

Erdemli, Candan; Vall Castelló, Judit

Article

Gender differences in online education

SERIEs - Journal of the Spanish Economic Association

Provided in Cooperation with:

Spanish Economic Association

Suggested Citation: Erdemli, Candan; Vall Castelló, Judit (2024) : Gender differences in online education, SERIEs - Journal of the Spanish Economic Association, ISSN 1869-4195, Springer, Heidelberg, Vol. 15, Iss. 4, pp. 349-388,
<https://doi.org/10.1007/s13209-024-00299-1>

This Version is available at:

<https://hdl.handle.net/10419/326960>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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



<https://creativecommons.org/licenses/by/4.0/>



Gender differences in online education

Candan Erdemli¹ · Judit Vall Castelló²

Received: 10 July 2023 / Accepted: 25 June 2024 / Published online: 5 August 2024
© The Author(s) 2024

Abstract

Online learning resources have become extremely popular, particularly after the restrictions caused by the Covid-19 outbreak. In this paper, we use data for Spain at the individual level from an online math learning platform which is used by children from over 100 countries, to document the gender differences in the context of online learning. We quantify the gender gaps in effort and relative performance outcomes and analyze whether the gaps differ by the gender of the parent who mainly supervises the children. Our main results point toward significant gender gaps in the relative performance outcomes in favor of boys, while the evidence for the effort gender gaps is only significant when we compare the siblings of the opposite gender (controlling for parent fixed effects). Further, we find that living in municipalities with more egalitarian gender norms is associated with narrower or positive gender gaps in effort outcomes, while we do not find such differences in the relative performance outcomes. Taking into account the increase in the use of online learning tools and their progressive integration into the regular educational system, our results provide important information to minimize gender biases in these new settings.

Keywords Education · Gender · Online learning

We are grateful to participants at the IWIP Seminars at IEB, University of Barcelona; the 34th Annual Conference of the European Society for Population Economics (ESPE); AEET - Labor Economics Meeting 2021; and the Applied Economics Conference: Labour, Health, Education, and Welfare; and to Libertad González for their helpful comments. We would like to express our appreciation to Smartick Company for generously providing us with the dataset used in this study. We would like to acknowledge financial support from Horizon Europe - European Research Executive Agency (REA) HORIZON-CL2-2022-DEMOCRACY-01; project number 101095106: “Education for Responsible Democratic Citizenship (DEMOCRAT).” Candan Erdemli acknowledges financial support from the FI-AGAUR. The replication material for the study is available at 10.5281/zenodo.11314201.

✉ Candan Erdemli
erdemlic@ceu.edu

Judit Vall Castelló
judit.vall@ub.edu

¹ Department of Economics and Business, Central European University and IEB, Quellenstraße 51, Wien 1100, Austria

² Department of Economics, University of Barcelona, IEB and CRES-UPF, J.M. Keynes, 1, 11, 08034 Barcelona, Spain

1 Introduction

During the last few years, the use of online learning tools in education has been on the rise in most developed countries. This trend was further accelerated by the Covid-19 outbreak and the subsequent school closures in 2020. Given the increasing prevalence of these tools and their likely integration into the mainstream education system, it is essential to understand the role played by parents and the efforts put forth by children. On the other hand, previous economics literature has documented gender differences in parental investment, not only in developing countries but also in the context of developed countries. In the USA, boys receive more paternal time than girls (Lundberg et al. 2007; Price 2008). In Canada, the UK, and the USA, parental time investments in teaching activities, such as reading, tend to favor girls, while fathers invest significantly more time in boys (Baker and Milligan 2016). Differential parental investment for boys and girls might vary between mothers and fathers as well. Mammen (2011) finds that in the USA, fathers allocate more time to their children if they have at least one boy, whereas mothers' total time investment is the same regardless of the gender composition of their children. Furthermore, not only the time invested may be different according to the gender match between parents and their children but also other elements affecting the educational outcomes may be different depending on this gender match. If similar parental gender bias and differences are present when using online learning tools, it may lead to a future gender gap in educational achievement and labor market outcomes.

This paper analyzes the gender gaps in online learning and how these gaps are correlated with the gender of the main supervisor. In particular, we quantify the effort and performance gaps between girls and boys when using an online learning platform in Spain and descriptively analyze whether the direction and the magnitude of these gaps differ depending on the gender of the parent who mainly supervises the child. Furthermore, we analyze whether the results are heterogeneous by the gender of the eldest sibling using the platform (to test for any role model effect), and the gender norms in the municipality of residence.

We find evidence of significant gender gaps in the relative performance outcomes in favor of boys, while the evidence for the effort gender gaps is only significant when we compare the siblings of the opposite gender. We also find suggestive evidence that the effort gaps are narrower—even positive in favor of girls for persistency outcome which indicates whether at least one session is completed in each month—when the main supervisor is the mother. For performance outcomes, on the other hand, we do not observe any differences in the gender gap depending on the gender of the main supervisor. We conduct the same analysis within a subsample of siblings of the opposite gender, including parent fixed effects. While gender gaps in performance outcomes in sign and magnitude are very similar to those in the full sample, we do not observe any significant differences in the gender gaps (neither for effort nor for performance outcomes) depending on the gender of the main supervisor in the siblings' subsample. We find heterogeneity by the gender of the eldest sibling in both fathers'

and mothers' samples. Specifically, while the gender gap in the extra time devoted to solving problems and the number of sessions completed per month come from families where the eldest is a girl, the gender gap in the ratio of correctly solved problems is more pronounced in families where the eldest is a boy. Additionally, we find that living in a municipality with more egalitarian gender norms is associated with gender gaps in favor of girls in effort outcomes. However, we do not observe such a difference in gender gaps in relative performance outcomes. The gender gap in delayed completion which represents the extra time devoted to completing the sessions decreases with age, while the gender gap in other outcomes does not significantly change across age groups. Lastly, we do not find significant differences in the gender gaps in outcomes by income levels of the municipality of residence.

This study contributes to the literature on gender differences in education in several ways. First, while previous research has documented gender gaps in traditional face-to-face education, the gender gap in online learning outcomes has been understudied. Our paper contributes to the literature by providing evidence from a contemporary online learning tool.¹ This allows us to document gender differences in completing the daily exercises, rather than the test outcomes, which is important in several ways. Balart et al. (2018) show that performance in cognitive tests, such as PISA, is influenced by non-cognitive skills (also named as personality traits, soft skills, or character skills), by using a decomposition in PISA test scores. Also, Anaya et al. (2022) show that the difficulty level at the beginning of tests may influence the success of the later questions. There is also some evidence in the psychology literature suggesting that under low-stakes testing conditions, some individuals try harder than others (Duckworth et al. 2011). This issue becomes more important if the non-cognitive effects differ by gender. An example of this is the study by Montolio and Taberner (2021), where they find that male university students outperform their female counterparts under high pressure. Considering these, it is essential to focus on outcomes that accumulate within a month as a result of daily exercises that are completed under no or little pressure. In addition, using data from an online learning platform helps us document gender gaps in student effort and motivation. Attending the test session, time spent on each test item, and self-reported effort have been used as primary measures of motivation in the previous literature documenting gender gaps in the motivation of students, where females are generally found to exert more effort compared to males (DeMars et al. 2013).² We add to this literature by analyzing the gender gap in effort when using an online learning platform, where we measure effort by the indicators created based on the number of completed online sessions in a month.

Second, understanding the source of gender discrimination or bias within the family is crucial for designing relevant and effective policies. Several studies show that moth-

¹ Online learning has become the interest of many researchers since the start of the Covid-19 outbreak, subsequent school closures, and the rise of online education. Among others, Chetty et al. (2020) find that children experienced a reduction in learning on a math learning platform used in US schools and Ikeda and Yamaguchi (2021) find that school closures during Covid-19 increased students' study time using an online learning service in Japan. However, Chetty et al. (2020) use school-level data, and they do not examine results by gender. On the other hand, Ikeda and Yamaguchi (2021) do not find any heterogeneous effect by gender.

² See DeMars et al. (2013) for a comprehensive literature review.

ers devote more time to childcare responsibilities than fathers, and this gap increased after the Covid-19 pandemic (Golin 2021; Andrew et al. 2020; Del Boca et al. 2020). We contribute to this part of the literature by comparing the gender gap in educational outcomes across two types of families: the ones where online learning is mainly supervised by the mother and those by the father. Although we cannot identify the causal impact of the division of family responsibilities and childcare in our setting, we descriptively show whether the gender gaps differ depending on the gender of the main supervisor of the children. Furthermore, our study differs from the previous literature by identifying the gender gap among siblings, while most of the existing studies focus on the gap among peers. This enables us to estimate the gender differences in learning outcomes between a boy and a girl raised in the same family.

Third, we contribute to the extensive literature on the relationship between the gender gap in education and gender norms. Several studies find that countries with more equal gender social norms tend to have smaller gender gaps in achievement or educational preferences (e.g., De San Román and De La Rica 2016; Nollenberger et al. 2016; Rodríguez-Planas and Nollenberger 2018; Gevrek et al. 2020). Although our data are specific to one country, we adapt this approach in our context to examine whether the gender gap in academic effort and performance is larger or smaller in municipalities with more egalitarian gender norms, which we proxy using the relative shares of females' and males' employment rates and contributions to household chores.

2 Data and methodology

2.1 Smartick

We use individual-level anonymous data from Smartick, an online learning platform that is used in over 100 countries, including but not limited to Spain, the UK, the USA, Mexico, Colombia, Peru, Brazil, and South Africa. The content is offered in different dialects of Spanish, English, and Portuguese, according to the user's choice. In our study, we use information on members residing in Spain since they constitute the largest proportion of Smartick users.

Smartick is one of the most widely used online learning platforms in Spain. It is a math learning tool that offers a 15-minute online math session every day.³ Students engage in four main areas: mental calculation, reasoning, logic, and programming. The platform is designed for children aged 4 to 14 and incorporates artificial intelligence to create personalized sessions tailored to the student's knowledge and abilities. These sessions are interactive and guided, with each exercise corrected immediately. If an answer is incorrect, it explains how it should have been done correctly. Therefore,

³ Although Smartick also offers reading exercises, we focus solely on the outcomes of math exercises due to data restrictions. Smartick released the reading section in September 2020. Therefore, we have a much smaller sample of children who completed the reading exercises compared to that of math, and we are not confident that the sample is large enough to derive any conclusions. In addition, one of our main outcomes to measure the speed of children in completing the exercises, namely the delayed completion variable for the reading section was not provided to us in the original data from Smartick.

parents' assistance is not required and Smartick encourages children to work on the sessions independently.

Since Smartick is mainly parent-based, either the father or the mother registers the child to the platform and he/she keeps track of the child's progress by checking the daily emails sent by Smartick regarding the child's attendance and performance in the sessions. In addition to reviewing the daily emails, parents can (and are encouraged by Smartick to) log in to the platform using their account to access a detailed performance analysis and the child's progress. This individual relationship established between one of the parents and Smartick automatically assigns responsibility for the child's online learning process to that parent.⁴

After a 7-day free trial period, parents are offered and choose from three types of paid contracts: one-month, three-month, and one-year contracts. We observe registration dates for free trial and paid contract, as well as monthly averages of the outcomes for each child starting from their contract date. We focus only on the duration of the first contract.⁵ Our dataset includes members registered for a paid contract between January 2019 and July 2021. We restrict our sample to the members who registered for a free trial period starting from January 2019 to observe their first contract outcomes, as well as those who registered for a paid contract before 20 June 2021, to ensure that we observe the outcomes of at least one complete month.

Our dataset includes children's age, gender, and the presence of a health condition.⁶ We also have information about the municipality of residence.⁷ Additionally, we observe the anonymous parent ID, which allows us to identify the siblings in the sample, as well as the gender of the parent who registered the child on the platform.

Since Smartick is a paid platform, the characteristics of its users are likely to differ from those of the overall Spanish population, raising concerns about the external validity of our results. Therefore, following Chetty et al. (2020), we present the demographic characteristics of Smartick users in our sample in Table 1 to show the extent of the selection. Since we do not have information on the characteristics of parents, we proxy the demographic characteristics using the income levels of the municipalities where they reside. Specifically, we compare the income quartiles of the Smartick municipalities that we define as those where at least 5 (and 1) Smartick members live, to the income quartiles of all Spanish municipalities. As we expect, the results reveal a selected sample in terms of income distribution.

⁴ Even though one parent registers the child and receives the emails about the learning process, we only use this relationship as a proxy of "being the main supervisor" since we cannot observe whether the actual supervisor is the one who registers on the platform.

⁵ We do not include the outcomes of the months of later contracts because we cannot clearly identify whether a missing value corresponds to a month without a paid contract or a month with no completed sessions.

⁶ The health conditions include high intellectual capacity, dyscalculia, dyslexia, intellectual disability, hearing disability, cerebral palsy, maturational delay, Down syndrome, attention deficit hyperactivity disorder (ADHD), autism, Asperger's syndrome, and pervasive developmental disorder.

⁷ The municipality of residence is detected by Smartick as the location where they first register on the platform. In some cases, we observe different locations for children registered by the same parent, or the location information is missing for a child but identified for his/her sibling. In these cases, we assign all children for a given parent the first identified location.

Moreover, the public school participation rate of the Smartick users⁸ is 49.4%, while the public school participation rate in the Spanish population at the primary school level is 68%. Overall, Table 1 shows the degree of selection in terms of income level and public school participation in our sample compared to the Spanish population. In this selected sample, we anticipate less gender bias since these parents are likely to be more educated and more aware of gender equality concerns compared to the average Spanish population. Therefore, we believe that our results represent a lower bound of gender bias in this context.

The outcomes of interest in our study include the total number of sessions completed in a month, the average ratio of time spent to complete the exercises to the expected time, and the average ratio of correctly solved problems. Using the total number of sessions completed in each month, we also create two binary outcome variables representing whether the child completes at least one session and at least twenty sessions in each month, respectively. We categorize these outcome variables into two groups: children's effort and relative performance. The measures of effort include persistency, completion, and sessions, while the measures of relative performance include delayed completion and accuracy. *Persistency* and *completion* are binary variables that take the value 1 if at least one session and twenty sessions are completed in each month, respectively. *Sessions* represents the average number of sessions completed per month. *Delayed completion* is the average ratio of time spent on problems to the expected time. Lastly, *accuracy* is the average ratio of correctly solved problems.

It is important to note that we only observe the accuracy and the delayed completion variables for the months with at least one completed session. We construct these outcomes as averages over the fully observed months of the first contract.⁹ For consistency, we exclude a monthly outcome if we do not observe the full month in our dataset. For example, if the first contract is for 3 months but our data only cover the period of the first month and a half, we only consider the outcomes of the first month.

The first panel in Table 4 shows the descriptive statistics in the full sample, which includes 28,236 children residing in Spain. Of these, 52% of them are girls, creating a balanced sample in terms of child gender. However, there is an imbalance in parent gender, with 66% of the parents being mothers. This ratio aligns with the difference in average minutes spent per day with children in teaching-related activities by mothers and fathers. According to the nationally representative Spain Time Use Survey 2009–2010, mothers and fathers spend an average of 5.26 and 2.68 min per day, respectively, on teaching-related activities with their children, indicating that 66% of the total teaching time is contributed by mothers.^{10, 11}

⁸ According to a survey conducted by Smartick in September 2021 on a representative sample of Smartick members, consisting of 2894 responders.

⁹ We conducted a separate analysis focusing on the outcomes of the first month, regardless of the contract type. The results are very similar to the baseline analysis and are available upon request.

¹⁰ A mother (father) is defined as a woman (man) who has a daughter or a son living in the same household, but not a grandparent.

¹¹ Although we expect our sample to be more egalitarian in terms of gender norms and the share of responsibilities than the average population because they reside in municipalities with higher than average income, we observe that the ratio of mothers registering their children to the platform is very similar to the mothers' share of time devoted to teaching activities with children on average. On the other hand,

Table 1 Comparison of demographic characteristics between the Smartick users and the Spanish population

	Smartick municipalities		Spanish population
	≥ 5 members	≥ 1 member	
<i>Average HH net income</i>			
25th percentile	27,442	25,248	21,864
Median	31,142	28,862	25,535
75th percentile	34,978	33,519	30, 411
Public school participation	49.40%	49.40%	68%
<i>Number of people</i>			
Number of Smartick members	26,648	28,236	–
2019 population	31,952,187	38,733,271	47,026,208
Number of municipalities	484	1302	8151

This table shows the demographic characteristics of Smartick users, following Chetty et al. (2020). Smartick municipalities are defined as municipalities where at least 1 or 5 Smartick members reside. Income and population information is based on 2018 statistics from the Spanish Statistical Office (INE)

The registered children in our dataset range in age from 4 to 16, with an average age of 8.54. Approximately 6% of the children in our sample have a health condition. Since we do not have information on household income levels, we use the average household net income in their municipality of residence as a proxy for income, with a mean value of 37,628 Euros. A total of 70% of our sample registered to the platform after the Covid-19 outbreak in Spain (after March 9, 2020), and 48% registered in the first 3 months of the outbreak when the schools were closed, and people had to stay at home. On average, each child has a total of 4.28 contracts, and 52% of the first contracts have a duration of 1 month. An average Smartick user completes 21.47 sessions in a month. Furthermore, 95% of the users complete at least one session, and 56% complete at least 20 sessions each month during their first contract period. Children, on average, spend 1.12 times the expected time to complete the sessions and answer 84% of the exercises correctly.

The second panel in Table 4 presents descriptive statistics for the sibling subsample, which consists of 7533 children who have at least one sibling of the opposite gender registered on the platform. We create this subsample to analyze gender differences among siblings, allowing us to control for family or household fixed effects while documenting the gender gaps. The distributions of variables in this sample closely resemble those in the full sample. In the siblings' subsample, the average number of children with the same parent is inherently higher, and we observe that 50% of the observations in this subsample are from households where the eldest child is a boy.

Table OA1 in the online appendix shows the mean differences of the variables by child gender. In the full sample, boys are on average 0.23 years younger, 4% less likely to be registered by their mothers, and 3% more likely to have a health

we observe that in the Time Use Survey 2009–2010, mothers' corresponding time share in higher-income families (more than 2000 Euros net household income per month) is around 60%, and mothers spend more than half of the total time spent in all income groups. Therefore, we believe that the Smartick users are not very different than the average population in terms of parental time investment to teaching-related activities.

condition compared to girls. In the siblings' subsample, similarly, boys are on average 0.25 years younger and 1% more likely to have a health condition than girls. Boys in the siblings' subsample completes on average 0.36 more sessions than girls per month. In both samples boys and girls come from similar backgrounds in terms of income level and gender-age composition. For the relative performance variables, we observe statistically significant gender gaps in favor of boys, without controlling for any characteristics; however, the magnitudes are small.

Table OA2 shows the mean differences by the gender of the parent. In line with the information in the previous table, fathers are 4% less likely to register girls on the platform compared to the mothers in the full sample. On average, children registered by their fathers are less likely to have a health condition and the number of children in their family is slightly lower compared to those registered by their mothers, in both the full sample and siblings' subsample. The probability of the eldest child being a boy is 2 pp lower in the families where the father registers children in the siblings' subsample. In the full sample, children registered by their fathers complete 0.23 fewer sessions per month and they have on average 2 pp lower completion rate than those registered by their mothers. However, we do not observe significant differences in the relative performance outcomes.

Figure 3 presents the evolution of the total number of new registrations per month. As indicated by the red line, the number of registrations increased disproportionately in the month when the Covid-19 outbreak started. However, starting from June 2020, the numbers returned to previous levels. Figures 4 and Fig. 5 show the total number of registrations per month, categorized by the gender of the child and by the gender of the parent, respectively. In terms of percentages, Fig. 6 does not show a clear pattern of gender pairs over the months. As indicated in Fig. 7, the majority of members are primary school students, representing a constant trend. However, there is an increase in registrations by preschool children along with a decrease in registrations by secondary school students over time. Figure 8 shows the percentage of new registrations by income categories, defined based on the quartiles of average income in the municipality of residence. The majority of members reside in high-income municipalities, and this trend remains stable over time.

These figures suggest that there is no clear pattern of change in the gender composition of children and parents, as well as in the age and income distribution over time, including the period after the Covid-19 outbreak. In other words, the observable characteristics of our sample do not exhibit significant changes over time, except for a slight shift in the age categories.¹²

¹² We are aware that the characteristics of new registrations in June 2021 seem different than in previous months. One potential reason is that we only include new registrations until June 21 since we restrict our sample to the fully observed months (see Sect. 2.1 for a detailed explanation), and related to this, we only observe 12 new registrations in this month.

2.2 Identifying the gender gaps

In order to identify the gender gap in effort and performance outcomes conditional on the main characteristics of children, we estimate the following linear regression:

$$Y_{ipm} = \beta_0 + \beta_1 G_{ipm} + \beta_2 X_{ipm} + \gamma_p + \theta_m + \varepsilon_{ipm} \quad (1)$$

where G_{ipm} is the dummy variable for the girl, X_{ipm} includes the child characteristics (age, presence of a health condition, total number of contracts, and type of the first contract), γ_p captures province fixed effects and θ_m captures contract year-month fixed effects. Standard errors are clustered at the province level. Our coefficient of interest, β_1 , captures the gender gap in the related outcome. We estimate this equation in the mothers' and fathers' samples separately, as well as in the full sample by controlling for the gender of the parent. We normalize the continuous outcome variables (sessions, delayed completion, and accuracy) before the analysis so that we can compare the estimated coefficients properly.¹³

Next, in order to control for the household fixed effects and identify the gender gap among siblings, we estimate the same regression on the siblings' subsample. As distinct from the previous regression, we include parent fixed effects instead of province fixed effects and we do not control for the number of total contracts, type of the first contract, and contract year-month since they are likely to be the same across children for a given parent. Formally, in Eq. 2, Z_{ipm} includes the child characteristics (age and presence of a health condition), α_p captures the parent fixed effects, and the rest of the variables are the same as in Eq. 1, except that p denotes parent instead of province.

$$Y_{ipm} = \beta_0 + \beta_1 G_{ipm} + \beta_2 Z_{ipm} + \alpha_p + \xi_{ipm} \quad (2)$$

For the first part of the heterogeneity analysis, we estimate the regression in which we interact the girl dummy (G_{ipm}) with the EldestBoy dummy. The EldestBoy dummy takes the value of 1 if the eldest sibling registered on the platform is a boy (or the siblings of the opposite gender are of the same age), and 0 otherwise. The caveat of this approach is that we can only observe the children who are registered on the platform. We cannot observe an elder brother or sister who is not registered on the platform.

Then, we aim to understand whether the gender gap in educational outcomes is more pronounced in municipalities where there is less gender equality. Therefore, as the second part of our heterogeneity analysis, we create two variables to proxy for gender equality at the municipality level, using information from the 2011 Spanish Census microdata obtained from INE (Instituto Nacional de Estadística).¹⁴ First, we

¹³ In our analysis, we normalize the continuous outcome variables within the full sample and siblings' sample separately before splitting them as mothers' and fathers' subsamples.

¹⁴ The 2011 Spanish Census is the most recent census available for our analysis. In the microdata, municipalities with a population of less than 20,000 are not identified for confidentiality reasons. We match information from 394 municipalities identified in the census to 24,896 observations (88%) in our dataset out of 28,236 observations in total.

define the employment gender ratio as the ratio of the female employment rate to the male employment rate in each municipality.¹⁵ Then, we create a dummy variable (E_m) which takes the value one if the employment gender ratio of a given municipality is equal to or above the median, and zero otherwise. Second, we define chores gender ratio as the ratio of the share of females reporting that they take care of most of the household chores to that of males in a given municipality. Then, we assign value one to variable C_m if the chores gender ratio in municipality m is below the median, and zero otherwise. We define both E_m and C_m variables in a way that value 1 reflects more egalitarian gender norms in municipality m . Then we estimate the previous equation by interacting the G_{ipm} variable with E_m , and C_m in separate regressions.¹⁶

3 Results

3.1 Main results

Before presenting our results for the mothers' and fathers' samples separately, we document the gender gaps in the full sample, conditional on the gender of the supervising parent, other characteristics of the child, province, and contract year-month fixed effects. We also analyze the gender gaps in the siblings' subsample, conditional on the characteristics of the child and parent fixed effects.

As shown in the first panel of Table 5, we do not observe gender differences in the effort outcomes in the full sample, but there are significant differences in relative performance outcomes including accuracy (8% of a standard deviation) and delayed completion (15% of a standard deviation). The first panel shows that children supervised by their mothers are less likely to be persistent, more likely to complete at least 20 sessions each month and complete more sessions per month on average, while they solve fewer problems correctly. In the second panel of Table 5, where we focus on the siblings of opposite gender and control for parent fixed effects, we observe that girls complete fewer sessions (4% of a standard deviation), solve fewer exercises correctly (5.4% of a standard deviation) and spend proportionally more time than expected in completing sessions (18% of a standard deviation) compared to their brothers.

Figure 1 shows estimated gender gaps in the full sample for the children registered by their mothers and fathers. The complete set of coefficient estimates is presented in Table 6 for both the mothers' and fathers' samples in two panels.

Regarding the effort variables (persistency, completion, and sessions), the gender gap is consistently positive when children are supervised by their mothers but negative when they are supervised by their fathers, although not all the coefficients

¹⁵ The female (male) employment rate in a municipality is defined as the ratio of the number of employed females (males) to the number of all females (males) living in the municipality and aged from 18 to 65 years old.

¹⁶ We also perform heterogeneity analysis where we use employment gender ratio, labor force participation gender ratio, and labor force participation ratio for the women with children under 16 years old at the province level instead of municipality level, by using data from 2019 employment statistics provided by INE. However, we do not find statistically significant differences between the provinces with more and less egalitarian gender norms. We believe that this is due to the lack of variation in the indicators of gender norms across provinces. Results are not shown in this paper but are available upon request.

are precisely estimated. Girls are 0.7% more likely to be persistent than boys when supervised by their mothers, and they complete fewer sessions (3.2% of a standard deviation) than boys when supervised by their fathers. Table 7 presents the results of full sample regressions where we include the interaction term of girl and father dummy variables. The third row indicates that the gender gap for children registered by their fathers is significantly different for persistency and sessions outcomes, although the coefficient estimate for the sessions outcome is not statistically significant after the multiple hypothesis correction. The gender gaps in the relative performance variables are negative in both the mothers' and fathers' samples. Specifically, girls solve fewer exercises correctly (8% and 9% of a standard deviation with mothers and fathers, respectively) and they spend proportionally more time than expected in completing sessions compared to boys (16% and 13% of a standard deviation with mothers and fathers, respectively).

We repeat the same analysis in the siblings' subsample by controlling for parent fixed effects.¹⁷ Figure 2 presents the estimated gender gaps among siblings, and Table 8 shows the full regression results. Similar to the results in the full sample, we observe opposite directions of gender gaps in the persistency variable in the mothers' and fathers' samples, though neither of the estimates is significantly different from zero. While the gender gap in completion is very close to zero and imprecisely estimated in the mothers' sample, it is 3% of a standard deviation and significant at the 90% level in the fathers' sample. Girls complete 2.3% and 7.4% of a standard deviation fewer sessions than boys in the mothers' and fathers' samples, respectively.

We observe significant gender gaps in the mothers' sample for both accuracy and delayed completion variables. However, in the fathers' sample, we only observe a significant gender gap in the delayed completion variable. These results for the siblings' sample are robust to including province fixed effects instead of parent fixed effects in the regressions, as presented in Table 16.

Some previous studies have found that females invest more effort under low-stakes conditions (DeMars et al. 2013; Segal 2012). We do not observe this pattern in our results, except for the persistency variable in mothers' full sample, for which we find a small gender gap in favor of girls.

3.2 Differences by gender composition of the siblings and gender norms

First, we explore whether the gender gaps are more pronounced in families where the eldest sibling is a boy. Table 2 shows that in the families where the eldest sibling is a girl, girls complete fewer sessions (10% of a standard deviation), solve more problems correctly (10% of a standard deviation), and are slower in completing sessions (32% of a standard deviation) than boys. In contrast, in families where the eldest sibling is a boy, girls complete more sessions (1.8% of a standard deviation), solve fewer problems correctly (17% of a standard deviation), and are slower in completing sessions (8.6%

¹⁷ We exclude the number of contracts and type of the first contract variables when including the parent fixed effects, as they are the same for children with the same parent in most cases.

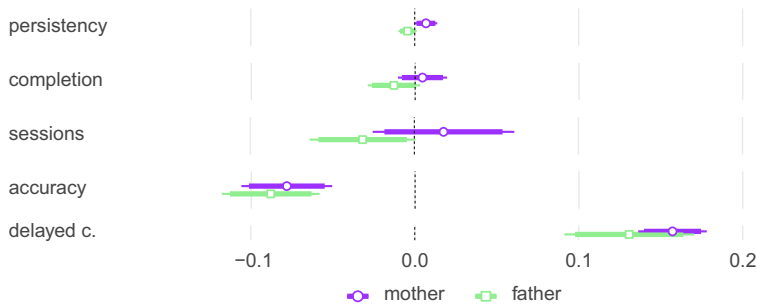


Fig. 1 Gender gap—full sample. The Figure shows the estimated gender gaps for different outcomes in mothers' and fathers' full samples separately, based on Eq. 1. Control variables are the age of the child, presence of a health condition, total number of contracts, type of the first contract, province, and contract year-month fixed effects. Standard errors are clustered at the province level. Thin and thick lines represent the 95% and 90% confidence intervals, respectively

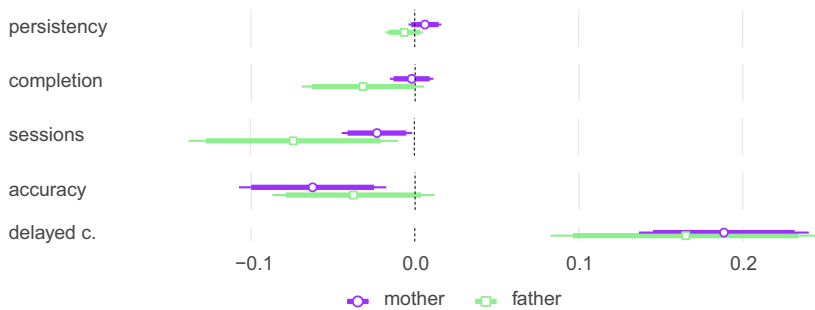


Fig. 2 Gender gap—siblings' sample. The Figure shows the estimated gender gaps for different outcomes in mothers' and fathers' siblings' samples separately, based on Eq. 2. Control variables are the age of the child, the presence of a health condition, and parent fixed effects. Standard errors are clustered at the province level. Thin and thick lines represent the 95% and 90% confidence intervals, respectively

of a standard deviation) than boys.¹⁸ Overall, these results suggest that the gender gap in accuracy is associated with the presence of an eldest brother using the platform, whereas the gap in sessions is associated with the presence of an elder sister. Although the delayed completion gap is consistently in favor of boys, we observe that it is narrower in the presence of an elder brother. This observation might be interpreted as girls taking more risks and being faster like their “role model” elder brothers.¹⁹

Next, we examine whether gender gaps in the outcomes are more pronounced in municipalities with lower gender equality. We interact the dummy variables constructed using employment gender ratio and chores gender ratio as proxies for gender equality in municipalities, with the girl dummy variable.²⁰ The first panel of Table 3

¹⁸ Calculations by using the coefficient estimates from Table 2: 1.8% of a sd: $-0.097 + 0.115 = 0.018$; 17% of a sd: $0.104 - 0.275 = 0.170$; 8.6% of a sd: $0.318 - 0.232 = 0.086$.

¹⁹ We acknowledge that, in principle, both elder brothers and younger brothers could potentially be seen as role models in our context. However, since we find narrower gender gaps in the presence of an elder brother, we interpret the results as a potential role model effect from the elder brothers.

²⁰ See Sect. 2 for definitions.

Table 2 Gender gap—differences by gender of the eldest child

	Persistency	Completion	Sessions	Accuracy	Delayed c
<i>Siblings' Sample</i>					
Girl	0.001 (0.005) [0.86] <i>[0.86]</i>	−0.021* (0.011) [0.05] <i>[0.07]</i>	−0.097*** (0.019) [0.00] <i>[0.00]</i>	0.104*** (0.021) [0.00] <i>[0.00]</i>	0.318*** (0.032) [0.00] <i>[0.00]</i>
EldestBoy	0.003 (0.006) [0.62] <i>[0.62]</i>	−0.013 (0.013) [0.32] <i>[0.40]</i>	−0.050 (0.030) [0.10] <i>[0.17]</i>	0.151*** (0.037) [0.00] <i>[0.00]</i>	0.172*** (0.028) [0.00] <i>[0.00]</i>
GirlxEldestBoy	0.001 (0.007) [0.86] <i>[0.86]</i>	0.021 (0.017) [0.22] <i>[0.27]</i>	0.115*** (0.035) [0.00] <i>[0.00]</i>	−0.275*** (0.034) [0.00] <i>[0.00]</i>	−0.232*** (0.062) [0.00] <i>[0.00]</i>
R ²	0.087	0.180	0.120	0.102	0.108
Num. obs	7533	7533	7533	7505	7502
N Clusters	50	50	50	50	50

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. Each regression includes a constant term, province, and contract year-month fixed effects. Control variables: age, health condition, number of contracts, type of the first contract

suggests that the gender gap in completion is not present in municipalities with more gender equality in employment. We observe similar patterns for the persistency and sessions outcomes, although the estimates are not precise. The second panel shows the differences in gender gaps by the “relative contribution to household chores” in the municipality of residence. Positive gender gaps in favor of girls in effort outcomes come from municipalities with higher gender equality in the relative contribution of females and males to the chores. However, we do not observe such differences in relative performance outcomes. Overall, these results suggest that municipalities with more egalitarian gender norms are associated with narrower or positive gender gaps in effort outcomes, while no such differences are found for relative performance outcomes.²¹

²¹ Table 9 shows the regression results where we use the continuous measures of the gender norms instead of binary measures. In this table, E_m is the standardized ratio of the female employment rate to the male employment rate in municipality m , and C_m is the ratio of the share of females reporting that they take care of most of the household chores to that of males in municipality m . The signs of the coefficient estimates are in line with the ones from the regressions with binary measures, although the estimates are less precise when we use continuous measures. (Note that lower values of the relative contribution to chores mean more egalitarian municipalities by definition of the continuous variable, while we had defined the binary variable in a way that higher values reflect more egalitarian norms. Therefore, the signs of the coefficient estimates for the interaction term of the girl and relative contribution to chores are the opposite in the regressions with the continuous measure.)

Table 3 Gender gap—differences by indicators of gender norms

	Persistency	Completion	Sessions	Accuracy	Delayed c
<i>Employment gender ratio</i>					
Girl	−0.010 (0.009) [0.27] <i>[0.27]</i>	−0.030** (0.014) [0.04] <i>[0.08]</i>	−0.067* (0.038) [0.08] <i>[0.10]</i>	−0.082** (0.041) [0.05] <i>[0.08]</i>	0.109** (0.041) [0.01] <i>[0.05]</i>
E_m	−0.006 (0.005) [0.24] <i>[0.30]</i>	−0.041 (0.030) [0.18] <i>[0.30]</i>	−0.113** (0.050) [0.03] <i>[0.14]</i>	0.024 (0.026) [0.36] <i>[0.36]</i>	−0.054 (0.033) [0.11] <i>[0.26]</i>
Girlx E_m	0.016 (0.010) [0.12] <i>[0.20]</i>	0.035* (0.018) [0.05] <i>[0.20]</i>	0.082 (0.049) [0.10] <i>[0.20]</i>	0.007 (0.047) [0.88] <i>[0.88]</i>	0.047 (0.040) [0.25] <i>[0.31]</i>
R ²	0.071	0.147	0.110	0.119	0.092
Num. obs	24896	24896	24896	24792	24783
N Clusters	50	50	50	50	50
<i>Relative contribution to chores</i>					
Girl	−0.005 (0.005) [0.34] <i>[0.34]</i>	−0.012 (0.010) [0.21] <i>[0.26]</i>	−0.036 (0.023) [0.13] <i>[0.22]</i>	−0.072** (0.027) [0.01] <i>[0.03]</i>	0.151*** (0.021) [0.00] <i>[0.00]</i>
C_m	−0.005 (0.006) [0.34] <i>[0.55]</i>	0.005 (0.009) [0.55] <i>[0.55]</i>	0.012 (0.016) [0.48] <i>[0.55]</i>	0.036* (0.020) [0.07] <i>[0.36]</i>	0.026 (0.019) [0.17] <i>[0.43]</i>
Girlx C_m	0.012** (0.006) [0.04] <i>[0.09]</i>	0.018* (0.011) [0.09] <i>[0.15]</i>	0.056** (0.021) [0.01] <i>[0.06]</i>	−0.005 (0.026) [0.85] <i>[0.99]</i>	−0.000 (0.019) [0.99] <i>[0.99]</i>
R ²	0.071	0.147	0.110	0.119	0.092
Num. obs	24896	24896	24896	24792	24783
N Clusters	50	50	50	50	50

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. Each regression includes a constant term, province, and contract year-month fixed effects, as well as the control variables: age, health condition, number of contracts, and type of the first contract

3.3 Differences by age and income levels

We then analyze whether gender gaps in effort and relative performance outcomes differ by the age of the children. To do so, we utilize both the continuous age variable and the categorical age variable that we define according to the school levels (preschool (age 4–5), primary school (6–11), and secondary school (12–16)). We interact the age variable with the girl dummy variable in the full sample and estimate the regression conditional on the gender of the parent who supervises the child.

In Table 10, Panel A shows that while the gender gap in the effort outcomes and accuracy variable does not change significantly by age, the gender gap in delayed completion decreases with age. In Table 10, Panel B shows that the gender gap in completion outcomes primarily comes from preschool kids and we observe a similar pattern in terms of the magnitudes for sessions outcome, although the estimations are not precise. While the accuracy gap comes from the primary school kids, the delayed completion gap is more pronounced for children in both preschool and primary school. Overall, Table 10 suggests that the gender gaps are stronger for younger girls, but most of them fade away with age.

Next, we create income categories within the Spanish population and among Smartick users based on the quartiles of average income in the municipality of residence. We then interact this categorical variable with the girl dummy variable to explore whether the sign and magnitude of the gender gaps differ depending on the income level. As shown in Tables 11 and 12, we do not observe any significant differences in the gender gaps by income categories. We believe that the main reason for this is the high homogeneity of our sample in terms of income level, as the majority of the users come from high-income families.²²

3.4 Robustness checks

One concern in the context of a paid learning platform is the potential negative selection on ability in our sample. In other words, families might be more likely to register their children on this platform if their children are not performing well academically, and this selection might also vary by the gender of the child, which could potentially affect our results. To address this concern, we repeat the analysis in the “covid sample”, which we define as the sample of children who registered on the platform between March 9 and May 31, 2020. This period corresponds to the first wave of the Covid-19 outbreak in Spain. During this period, parents were seeking educational tools to help their children stay on track because schools were closed unexpectedly, and distance

²² We present the heterogeneity analysis by the gender of the eldest sibling, age, income at the municipality level and gender norms measures at the municipality level in the full sample only. We do not run the heterogeneity analysis by gender norms and income in the siblings’ subsample and including the parent fixed effects. The main reason is that in the siblings’ subsample many municipalities have a few families with at least one girl and one boy using the Smartick platform, making it difficult to estimate the heterogeneity by municipality-level information, within families. We also cannot run the heterogeneity analysis by gender of the eldest sibling in the siblings’ subsample and including parent fixed effects, since there is no variation in the gender of the eldest sibling for a given parent. However, we explore the heterogeneity by age in the siblings’ subsample by including parent fixed effects. The results are in line with those in the full sample and are available upon request.

education did not start immediately after the school closures. Therefore, we believe that potential selection on ability and gender in registrations plays a minimal role in this time period. This idea is supported by the increase in the total number of registrations and differences in the characteristics and outcome variables of the children registered during this period, as shown in Fig. 3 and Table OA3.

When we compare the characteristics and outcome variables of children in the covid sample to the rest of the sample (as shown in Table OA3), we find some notable differences. In this sample, the share of girls registered to the platform is 3% lower, while the share of mothers is 4% higher. The children in the covid sample are, on average, about half a year younger, less likely to have a health condition and live in municipalities with, on average, lower income levels. While in the full sample, the average number of children in the families is slightly higher, in the siblings' sample it is slightly lower. In addition to the characteristics, we also observe differences in the distribution of outcome variables. Children in the covid sample complete, on average, 2.27 more sessions per month, are 1% more likely to be persistent and 12% more likely to complete at least 20 sessions per month compared to the others. They also spend less time on completing the sessions and solve slightly more exercises correctly. These differences suggest that if there is any selection on ability in our sample, this selection would be the smallest in this period. Furthermore, the income level of the municipalities where the children live is closer to the average Spanish population compared to the rest of the sample.

We estimate the gender gaps in the “covid sample” to examine whether our results are affected by a potential selection on ability and gender. As shown in Fig. 9 and Fig. 10, the magnitudes of estimated gender gaps in the covid sample are very similar to those in our baseline analysis. However, the estimates are not as precise as the baseline estimates in the siblings' subsample, which might be due to the lower number of observations. The magnitudes of the coefficient estimates are presented in Tables 13 and 14.

Next, we divide the full sample into two groups: those who registered before and after the Covid-19 outbreak in Spain. In Table 15, we present the regression results where we interact the girl dummy variable with the father dummy variable in these two subsamples.

We find that the gender gap in the persistency outcome and its negative association with the fathers' supervision come from the registrations made before the outbreak. Although the signs of the estimates are mostly in the same direction in the sample of registrations after the outbreak for the effort outcomes, they are not significantly different from zero. Regarding the relative performance outcomes, we observe that in the sample of registrations after the outbreak, the gender gap in the delayed completion outcome is slightly lower for the children with fathers.

We further examine whether the estimated gender gaps in the siblings' subsample are robust to including province fixed effects instead of parent fixed effects in the regressions, as shown in Table 16. While the magnitudes of the estimates slightly change, they remain very similar to the baseline estimates and maintain the same statistical significant levels.

For the binary outcome variables (persistency and completion), we estimate logit regressions as an alternative to the linear probability model. Table 17 presents the

results for the full sample and siblings' sample, respectively. The estimated gender gaps are consistent in direction with those estimated using linear probability models.

Next, we exclude children with a health condition from the sample to assess whether our results are affected by the presence of a health condition. Table 18 shows the estimated gender gaps in the full sample of children without a health condition. The results closely resemble the baseline results reported in Table 7.

We construct persistency and completion variables based on the number of sessions completed per month. We find that girls are 0.7% more likely to complete at least one session per month (persistency) than boys when the main supervisor is the mother. However, we do not observe gender gaps in the completion variable (see Table 6).

To assess the stability of our results to different cut-offs of the number of sessions completed per month, we construct additional outcome variables with various cut-offs (1, 6, 11, ..., 26), where the outcome of the first cut-off is the same as the persistency variable. The results are presented in Fig. 11, where we estimate Eq. 1 for different cut-off outcomes. In the mothers' sample, there are no significant gender gaps for different cut-offs, except for the cut-off of 1. However, in the fathers' sample, we observe that girls are 1.8% (2%) less likely than boys to complete at least 11 sessions (16 sessions) per month.

In our main specification, we use the continuous age variable to control for the differences in the gender gap across different ages. Additionally, in Fig. 12, we present our main results by including the age fixed effects instead of the continuous age variable. The significance levels and the magnitudes of the coefficient estimates in this model are very similar to those of our main specification.

To further control for differences by age, we repeat our analysis by normalizing all the outcome variables within each age. This is particularly important for the siblings' sample, where we compare siblings at different ages, and the difficulty level of exercises for different ages is likely to be different. The results presented in Figs. 13 and 14 are very similar to our baseline results. If anything, we observe larger magnitudes of coefficient estimates for the persistency variable, with the signs remaining the same as in the baseline.

3.5 Potential mechanisms and discussion

Our results suggest gender gaps in online math learning in favor of boys in performance outcomes. For the effort outcomes, we do not find gender gaps in the full sample on average; however, we find a gender gap in favor of boys for the sessions variable when we focus on the siblings' subsample. Further analyzing the gender differences in the mothers' and fathers' samples, we find that in the full sample, the gender gap in the persistency variable is positive for those supervised by their mothers but negative for those supervised by their fathers, with a significant difference robust to Benjamini and Hochberg (1995) multiple hypothesis testing. However, this difference is not statistically significant in the siblings' subsample and is not observed for other variables measuring effort.²³ For this reason, we interpret our findings as suggestive evidence

²³ Although we find a significant difference in the gender gap by the gender of the parent in the full sample for the sessions variable, this difference is not robust to multiple hypothesis testing.

for narrower/positive gender gaps in effort outcomes when the main supervisor is the mother. While data limitations prevent us from analyzing the mechanisms behind these results, in this section, we discuss the potential explanations for our findings.

In addition to the better math performance of boys, which is also evident in traditional education, another contributing factor may be that boys tend to be more interested in video games compared to girls. This interest could explain differences in effort outcomes, as boys might be more motivated to complete sessions they find enjoyable. However, it is important to note that we do not observe this pattern for children with their mothers as the main supervisors. This difference becomes more pronounced when examining the probability of completing at least 1, 11, and 16 sessions per month.

One possible explanation for our suggestive results on the differences in effort gender gaps by the gender of the main supervisor is that mothers might provide greater support to their daughters in engaging with math learning activities compared to fathers, which could lead to the disappearance of the difference in effort gaps. This may also explain why we do not observe such differences in performance gaps.

Since Smartick is a platform where parental assistance is not required during the sessions, parents might mainly influence the child's decision to start the exercises. However, due to data limitations, we cannot directly observe the actual support provided by parents. Both parents may provide support even if only one has registered the child. Therefore, we cannot rule out the possibility that the differences in gender gaps depending on the gender of the supervisor might be driven by differences in family and household characteristics that are not observable to us.

One might expect that in more gender-egalitarian municipalities, fathers are more likely to supervise children. This expectation could create tension in our interpretation of the results since we find (i) suggestive evidence for larger gender gaps in effort when the father is the main supervisor, and (ii) living in a municipality with more egalitarian gender norms is associated with positive gender gaps in effort outcomes in favor of girls. To investigate this, we estimate a linear regression of the dummy variable, which takes the value of 1 if the mother supervises, on the gender norm measures we construct (as detailed in Sect. 3.2). We control for the gender of the child and include other control variables in our main specification (Eq. 1). Table 19 shows that there is no evidence to support a higher likelihood of fathers' supervision in more egalitarian municipalities.

Finally, it is important to emphasize that our results are specific to mathematics training. Given that performance and effort by gender have been observed to differ significantly across traditionally male or female subjects, the outcomes may vary when examining online learning in other subjects.

4 Conclusion

This paper investigates gender gaps in effort and relative performance outcomes in the context of online education, focusing on children primarily supervised by their mothers and fathers. Utilizing data from a widely used online learning platform in Spain, we find significant gender gaps in performance outcomes in favor of boys, and these results hold when we compare siblings of the opposite gender. Moreover, we find

suggestive evidence that the gender gaps in the effort outcomes are more pronounced for the children who are supervised by their fathers.

Since our study focuses on the context of online learning, our outcome variables may not directly correspond to those commonly used in previous literature, such as standardized test scores. Among our outcome variables, the most comparable to standardized test scores is accuracy, which represents the average ratio of correctly solved problems when using the online platform for each child. We observe an accuracy gender gap of 8% of a standard deviation²⁴ favoring boys. This estimated gap is of a similar magnitude to the math gender gap in PISA 2018, where boys outperformed girls with a difference of 6.8% of a standard deviation.²⁵

We find that living in a municipality with more egalitarian gender norms is associated with gender gaps in effort outcomes in favor of girls, which is in line with the previous findings in the literature (e.g., by De San Román and De La Rica 2016; Nollenberger et al. 2016; Rodríguez-Planas and Nollenberger 2018; Gevrek et al. 2020). However, the difference in effort outcomes does not translate into the gender gap in relative performance outcomes.

We contribute to the existing literature on gender gaps in learning outcomes by providing new evidence from an online learning context while most previous studies focus on traditional education settings.

The increasing adoption of online learning technologies by both parents and schools suggests that these tools will become an important part of regular education systems in the near future. In light of these developments, we believe that it is very important to document gender differences in the use of these tools, as these are likely to translate into differences in further education and labor market outcomes.

While our analysis focuses on Spain, it is important to note that the Smartick platform is utilized by children in many other countries, with the option for three different languages. Therefore, we believe that our study is relevant not only in the Spanish context but also in the context of other countries.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13209-024-00299-1>.

Declarations

Conflict of interest None.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

²⁴ In the siblings' subsample, where we compare opposite-gender siblings, the corresponding gender gap is 5.4% of a standard deviation.

²⁵ According to the authors' own calculations based on PISA 2018 descriptive statistics reported in OECD (2019).

Appendix

See Tables 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, Figs. 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 and 14.

Table 4 Descriptive statistics

Variable	N	Mean	Std. Dev.	Min	Max
<i>1. Full sample</i>					
Female	28236	0.52	0.5	0	1
Mother	28236	0.66	0.48	0	1
Age	28236	8.54	2.33	4	16
Health condition	28236	0.06	0.24	0	1
Income	27985	37628	9587	16692	90902
Registered after covid	28236	0.7	0.46	0	1
Covid sample	28236	0.48	0.5	0	1
Number of children	28236	1.72	0.88	1	9
Total contracts	28236	4.28	4.16	1	31
First contract	28236				
1	14615	52%			
3	10806	38%			
12	2815	10%			
Sessions	28236	21.47	7.91	0	32
Persistence	28236	0.95	0.22	0	1
Completion	28236	0.56	0.5	0	1
Delayed completion	28105	1.12	0.28	0.3	3.26
Accuracy	28116	0.84	0.06	0.22	1
<i>2. Siblings' sample</i>					
Female	7533	0.5	0.5	0	1
Mother	7533	0.65	0.48	0	1
Age	7533	8.56	2.37	4	16
Health condition	7533	0.05	0.22	0	1
Income	7459	38137	9987	17354	90902
Registered after covid	7533	0.71	0.45	0	1
Covid sample	7533	0.51	0.5	0	1
Number of children	7533	2.45	0.83	2	9
Eldest is a boy	7533	0.5	0.5	0	1
Total contracts	7533	4.36	4.29	1	30
First contract	7533				
1	3926	52%			
3	2782	37%			
12	825	11%			
Sessions	7533	22.14	7.76	0	32
Persistence	7533	0.95	0.21	0	1

Table 4 continued

Variable	N	Mean	Std. Dev.	Min	Max
Completion	7533	0.58	0.49	0	1
Delayed completion	7502	1.09	0.26	0.3	2.84
Accuracy	7505	0.84	0.06	0.42	1

The table shows the descriptive statistics for the full sample and siblings' sample. The income variable is measured as the average household net income of the municipality of residence, and the covid sample corresponds to the sample of users who registered between March 9, 2020, and May 30, 2020

Table 5 Gender gaps in the full sample and siblings' sample

	Persistency	Completion	Sessions	Accuracy	Delayed c.
<i>Full sample</i>					
Girl	0.003 (0.003) [0.25] [0.41]	−0.002 (0.005) [0.77] [0.97]	0.001 (0.014) [0.97] [0.97]	−0.080*** (0.009) [0.00] [0.00]	0.148*** (0.011) [0.00] [0.00]
Mother	−0.005* (0.003)	0.011** (0.005)	0.024** (0.012)	−0.048*** (0.014)	0.007 (0.010)
Age	−0.004*** (0.000)	0.002* (0.001)	0.004* (0.003)	−0.124*** (0.004)	−0.068*** (0.002)
Health Cond.	−0.015* (0.007)	−0.017* (0.008)	−0.044** (0.021)	−0.174*** (0.033)	0.346*** (0.026)
Num of contr	0.004*** (0.000)	0.023*** (0.001)	0.058*** (0.002)	0.008*** (0.001)	0.002 (0.002)
1st C:12	−0.143*** (0.007)	−0.241*** (0.011)	0.092*** (0.022)	−0.357*** (0.019)	0.538*** (0.017)
1st C:3	−0.056*** (0.004)	−0.205*** (0.013)	−0.044** (0.020)	−0.118*** (0.012)	0.112*** (0.013)
Constant	0.985*** (0.013)	0.476*** (0.039)	−0.418*** (0.081)	1.472*** (0.103)	0.631*** (0.051)
Province FE	Yes	Yes	Yes	Yes	Yes
Cont. Month FE	Yes	Yes	Yes	Yes	Yes
R ²	0.068	0.144	0.107	0.121	0.092
Num. obs.	28236	28236	28236	28116	28105
N Clusters	50	50	50	50	50
<i>Siblings' sample</i>					
Girl	0.002 (0.003) [0.53] [0.53]	−0.013 (0.008) [0.11] [0.13]	−0.041*** (0.011) [0.00] [0.00]	−0.054*** (0.016) [0.00] [0.00]	0.181*** (0.020) [0.00] [0.00]
Age	−0.005*** (0.001)	−0.007*** (0.002)	−0.025*** (0.004)	−0.037*** (0.011)	−0.031*** (0.011)

Table 5 continued

	Persistency	Completion	Sessions	Accuracy	Delayed c.
Health Cond.	−0.016 (0.012)	−0.022 (0.029)	−0.028 (0.064)	−0.348*** (0.083)	0.276** (0.127)
Constant	1.069*** (0.018)	1.096*** (0.032)	0.975*** (0.050)	0.162 (0.141)	−0.650*** (0.142)
Parent FE	Yes	Yes	Yes	Yes	Yes
Cont. Month FE	No	No	No	No	No
R ²	0.718	0.814	0.869	0.685	0.635
Num. obs.	7533	7533	7533	7505	7502
N Clusters	50	50	50	50	50

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. The coefficient estimates are from the regressions of five different outcome variables given in different columns, on the indicator variables of the child's gender, parent's gender, and the control variables shown in the table

Table 6 Gender gaps in mothers' and fathers' (full) sample

	Persistency	Completion	Sessions	Accuracy	Delayed c.
<i>Mothers' sample</i>					
Girl	0.007* (0.003) [0.05] <i>[0.09]</i>	0.004 (0.007) [0.55] <i>[0.55]</i>	0.018 (0.021) [0.41] <i>[0.51]</i>	−0.078*** (0.014) [0.00] <i>[0.00]</i>	0.157*** (0.010) [0.00] <i>[0.00]</i>
R ²	0.069	0.151	0.115	0.123	0.098
Num. obs.	18516	18516	18516	18435	18430
N Clusters	50	50	50	50	50
<i>Fathers' sample</i>					
Girl	−0.004 (0.003) [0.12] <i>[0.12]</i>	−0.013 (0.008) [0.11] <i>[0.12]</i>	−0.032* (0.016) [0.06] <i>[0.09]</i>	−0.088*** (0.015) [0.00] <i>[0.00]</i>	0.131*** (0.020) [0.00] <i>[0.00]</i>
R ²	0.074	0.141	0.103	0.126	0.088
Num. obs.	9720	9720	9720	9681	9675
N Clusters	50	50	50	50	50
<i>Two sample t-test</i>					
Difference	0.011	0.017	0.05	0.01	0.03
p-value	(0.01)	(0.11)	(0.07)	(0.63)	(0.24)

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. Each regression includes a constant term, province, and contract year-month fixed effects. Control variables: age, health condition, number of contracts, type of the first contract

Table 7 Gender gaps in the full sample

	Persistency	Completion	Sessions	Accuracy	Delayed c.
Girl	0.007* (0.003) [0.05] <i>[0.09]</i>	0.005 (0.007) [0.54] <i>[0.54]</i>	0.017 (0.021) [0.41] <i>[0.51]</i>	−0.078*** (0.014) [0.00] <i>[0.00]</i>	0.157*** (0.011) [0.00] <i>[0.00]</i>
Father	0.011*** (0.004) [0.01] <i>[0.04]</i>	−0.002 (0.007) [0.71] <i>[0.89]</i>	0.000 (0.020) [1.00] <i>[1.00]</i>	0.051** (0.022) [0.03] <i>[0.06]</i>	0.005 (0.014) [0.71] <i>[0.89]</i>
GirlxFather	−0.011*** (0.004) [0.01] <i>[0.04]</i>	−0.018 (0.011) [0.12] <i>[0.17]</i>	−0.049* (0.027) [0.08] <i>[0.17]</i>	−0.007 (0.023) [0.76] <i>[0.76]</i>	−0.025 (0.017) [0.14] <i>[0.17]</i>
Age	−0.004*** (0.000)	0.002* (0.001)	0.004* (0.003)	−0.124*** (0.004)	−0.068*** (0.002)
Health Cond.	−0.014* (0.007)	−0.016* (0.008)	−0.044** (0.022)	−0.173*** (0.033)	0.346*** (0.026)
Num of contr	0.004*** (0.000)	0.023*** (0.001)	0.058*** (0.002)	0.008*** (0.001)	0.002 (0.002)
1st C:12	−0.143*** (0.007)	−0.241*** (0.012)	0.092*** (0.022)	−0.357*** (0.019)	0.538*** (0.017)
1st C:3	−0.056*** (0.004)	−0.205*** (0.013)	−0.044** (0.020)	−0.118*** (0.012)	0.112*** (0.013)
Constant	0.978*** (0.014)	0.484*** (0.039)	−0.402*** (0.087)	1.423*** (0.092)	0.633*** (0.047)
Province FE	Yes	Yes	Yes	Yes	Yes
Cont. Month FE	Yes	Yes	Yes	Yes	Yes
R ²	0.068	0.144	0.107	0.121	0.092
Num. obs.	28236	28236	28236	28116	28105
N Clusters	50	50	50	50	50

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. The coefficient estimates are from the regressions of five different outcome variables given in different columns, on the indicator variables of the child's gender, parent's gender, their interaction, and the control variables

Table 8 Gender gaps in mothers' and fathers' (siblings') sample

	Persistency	Completion	Sessions	Accuracy	Delayed c.
<i>Mothers' sample</i>					
Girl	0.006 (0.005) [0.22] <i>[0.27]</i>	−0.002 (0.007) [0.75] <i>[0.75]</i>	−0.023** (0.011) [0.04] <i>[0.06]</i>	−0.063*** (0.022) [0.01] <i>[0.02]</i>	0.189*** (0.026) [0.00] <i>[0.00]</i>
R ²	0.703	0.824	0.870	0.687	0.634
Num. obs.	4889	4889	4889	4868	4867
N Clusters	49	49	49	49	49
<i>Fathers' sample</i>					
Girl	−0.006 (0.006) [0.27] <i>[0.27]</i>	−0.032* (0.019) [0.09] <i>[0.16]</i>	−0.074** (0.032) [0.02] <i>[0.06]</i>	−0.038 (0.025) [0.13] <i>[0.17]</i>	0.165*** (0.041) [0.00] <i>[0.00]</i>
R ²	0.753	0.796	0.868	0.684	0.636
Num. obs.	2644	2644	2644	2637	2635
N Clusters	50	50	50	50	50
<i>Two sample t-test</i>					
Difference	0.012	0.030	0.051	−0.025	0.023
p-value	(0.10)	(0.13)	(0.13)	(0.45)	(0.63)

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. Each regression includes a constant term, province, and contract year-month fixed effects. Control variables: age, health condition, number of contracts, type of the first contract

Table 9 Gender gap—differences by continuous indicators of gender norms

	Persistency	Completion	Sessions	Accuracy	Delayed c.
<i>Employment gender ratio</i>					
Girl	0.003 (0.006) [0.62] <i>[0.62]</i>	−0.011 (0.008) [0.16] <i>[0.27]</i>	−0.019 (0.019) [0.34] <i>[0.42]</i>	−0.078*** (0.015) [0.00] <i>[0.00]</i>	0.143*** (0.019) [0.00] <i>[0.00]</i>
E_m	0.002 (0.005) [0.74] <i>[0.74]</i>	−0.011 (0.016) [0.50] <i>[0.74]</i>	−0.051* (0.028) [0.08] <i>[0.38]</i>	0.005 (0.012) [0.68] <i>[0.74]</i>	−0.013 (0.017) [0.46] <i>[0.74]</i>
Girl $\times E_m$	0.002 (0.007) [0.80] <i>[0.87]</i>	0.016* (0.009) [0.07] <i>[0.34]</i>	0.033 (0.027) [0.22] <i>[0.56]</i>	0.003 (0.018) [0.87] <i>[0.87]</i>	0.010 (0.019) [0.60] <i>[0.87]</i>

Table 9 continued

	Persistency	Completion	Sessions	Accuracy	Delayed c.
R ²	0.070	0.147	0.110	0.119	0.092
Num. obs.	24896	24896	24896	24792	24783
N Clusters	50	50	50	50	50
<i>Relative contribution to chores</i>					
Girl	−0.000	−0.005	−0.007	−0.076***	0.138***
	(0.004)	(0.007)	(0.018)	(0.018)	(0.018)
	[1.00]	[0.43]	[0.70]	[0.00]	[0.00]
	<i>[1.00]</i>	<i>[0.72]</i>	<i>[0.88]</i>	<i>[0.00]</i>	<i>[0.00]</i>
<i>C_m</i>	0.003	−0.006	−0.000	−0.010	0.012
	(0.004)	(0.006)	(0.017)	(0.015)	(0.028)
	[0.39]	[0.32]	[0.98]	[0.51]	[0.68]
	<i>[0.85]</i>	<i>[0.85]</i>	<i>[0.98]</i>	<i>[0.85]</i>	<i>[0.85]</i>
Girlx <i>C_m</i>	−0.008*	−0.013	−0.024	0.000	−0.023
	(0.005)	(0.009)	(0.017)	(0.019)	(0.021)
	[0.10]	[0.15]	[0.15]	[0.99]	[0.29]
	<i>[0.25]</i>	<i>[0.25]</i>	<i>[0.25]</i>	<i>[0.99]</i>	<i>[0.36]</i>
R ²	0.071	0.147	0.109	0.119	0.092
Num. obs.	24896	24896	24896	24792	24783
N Clusters	50	50	50	50	50

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. Each regression includes a constant term, province, and contract year-month fixed effects, as well as the control variables: age, health condition, number of contracts, and type of the first contract

Table 10 Gender gap in the full sample—differences by age

	Persistency	Completion	Sessions	Accuracy	Delayed c.
<i>A. Continuous age variable</i>					
Girl	0.012*	−0.004	−0.003	−0.091***	0.451***
	(0.006)	(0.016)	(0.038)	(0.031)	(0.050)
	[0.05]	[0.82]	[0.94]	[0.01]	[0.00]
	<i>[0.09]</i>	<i>[0.94]</i>	<i>[0.94]</i>	<i>[0.01]</i>	<i>[0.00]</i>
Age	−0.003***	0.002	0.004	−0.125***	−0.050***
	(0.001)	(0.001)	(0.003)	(0.004)	(0.003)
	[0.00]	[0.14]	[0.16]	[0.00]	[0.00]
	<i>[0.00]</i>	<i>[0.16]</i>	<i>[0.16]</i>	<i>[0.00]</i>	<i>[0.00]</i>

Table 10 continued

	Persistency	Completion	Sessions	Accuracy	Delayed c.
Girl*Age	−0.001 (0.001) [0.16] <i>[0.41]</i>	0.000 (0.002) [0.91] <i>[0.94]</i>	0.000 (0.005) [0.94] <i>[0.94]</i>	0.001 (0.003) [0.69] <i>[0.94]</i>	−0.035*** (0.006) [0.00] <i>[0.00]</i>
R ²	0.068	0.144	0.107	0.121	0.094
Num. obs.	28236	28236	28236	28116	28105
N Clusters	50	50	50	50	50
<i>B. Age categories</i>					
Girl	0.005 (0.006) [0.39] <i>[0.48]</i>	−0.041*** (0.014) [0.01] <i>[0.03]</i>	−0.051* (0.028) [0.08] <i>[0.13]</i>	0.026 (0.040) [0.52] <i>[0.52]</i>	0.127** (0.058) [0.03] <i>[0.08]</i>
Primary	−0.007* (0.004) [0.10] <i>[0.11]</i>	0.018 (0.011) [0.11] <i>[0.11]</i>	0.079*** (0.024) [0.00] <i>[0.00]</i>	−0.760*** (0.052) [0.00] <i>[0.00]</i>	0.215*** (0.047) [0.00] <i>[0.00]</i>
Secondary	−0.025*** (0.008) [0.00] <i>[0.00]</i>	0.009 (0.012) [0.46] <i>[0.46]</i>	0.038 (0.028) [0.19] <i>[0.23]</i>	−1.101*** (0.054) [0.00] <i>[0.00]</i>	−0.160*** (0.049) [0.00] <i>[0.00]</i>
Girl*Primary	−0.003 (0.008) [0.70] <i>[0.70]</i>	0.043** (0.016) [0.01] <i>[0.03]</i>	0.054 (0.034) [0.12] <i>[0.20]</i>	−0.141*** (0.047) [0.00] <i>[0.02]</i>	0.030 (0.066) [0.65] <i>[0.70]</i>
Girl*Secondary	−0.005 (0.009) [0.60] <i>[0.60]</i>	0.040* (0.024) [0.10] <i>[0.24]</i>	0.065 (0.046) [0.16] <i>[0.27]</i>	−0.044 (0.041) [0.28] <i>[0.35]</i>	−0.176** (0.071) [0.02] <i>[0.08]</i>
R ²	0.068	0.145	0.108	0.109	0.096
Num. obs.	28236	28236	28236	28116	28105
N Clusters	50	50	50	50	50

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. Each regression includes a constant term, province, and contract year-month fixed effects. Control variables: mother dummy, health condition, number of contracts, type of the first contract. For the regressions in Panel B, the baseline category is preschool (preschool age: 4–5, primary school age: 6–11, secondary school age: 12–16.)

Table 11 Gender gap—differences by income categories

	Persistency	Completion	Sessions	Accuracy	Delayed c.
Girl	−0.024 (0.032) [0.45] <i>[0.77]</i>	−0.022 (0.076) [0.77] <i>[0.77]</i>	−0.047 (0.145) [0.75] <i>[0.77]</i>	−0.135 (0.100) [0.18] <i>[0.77]</i>	0.057 (0.130) [0.66] <i>[0.77]</i>
25th–50th	−0.016 (0.018) [0.36] <i>[0.57]</i>	−0.063 (0.042) [0.14] <i>[0.47]</i>	−0.123 (0.092) [0.19] <i>[0.47]</i>	0.045 (0.110) [0.68] <i>[0.68]</i>	−0.093 (0.124) [0.46] <i>[0.57]</i>
50th–75th	0.001 (0.018) [0.96] <i>[0.96]</i>	−0.058 (0.040) [0.15] <i>[0.39]</i>	−0.150** (0.069) [0.04] <i>[0.18]</i>	−0.062 (0.097) [0.53] <i>[0.66]</i>	−0.115 (0.107) [0.29] <i>[0.48]</i>
>75th	−0.006 (0.017) [0.74] <i>[0.92]</i>	−0.057 (0.043) [0.19] <i>[0.32]</i>	−0.165** (0.080) [0.04] <i>[0.22]</i>	0.006 (0.094) [0.95] <i>[0.95]</i>	−0.155 (0.112) [0.17] <i>[0.32]</i>
Girl*25th–50th	0.030 (0.036) [0.41] <i>[0.82]</i>	0.016 (0.073) [0.82] <i>[0.82]</i>	0.050 (0.147) [0.73] <i>[0.82]</i>	−0.064 (0.121) [0.60] <i>[0.82]</i>	0.127 (0.171) [0.46] <i>[0.82]</i>
Girl*50th–75th	0.023 (0.036) [0.53] <i>[0.89]</i>	0.011 (0.074) [0.89] <i>[0.89]</i>	0.022 (0.141) [0.88] <i>[0.89]</i>	0.065 (0.108) [0.55] <i>[0.89]</i>	0.059 (0.131) [0.65] <i>[0.89]</i>
Girl*>75th	0.027 (0.032) [0.40] <i>[0.84]</i>	0.023 (0.076) [0.76] <i>[0.84]</i>	0.054 (0.145) [0.71] <i>[0.84]</i>	0.020 (0.101) [0.84] <i>[0.84]</i>	0.076 (0.131) [0.56] <i>[0.84]</i>
Province FE	Yes	Yes	Yes	Yes	Yes
Cont. Month FE	Yes	Yes	Yes	Yes	Yes
R ²	0.067	0.144	0.107	0.047	0.070
Num. obs.	27985	27985	27985	27865	27854
N Clusters	49	49	49	49	49

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. Each regression includes a constant term, province, and contract year-month fixed effects. Control variables: mother dummy, health condition, number of contracts, type of the first contract. Baseline category is “< 25th percentile”

Table 12 Gender gap—differences by income categories within Smartick

	Persistency	Completion	Sessions	Accuracy	Delayed c.
Girl	−0.006 (0.012) [0.64] <i>[0.86]</i>	−0.005 (0.030) [0.86] <i>[0.86]</i>	−0.014 (0.060) [0.81] <i>[0.86]</i>	−0.195*** (0.062) [0.00] <i>[0.02]</i>	0.143* (0.082) [0.09] <i>[0.22]</i>
25th–50th	0.015 (0.012) [0.24] <i>[0.34]</i>	−0.023 (0.022) [0.29] <i>[0.34]</i>	−0.081* (0.048) [0.10] <i>[0.24]</i>	−0.128** (0.049) [0.01] <i>[0.06]</i>	−0.059 (0.060) [0.34] <i>[0.34]</i>
50th–75th	0.016 (0.010) [0.12] <i>[0.26]</i>	−0.002 (0.026) [0.94] <i>[0.94]</i>	−0.037 (0.049) [0.46] <i>[0.57]</i>	−0.078** (0.034) [0.03] <i>[0.13]</i>	−0.101 (0.070) [0.16] <i>[0.26]</i>
>75th	0.009 (0.012) [0.47] <i>[0.47]</i>	−0.050* (0.029) [0.09] <i>[0.13]</i>	−0.151*** (0.051) [0.00] <i>[0.02]</i>	−0.067* (0.038) [0.09] <i>[0.13]</i>	−0.126 (0.076) [0.11] <i>[0.13]</i>
Girl*25th–50th	0.012 (0.016) [0.48] <i>[0.84]</i>	−0.016 (0.037) [0.67] <i>[0.84]</i>	−0.004 (0.073) [0.96] <i>[0.96]</i>	0.133 (0.083) [0.11] <i>[0.57]</i>	−0.037 (0.078) [0.63] <i>[0.84]</i>
Girl*50th–75th	−0.001 (0.013) [0.91] <i>[0.99]</i>	−0.005 (0.032) [0.89] <i>[0.99]</i>	−0.009 (0.064) [0.88] <i>[0.99]</i>	0.090 (0.058) [0.13] <i>[0.63]</i>	0.002 (0.084) [0.99] <i>[0.99]</i>
Girl*>75th	0.012 (0.012) [0.33] <i>[0.82]</i>	0.012 (0.030) [0.70] <i>[0.86]</i>	0.032 (0.060) [0.59] <i>[0.86]</i>	0.077 (0.063) [0.23] <i>[0.82]</i>	−0.015 (0.084) [0.86] <i>[0.86]</i>
Province FE	Yes	Yes	Yes	Yes	Yes
Cont. Month FE	Yes	Yes	Yes	Yes	Yes
R ²	0.067	0.145	0.108	0.047	0.070
Num. obs.	27985	27985	27985	27865	27854
N Clusters	49	49	49	49	49

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. Each regression includes a constant term, province, and contract year-month fixed effects. Control variables: mother dummy, health condition, number of contracts, type of the first contract. Baseline category is “< 25th percentile”

Table 13 Gender gaps in the mothers' and fathers' covid sample

	Persistency	Completion	Sessions	Accuracy	Delayed c.
<i>Mothers' sample</i>					
Girl	0.010*** (0.004) [0.01] <i>[0.01]</i>	0.007 (0.012) [0.56] <i>[0.56]</i>	0.027 (0.025) [0.30] <i>[0.37]</i>	−0.066*** (0.015) [0.00] <i>[0.00]</i>	0.139*** (0.013) [0.00] <i>[0.00]</i>
R ²	0.078	0.183	0.138	0.098	0.098
Num. obs.	9116	9116	9116	9088	9088
N clusters	50	50	50	50	50
<i>Fathers' sample</i>					
Girl	−0.005 (0.004) [0.20] <i>[0.33]</i>	−0.006 (0.012) [0.60] <i>[0.60]</i>	−0.018 (0.023) [0.43] <i>[0.54]</i>	−0.103*** (0.029) [0.00] <i>[0.00]</i>	0.096*** (0.021) [0.00] <i>[0.00]</i>
R ²	0.096	0.197	0.136	0.099	0.100
Num. obs.	4397	4397	4397	4381	4381
N Clusters	50	50	50	50	50

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. Each regression includes a constant term, province, and contract year-month fixed effects. Control variables: age, health condition, number of contracts, type of the first contract

Table 14 Gender gaps in the mothers' and fathers' covid sample (siblings)

	Persistency	Completion	Sessions	Accuracy	Delayed c.
<i>Mothers' sample</i>					
Girl	0.004 (0.011) [0.74] <i>[0.74]</i>	0.004 (0.011) [0.71] <i>[0.74]</i>	−0.027* (0.015) [0.07] <i>[0.18]</i>	−0.053 (0.041) [0.20] <i>[0.34]</i>	0.160*** (0.047) [0.00] <i>[0.01]</i>
R ²	0.720	0.846	0.889	0.713	0.673
Num. obs.	2590	2590	2590	2581	2581
N Clusters	49	49	49	49	49
<i>Fathers' sample</i>					
Girl	−0.009 (0.008) [0.23] <i>[0.25]</i>	−0.017 (0.015) [0.25] <i>[0.25]</i>	−0.071* (0.035) [0.05] <i>[0.25]</i>	−0.058 (0.043) [0.18] <i>[0.25]</i>	0.096 (0.060) [0.11] <i>[0.25]</i>
R ²	0.799	0.841	0.887	0.715	0.668
Num. obs.	1215	1215	1215	1213	1213
N clusters	48	48	48	48	48

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. Each regression includes a constant term and parent fixed effects. Control variables: age, health condition

Table 15 Gender gaps in the full sample—registrations before and after Covid-19 outbreak

	Persistency	Completion	Sessions	Accuracy	Delayed c.
<i>Registrations before Covid-19 outbreak</i>					
Girl	0.011* (0.006) [0.07] <i>[0.11]</i>	0.003 (0.010) [0.77] <i>[0.77]</i>	0.026 (0.021) [0.21] <i>[0.27]</i>	−0.061** (0.028) [0.04] <i>[0.09]</i>	0.161*** (0.025) [0.00] <i>[0.00]</i>
Father	0.018*** (0.005) [0.00] <i>[0.00]</i>	0.005 (0.018) [0.79] <i>[0.79]</i>	0.026 (0.030) [0.40] <i>[0.49]</i>	0.050** (0.024) [0.04] <i>[0.11]</i>	−0.035 (0.031) [0.27] <i>[0.45]</i>
GirlxFather	−0.016** (0.006) [0.01] <i>[0.07]</i>	−0.002 (0.014) [0.90] <i>[0.90]</i>	−0.029 (0.035) [0.42] <i>[0.90]</i>	0.015 (0.030) [0.62] <i>[0.90]</i>	0.011 (0.035) [0.75] <i>[0.90]</i>
Province FE	Yes	Yes	Yes	Yes	Yes
Cont. Month FE	Yes	Yes	Yes	Yes	Yes
R ²	0.103	0.136	0.097	0.116	0.115
Num. obs.	8520	8520	8520	8477	8476
N Clusters	50	50	50	50	50
<i>Registrations after Covid-19 outbreak</i>					
Girl	0.005 (0.004) [0.15] <i>[0.25]</i>	0.006 (0.010) [0.56] <i>[0.56]</i>	0.016 (0.027) [0.55] <i>[0.56]</i>	−0.082*** (0.013) [0.00] <i>[0.00]</i>	0.153*** (0.011) [0.00] <i>[0.00]</i>
Father	0.008 (0.005) [0.13] <i>[0.32]</i>	−0.006 (0.011) [0.62] <i>[0.66]</i>	−0.012 (0.029) [0.66] <i>[0.66]</i>	0.055* (0.028) [0.05] <i>[0.26]</i>	0.020 (0.017) [0.26] <i>[0.44]</i>
GirlxFather	−0.009 (0.006) [0.14] <i>[0.24]</i>	−0.024 (0.019) [0.19] <i>[0.24]</i>	−0.057 (0.043) [0.19] <i>[0.24]</i>	−0.022 (0.028) [0.42] <i>[0.42]</i>	−0.037* (0.021) [0.09] <i>[0.24]</i>
Province FE	Yes	Yes	Yes	Yes	Yes
Cont. Month FE	Yes	Yes	Yes	Yes	Yes
R ²	0.057	0.152	0.118	0.130	0.090
Num. obs.	19716	19716	19716	19639	19629
N Clusters	50	50	50	50	50

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. Each regression includes a constant term, province, and contract year-month fixed effects. Control variables: age, health condition, number of contracts, type of the first contract

Table 16 Gender gaps in the mothers' and fathers' (siblings') sample—province FE

	Persistencey	Completion	Sessions	Accuracy	Delayed c.
<i>Mothers' sample</i>					
Girl	0.005 (0.004) [0.24] <i>[0.31]</i>	−0.003 (0.005) [0.54] <i>[0.54]</i>	−0.028*** (0.009) [0.00] <i>[0.01]</i>	−0.047*** (0.016) [0.01] <i>[0.01]</i>	0.205*** (0.020) [0.00] <i>[0.00]</i>
R ²	0.097	0.204	0.143	0.104	0.121
Num. obs.	4889	4889	4889	4868	4867
N Clusters	49	49	49	49	49
<i>Fathers' sample</i>					
Girl	−0.005 (0.004) [0.21] <i>[0.26]</i>	−0.027* (0.015) [0.07] <i>[0.12]</i>	−0.066** (0.027) [0.02] <i>[0.05]</i>	−0.015 (0.019) [0.43] <i>[0.43]</i>	0.189*** (0.029) [0.00] <i>[0.00]</i>
R ²	0.108	0.196	0.149	0.128	0.126
Num. obs.	2644	2644	2644	2637	2635
N Clusters	50	50	50	50	50

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. Each regression includes a constant term, province, and contract year-month fixed effects. Control variables: age, health condition, number of contracts, type of the first contract

Table 17 Logit regressions

	Mothers Persistencey	Completion	Fathers Persistencey	Completion
<i>A. Full sample</i>				
Girl	0.136 (0.086)	0.022 (0.034)	−0.117* (0.069)	−0.062* (0.037)
AIC	5723.798	22488.430	2975.372	12078.474
BIC	6389.041	23153.673	3585.837	12688.939
Log Likelihood	−2776.899	−11159.215	−1402.686	−5954.237
Deviance	5553.798	22318.430	2805.372	11908.474
Num. obs.	18516	18516	9720	9720
<i>B. Siblings' sample</i>				
Girl	0.423 (0.277)	−0.079 (0.124)	−0.852 (0.574)	−0.681** (0.320)

Table 17 continued

	Mothers Persistency	Completion	Fathers Persistency	Completion
AIC	4649.769	5455.443	2476.539	3073.614
BIC	18542.024	19347.699	9362.076	9959.150
Log Likelihood	−185.884	−588.722	−67.270	−365.807
Deviance	371.769	1177.443	134.539	731.614

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Control variables for the regressions in Panel A: age, health condition, number of contracts, type of the first contract, province, and contract year-month fixed effects. Control variables for the regressions in Panel B: age, health condition, and parent fixed effects

Table 18 Gender gaps in the full sample of children without a health condition

	Persistency	Completion	Sessions	Accuracy	Delayed c.
Girl	0.008* (0.004) [0.06] <i>[0.10]</i>	0.004 (0.008) [0.60] <i>[0.60]</i>	0.019 (0.025) [0.46] <i>[0.57]</i>	−0.071*** (0.015) [0.00] <i>[0.00]</i>	0.159*** (0.012) [0.00] <i>[0.00]</i>
Father	0.011*** (0.004) [0.01] <i>[0.04]</i>	−0.004 (0.008) [0.64] <i>[0.88]</i>	0.001 (0.024) [0.98] <i>[0.98]</i>	0.041* (0.023) [0.08] <i>[0.19]</i>	0.006 (0.015) [0.70] <i>[0.88]</i>
GirlxFather	−0.011** (0.005) [0.04] <i>[0.22]</i>	−0.016 (0.013) [0.20] <i>[0.26]</i>	−0.046 (0.033) [0.17] <i>[0.26]</i>	0.009 (0.025) [0.73] <i>[0.73]</i>	−0.025 (0.019) [0.19] <i>[0.26]</i>
R ²	0.067	0.144	0.107	0.117	0.091
Num. obs.	26476	26476	26476	26373	26363
N Clusters	50	50	50	50	50

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are (in parenthesis) clustered at the province level. Original p -values and Benjamini and Hochberg adjusted p -values (*italic*) are given in brackets. Each regression includes a constant term, province, and contract year-month fixed effects. Control variables: age, number of contracts, type of the first contract

Table 19 Full sample—registration probability by gender norms municipalities

	Mother	Mother
Chores	−0.005 (0.007)	
EmpGendRatio		0.005 (0.013)
Girl	0.044*** (0.011)	0.044*** (0.011)
Age	−0.002 (0.001)	−0.002 (0.001)
Health cond.	0.073*** (0.011)	0.074*** (0.011)
Num of contr	−0.004*** (0.001)	−0.004*** (0.001)
1st C:12	−0.086*** (0.011)	−0.086*** (0.011)
1st C:3	−0.010 (0.009)	−0.010 (0.009)
Constant	1.519*** (0.032)	1.514*** (0.032)
Province FE	Yes	Yes
Cont. Month FE	Yes	Yes
R ²	0.013	0.013
Num. obs.	24896	24896
N Clusters	50	50

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Standard errors are clustered at the province level. The coefficient estimates are from the regressions of the dummy variable which takes a value of 1 if the mother registers the child, and 0 otherwise on the gender norms measurement variable (chores and EmpGendRatio in two columns) and the control variables

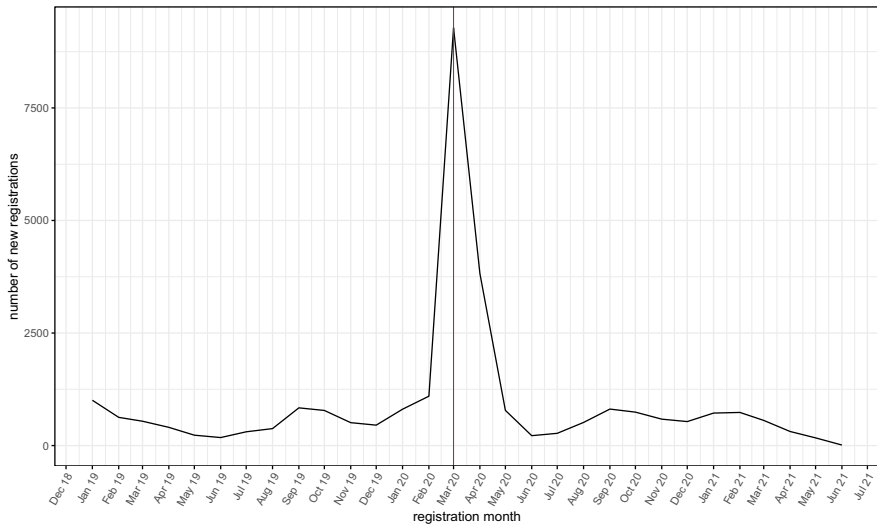


Fig. 3 Total number of new registrations—monthly. This figure shows the evolution of the total number of new registrations to the Smartick platform per month, by people residing in Spain

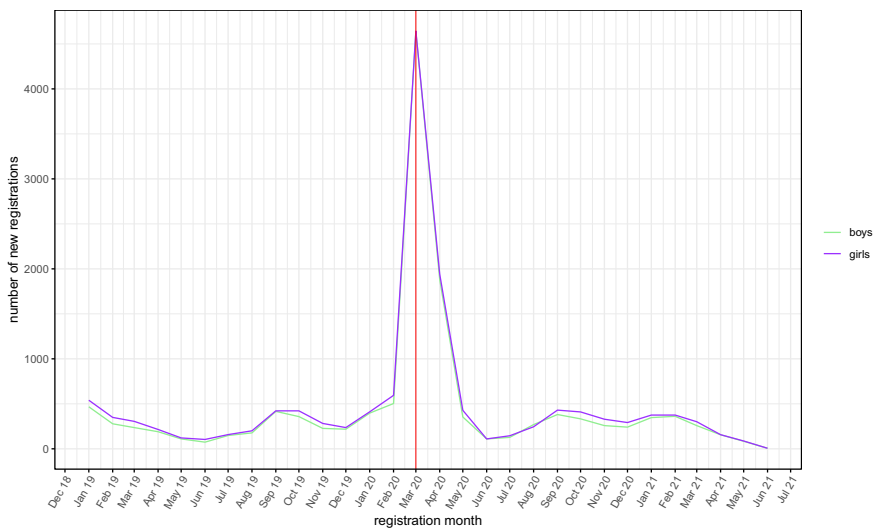


Fig. 4 Total number of new registrations by child gender—monthly. This figure shows, by child's gender, the evolution of the total number of new registrations to the Smartick platform per month, by people residing in Spain

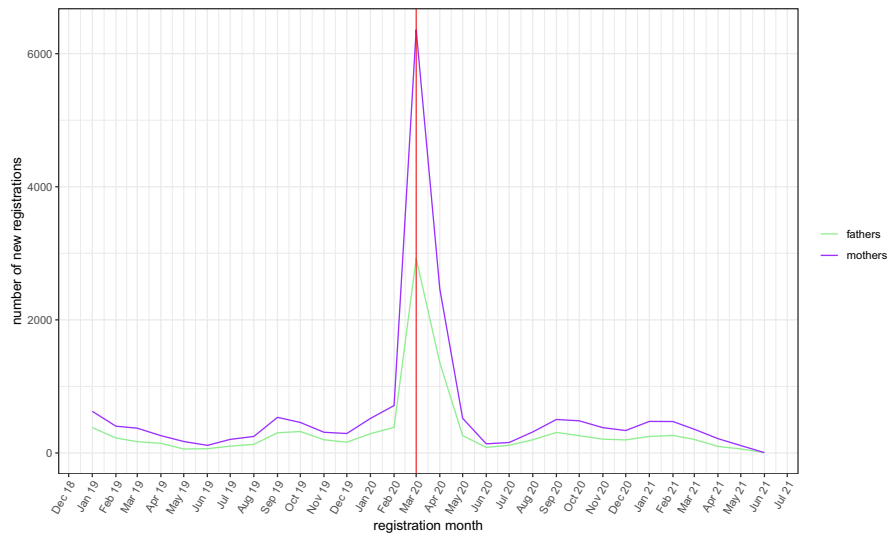


Fig. 5 Total number of new registrations by parent gender—monthly. This figure shows, by parent’s gender, the evolution of the total number of new registrations to the Smartick platform per month, by people residing in Spain

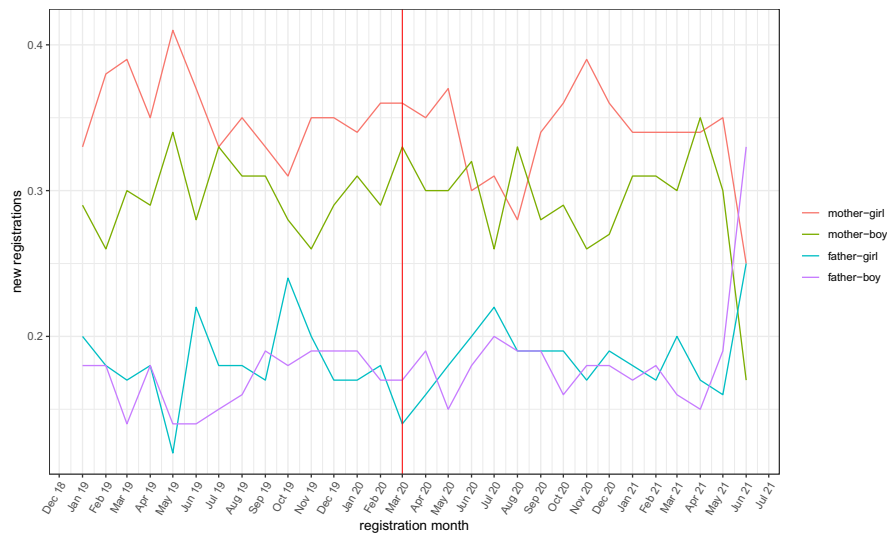


Fig. 6 Percentage of new registrations by gender pairs—monthly. This figure shows the evolution of the percentages of new registrations by gender pairs of children and parents to the Smartick platform per month, by people residing in Spain

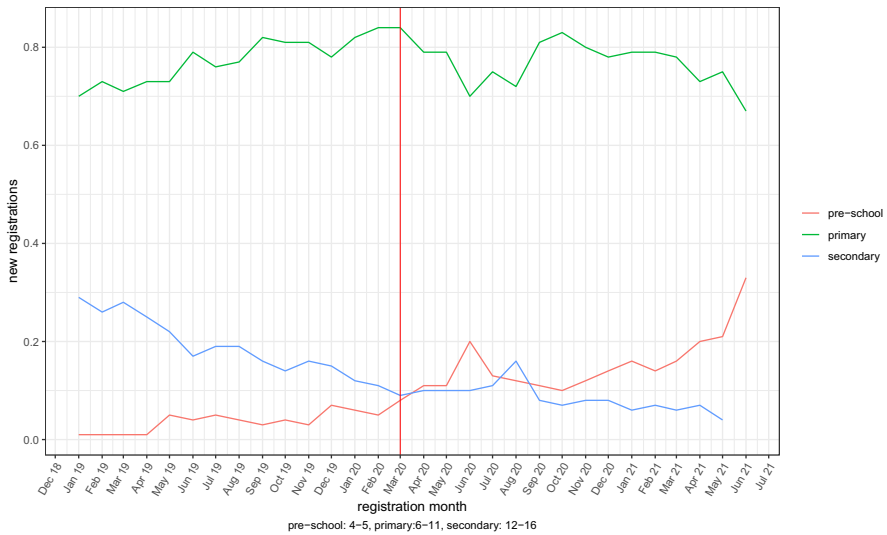


Fig. 7 Percentage of new registrations by age categories—monthly. This figure shows the evolution of the percentages of new registrations by age categories of children to the Smartick platform per month, by people residing in Spain

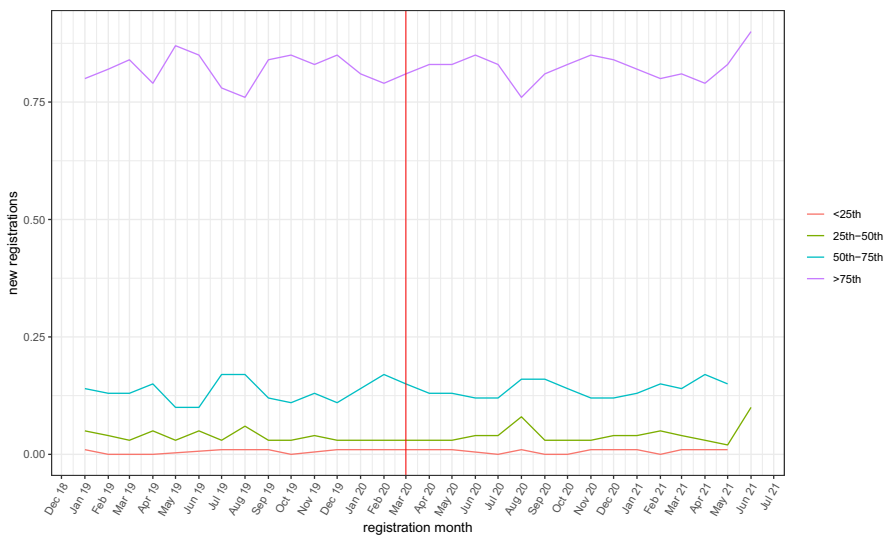


Fig. 8 Percentage of new registrations by income categories—monthly. This figure shows the evolution of the percentages of new registrations by income categories to the Smartick platform per month, by people residing in Spain

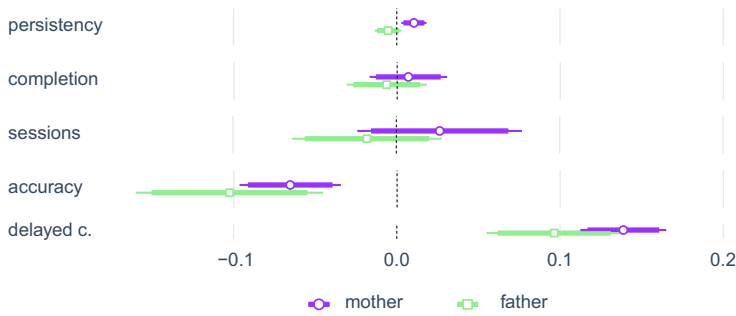


Fig. 9 Gender gap—covid sample. The Figure shows the estimated gender gaps for different outcomes in mothers and fathers full covid samples separately, based on Equation 1. Covid sample includes the members registered between 9 March 2020 and 31 May 2020, which corresponds to the first wave of the Covid-19 outbreak in Spain. Control variables are age of the child, presence of a health condition, total number of contracts, type of the first contract, province and contract year-month fixed effects. Standard errors are clustered at the province level. Thin and thick lines represent the 95% and 90% confidence intervals, respectively

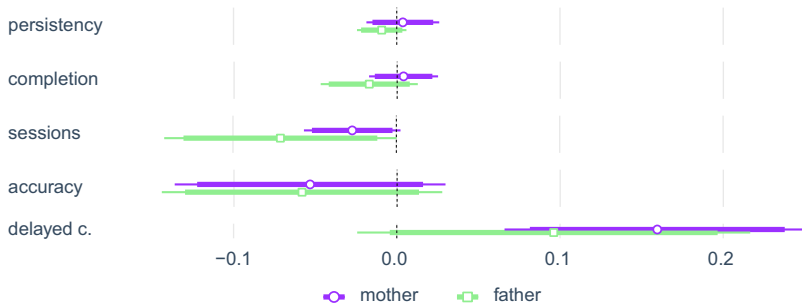


Fig. 10 Gender gap—covid sample (siblings). The Figure shows the estimated gender gaps for different outcomes in mothers and fathers siblings covid samples separately, based on Equation 2. Covid sample includes the members registered between 9 March 2020 and 31 May 2020, which corresponds to the first wave of the Covid-19 outbreak in Spain. Control variables are age of the child, presence of a health condition and parent fixed effects. Standard errors are clustered at the province level. Thin and thick lines represent the 95% and 90% confidence intervals, respectively

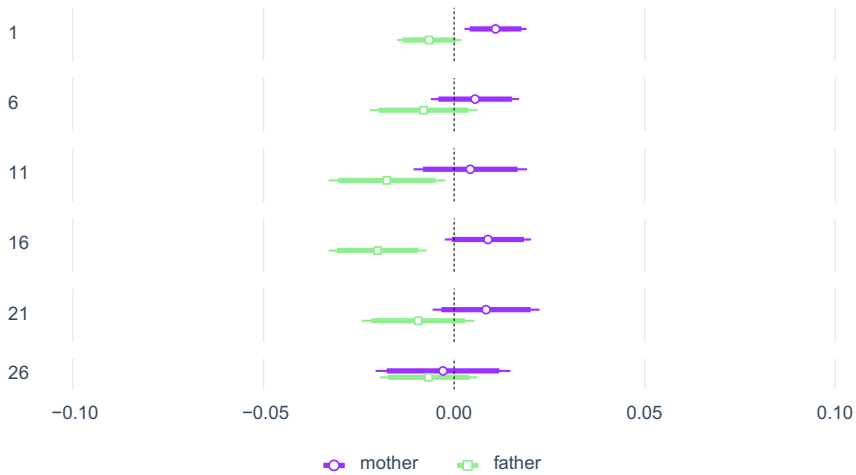


Fig. 11 Different number of session cut-offs for completion variable. The Figure shows the estimated gender gaps in completion outcome for different numbers of session cut-offs in mothers' and fathers' full samples separately, based on Equation 1. Control variables are the age of the child, presence of a health condition, total number of contracts, type of the first contract, province, and contract year-month fixed effects. Standard errors are clustered at the province level. Thin and thick lines represent the 95% and 90% confidence intervals, respectively

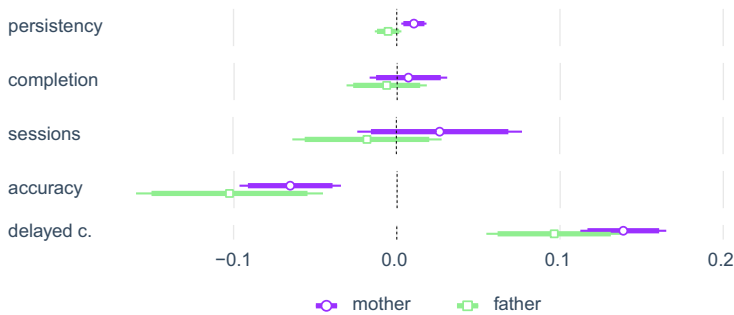


Fig. 12 Age fixed effects in the main specification—full sample. The Figure shows the estimated gender gaps for different outcomes in mothers' and fathers' full samples separately, based on Equation 1, where the regression includes age fixed effects instead of the continuous age variable. Control variables are the age fixed effects, presence of a health condition, total number of contracts, type of the first contract, province, and contract year-month fixed effects. Standard errors are clustered at the province level. Thin and thick lines represent the 95 and 90% confidence intervals, respectively

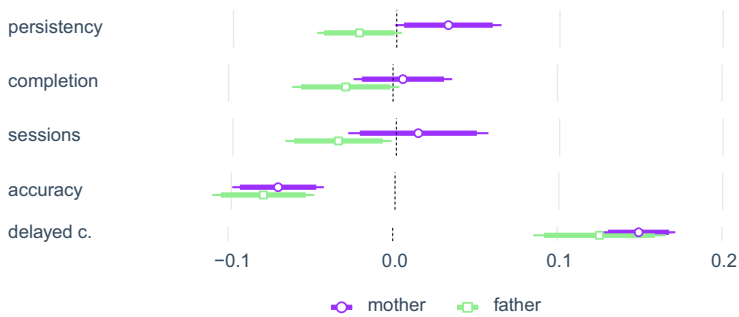


Fig. 13 Normalization within age—full sample. The Figure shows the estimated gender gaps for different outcomes in mothers’ and fathers’ full samples separately, based on Equation 1, where the outcome variables are normalized within each age. Control variables are the age of the child, presence of a health condition, total number of contracts, type of the first contract, province, and contract year-month fixed effects. Standard errors are clustered at the province level. Thin and thick lines represent the 95% and 90% confidence intervals, respectively

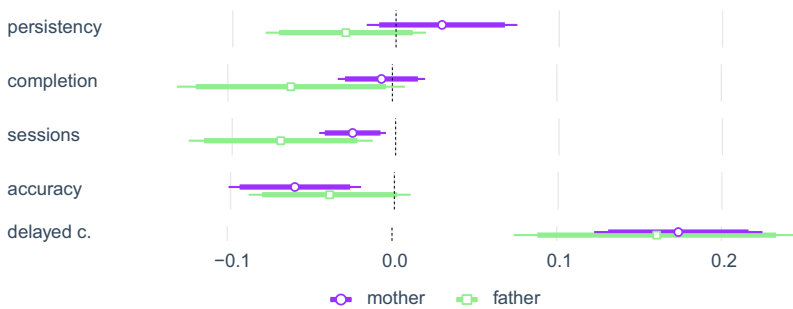


Fig. 14 Normalization within age—siblings’ sample. The Figure shows the estimated gender gaps for different outcomes in mothers’ and fathers’ siblings’ samples separately, based on Equation 2, where the outcome variables are normalized within each age. Control variables are the age of the child, the presence of a health condition, and parent fixed effects. Standard errors are clustered at the province level. Thin and thick lines represent the 95% and 90% confidence intervals, respectively

References

- Anaya L, Iriberrí N, Rey-Biel P, Zamarro G (2022) Understanding performance in test taking: the role of question difficulty order. *Econ Educ Rev* 90:102293
- Andrew A, Cattán S, Costa Dias M, Farquharson C, Kraftman L, Krutikova S, Sevilla A (2020) Inequalities in children’s experiences of home learning during the COVID-19 lockdown in England. *Fisc Stud* 41(3):653–683
- Baker M, Milligan K (2016) Boy-girl differences in parental time investments: evidence from three countries. *J Hum Cap* 10(4):399–441
- Balart P, Oosterveen M, Webbink D (2018) Test scores, noncognitive skills and economic growth. *Econ Educ Rev* 63:134–153
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc Ser B (Methodol)* 57(1):289–300
- Chetty R, Friedman JN, Stepner M (2024) The economic impacts of COVID-19: evidence from a new public database built using private sector data. *Q J Econ* 139(2):829–889
- De San Román AG, De La Rica S (2016) Gender gaps in PISA test scores: the impact of social norms and the mother’s transmission of role attitudes. *Estudios de Economía Aplicada* 34(1):79–108

- Del Boca D, Oggero N, Profeta P, Rossi M (2020) Women's and men's work, housework and childcare, before and during COVID-19. *Rev Econ Household* 18(4):1001–1017
- DeMars CE, Bashkov BM, Socha AB (2013) The role of gender in test-taking motivation under low-stakes conditions. *Res Pract Assess* 8:69–82
- Duckworth AL, Quinn PD, Lynam DR, Loeber R, Stouthamer-Loeber M (2011) Role of test motivation in intelligence testing. *Proc Natl Acad Sci* 108(19):7716–7720
- Gevrek ZE, Gevrek D, Neumeier C (2020) Explaining the gender gaps in mathematics achievement and attitudes: the role of societal gender equality. *Econ Educ Rev* 76:101978
- Golin M (2021) Gender gaps in home schooling time. mimeo
- Ikeda M, Yamaguchi S (2021) Online learning during school closure due to COVID-19. *Jpn Econ Rev* 72(3):471–507
- Lundberg S, Pablonia SW, Ward-Batts J (2007) Time allocation of parents and investments in sons and daughters. Unpublished paper
- Mammen K (2011) Fathers' time investments in children: do sons get more? *J Popul Econ* 24(3):839–871
- Montolio D, Taberner PA (2021) Gender differences under test pressure and their impact on academic performance: a quasi-experimental design. *J Econ Behav Org* 191:1065–1090
- Nollenberger N, Rodríguez-Planas N, Sevilla A (2016) The math gender gap: the role of culture. *Am Econ Rev* 106(5):257–61
- OECD (2019) PISA 2018 results (Volume I): what students know and can do. PISA. OECD Publishing, Paris
- Price J (2008) Parent-child quality time does birth order matter? *J Hum Resour* 43(1):240–265
- Rodríguez-Planas N, Nollenberger N (2018) Let the girls learn! It is not only about math... it's about gender social norms. *Econ Educ Rev* 62:230–253
- Segal C (2012) Working when no one is watching: motivation, test scores, and economic success. *Manage Sci* 58(8):1438–1457

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.