, and 31% in Round 2.

According to the 'greater male variability hypothesis', males display greater variability in traits such as cognitive ability than females do. In other words, human males are more likely than females to have very high or very low intelligence.[9] If this were so for our *Janghak* data, we would expect to see the proportion female increasing from Round 1 to Round 2, as the boys in the left tail of the cognitive ability distribution got knocked out of the competition. But this is not what we observe. Instead, we find that girls are less likely than boys to advance to Round 2. We shall return to this issue later in our regression analysis of Round 2 performance.

Table 2 shows that the average age of contestants is just over 17 years.[10] Proportionately more girls than boys in Round 1 attend general schools (rather than the specialist schools that are the alternative).[11] Just over half the sample attend private schools.[12] Most of the sample lives in metropolitan areas and medium-size cities. The variable 'Round 1 top 1' refers to being first in Round 1, and girls are less likely than boys to be in this group. 'Round 1 top 2' refers to being first or second in the first round, and again girls are less likely to be in this group. Of the contestants advancing into Round 2, boys have a higher Round 1 score than girls, though this difference is significant only at the 10 percent level.

We will be using these variables in subsequent regression analyses below, where we control for other factors likely to affect performance and investigate if gender gaps in performance vary across games with different rules.

4. Regression Results

[9] There is considerable controversy around this (see for example Irwing and Richard, 2005; Lindberg et al, 2019). Nonetheless we consider this in the context of our *Janghak* data.

[10] We have converted high school grade to age, since contestants do not reveal their age but only their high school grade. To do this conversion, we assumed high school grade 1 = age 16 and so on. Note further that school grade is missing for 371 contestants (41%). This is because it is up to them whether to introduce their grade or not.

[11] General schools are "regular" schools. South Korea also has some special-purpose schools like arts, music, science, or foreign language. Special-purpose schools are selective, while students are assigned to general schools based on their residence (school district).

[12] Private schools in Korea are privately-owned, but under the so-called "equalization policy" they are not very different from public schools. High schools governed by the equalization policy receive equal government funding, charge the same fees, and follow the same national curriculum. However, private schools maintain autonomy over their personnel decisions, while public schools do not (Hahn et al. 2018).

In Sections 4.1 and 4.2 below, we report results from estimation of a simple linear equation as follows:

$$S_{ij} = \beta_0 + \beta_1 Girl_{ij} + X_{ij}\gamma + \delta_j + \varepsilon_{ij} \quad (1)$$

where subscript i denotes individual contestant and j the episode. Recall that there are five contestants per episode. Thus, we can control for episode-specific fixed effects δ_j, which capture any effects of episode-level unobservable factors that equally affect the five contestants in an episode. X_{ij} includes other control variables such as age, indicators for school types (general and private schools), and residential areas (metropolitan and medium-sized cities).

The dependent variable S_{ij} is the standardized score for Round 1 and sometimes for Round 2. We standardize the raw score by dividing it by each episode's mean score because questions are different across shows in terms of the level of difficulty. Furthermore, the game rules are different (such as whether or not there is a point deduction for a wrong answer).[13] For robustness, we use as the dependent variable two alternative measures for contestant performance: whether the contestant is the top performer and whether he or she is within the top two. In Section 4.2 below, we will also present estimates of the between-round score difference $(S_{i2} - S_{i1})$.

4.1 Stage 1 Performance

The first three columns of Table 3 give the results from estimating the correlates of the standardized score. Note that robust standard errors, clustered by episode, are presented in parentheses. Here we see that girls have a significantly lower standardized score than boys, with or without the other controls. The estimates are robust to controlling for episode-specific fixed effects in column (3). The gender gap amounts to about 5% compared to the mean score.

[13] Any effects of the game rules should be subsumed by the episode-specific fixed effects. For comparison, we present the results with or without the fixed effects.

The only other control that is statistically significant is attendance at a *general* school, which is associated with a decline in score, as expected (the base group is attending a selective specialist school).

[INSERT TABLE 3]

Columns (4) to (6) report the results where the dependent variable is being the top performer, while columns (7) to (9) estimate being one of the top two performers. Being female is associated with a lower likelihood of being either a top performer or one of the top two performers, *ceteris paribus*. Girls are slightly less likely to be one of the top two performers than they are of being a top performer. This is as expected. If girls and boys are equally likely to be in either of the top two places, the fact that boys are in the majority makes them more likely to be top-placed.

To summarize, the estimates reported in Table 3 show that the performance of girls in the first round of the quiz show is worse than that of boys. We next see if the data can reveal anything about the underlying mechanism for this stylized fact. To explore this, in Table 4, we further disaggregate the Round 1 data into various game-types, following the taxonomy given in Table 1; knock-out, buzzer, and penalty.

[INSERT TABLE 4]

In column (1) of Table 4 we report estimates of games in which there is no knock-out. Our conjecture here is that, in games with no knock-outs, there will be less psychological pressure since there is no risk of 'sudden death.' We have 195 contestants participating in this type of game, and our estimates show that there is no gender gap in Round 1 score for this subsample. The estimate is not only statistically insignificant but also close to zero. In panel B, we further control for episode-specific fixed effects and find that the results are similar.

Column (2) of Table 4 presents estimates of games in which there is at least one knock-out, that is, where at least one contestant is eliminated. Here, in principle, each individual faces a 20% (one knock-out) or 40% (two) probability of being thrown out. We have 705 contestants participating in this type of game. Numerous studies show that on average girls are more risk averse than boys (see for example Eckel and Grossman, 2002; Dohmen et al, 2011; Booth and Nolen, 2012). We would therefore expect that the gender gap in performance would increase in games where psychological pressure is raised as risk is made more salient. Indeed, we find the magnitude of the estimated coefficient to gender is now larger in absolute terms. Moreover, it is now statistically significant at the 1 percent level, with or without episode fixed effects. In sum, the results from these first two specifications show that a game show of the type where losing contestants are flung out of the game is associated with a greater gender gap in performance than in a game show where no one is thrown out. This is consistent with our conjecture that girls, who are typically more risk averse than boys, face greater psychic pressure than boys in this form of the game. This 'choking under pressure' likely explains some of the observed gender gap.[14]

Can the buzzer-press format or a penalty also have an effect? We continue to explore game-types following the episode summary given in Table 1. In column (3) we report estimates of the gender gap in Round 1 score when there is no buzzer to press when the contestant is ready to answer, and in column (4) when there is a buzzer to press.[15] There is no statistically significant gender gap in scores without the buzzer, while in contrast scores with the buzzer exhibit a pronounced and statistically significant negative gender gap.

In column (5) we report estimates of the gender gap in Round 1 score when there is no penalty, and in column (6) when there is a penalty. A penalty of a points loss is likely to be perceived differently across contestants depending on their loss aversion. Studies have shown

[14] Moreover, according to psychologists such as Arch (1993) and Block (1983), men are more likely to see a risky situation as challenges to action, whereas women view it as a threat to be avoided.

[15] Note that there are alternatives to the buzzer-press method of responding to questions. For instance, contestants can hand-write responses on a screen that is seen by viewers but not by other contestants. These are included as non-buzzer games.

that women exhibit greater loss aversion than men (see for example Schmidt and Traub, 2002). We might therefore expect that our girls might perform worse than boys in situations where there is a probability of losing points. Confidence might also play a role, since these penalties represent not only a points loss but also a public humiliation, and it is well known from other studies that girls are less confident than boys.[16] We find here that there is a larger and statistically significant gender gap in score with the penalty than without it.

So far we have found that girls typically perform worse than boys in Round 1 in terms of score, especially in the settings where psychological pressure is higher. Next we check whether the score deficit of girls is significantly large enough to affect the probability of their advancing to the next round when there is any knock-out. Table 5 presents the results from estimation of the linear probability model where the dependent variable is the indicator of whether the contestant successfully advances to Round 2. Column (1) presents the results for all contestants who play Round 1 with knock-outs (705 observations). We omit the estimates for control variables and present only the results for the indicator of gender. We find that, regardless of whether or not we control for episode fixed effects, girls are less likely to advance to Round 2 than boys. The probability gap is about 12 percentage points. If there is one knock-out, then the probabilty of advancing to Round 2 is 80% when it is randomly decided. And the probability is 60% when there are two knock-outs. Given this, the deficiency of 12 percentage point is substantial.

[INSERT TABLE 5]

[16] See for example Jakobsson, Levin and Kotsadam (2013), Sarsons and Xu (2016), and references therein. As an aside, Benabou and Tirole (2002) list three main benefits of 'optimistic self-views' rather than accurate ones. These benefits encompass a consumption value, a signaling value, and a motivation value. These authors focus on the motivational value – that overconfidence or self-deception can propel an individual along a path where willpower might fail. We shall briefly return to this in Section 4.3 below.

In the next four columns, from (2) to (5), we separate the sample by game type; whether there is a buzzer and whether there is a point reduction. Here we find, consistent with the results in Table 4, that girls' underperformance is present or more salient in the game settings where psychological pressure is higher, i.e., where contestants press the buzzer fastest to win the right to answer or there is a point reduction to a wrong answer.

4.2 Performance in Round 2

Our Round 1 estimates suggested that girls are more responsive to psychological stress than boys. This emerged from the 'natural experiment' that we have at our disposal, arising from contestants' responses to changes over time in the rules of the game.

We next examine contestants' behavior at the second stage of the game. Our conjecture is that the gender gap in performance at the second stage will be bigger than at the first, because the level of stress will be increasing as the competition proceeds. The stakes will also be increasing, as potential earnings grow.[17]

We already know from the results in Table 5, as well as from inspection of the means in Table 2, that girls are less likely than boys to advance to Round 2 and that the 'greater male variability hypothesis' is not supported by our data. But perhaps the 'better' girls have survived to Round 2. If so, this would drive down the gender gap and swamp the psychological stress effect.

[INSERT TABLE 6]

[17] Cai et al (2018) find that, compared to male students, females underperformed on the highly competitive Chinese entrance exam - called *Gaokao* - relative to their performance in the low stakes mock examination. They attribute this to female's relatively lower tolerance for psychological pressure as well as their weaker incentives to perform in such a high-stakes situation.

To address these issues, panel A of Table 6 reports estimates from an equation where the dependent variable is Round 2 score. In column (1) we have the gender gap with no other controls. Here we see that the gender gap in performance is 20 percent, substantially larger than the comparison of just 5.5 percent from column (1) of Table 3. Thus the gender gap in performance in Round 2 is much bigger than at Round 1. The estimate in column (2) where the other control variables are included is similar and that in column (3) with episode fixed effects added is even larger, being about 25 percent. This is likely because the level of stress is increasing as the competition proceeds. This result is similar to that found by Cai et al (2019) and Iriberri and Rey-Biel (2018).[18]

In panel B of Table 6, we report results from estimating an equation of the form $(S_{i2} - S_{i1})$, where the first difference allows us to difference out individual fixed effects that are likely to affect performance. This is estimated on the subsample of 454 contestants for whom we have complete data in Rounds 1 and 2 (i.e., 454*2 person-round observations). Here we also find that differenced performance is significantly lower for girls than boys, again likely illustrating that the level of stress is increasing and self-confidence eroding as the competition proceeds.

Figure 1 shows that the gender gap in Round 2 is mainly driven by high-performing contestants. To draw the graph, we divide the sample of contestants based on their Round 1 score into five quintile groups and we plot the average Round 2 score for each quintile group. The graph shows that those contestants who scored higher in Round 1 also score higher in Round 2 but the correlation is higher among boys than among girls. In fact, girls in the bottom quintile perform better than their counterpart boys.

4.3 Play-by-Play Performance

---

[18] Iriberri and Rey-Biel (2018) analyse two-stage elimination math contests, in which participants compete to pass from stage 1 to stage 2 and later to be among the winners. They find that the gender gap in maths performance increases from stage 1 to stage 2 of a maths competition. They attribute the increase in female underperformance to higher competitive pressure.

We are able to undertake play-by-play panel data estimation using data from some of the shows. As can be seen in Table 1, in episodes 577 to 592, there were no rounds or knock-outs, and each show's five contestants answered all 25 questions until the end of the game. The contestant with the highest score becomes the winner. Two of these 15 shows were special format, and we therefore dropped those and use data from 13 of these shows. From these 13 shows we construct individual play-by-play performance data, with a panel structure of 25 periods (questions) per contestant. This yields an estimating subsample of 1,625 person-round observations. For each question, we record detailed information about each contestant's behavior/outcome: whether he or she successfully pressed the buzzer (that is, pressed it faster than others) and whether his or her answer was correct. We will use these as the dependent variable in the following regression equation.

$$Y_{ijt} = \beta_0 + \beta_1 Girl_{ij} + \beta_2 Win_{ijt} + \beta_3 Girl_{ij} \times Win_{ijt} + \gamma PT_{jt} + \tau_t + \delta_j + \varepsilon_{ijt} \qquad (2)$$

where subscripts indicate individual contestant i and question number t (i.e., t-th question, from 1 to 25) in episode j. Notations are the same as those in equation (1) except for subscript t, as we now observe individuals repeatedly in 25 questions. We control for the fixed effect (τ_t) for the question number to capture any effects that might arise as the game approaches the end. As before, δ_j is the episode fixed effect. Since the sample is a balanced panel dataset at the individual level, we can further control for individual contestant-specific fixed effects instead of the episode fixed effect. For robustness, we will present the results from both specifications.[19]

Unlike cross-sectional analysis using equation (1), we control for the (predicted) winning probability of each contestant at the moment of question t (Win_{ijt}). It is obvious that

[19] In the panel data analysis, we did not control for the control variables included in equation (1). They are individual-specific constant variables, such as age and school type. Instead, as a robustness check, we try to control for individual-specific fixed effects, which should absorb all the effects from those time-invariant individual characteristics.

the winning probability matters for contestants' behavior or strategy, but it is not directly observed even to contestants themselves. We assume that contestants, like econometricians, estimate their winning probability at the moment of question t by using the same Probit model we estimate, in which the dependent variable is whether the contestant is the final winner. Predictors include the contestant's score at question t, the gap between the contestant's score and the highest score at the moment, and the maximum remaining points afterwards. At the beginning of the show, we assume that all contestants have an equal probability of winning (0.2). In addition, to check whether the effect of winning probability differs by gender, we include the interaction term between the winning probability (in the mean deviation form) and the indicator for girls ($Girl_{ij} \times Win_{ijt}$). Lastly, we control for the number of points of the question (PT_{jt}). All contestants know these points before responding. We expect that the larger the points are, the more likely contestants take a risk, that is, pressing the buzzer even without full confidence in their answer. Thus, the likelihood of pressing the buzzer will be higher but the accuracy rate will be lower.

Note that when we examine whether the contestant's answer is correct or not, we condition the sample to those who obtain the right to answer by pressing the buzzer faster than others. There might be multiple contestants who obtain the right to answer because the fastest one got the wrong answer, and then the next one is determined again depending on who presses the buzzer fastest excepting the one who got the wrong answer in the first place.

The results in Table 7 reveal some intriguing gender differences. First, girls are less likely to press the buzzer fastest. The results in the left panel titled as 'Buzzer' show that their probability of obtaining the right to answer is about 4.5 percentage points lower than boys'. The estimates are robust to including various control variables. In columns (2) and (3), we included the interaction term between the winning probability and the indicator for girls. The results in column (2) show that girls are less likely to press the buzzer fastest and this tendency is more salient when their winning probabiltiy is higher. The results are robust to controlling for individual-specific fixed effects in column (3).

[INSERT TABLE 7]

The right panel, titled as ‘Accuracy’, presents the results for accuracy, that is, whether or not the contestant earns points. In other words, this is whether or not the contestant presses the buzzer fastest *and* his or her answer is correct.[20] Thus the outcome variable is the muliplication of the probabiltiy of obtaining the right to answer and that of the answer being correct. In columns (4) and (5), we find significant gender differences; girls are less likely to earn points than boys. However, the magnitude of the estimates is absolutely smaller than that of those in columns (1) and (2) where the dependent variable is just the probability of obtaining the right to answer. Comparing the estimates for the interaction term between columns (2) and (5), we also find that the estimate in column (5) is smaller in absolute terms than that in column (2). This means that while girls are less likely to press the buzzer fastest, the gender gap is weaker because there is a smaller gender gap in terms of accuracy. The estimate for the interaction term in column (6) is also smaller and statistically insignificant.[21]

The results suggest that girls are more passive than boys, especially when they are closer to winning the game. To see this pattern from a dynamic perspective, we estimate equation (2) by restricting the sample to that closer to the end of the game gradually, i.e., restricting the sample by $t \geq 1, 2, \ldots, 25$. We use specifications in columns (2) and (5) in Table 7, where we control for episode fixed effects and present only the estimates for the gender dummy ($Girl_{ij}$) and the interaction term ($Girl_{ij} \times Win_{ijt}$). Figure 2 plots the point estimates and 95 percent confidence intervals. Also for comparison, we present the estimates for the average effects from columns (2) and (5) in Table 7 by horizontal dashed lines.

Figure 2 reveals intriguing dynamics of gender differences. In particular, considering Panel A, the female gender dummy effect is improving as games wear on, but only up to around the 20th game; after this point both buzzer-pressing and accuracy decline for girls. This suggests that girls may be losing tenacity or confidence as the end of the 25-stage game approaches, and

---

[20] In order to estimate the gender effect on accuracy per se, we need to estimate a two-stage system of equations where the selection equation for obtaining the right to answer is jointly estimated.

[21] In the results for accuracy, it is notable that accuracy is lower when the points are larger. This is probably because contestants are more likely to take a risk when the stake is larger.

that as a consequence they reduce their effort.[22] In other words, the girls are quitting too easily. Panel B shows the dynamics of accuracy rates. The trends are similar to those in Panel A, which means that the dynamics are basically driven by the behavior of buzzer-pressing. It is interesting that Benabou and Tirole (2002) mention that investing in self-confidence is important for individuals, since it is a way for them to ensure time-consistent behaviour that will counter any natural tendency to quit too easily.

5. Conclusions

In this paper we analysed performance data from a long-running Korean television quiz show, whose contestants are on average 17 years old. First, we found there is typically a gender gap in performance across all episodes of the quiz show game. Second, to investigate underlying mechanisms that might explain these gender gaps, we explored how male and female performance varied under different rules of the game. We found that there are no gender gaps when stress is kept to a minimum – that is, in games without knock-outs, penalties, or 'exhibitionist' ways of pressing the response buzzer. However, in games with these features, there are significant gender gaps in performance. Third, we explored performance in Round 2 of the shows, where we found larger gender gaps. These are consistent with girls being increasingly hindered by psychological stress and risk aversion as the competition proceeds. Finally, we used panel data to estimate performance in the games in which players stay in for 25 questions. Here we found that girls are less likely to respond faster even when their winning probability is higher. We also found that their probability of answering correctly is lower. These panel data estimates are consistent with boys' over-confidence and girls' under-confidence.

It is interesting that we have found these gender gaps in performance in Korea, whereas they have not been found in the US game shows (see Jetter and Walker, 2016, 2017). This may

[22] An analogous effect was found by Cai et al (2019) and Iriberri and Rey-Biel (2018), whose females' average performance dropped off at the final stage of their 2-stage competitions.

be because of different cultural values between the two countries, as highlighted in OECD (2017), or because of other unobserved differences between the shows.

References

Arch, Elizabeth. 1993. "Risk-Taking: A Motivational Basis for Sex Differences." *Psychological Reports*, 73(3): 6–11.

Bénabou, Roland and Jean Tirole (2002). 'Self-confidence and personal motivation'. *The Quarterly Journal of Economics,* 117 (3), 871–915.

Block, Jeanne H. 1983. "Differential Premises Arising from Differential Socialization of the Sexes: Some Conjectures." *Child Development,* 54(6): 1335–54.

Booth, Alison L. and Patrick Nolen (2012). 'Gender Differences in Risk Behaviour: Does Nurture Matter?' (2012) *The Economic Journal*, 122 (February), F56–F78. Doi: 10.1111/j.1468-0297.2011.02480.x.

Booth, Alison and Eiji Yamamura (2018). 'Performance in Mixed-sex and Single-sex Competitions: What We Can Learn from Speedboat Races in Japan'. *Review of Economics and Statistics.,* October, 100(4): 581-593.

Cai, Xiqian, Yi Lu, Jessica Pan and Songfa Zhong (2019). 'Gender Gap under Pressure: Evidence from China's National College Entrance Examination'. Forthcoming *Review of Economics and Statistics.*

Dohmen, T., Falk, A., Huffman, D., Schupp, J., Sunde, U. and Wagner, G. (2011). 'Individual risk attitudes: new evidence from a large, representative, experimentally-validated survey', *Journal of the European Economic Association*, vol. 9, pp. 522–50.

Eckel, C. and Grossman, P. (2002). 'Differences in the Economic Decisions of Men and Women: Experimental Evidence', Amsterdam: Elsevier Science B.V. North-Holland.

Hahn, Youjin, Liang Choon Wang and Hee-Seung Yang (2018). "Does Greater School Autonomy Make a Difference? Evidence from a Randomized Natural Experiment in South Korea." *Journal of Public Economics* 161, pp. 15-30.

Iriberri, Nagore and Pedro Rey-Biel (2018). 'Competitive Pressure Widens the Gender Gap in Performance: Evidence from a Two-Stage Competition in Mathematics', forthcoming *The Economic Journal*.

Irwing, Paul. Lynn, Richard (2005). "Sex differences in means and variability on the progressive matrices in university students: A meta□analysis." *British Journal of Psychology*. 96 (4), 505-524. https://doi.org/10.1348/000712605X53542 .

Jakobsson Niklas, Minna Levin, and Andreas Kotsadam (2013). 'Gender and Overconfidence: Effects of Context, Gendered Stereotypes, and Peer Group'. *Advances in Applied Sociology*. Vol.3 (2), 137-141.

Jetter, Michael and Jay K. Walker (2016) ‘Gender in Jeopardy!: The Role of Opponent Gender in High-Stakes Competition’. IZA Discussion Paper No. 9669 January 2016

Jetter, Michael and Jay K. Walker (2017) ‘Gender Differences in Competitiveness and Risk-Taking among Children, Teenagers, and College Students: Evidence from Jeopardy!’ IZA DP No. 11201.

Jetter, M., Walker, J. (2017), 'Anchoring in financial decision-making: Evidence from Jeopardy!', *Journal of Economic Behavior & Organization*, 141, pp. 164-176.

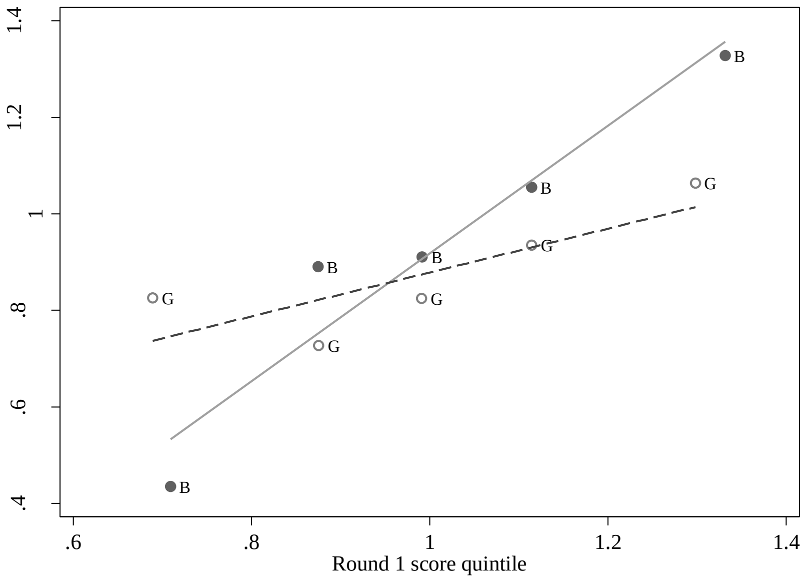
Lindberg, Sara M., Hyde, Janet Shibley, Petersen, Jennifer L.,Linn, Marcia C. (2010). "New trends in gender and mathematics performance: A meta-analysis." *Psychological Bulletin*,136(6), 1123-1135. http://psycnet.apa.org/doiLanding?doi=10.1037%2Fa0021276.

Lindquist, Gabriella Sjögren, Jenny Säve-Söderbergh (2011). ‘“Girls will be Girls”, especially among Boys: Risk-taking in the “Daily Double” on *Jeopardy.’ Economics Letters,* Vol. 112, Issue 2, Aug., pp 158-160.

Sarsons, Heather and Guo Xu (2016). ‘Confidence Men? Gender and Confidence: Evidence among Top Economists.’ Mimeo, Harvard.

Save-Soderbergh, J. and Lindquist, G. S. (2017). Children do not behave like adults: Gender gaps in performance and risktaking in a random social context in the high-stakes game shows Jeopardy and Junior Jeopardy. *The Economic Journal*, 127(603):1665–1692.

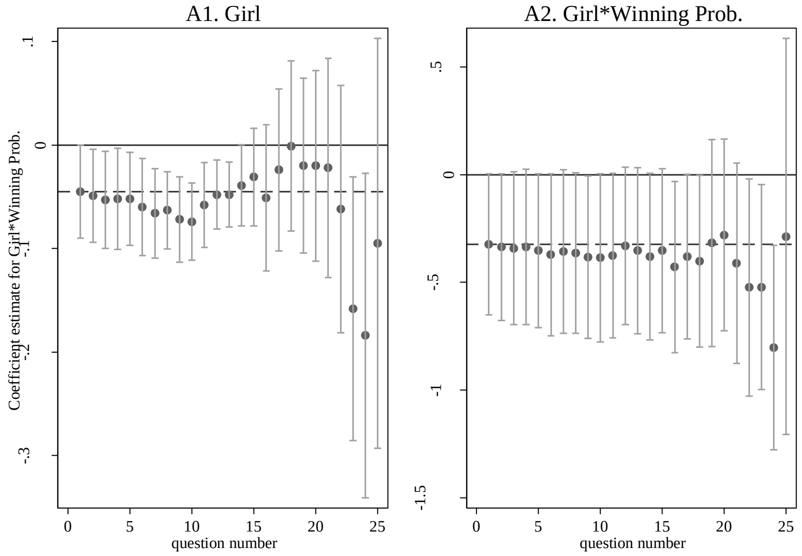
Figure 1. Gender Differences in Score between Round 1 and 2



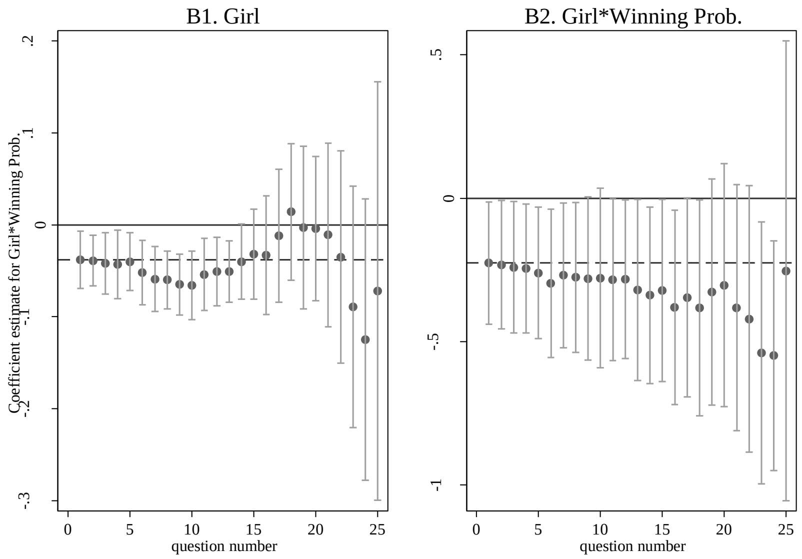
Note: The horizontal axis represents Round 1 score quintiles and the vertical axis the average score of Round 2 for each quintile. B represents boys and G girls. The lines are simple regression lines.

Figure 2. Gender Differences in End-of-Game Behavior

A. Buzzer



B. Accuracy



Note: The graphs are based on the estimates from equation (2) for the subsample where question number is greater or equal to t. The point estimates and 95% confidence intervals are presented. The horizontal dashed line represents the average effect presented in columns (2) and (5) of Table 7.

Table 1. Game Formats

Episode	Round 1			Round 2			Round 3	
	Buzzer	Knock-out	Penalty	Buzzer	Knock-out	Penalty	Buzzer	Penalty
Annual								
577-592	Y	N	Y	--	--	--	--	--
593-627	N	N	N	N	N	N	Y	Y
Survival								
629-635	Y	1	N	N	3	N	Y	N
636-650	Y	1	N	Y	3	N	Y	N
651-663	Y	1	N	Y	3	Y	Y	N
664-679	Y	2	N	Y	2	Y	Y	N
681-703	Y	2	Y	Y	2	Y	Y	N
704-714	N	2	N	Y	2	Y	Y	N
715-777	N	2	N	Y	2	N	Y	N

Note: Some episodes are not included in our sample because they are special shows or their video files are missing. In episodes 613 to 627, when no contestant obtained 500 points until round 3, there was a final round where the best two contestants competed for 500 points.

Table 2. Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All				Round 2 Advances			
	All	Boys	Girls	B-G	All	Boys	Girls	B-G
Girl	0.361				0.306			
	(0.481)				(0.461)			
Age	17.388	17.448	17.281	0.167	17.415	17.461	17.294	0.167
	(0.532)	(0.533)	(0.516)	[0.000]	(0.556)	(0.543)	(0.575)	[0.035]
Coed	0.407	0.390	0.437	-0.047	0.432	0.416	0.468	-0.052
	(0.491)	(0.488)	(0.497)	[0.165]	(0.496)	(0.494)	(0.501)	[0.306]
General school	0.839	0.826	0.862	-0.035	0.806	0.794	0.835	-0.041
	(0.368)	(0.379)	(0.346)	[0.165]	(0.396)	(0.405)	(0.373)	[0.311]
Private school	0.542	0.569	0.495	0.073	0.555	0.578	0.504	0.074
	(0.498)	(0.496)	(0.501)	[0.034]	(0.498)	(0.495)	(0.502)	[0.143]
Metropolitan	0.514	0.515	0.514	0.001	0.522	0.521	0.525	-0.005
	(0.500)	(0.500)	(0.501)	[0.978]	(0.500)	(0.500)	(0.501)	[0.929]
Medium-size cities	0.394	0.383	0.415	-0.033	0.385	0.381	0.396	-0.015
	(0.489)	(0.486)	(0.494)	[0.334]	(0.487)	(0.486)	(0.491)	[0.767]
Round 1 score	1.000	1.020	0.965	0.055	1.114	1.127	1.085	0.042
	(0.224)	(0.231)	(0.206)	[0.000]	(0.175)	(0.183)	(0.151)	[0.018]
Round 1 rank	3.000	2.904	3.169	-0.265	2.175	2.114	2.313	-0.199
	(1.376)	(1.393)	(1.331)	[0.005]	(0.948)	(0.962)	(0.903)	[0.039]
Round 1 top 1	0.210	0.242	0.154	0.088	0.335	0.375	0.245	0.130
	(0.408)	(0.429)	(0.361)	[0.002]	(0.472)	(0.485)	(0.431)	[0.007]
Round 1 top 2	0.439	0.468	0.388	0.080	0.670	0.695	0.612	0.084
	(0.497)	(0.499)	(0.488)	[0.020]	(0.471)	(0.461)	(0.489)	[0.081]
Advance to round 2	0.643	0.673	0.591	0.082				
	(0.479)	(0.470)	(0.492)	[0.001]				
Round 2 score					1.000	1.060	0.863	0.197
					(0.772)	(0.752)	(0.800)	[0.012]
N =	900	575	325		454	315	139	

Note: Standard deviations are presented in parentheses. "Advance to round 2" is defined for the shows with round 1 knock-outs (episode no. ≥ 629 in Table 1, N = 705). Age is missing for 371 contestants in the all sample and 208 for the round 2 advances sample. Columns (4) and (8) present the gender differences in sample means and p-values in brackets.

Table 3. Round 1 Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Standardized Score			Top 1			Top 2	
Girl	-0.055***	-0.051***	-0.054***	-0.088***	-0.082***	-0.086**	-0.080**	-0.077**	-0.080*
	(0.016)	(0.016)	(0.019)	(0.028)	(0.029)	(0.035)	(0.035)	(0.035)	(0.043)
General school		-0.049**	-0.063**		-0.091**	-0.109*		-0.026	-0.040
		(0.021)	(0.030)		(0.040)	(0.056)		(0.046)	(0.064)
Private school		0.007	0.008		-0.026	-0.029		0.002	-0.005
		(0.015)	(0.021)		(0.026)	(0.034)		(0.034)	(0.047)
Metropolitan		0.001	0.004		-0.018	-0.043		-0.005	-0.039
		(0.024)	(0.034)		(0.052)	(0.074)		(0.061)	(0.084)
Medium-size cities		-0.004	-0.004		-0.018	-0.029		-0.005	-0.050
		(0.023)	(0.033)		(0.053)	(0.074)		(0.061)	(0.084)
Constant	1.020***	1.056***	1.076***	0.242***	0.342***	0.369***	0.468***	0.485***	0.499***
	(0.006)	(0.026)	(0.041)	(0.012)	(0.052)	(0.080)	(0.014)	(0.062)	(0.095)
Episode FE	N	N	Y	N	N	Y	N	N	Y
Observations	900	900	900	900	900	900	900	900	900
R-squared	0.014	0.023	0.027	0.011	0.021	0.070	0.006	0.007	0.046

Note: Robust standard errors, clustered by episode, are presented in parentheses. Age dummies with missing age as the reference group are included in all specifications. * 10% significance; ** 5% significance; *** 1% significance.

Table 4. Gender Differences in Round 1 Score by Game Type

	(1) No knock-out	(2) Knock-out	(3) No buzzer	(4) Buzzer	(5) No penalty	(6) Penalty
A. Without Episode FE						
Girl	-0.004	-0.065***	0.002	-0.108***	-0.040**	-0.088**
	(0.033)	(0.018)	(0.019)	(0.024)	(0.018)	(0.034)
Observations	195	705	470	430	715	185
R-squared	0.017	0.036	0.024	0.059	0.018	0.065
B. With Episode FE						
Girl	-0.002	-0.069***	0.006	-0.114***	-0.042*	-0.095**
	(0.040)	(0.022)	(0.025)	(0.028)	(0.022)	(0.041)
Observations	195	705	470	430	715	185
R-squared	0.021	0.043	0.032	0.069	0.022	0.078

Note: Robust standard errors, clustered by episode, are presented in parentheses. All control variables of the full specification in Table 3 are included. * 10% significance; ** 5% significance; *** 1% significance.

Table 5. Probability of Advancing to Round 2

	(1) All	(2) No buzzer	(3) Buzzer	(4) No penalty	(5) Penalty
A. Without Episode FE					
Girl	-0.120***	-0.025	-0.205***	-0.082*	-0.280***
	(0.042)	(0.064)	(0.054)	(0.047)	(0.090)
Observations	705	345	360	590	115
R-squared	0.024	0.019	0.063	0.021	0.132
B. With Episode FE					
Girl	-0.113**	0.002	-0.211***	-0.069	-0.302**
	(0.052)	(0.083)	(0.066)	(0.057)	(0.110)
Observations	705	345	360	590	115
R-squared	0.061	0.032	0.111	0.057	0.154

Note: Linear probability models estimated by OLS. Robust standard errors, clustered by episode, are presented in parentheses. All control variables of the full specification in Table 3 are included. * 10% significance; ** 5% significance; *** 1% significance.

Table 6. Round 2 Performance

	(1)	(2)	(3)
A. Round 2 score			
Girl	-0.197**	-0.193**	-0.258*
	(0.083)	(0.086)	(0.132)
Episode FE	N	N	Y
Observations	454	454	454
R-squared	0.014	0.020	0.039
B. Score difference			
Girl	-0.155*	-0.155*	-0.208*
	(0.079)	(0.082)	(0.126)
Episode FE	N	N	Y
Observations	454	454	454
R-squared	0.009	0.016	0.042

Note: Robust standard errors, clustered by episode, are presented in parentheses. All control variables of the full specification in Table 3 are included. * 10% significance; ** 5% significance; *** 1% significance.

Table 7. Individual Play-by-Play Performance

	(1)	(2) Buzzer	(3)	(4)	(5) Accuracy	(6)
Girl	-0.045*	-0.045*		-0.039**	-0.038**	
	(0.022)	(0.023)		(0.016)	(0.016)	
Points		0.002	0.002		-0.002**	-0.002
		(0.002)	(0.002)		(0.001)	(0.002)
Winning prob.		0.131	-0.119		0.142	-0.238*
		(0.135)	(0.127)		(0.124)	(0.120)
Girl*Winning prob.		-0.323*	-0.369*		-0.226*	-0.156
		(0.167)	(0.194)		(0.109)	(0.152)
Constant	0.290***	0.199***	0.246***	0.198***	0.217***	0.297***
	(0.008)	(0.040)	(0.065)	(0.006)	(0.036)	(0.062)
Question number FE	Y	Y	Y	Y	Y	Y
Episode FE	Y	Y		Y	Y	
Individual FE			Y			Y
Observations	1,625	1,625	1,625	1,625	1,625	1,625
R-squared	0.008	0.021	0.065	0.003	0.008	0.057

Note: N = 1,625 (= 13 episodes*5 contestants*25 questions). Robust standard errors, clustered by episode in columns (1), (2), (4) and (5) and by individual in columns (3) and (6), are presented in parentheses. Columns (3) and (6) control for individual-specific fixed effects. Question number fixed effects are 24 dummies. * 10% significance; ** 5% significance; *** 1% significance.