Should the children turn off the video game?

Measuring the impact of playing video games on educational achievement thanks to three international assessment studies

Alice Dauriach*

August 30, 2013

Master thesis defended in the academic year 2012/2013
in fulfilment of the requirements for the Master's degree "Economics & Public Policy"

Institut d'Études Politiques de Paris
École Polytechnique
École Nationale de la Statistique et de l'Administration Économique

Supervisor: Yann Algan, professor of economics

Institution: Institut d'Études Politiques de Paris
27 rue Saint-Guillaume
75 007 Paris, France

*Student at the Institut d'Études Politiques de Paris. Contact: alice.dauriach@sciencespo.fr. All views expressed and all errors remain my responsibility.

Should the children turn off the video game?

Measuring the impact of playing video games on educational achievement thanks to three international assessment studies

Alice DAURIACH

Abstract

This Master's thesis assesses the extent to which playing video games affects pupils' educational achievement in the countries that participated in the international studies PISA (OECD) in 2009, TIMSS (IEA) in 2007 and PIRLS (IEA) in 2006. The main challenge is to identify a causal relationship in a situation where pupils themselves choose how much they play video games (endogenous treatment). Building on a classical model of education production function, the behaviour of pupils is modelised at the micro-level in order to predict a fitted value of their amount of video game play. The impact of playing on the scores obtained in reading, mathematics and science is then estimated with ordinary least squares. We find an inverted-U relationship between game playing and scores, with low amounts and reasonable frequency of playing actually having a good impact on academic performance (the order of magnitude ranging from a bonus of 20 to 50 score points, compared to not playing at all, or even more depending on the specification of the model). However, playing several hours each day (typically more than 3 to 5 hours) can diminish the scores by up to 80 points. Evidence from PIRLS for reading, and PISA for math, shows that the difference in the time spent by boys and girls playing video games might explain the average gender gap observed in academic scores.

Acknowledgements. I thank Yann Algan for supervising this Master's thesis and providing me guidance and ideas. I thank Alexis for supporting during all my research.

Contents

1 Introduction

Video games are a prominent issue because they are greatly and increasingly popular, especially among children. Such a widespread and sometimes intense activity araises serious concerns. Might video games have harmful effects on the development on children? Should public authorities advice to reduce pupils' time in front of computer games? Should the parents ask their children to turn off the video game? Indeed, some studies have highlighted that, like watching television, playing video games may damage the health, possibly encouraging obesity (see for example Vandewater et al., 2004) and attention deficit (see Tahiroglu et al., 2010). In the present thesis, we aim at measuring the impact of playing video games on children's achievement at school, proxied by their scores. Is playing everyday likely to increase or decrease one's scores, everything else being equal? And by how much? Is the effect of playing video games linear, or should we expect that the first hours of play can be beneficial, while too much gaming can be detrimental? Does the effect vary with the age of the pupils considered (children or adolescents)? And with the gender? These are the few research questions we would like to answer.

The goal of this paper is to assess whether playing computer games affects a pupil's educational achievement, but it is *not* to review all kinds of video games, all types of use, and all educational outcomes. We choose to focus on video games that are played at home, for recreational purposes, mostly on a computer. The video games we consider are not specifically educational or "serious" video games. They are all types of commercial games. We disregard the use of computer games in the classroom and at school in general. Of course, it is interesting to know if games can be part of the curriculum, and if they are a good way to teach some skills and knowledge, as many other studies already did (see for instance Adams, 2009 about using role-playing games in the reading class). Instead, we focus on the following question: can the recreational use of video games that are not necessarily educational still have a positive impact on children's learning? And we focus on one dimension only of children's learning: the cognitive achievement, approximated by the scores obtained by the pupils in several major disciplines like reading and mathematics. Other dimensions like non-cognitive achievements and social behaviour are left to future research.

Our approach first and foremost relates to the literature in economics trying to find the determinants of educational achievements. The study about educational attainment by Barro and Lee [2001] is still a landmark today. We use an adaptation of their model as our theoretical framework. Many authors who took an economics perspective about education studied the international differences across education systems, and many of them used international assessments, as we plan to do. An excellent review of this field is proposed by Hanushek and Woessmann [2010]. Our empirical method also draws from the field of economics and uses the tools of econometrics. We choose and analyse our data in a fashion similar to Algan et al. [2011]. Our unusual identification strategy is inspired by a paper by Spiezia [2010], who seeks to measure the impact of using a computer on test scores obtained by the pupils in the PISA study ("Programme for International Student Assessment"). He accounts for the possible endogeneity of computer use by including in his model the predicted value of computer use, in a way that resembles IV (instrumental variable method) but differs from it on a crucial point: he faces no exclusion restriction.

Another important base for our reflection is of course the literature about video games. Yet, the research questions specifically related to video games have so far been treated mainly within the fields of developmental psychology and neuroscience. To this day, most of the studies consist in small-scale "clinical-like" experiments and the research effort is particularly strong in the United States of America. Gentile [2011], in a special issue of *Child development perspectives* devoted to the ramifications of video game play for academic learning and cognitive skill acquisition, provides a synthetical account of the dimensions along which video games can exert an effect on children. He identifies five dimensions in the existing scientific literature: "the *amount* of play, the *content* of play, the game *context*, the *structure* of the game, and the *mechanics* of game play". He stresses the fact that many studies focus on the amount of play, but that the effects they find might actually be artifacts of the relation between amount and other dimensions. Playing more exacerbates the other effects. He produces an analysis that is very relevant for the present thesis in a 2007 co-authored work about violent games [Anderson et al., 2007]. They basically find that: 1) when isolating the amount of play and the violent content, amount directly predicts poorer school performance (but not aggressive behaviour), while it is the contrary for violent content; 2) this relationship may be due to the children themselves: the ones who perform poorly at school are likely to play more, and they show this by studying the amount of video game play early in a school year: they find that it negatively predicts school performance later in the school year. As a matter of fact, the literature about gaming gives plenty of evidence about various negative effects on the children, which may affect educational achievement (see Dorman, 1997 for a review): cardiovascular implications, seizures, nintendinitis, pathological preoccupation and aggression...

The literature focusing on the potential positive effects is somewhat less abundant but has been fastly growing in the recent years. To our knowledge, the impact of recreational video games on academic performance has not yet been assessed. Boot et al. [2011] review the studies about the impacts of gaming on perception and cognition. Importantly, they remind that training on one task rarely improves performance on others, and point out the pitfalls of clinical trials design. Though, many studies emphasize the improvement of visual attention [e.g. Green and Bavelier, 2007], spatial cognition [Feng et al., 2007] and speed of processing [Dye et al., 2009] that can arise from gaming. The publication by Feng et al. is of particular interest because they find evidence for a gender gap (in the favour of boys) in spatial attention (e.g. mental rotation ability). And they manage to overcome this gap by making the participants train 10 hours on an action video game. A control group plays a non-action video game and shows no improvement in spatial cognition ability. Their findings have practical implications because, as they argue, having spatial skills is essential in mathematics and engineering.

This leads us to the last field of literature of interest for our study: gender. It is acknowledged that boys and girls have different attitudes toward video games. Girls play on average less than boys. They also tend to play different kinds of games, the boys leaning more to the side of games imitating real physical competition (action, sports and racing games), and the girls to traditional video games (classic board games, puzzles) according to Greenberg et al. [2010]. This is confirmed by Phan et al. [2012] who, in their study of an online survey targeted at adults, find that women prefer social games, puzzles, educational and simulation games, whereas men are more drawn to strategy, role-playing, action and fighting games. Such empirical differences are worth comparing with another empirical stylised fact: the gender gap in academic performance, as highlighted by Fortin et al. [2012].

In the present thesis, the simple theoretical framework allows to formulate two hypotheses. The first one asserts that there is an inverted-U relationship between video games and scores, and the second one that the gender gap in scores can be explained by the gender gap in gaming. To test these hypotheses, and measure the impact of video games on educational achievement, we propose to use the most recent versions of three well-known international assessments programmes: PISA (conducted by the OECD) in 2009; TIMSS ("Trends in International Mathematics and Science Study", conducted by the International Association for the Evaluation of Educational Achievement, hereafter IEA) in 2007; and PIRLS ("Progress in International Reading Literacy Study", also by the IEA) in 2006. These three studies all submit standardised tests in various course subjects as well as background questionnaires to the pupils sampled. Thus the studies collect a lot of information about the students' habits, characteristics and attitudes.

Of course, our analysis is bounded by the data available. First, there are several distinctions that we are not able to make. The international questionnaires at hand do ask whether the pupils own a video game console and/or a personal computer at home. Yet, they do not ask specific questions about the time spent playing video games on a console. They usually either target specifically computer games or do not make a difference between the two types of games. Similarly, we usually cannot distinguish among the different types of computer games, except in the PISA study, where one-player computer games and collaborative online computer games are treated separately. But all the subtleties in the content, structure and mechanics of the games, which are achieved by the psychological literature and experiments, are unfortunately out of our reach.

The paper is organised as follows. The theoretical framework for the modelisation is explained in the next section. Section 3 presents the design of the three datasets we use, as long as some descriptive statistics. Given the model and the data at hand, we implement an empirical strategy for identification of the causal relationship of interest in section 4. The results are presented and discussed in section 5. Section 6 concludes.

2 Theoretical framework

2.1 Modelising how educational achievement is determined

We take as base the simple theoretical model by Lee and Barro [2001]. The educational outcome (in their case, schooling quality) A is obtained thanks to a production function which takes family factors F (such as the level of education of the parents) and school resources R (e.g. pupil/teacher ratio) as inputs. The error term ϵ corresponds to unobserved determinants of schooling quality.

$$A = A(F, R) + \epsilon$$

Additional inputs to the education production function are of special interest here, therefore the baseline model is rather:

$$A = A(F, R, G, P) + \epsilon \tag{1}$$

where A stands for educational achievement and can be measured through scores obtained in standardised international tests, G represents the frequency of use of computer games, and P the other pupil factors (e.g. gender, age). In the present essay, the focus is on the input G. The three other sets of variables will be studied and included as controls in the estimations, but their effect will not be analysed *per se*.

At a micro level, we postulate that the educational achievement of a pupil is a function of two arguments. The first one contains the pupils' exogenous characteristics, i.e. the ones that are beyond their grasp. A good example of such characteristic is gender, which is completely exogenous to the pupils but determines the way in which they are socialised, encouraged or prejudiced against. It is also necessary to mention the potential role of innate capacities and tastes, which are difficult to observe. Another important component would be the social environment of the pupil, which we approximate with the sets of parent and school variables mentioned above.

The second argument relates to the pupils' learning effort, which we measure thanks to their involvement in out-of-class-time activities which are beneficial to learning. They choose to allocate their total time to various activities $j = 1, ..., J$. The total amount of out-of-class time available to the pupil is normalised to 1 for simplicity. We note t_j the time spent on activity j. The activities include for example doing homework, reading, playing sports, being with friends, watching television and, of course, playing computer games.

$$\begin{aligned} A &= f(F, R, P) + \sum_{j=1}^{J} h_j(t_j) \\ \sum_{j=1}^{J} t_j &= 1 \end{aligned}$$

The functions $h_j(.)$ are concave, increasing for low values of t_j but decreasing for values of t_j superior to some tipping point p_j where they reach their maximum. The functions h_j depend on the activity j because some activities are more "educational" than others. They denote the fact that all activities can have some benefit for the pupils' educational achievement when a reasonably low amount of time is spent on them. But for higher levels of t_j, the marginal benefit of one activity becomes lower and lower, until the point where an additional minute spent on the activity will actually be harmful for the pupil's score. The tipping point may be very high in the case of homework, for example. Still, it is straightforward to assume that it exists (because the children can get tired and unhappy when spending all their time on homework). Given what we know about the effects of gaming from the existing literature, we suspect that the tipping point is quite low for the activity "playing video games". As Gentile [2011] puts it, "[...] each hour a child spends playing entertainment games is an hour not spent on homework, reading, exploring, creating, or other things that might have more educational benefit".

Theoretical hypothesis 1: *Devoting a small amount of time to playing computer games (or playing not frequently) raises educational achievement, whereas playing a large amount of time (or frequently) diminishes it.*

2.2 Modelising how pupils behave

The main problem in our study is that the pupils themselves choose how much time they allocate to playing computer games. The frequency of play G may be seen as an endogenous treatment variable, following the denomination of Vella [1998]. It is very likely influenced by a whole set of observable and unobservable pupils' characteristics which also determine the scores they obtain to international tests. For example, we can think of innate abilities, preferences, or the relationship with the parents, etc... The unobservable characteristics present a problem if they are correlated with G (omitted variables bias). The observable characteristics are less problematic but still have to be taken into account. They may overlap (possibly entirely) the control variables in equation 1. Hence we must modelise the behaviour of pupils and consider a second equation, which is simultaneous to the first one:

$$G = G(F, R, P)$$

What might be even more important is that playing computer games may itself be influenced by the scores obtained by the pupil. This situation is known as reverse causality. Evidence of negative reverse causality has been produced by Anderson et al. [2007]. We will have to check whether such a problem arises in our samples.

$$G = G(A, F, R, P)$$

2.3 Explaining the gender gap in educational achievement

Our last hypothesis builds on the literature about the sex differences in children's play and considers that the main source of sex differences in children's cognitive skills is the different attitude of boys and girls toward video games. Because we are considering *differences* between boys and girls, it is not possible to elaborate a micro model. We will just test this hypothesis empirically by means of predictions.

Theoretical hypothesis 2: *The gender gap in test scores can be entirely explained by the gender gap in the amount (or frequency) of gaming.*

3 Data

3.1 The databases: PISA, TIMSS and PIRLS

There is a long tradition of international educational assessment that dates back to the early 1960s. The first major international test was conducted in 1964 and brought together twelve countries: it was the "First International Mathematics Study" (FIMS), developed by the International Association for the Evaluation of Educational Achievement (IEA). A number of assessments were conducted in the next decades, and two of them became regular from the mid-1990s onwards: TIMSS and PIRLS.

Table 1: Sample populations, disciplines tested and background questions in each dataset

	PISA	**TIMSS**	**PIRLS**
Year	2009	2007	2006
4th grade		x	x
8th grade	x	x	
Reading test	x		x
Math test	x	x	
Science test	x	x	
Frequency of gaming	x		
Gaming hours per day		x	x

The OECD lauched its own programme, PISA, in 2000. PISA tests more applied knowledge and skills compared to TIMSS and PIRLS, which have rather a curricular focus. But the results of the three tests at the country-level are highly correlated. Hanushek and Woessmann [2010] report a correlation of 0.87 between the math scores obtained in TIMSS 2003 and PISA 2003 by the countries participating in both. The correlation is even 0.97 for science scores. Nonetheless, we will use all three tests in our study, because their background questionnaires and their sample population are complementary, as presented in table 1. We want to use the most recent databases for the obvious purpose of making up-to-date policy recommendations. Alas, neither TIMSS 2011 or PIRLS 2011 ask to the students how much time they spend playing computer or video games. This question has simply been withdrawn from the student background questionnaires. We can only hope that the next versions of the tests will include it again, and we use instead the previous occurrences of the tests.

PISA data allow to estimate the effect of rather low frequencies of playing computer games, comparing students who declare never playing them to students who declare to playing them every month, every week, up to every day (see table 2 for the detail of the questions). Due to these low frequencies, we expect to find rather positive effects of gaming on scores in PISA. We can differentiate this effect on scores in three disciplines: reading skills, mathematics skills and science skills. But among the students who play video games every day, we cannot distinguish who is a hard gamer and who plays little but often. This is where TIMSS and PIRLS become useful, because their question about computer games is phrased differently: it asks how many hours the pupil plays computer games every school day. It is thus more difficult to know exactly the frequency and regularity of playing, but we have a better idea of the mean value of the amount of play every day. We can distinguish those who play really a lot, for instance those who declare playing more than 5 hours a day in PIRLS. Thus we expect to find rather negative effects of gaming in TIMSS and PIRLS. We are also able to differentiate the effect of gaming on reading (PIRLS test), mathematics and science skills (TIMSS tests). Finally, there is another complementarity that we will exploit. TIMSS 2007 is delivered to students in the 4th grade and the 8th grade, allowing to compare the effect of playing computer games on younger children (age 11) and older ones (age 15) for the same disciplines (maths and science) and the same measurement method.

3.1.1 Advantages and disadvantages of international student achievement tests

International achievement tests are of particular interest when studying the determinants of cognitive skills. Their main benefit comes from their design: international tests are generally more useful than national ones because they provide *large* and *representative* samples of data. The external validity of the results is thus easier to defend. In addition, Hanushek and Woessmann [2010] highlight no less than six advantages of international achievement tests over data restricted to single countries. First, international data allow to exploit cross-country variation in institutional features. Even where there is some within-country variation, cross-country variation is likely to be larger (e.g. in the characteristics of the schools and the population). International data make possible the comparison of estimates across countries, to find if a result is country-specific or general. Such data help identify systematic heterogeneity in these estimates, relating them to other factors. Aggregating at the country level may help circumventing selection problems. Finally, cross-country studies may be more efficient at uncovering general equilibrium effects. In our case though, the focus is on micro behaviours rather than institutional, national-level features, so cross-country analysis will be of lesser interest.

International tests are not perfect, of course. As Hanushek and Woessmann [2010] also point out, they are limited to a rather small number of countries, generally the most "developed" or industrialised ones. For instance, 59 countries participated in TIMSS in 2007. Overall in 2011, a total of 102 countries ever participated in one of the three major testing programmes (TIMSS, PIRLS or PISA). Furthermore, almost all the assessments have a cross-sectional design which prevents the researcher from tracking individuals through time (schools that are randomly selected for example for PISA 2006, may not be selected in PISA 2009 and cannot be identified anyway). This is the main drawback compared to small-scale experiments which track individuals. There also might be some cultural factors entering the processes of interest, but we cannot observe them. Besides, international tests focus on measuring educational achievement through test scores, and they have yet to develop internationally comparable and sound measures of non-cognitive skills. This is why our study will focus on cognitive skills, even though we can expect that video game playing also has an effect on non-cognitive ones. Last but not least, the population studied is restricted to primary and secondary school, where the basic general skills are generally learned. This ensures international comparability but unfortunately leaves aside both vocational training and higher education.

3.1.2 The PISA survey

The Programme for International Student Assessment (PISA), conducted by the OECD, submits 15-year-old students, regardless of their grade, to a standardised test. It aims at measuring how well students apply the skills they learned both in and out of school. The tests last up to two hours and are taken with paper and pencil. They cover a wide scope of skills and levels of difficulty. The results are scaled to achieve a mean of 500 points and a standard deviation of 100 across the OECD countries. Students also answer background questionnaires about their family context and educational experiences. School principals are also asked to fill in a questionnaire about their school. A PISA wave is held every three years since 2000. In each wave, three areas of knowledge are tested: reading, mathematics, and science, but one of them is assessed particularly in depth. In 2009, during the most recent wave of PISA, reading was the subject tested more in depth. 65 countries participated in it (a detailed list and a map are available in Appendix, figure 5).

PISA is designed to ensure that each country tests a representative sample of its students, and the procedures are monitored accordingly. Data which are not meeting PISA standards are dropped out from subsequent analyses (that was the case of the Netherlands in 2000, for instance). Most countries adopt a two-stage sampling design. In a first step, a sample of schools receiving 15-year-olds is drawn randomly. The probability for a school to be selected is proportional to its size. On average 250 schools are selected in each country. Then, a sample of on average twenty-seven 15-year-old students is drawn randomly within each school. Each student has an equal probability of being selected. The selected students are then given the tests and questionnaires. Because the tests are very long and comprise many questions to assess a large scope of skills, each student is given only a part of the test. Thus, we cannot know exactly what their overall score is. A set of five plausible values for their score has been constructed in order to take into account this limit of the data. Estimations have to be carried out once separately for each one of the five plausible values and the variance between these values makes up for the incertitude about the score, thus inflating the standard errors obtained. Last but not least, each individual in the dataset is attributed a weight that compensates for the probability of being sampled (all subsequent analyses in this thesis are done with weighted observations). To take account of the complex two-stage sampling design, of the plausible values method and of the particular use of replicate weights to compute standard errors (BRR technique), we use the Stata function `pv` written by Kevin Macdonald for the explicit purpose of exploiting PISA, TIMSS and PIRLS data.

3.1.3 The TIMSS and PIRLS surveys

The International Association for the Evaluation of Educational Achievement (IEA) conducts both the TIMSS and PIRLS tests. TIMSS (Trends in International Mathematics and Science Study) targets pupils enrolled in the fourth or eighth grade of formal schooling and tests them on several mathematical and scientifical topics, e.g. algebra, geometry, biology, chemistry. We will focus on the overall grades obtained in mathematics and science. Background questionnaires are also submitted to students, school principals and teachers. 67 countries and benchmarking entities participated in TIMSS 2007 in total, and divide up as follows: 56 countries tested 8th-grade students, and 44 countries tested 4th-grade students. The basic sampling design is a two-stage stratified cluster design, which is very similar to PISA's. In the first stage a sample of schools is drawn (typically 150 schools in each country), and in the second stage a sample of one or two intact classrooms is drawn from each of the sampled schools. Student scores in mathematics and science are scaled according to the item response theory (IRT) scaling methods [see Olson et al., 2008]. Plausible values are used for the same purpose as in PISA.

PIRLS (Progress in International Reading Literacy Study) targets fourth graders because, as the IEA argues, "The fourth grade is an important transition point in children's development as readers, because most of them should have learned to read, and are now reading to learn" [Martin et al., 2007, page 1]. The test covers several reading skills going from reading for literary experience to reading to acquire and use information. PIRLS includes the largest set of questionnaires that gather information from the student, the parents, the teacher, the school principal, and the national research coordinator. 40 countries and 5 Canadian provinces participated in PIRLS 2006. The sampling design and scaling method are exactly the same as in TIMSS. Again, we use the function `pv` to take account of these characteristics and of the jackknife method used to compute standard errors in estimations.

Table 2: Questions related to video games in the student background questionnaires

Test	***Question code***	***Question***	***Answers proposed***
PISA	IC04Q01 IC04Q02	How often do you use a computer for following activities <u>at home</u>? a) Play one-player games ; b) Play collaborative online games	1. Never or hardly ever 2. Once or twice a month 3. Once or twice a week 4. Every day or almost every day
TIMSS	AS4GPLCG BS4GPLCG	On a normal school day, how much time do you spend before or after school doing each of these things? – I play computer games	1. No time 2. Less than 1 hour 3. 1-2 hours 4. More than 2 but less than 4 hours 5. 4 or more hours
PIRLS	ASBGTSP2	About how much time do you spend doing the following things outside of school on a normal school day? – Playing video or computer games (including Nintendo, Gameboy, or Play Station)	1. No time 2. Up to 1 hour 3. From 1 hour up to 3 hours 4. From up to 3 hours up to 5 hours 5. 5 hours or more

3.2 Descriptive statistics

3.2.1 Measuring video game playing

In PISA, students are questioned about their familiarity with information and communication technologies (ICT) in a separate questionnaire, which is administered only in the countries who chose to do so (44 countries on a total of 65). Two questions deal with computer games: the first one concerns playing one-player computer games (which we will abbreviate POPCG), and the second one, playing collaborative online computer games (hereafter PCCG). Playing one type of game is positively correlated with playing the other type (0.52 correlation). But the distinction between the two allows to measure the respective effect on educational achievement of games with and without a social component. As mentioned before, the questions in PISA are phrased in terms of frequency (see table 2). The PISA student questionnaire also contains a question about whether the student owns a console at home, but not whether he or she plays video games with it. There is only a mild correlation (0.16) between owning a console at home and playing any type of computer games, which tends to show that either computer-game players and console-game players are different, or that someone else in the household (possibly no one) uses the console.

In TIMSS and PIRLS, the question about computer games is part of the general student background questionnaire and so it is administered by all participating countries. It is phrased in terms of amount of time rather than frequency. The TIMSS question specifically targets computer games while PIRLS combines in a single question computer games and video games played on a console. However, there is no information on whether the student owns a console. In a nutshell, the difference in the scope and phrasing of the background questions makes it more difficult for us to compare the estimates obtained across the three studies, and we should bear it in mind when interpreting the results. Table 3 illustrates this problem. It summarizes how students answered the question about their gaming habits, by sex. If we compare the distribution of answers in PIRLS and in TIMSS for 4th graders, we observe that the

Table 3: Percentage of students by answer to the questions about playing computer games, by sex

	PISA				**TIMSS**				**PIRLS**	
	POPCG		PCCG		Grade 8		Grade 4			
	Boys	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*
1	25.4	44.5	39.7	68.4	34.8	51.8	35.1	41.0	20.4	33.2
2	18.5	26.0	14.8	16.1	24.7	24.7	28.3	34.7	23.8	31.5
3	27.5	19.5	19.7	10.1	21.7	14.3	19.4	15.2	21.0	16.1
4	28.6	10.0	24.8	5.4	10.1	5.6	7.2	4.6	14.4	9.5
5	–	–	–	–	8.7	3.6	10.0	4.5	20.4	9.7
Total	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
Pearson χ^2	7.5e+05		1.1e+06		8.5e+05		2.8e+05		7.8e+05	

N.B. The categories 1 to 5 correspond to the answers proposed to the students (see table 2).

students declared playing longer each day in PIRLS, whatever their sex. For example, about twice as many students declared playing 5 hours a day or more in PIRLS compared to students who declared playing 4 hours a day or more in TIMSS. If the countries sampled in the two tests are comparable, we can attribute this difference to the fact that the PIRLS question includes explicitly console-based video games. Thus, we might be underestimating the number of students playing video games in TIMSS and PISA, and their amount and frequency of play. It is important not to extrapolate our results to all types of games.

There is one last important observation to make about the measurement of video game playing. As the reader has noticed, the questions are multiple-choice questions, with categories to pick from. We therefore cannot know accurately how often or how much time a pupil plays video games, we only know an approximation (PISA) or an interval (TIMSS and PIRLS). The upper values are censored to the highest category value. This feature will determine the way we will include the variables concerning video games in our estimations.

3.2.2 Who are the game-players?

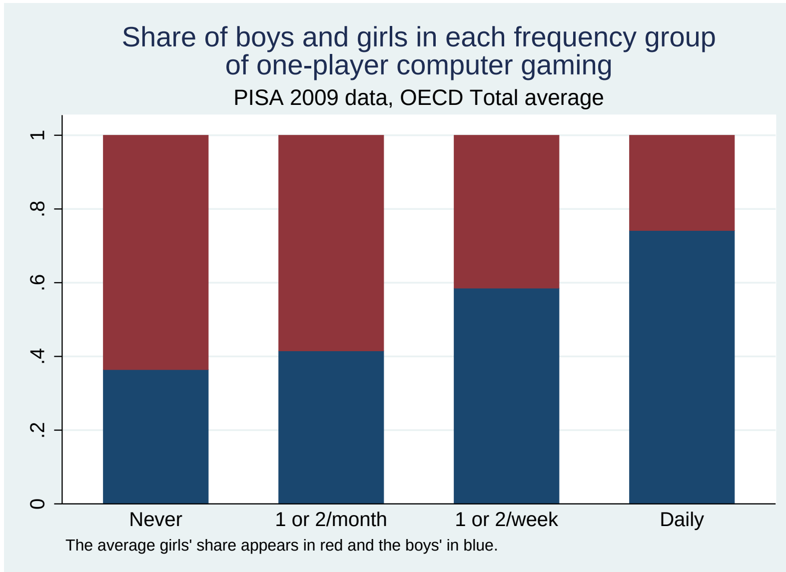
Gender. The first and foremost feature of the data is, unsurprisingly, that boys play video games more than girls (see table 3). Pearson chi-square tests all lead to rejecting the null hypothesis that sex and time spent playing video games are independent. The relationship is recurrent and may be best illustrated by figure 1 on the next page, which represents the respective share of boys and girls in each category of POPCG (PISA 2009 data[1]).

Computer. Another unsurprising fact is that students are more likely to play computer games if their household owns a computer, which is actually the case for an overwhelming share of students (83.7 % in PISA 2009). Among the ones who do not have access to a computer at home, 36.2 % play computer games[2] at least once a month, probably at a friend's or relative's place. This is to compare

[1] For the sake of concision, we do not show the results obtained from the same computations made on each of the three data sets, and generally choose the most recent one as illustration (PISA 2009). But obviously, the patterns highlighted in this section are broad patterns that can also be found in TIMSS 2007 and PIRLS 2006.

[2] Either one-player computer games or collaborative online computer games (PISA 2009 data).

Figure 1: Video games and gender



with the 77.2 % of students owning a computer who play at least once a month. Playing computer games is indeed a wide-spread activity among the young students surveyed.

Internet. Similarly, we find that having an Internet connection at home increases the likeliness of playing computer games. But, while the difference it makes for one-player games is mild, it is greater for collaborative online computer games: 78.4 % of students without Internet at home never or hardly ever play online games (against 46.1 % of students with the Internet). Again, the PISA sample is comprised of students who in vast majority (75.6 %) have access to the Internet at home.

Other ICT-related activities. The ICT familiary component for the PISA student questionnaire includes questions about other activities such as doing homework on the computer, browsing the Internet for fun, writing e-mails, chatting, maintaining a blog and participating in a forum. These activities all appear to be positively correlated with playing computer games. The most correlated activity is downloading music, which is correlated at 29 % with playing one-player computer games, and at 42 % with collaborative online computer games.

Socio-economic background. Many other characteristics related to the socio-economic background display no major difference between categories of players. For example, in PISA 2009, playing any kind of computer games is almost not at all correlated with age, immigration status (first and second generation), education level of the parents, and the number of inhabitants in the community where the school is located (see table 4). It is somewhat positively correlated to parents' wealth and the index of educational resources possessed at home (books, dictionaries, manuals, educational software...). It is negatively correlated with the student's perception of the quality of the disciplinary climate in class (the better the climate, the less the pupils play computer games on average).

Table 4: Computer games and socio-economic background, PISA 2009 data

Pairwise correlation between playing computer games and...			
Age	-0.02	Highest parental level of education	0.08
Immigrant	0.03	Public school	0.07
Student perception of class disciplinary climate	-0.12	School community	0.01
Parents' wealth index	0.16	School's student/teacher ratio	-0.06
Parents' educational resources index	0.20	School's educational resources index	-0.02

Scores. Because our goal is to find if playing computer games has an effect on educational achievement, it is of paramount interest to see if such a pattern appears when exploring the data. In PISA data, we observe a sort of inverted-U shape in the relationship between scores and the frequency of play, as displayed in figure 2 for POPCG (and figure 8 in Appendix for PCCG). On average in all the OECD countries, the highest scores are obtained by students who play computer games reasonably often. The gender gap is remarkable, with on average girls performing better than boys at reading tests but worse at math tests. The gap in science is unclear. Disaggregating the data by country (figure 3) tends to confirm this observation (see table 8 for the list of country names and abbrevations). The relationship between scores and game playing is not strikingly clear at this level, even though we can guess that it is negative for high levels of gaming. Indeed, the countries where the percentage of pupils declaring to play at least once a week is the highest tend to be lower-performing countries (for example, Serbia or Bulgaria). The highest-performing countries appear to have moderately high rates of pupils playing once a week or more. Similar patterns are obtained for the other disciplines (mathematics, science). For collaborative online computer games, the picture is even less clearcut (see figure 9 in appendix). The countries where the boys are the highest performing also tend to play more online games. Girls do not play online games often, and the lack of variance across countries conceals any potential relationship with scores.

Historical perspective. Gender gaps are not something new. Girls have historically performed better in reading tests, whereas boys have more played video games than girls, and performed better in math tests (though this last gap has been reducing, and even inverting recently). Both girls and boys play video games more often as years pass, but we observe that boys still significantly play more in all studies. As for scores, back in 1995, boys significantly outperformed girls in maths by 5.3 score points on average (total of countries participating to TIMSS). This difference plummeted to zero in 2003, and girls outperformed boys in 2007, by 3.5 score points. This is a small, yet significant, difference compared to the average of around 500 score points. The gender gap in reading is somewhat larger, though slowly shrinking. It is estimated in PIRLS to be on average 19.7 in 2001, and 16.5 in 2006 (the difference between the two dates is significant). Over a broader time span and in a slightly different set of countries, PISA data uncover a significant increase in the gender gap in reading in a minority of countries, and no significant change in the other countries (see figure 10 in Appendix).

Figure 2: Scores and computer games (POPCG)

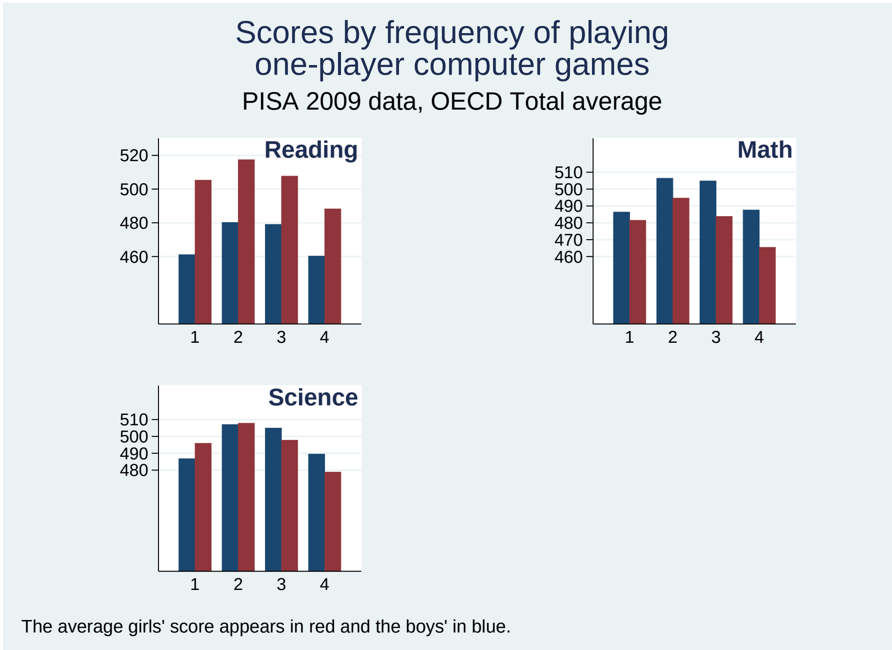
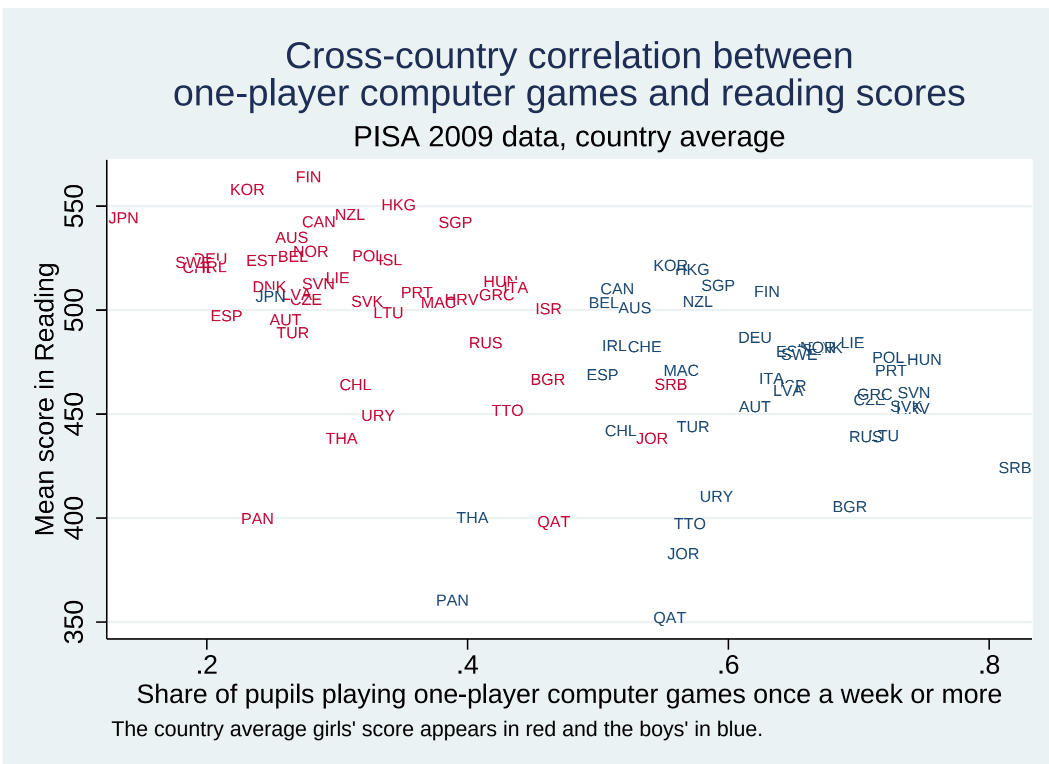


Figure 3: Reading scores and computer games (POPCG) by country



4 Empirical strategy

4.1 Baseline econometric model

We use a linear regression to model the relationship between scores and playing computer games. We take account of possible non-linearities (we suspect a quadratic relationship) by including separately each category of intensity of playing, as a dummy. Hence, we seek to estimate by ordinary least squares (OLS) the parameters β and δ_j of the following equation:

$$Score_i = \sum_{j=1}^{J} \delta_j^{'} PCGSEX_{ij} + \beta^{'} X_i + \varepsilon_i \tag{2}$$

where:

$Score_i$ is the result obtained by student i on the test,

$PCGSEX_{ij}$ are the dummies indicating which answer the student gave to the question concerning computer games, interacted with the sex of the student,

X_i is the set of exogenous control variables,

ε_i is the error term.

This first model estimates the effect of playing computer games across countries. The model is estimated separately for each dataset (PISA, TIMSS and PIRLS) and for each grade (TIMSS grade 4 and TIMSS grade 8), for obvious reasons regarding the consistency of the sample. It is also estimated separately for each discipline (reading, mathematics and science), which translate into three distinct dependent variables.

The variable PCG is interacted with the sex of the student because we expect the effect of playing computer games to be different if the player is a girl or a boy. This has nothing to do with an intrinsical difference between girls and boys, but rather with the differences observed in the types of games most played by the two sexes. Greenberg et al. [2010], notably, indeed show that girls tend to prefer more "traditional" games (such as board games, puzzles...) and boys more "physical" games (like action, racing and sports games). By distinguishing the effect on the two sexes, we are trying to capture this difference in the average patterns of playing in order to approximate the effect of different types of games on scores.

The baseline model can be enriched by including country fixed effects (the dummies C_k), which allow to control for all country-specific characteristics that might impact scores, for example the national institutional framework. It then becomes:

$$Score_i = \sum_{j=1}^{J} \delta_j^{'} PCGSEX_{ij} + \beta^{'} X_i + \sum_{k=1}^{K} \zeta_k C_k + \varepsilon_i \tag{3}$$

4.2 Dealing with the potential sources of bias

4.2.1 Omitted variables

The baseline model is difficult to estimate without bias due to the endogeneity issue mentioned in section 2.2: there might be omitted variables in equation 2 or 3 that would also be correlated to $PCGSEX_{ij}$. We could use an instrumental variable strategy if a valid exogenous instrument would be found. Unfortunately, it is difficult to imagine any instrument that would comply with the exclusion restriction: all variables that affect the choice of playing computer games can, to some extent, also affect the scores directly. Therefore, we choose to adopt a different approach, which uses the so-called "generalised residual" [elaborated by Gourieroux et al., 1985]. We adapt the methodology used in a similar framework by Spiezia [2010].

The first step consists in estimating a model to explain the time spent playing computer games, noted PCG^*, with a set of background characteristics X_i, that we think are exogenous to the pupils.

$$PCG_i^* = \lambda' X_i + \nu_i \tag{4}$$

However, because the continuous PCG^* is unobservable and latent, and we only have access to discrete categories of PCG, we have to estimate this model by ordered probit (assuming the errors ν_i are normally distributed).

In the second step, we compute the generalized residual $\hat{\nu_i}$ according to the formula provided by Gourieroux et al. [1985, page 77]. We then include this generalized residual as an explanatory variable in our model of interest. The other explanatory variables are the same as in our baseline equation.

$$Score_i = \sum_{j=1}^{J} \delta_j^{'} PCGSEX_{ij} + \gamma^{'} \hat{\nu_i} + \beta^{'} X_i + \sum_{k=1}^{K} \zeta_k C_k + u_i \tag{5}$$

The reasoning behind such a method is the following. By introducing the estimated generalized residual of the first equation in the explanatory variables of the equation for the score, we control for all the unobservable determinants of the use of computer games, and thus avoid the omitted variable problem. The only condition for the results to be unbiased is that equation 4 is correctly specified, meaning that the generalised residual $\hat{\nu_i}$ is sufficiently close to a normal distribution.

4.2.2 Reverse causality

There is a major problem if indeed scores influence how much pupils play video games, rather than (or simultaneously to) the contrary. If reverse causality happens, all the coefficients we try to estimate in our equations will be biased. The problem is even more serious in equation 5, because if score determines $PCG*$, then score is an omitted variable in the first-step equation 4, and it ends up in the generalised residual, which in the second step we introduce as an explanatory variable for... score. However, one key element is comforting us into thinking that reverse causality should not be an issue here. The scores we are studying are not the scores the pupil got in the past, and that might have influenced her attitude toward computer games. The scores we are studying are scores obtained to an international test, unknown at the time the student reported their actual consumption of video games, and that has no impact whatsoever on the overall grades of the student. We can expect that, for all students, there is some discrepancy, some variance due to the test's conditions and specific questions, between the scores generally obtained at school, and the scores obtained for the international standardised test. As small as it might be, this discrepancy is key to our reasoning. If there is reverse causality, what might be an omitted variable in equation 4 (step 1 of the model) are definitely past scores, which might influence the choice of playing video games or not. Then, what enter the generalised residual are also past scores. But what lies on the left-hand-side of the equation is actually the PISA or TIMSS or PIRLS score, which is not perfectly colinear to past scores. Therefore, the coefficient δ we get is an estimate of the causal impact of playing computer games, with past scores (among others) maintained equal. Because the international test score cannot cause the amount of playing, and we control for past scores (if they cause $PCG*$), the endogeneity issue is dealt with.

4.2.3 Measurement errors and selection

The international tests are subject to the same pitfalls as traditional surveys. In the questionnaires, self-declaration might bias the results, if pupils have a distorted idea of how much time they spend playing computer games on average (which as a matter of fact *is* difficult to know). The order and wording of the questions and answers proposed might also influence the results, even though both the OECD and the IEA are very careful when formulating the questionnaires and their national adaptations. The most sizeable source of bias typically comes from non-response, which takes many forms. On the one hand, some individuals might be missing completely because they were absent on the day the test was administered, or because they refused to or could not participate. At our stage of study, we cannot deal with this problem, and must rely on the efforts of the OECD and IEA to ensure that the final sample is representative of the population (possibly with some correction weights). On the other hand, some items may be omitted by a student, some not reached (the survey was abandoned before completion), or not administered (sheets were missing). The problem is that, when any variable in our model is missing, the entire individual is withdrawn from the sample on which the estimation is computed. There is a high risk that pupils not answering to a particular question are not comparable to other pupils. Generally, they clearly tend to get lower grades on the tests.

Therefore, we must be careful when choosing the control variables we include in our models. We want to include them because we expect them to influence both scores and computer use, according to the literature, and because we need to limit the risk of omitted variable bias. But at the same time, we face two requirements to keep from introducing endogeneity: first, avoid sample selection; and second, avoid reverse causality and other endogenous treatments issues. Hence, variables with too many missing values (more than 15%, which is the threshold chosen by the OECD and IEA) are removed. This is why no variable concerning teachers' background characteristics are included: the response rates were far too low. In addition, only the variables that can soundly be considered exogenous are included. This is why other activities, like watching TV, reading or chatting on the Internet, were not included, because they are themselves endogenous treatments chosen by the students themselves. The variables included in the matrix of control variables X are presented in table 8 in Appendix. It is important to notice that the matrix of control variables X_i is the same in the step-1 and step-2 equations. There is no reason to think that a variable that could impact gaming would not impact scores, and vice-versa.

5 Results

5.1 Baseline cross- and within-country models

We first present the results obtained when estimating the two baseline models (simple OLS, with and without country fixed effects) for PIRLS data in table 5, columns 1 and 2 (we will comment column 3 in the next section). The coefficients δ_j, associated with the explicative variables $PCGSEX_{ij}$, are reported in this table, as long as their standard errors (hereafter, s.e.). They are to be interpreted as the effect on the score of a pupil switching from the first category of PCG (never or almost never playing computer games) to the j^{th} category of PCG, ceteris paribus.

For the sake of concision, we do not report the full results from the models estimated on the other datasets in this section, but we will describe and analyze their main features. We chose to present in details the results of PIRLS because, as mentioned in section 3.2.1, the PIRLS student questionnaire includes both console games and computer games, thus giving a better approximation of how much students really play video games. About twice as many students declare playing the highest time category in PIRLS as in TIMSS and PISA. It is likely that the two latter datasets actually underestimate the amount of video game playing among the children, and this might seriously bias our results, which should therefore be interpreted with care.

As a reminder, PIRLS data concern 4th-graders undertaking a test in reading. A first important observation is that in PIRLS, the results of the cross-country (column 1) and within-country (column 2) models are pretty different. In the cross-country model, we find a significant[3] positive impact of playing video games on reading scores, whatever the time spent playing for boys, and for periods of less than 3 hours per day for girls. When adding country fixed effects, the picture is more nuanced: playing is beneficial to girls only up to 1 hour of play per day, and for boys up to 3 hours. As expected, the inclusion of country fixed effects raises the average[4] R^2 from a respectable 0.24 to a more satisfying 0.52, which means that actually a lot of the variance observed in scores patterns can be explained with only country characteristics, and that there is a potentially high risk of bias in cross-country estimates.

The results so far can be interpreted as confirming our theoretical hypothesis 1, namely: reasonable amounts of playing are beneficial to achievement in reading, but too much time spent playing (more than 3 hours for girls and more than 5 hours for boys) is actually harmful. The "tipping point" seems to be situated much higher than we expected, since spending 3 hours on one activity on each school day is already a lot. The difference we observe between boys' and girls' coefficients might come from two channels: an intrinsic difference (be it innate or socially constructed) in the way girls and boys learn from video games, and transfer this knowledge in the area of reading skills; or an empirical difference in the types of games they play and in their attitudes toward these video games. A conjecture that we

[3] All coefficients declared to be significant in this thesis are statistically different from zero at least at the 5 % confidence level.

[4] The R^2 is "average" because it is the mean of the five R^2 obtained during the estimations run for each of the five plausible values.

Table 5: Models estimated with PIRLS 2006 data: OLS regression of P.V. in Reading on the variable "Playing computer games" intersected with sex

	(1) Eq. 2	(2) Eq. 3	(3) Eq. 5
PCG2 x GIRL	18.98***	7.451***	-10.92
	(2.421)	(1.820)	(14.99)
PCG3 x GIRL	8.282**	3.238	-26.26
	(2.588)	(1.909)	(24.03)
PCG4 x GIRL	-22.79***	-17.43***	-55.24
	(2.840)	(2.381)	(31.11)
PCG5 x GIRL	-30.00***	-34.30***	-83.99*
	(3.501)	(2.860)	(41.05)
PCG2 x BOY	32.22***	17.38***	0.646
	(3.621)	(2.776)	(13.67)
PCG3 x BOY	32.72***	17.58***	-9.975
	(3.508)	(3.251)	(22.58)
PCG4 x BOY	11.21**	-3.284	-38.99
	(4.228)	(3.308)	(29.16)
PCG5 x BOY	11.90***	-14.60***	-63.79
	(3.496)	(2.998)	(40.37)
Generalised residual	–	–	16.11
	–	–	(13.28)
Controls X	Yes	Yes	Yes
Country fixed effects	No	Yes	Yes
N	191,316	191,316	191,316
Average R^2	0.24	0.52	0.52

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

can make is that action, racing and sports games may demand more reading than "traditional" video games like puzzles.

The results for TIMSS 4th-graders are quite comparable to PIRLS', except for the fact that cross-country and within-country estimates are very similar, therefore we will report only the within-country estimates hereafter. For girls, playing computer games starts to have a significant negative impact both on math and science results at 1 hour of play per day (compared to never playing). The effect is mild at the threshold of 1 hour of play, the score being reduced by 7 points for math (standard error: 1.8) and 10 points for science (s.e.: 2.0), everything else equal. But it becomes large and highly significant ($p < 0.001$) when girls play more than 4 hours a day: the cut in score is then around -40 points. As for boys, playing a little bit each day (less than 1 hour) actually has a small but significant positive effect on scores in both disciplines (around +5 points). The coefficients start to be significantly negative at 3 hours of play and more, leading to a maximum estimated decrease in score of -26 points when boys play 4 hours rather than none. Again, the difference between girls and boys might come from the different types of games they play, and is consistent with the literature. The boys tend to play more action and racing game, which demand skills in spatial vision, and role-playing and strategy games, which reward reasoning.

Even though the model specification is similar, and the target sample comparable (4th graders), it is difficult to compare the coefficients from PIRLS and TIMSS data because the discrepancies add up: between disciplines, between measurement of playing (computer games only or all video games), and, last but not least, between survey designs. Still, at this point we can safely affirm that there is no contradiction between the two results.

For female 8th-graders in TIMSS, the findings are also very similar to 4th-graders'. But for male 8th-graders, playing computer games appears to have no impact significantly different from zero until very high amounts of time spent each day. It is only when boys play more than 4 hours a day that the effect becomes significantly negative, at -19.6 points (s.e.: 2.7).

As for PISA, the analysis is divided into two parts: equations with POPCG (one-player computer games), and equations with PCCG (collaborative online games). In each part, the analysis is conducted three times, one for each test subject as dependent variable. We will first review the findings related to the impact of one-player games. The coefficients of the cross-country models clearly tilt to the negative compared to within-country: POPCG has either no detectable impact on the three test scores, or a negative one. It starts being significantly negative at a frequency of once a week or more for girls, and almost every day or more for boys. But the PISA within-country analysis leads to coefficients which patterns, if not orders of magnitude, are very comparable to TIMSS' and PIRLS'. Playing once or twice a month rather than never lightly but significantly improves girls' scores in all disciplines (about +2.5 points), and boys' scores in science (+3.5 points, s.e. 1.5). There is then no discernable effect of playing between once a month and twice a week. Finally, playing every day or almost every day has a clear negative impact on all scores ($p < 0.001$), and for both sexes. The impact ranges between -5.5 and -9.4 points.

Playing collaborative online computer games has *always* a negative impact on scores in the baseline model, whatever the method (cross- or within-country), test subject, the sex, and the frequency of play. All the coefficients associated to it are significant at the 0.001 % level. Here again, the estimates from the cross-country model are larger in absolute value than the ones from the within-country model (meaning that the estimated impact on scores is more harmful). Girls' scores are particularly responsive to their frequency of PCCG. The effect of playing seems linear for girls: the more often they play, the more their scores are reduced, everything else being equal. The estimates range from -8.6 (s.e. 1.1) to -20.9 (s.e. 2.3) in the model with mathematics as the dependent variable, from -11.2 (s.e. 1.1) to -23.8 (s.e. 2.1) for reading, and from -10.5 (s.e. 1.0) to -23.2 (s.e. 2.3) for science (within-country model). Boys suffer a little less from their playing, but the effect is linear as well. Overall, math is the subject where the scores are the less damaged by the frequent play of online games.

Cross-country or within-country estimates? The advantages of cross-country comparisons were enumerated in section 3.1.1. But both kinds of estimations have their pros and cons. Cross-country estimation undoubtedly arises great concerns about omitted variables because of the unobserved heterogeneity at the country level. There are ways to deal with this problem without resorting to country fixed effects, though. But we argue that in our case, within-country estimations are more relevant than cross-country ones. Indeed, the drawback of within-country comparisons is that they do not deal with selection issues at the institutional level (for instance, pupils who attend private schools may differ from the others both on the observable and unobservable dimensions). We do face a selection problem because pupils self-select their level of video game playing. But we will deal with selection, at the micro level, with the two-step model, in next section. Therefore, within-country estimations are more interesting because they allow us to still have country fixed effects and to reduce the potential omitted variable bias. But the within-country method has another advantage, namely, dealing with the problem of sample selectivity being different across countries. Indeed, several authors, for example Rotberg [1995] about IEA studies and Prais [2003] about PISA, expressed the concern that some countries may sample their student populations differently, and that this difference may bias the results and prevent from drawing any policy conclusion. However, it is important to remind, as Hanushek and Woessmann [2010, page 13] did, that such bias occurs only if sample selectivity is systematic, i.e. correlated with the error term of the estimation equation. Moreover, including country fixed effects in our estimation equations removes all time-invariant factors for each country, hence limiting the possibility of bias related to varied sample selectivity.

5.2 Two-step model with the generalised residual method

5.2.1 Step 1: explaining and predicting gaming

The results of the first step of our model, estimating equation 4 by ordered probit, are not of paramount interest for our study, but for information they are available in Appendix in table 9 for PISA data (POPCG). The findings are similar in the other models. Confirming the correlations found during the exploratory analysis, we notice that the younger the pupil, the more likely s/he plays computer games. Household wealth and home educational resources are also positive determinants of playing frequency. On the contrary, a better perception of the disciplinary climate in class and a smaller student/teacher ratio are associated with less frequent videogame play. Having a computer at home is obviously a major determinant of POPCG. Curiously, having an Internet connection discourages playing one-player computer games, probably because online games are substitutes to them.

It is important to check that the model specification is valid, otherwise our second step of the reasoning (equation 5) will not give fruitful and reliable results. The first observation to make is that the goodness of fit is particularly absent, with a pseudo-R^2 as low as 0.08. Our model explains very little of the empirical variance in video game playing (and this is true for models applied to all three datasets). This is a serious limit, but that in some way we expected, because whether a child plays or not computer games is above all a matter of personal preferences, that are unobservable. As long as our model is correctly specified, we will be able to capture all the unobserved heterogeneity of playing in the generalised residual $\hat{\nu}_i$.

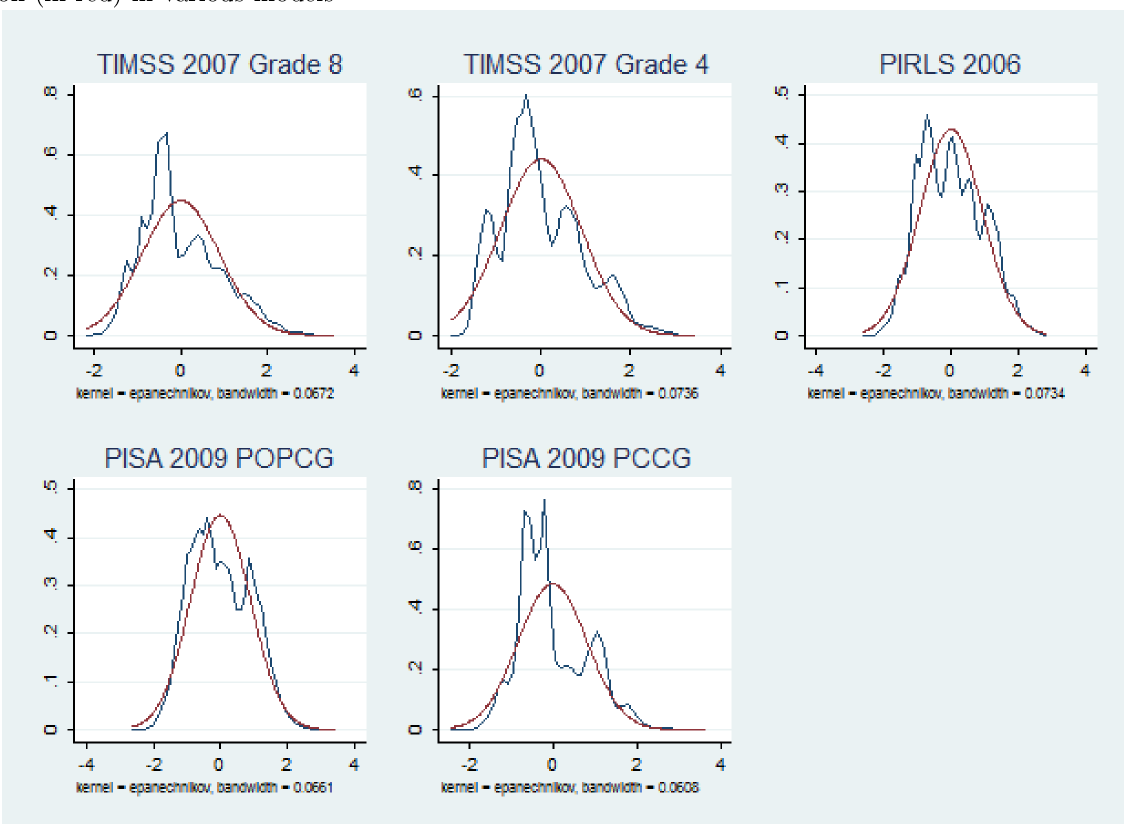
The residual, if the model is correct, should be standard-normally distributed[5]. Figure 4 shows the distribution of the generalised residual in our several models. It is never completely normal but approaches normality in the case of PIRLS and PISA POPCG. We will have to take with a pinch of salt the results from the models where the generalised residual departs from normality.

5.2.2 Step 2: explaining scores and introducing the generalised residual

We are now moving to the second step of our reasoning, the estimation of equation 5 using the generalised residual $\hat{\nu}_i$ to try to account for the self-selection of students' amount and frequency of

[5]We also tried an ordered logit, but the generalised residual did not even approach the logistic distribution.

Figure 4: Distribution of the generalised residual (in blue) compared to the standard normal distribution (in red) in various models



play. For PIRLS 2006 data, the results of the second step are presented in table 5, column (3). Compared to the baseline OLS framework, the coefficients now lose all significance, except for girls playing more than 5 hours each day, who lose 84 points (s.e. 41). The average fit of the model is not improved, with $R^2 = 0.52$ again. This suggests that the introduction of the generalised residual might not be as relevant as we thought. The generalised residual itself does not differ from 0 at the 5 % confidence level. Even so, we can cautiously observe that the point estimates are much larger in absolute value than in the baseline model. But this is also true of their standard errors, which makes the interpretation tricky.

The analysis is easier for PISA results concerning one-player games. Results are presented in table 6. The coefficients associated to playing one-player games intersected with sex are all positive and highly significant. They are much larger in absolute value than all the coefficients we encountered so far. However, we will never stress enough the difference in the way playing is measured (either through frequency or amount of time) between PISA and TIMSS/PIRLS. This makes the comparison between PISA's and the other datasets' estimates quite incorrect. In PISA, the most beneficial impact on scores is obtained when students play every day or almost every day (rather than never). It reaches the level of +56 points in reading when a student switches from not playing to playing every day, ceteris paribus, +69 points in mathematics, and around +73 points in science. It looks like playing computer games develops more the skills used in science than in math, and in math than in reading. Boys and girls draw about the same gain from playing: the differences are not significant, except for the first category. Girls benefit slightly more than boys from playing once or twice a month rather than never. Unfortunately, we lack information to interpret the cause of such peculiarity. The model has a moderate goodness-of-fit, with an average R^2 ranging between 0.29 and 0.32, which is not bad in our case of a microeconomic dataset.

We do not report the full estimates for PISA (PCCG) and TIMSS (grades 4 and 8) because of the restrictions on the normality of the generalised residual, but here are the main features, to be interpreted with the appropriate care. Let us first review PCCG. As in the previous paragraph, and contrarily to what we found in the baseline model, the coefficients associated with playing collaborative online computer games (intersected with sex) are all positive. But they are smaller (the maximum being a coefficient of 34 for boys playing every day, with the plausible value in science as dependent variable), and many are not significantly different from zero, suggesting that online games are not as good as one-player games in terms of educational achievement. Here again, the gains obtained in the science test are higher than in the math test, which are themselves higher than in the reading test. And the girls still get higher scores if they play once a month rather than never, while this has no impact on the boys' marks.

Estimating the same model with TIMSS data confirms the results we observe in PISA. In grade 4, we estimate that students playing computer games get a huge reward in terms of math scores, ranging from +50 points (s.e. 9.8) at least, for both sexes, to up to +87 (s.e. 26.23) for girls and +99 (s.e. 25.8) for boys. The impact of playing is monotonically increasing. The gains in science scores are more moderate, ranging from +30 to +50 points, and, remarkably, there is no significant gain for amounts of play superior to 4 hours a day.

Finally, the model estimated with the sample of 8th-graders in TIMSS gives some strikingly similar results. The main observations as for 4th-graders can be made. The only difference is that the

Table 6: Equation 5 with PISA 2009 data. OLS regression of P.V.s on the variable "Playing one-player computer games" intersected with sex, with generalised residual

	Reading	Math	Science
POPCG2 x GIRL	29.02***	34.79***	35.74***
	(5.387)	(5.265)	(5.623)
POPCG3 x GIRL	42.86***	51.34***	53.41***
	(8.616)	(8.348)	(8.794)
POPCG4 x GIRL	56.23***	69.70***	71.90***
	(12.89)	(12.53)	(13.00)
POPCG2 x BOY	24.17***	29.52***	31.43***
	(4.832)	(4.768)	(4.889)
POPCG3 x BOY	42.01***	49.95***	53.49***
	(8.015)	(7.882)	(7.831)
POPCG4 x BOY	55.53***	68.35***	74.02***
	(12.65)	(12.42)	(12.73)
Generalised residual	-23.68***	-28.03***	-29.08***
	(4.659)	(4.519)	(4.715)
Constant	166.8	193.3	199.4
	(118.1)	(138.2)	(155.8)
Controls X	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes
N	266,075	266,075	266,075
Average R^2	0.31	0.31	0.28

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

coefficients are somewhat lower, for example the maximum coefficient is found for boys playing more than 4 hours a day rather than none, and it is equal to 62 (s.e. 18.9), to compare with the 99 from the previous paragraph. Because the only difference between the two versions of TIMSS 2007 is the grade of the students (plus some minor discrepancies in the sample surveyed), these results suggest that age does not play a key role in determining how video games impact pupils' academic learning.

Interpreting the generalised residual. In every model we run, we notice that the coefficient associated to the generalised residual is always of the opposite sign of the coefficient associated to $PCGSEX_i$. Here is a possible interpretation: the pupils who actually play computer games a lot may differ significantly from the other students in their unobservable characteristics. When the generalised residual is significantly negative, like in TIMSS and PISA[6], we can infer that, were the gamer-children not to play computer games, they would get lower grades than other non-gamer-students. However, their playing computer games actually makes them overcome this handicap and get better grades. This may be why, when simply regressing the score on the time playing computer games (baseline model), we find a much less beneficial effect of playing.

[6] Reminder: the coefficient associated to the generalised residual is positive but not significantly different from zero in PIRLS.

5.2.3 Does the gender gap in gaming explain the gender gap in scores?

We now turn to the test of theoretical hypothesis 2: were girls to play as much as boys (or vice-versa), would they get the same grades as well? To verify this, we adopt a simple strategy. First, our two-step model is modified: we drop all boys from the sample and estimate the two steps of the model only for girls. Then, we store the estimated coefficients, and run a "what if" prediction: we replace the actual value of each girl's PCG_i with the mean value of boys' PCG_i, and predict an adjusted value for their score. In a last step, we compare the average predicted girls' score with the true average girls' score and the true average boys' score. We repeat this procedure for all the model specifications using the generalised residual mentioned above.

As explained in section 3.2.2, in recent years there has been a gender gap in international scores. The gap is in favour of girls in the case of reading, and typically in favour of boys in the case of mathematics. We observe no recurrent gender gap in science. Therefore, let us focus on math and reading. With PISA data, we find the following: the girls score on average 505 in reading, and the boys 466. If the girls were to play one-player games as often as boys, they would get a score of... 517! This actually leads to a larger gender gap, which is normal since, remember, we found a positive effect of POPCG on reading scores (table 6). Therefore, PISA results suggest that video games do not explain the gender gap in reading. As for math, the PISA girls get a mean score of 483, far below the 495 score obtained by boys on average. But their predicted score is 498, meaning that the gender gap in math would be filled if girls played video games as often as boys. In addition, PIRLS data suggest that the gender gap in reading can be explained by video games. The girls' average score is 508, against a mere 492 for boys. Since in PIRLS, the coefficient associated with PCG_i is negative, we find that the predicted girl scores is lower than the true one: 494, which is much closer to the true boys' mean score. Finally, in TIMSS, the results are not conclusive, and do not succeed in explaining the gender gap in scores by the gender gap in gaming.

5.2.4 Robustness checks

Allowing coefficients to vary across countries. For many reasons, interpreting coefficients that vary across countries is a difficult task. First of all, it is theoretically difficult to explain why the effect of computer games would be different from a country to another, but for the intervention of unobservable country characteristics like culture and preferences in types of video games. Even if we observe indeed a difference in the estimates obtained, we can only make conjectures about its cause, observing similarities and patterns across countries. Second, international assessment tests are not designed to give estimates for one country only. Third, comparing estimates of a same model performed in several countries separately is difficult because we have to assume that any bias is constant across countries. Generally, it does not make sense to interpret the size of each estimate [Hanushek and Woessmann, 2010, page 15]. Yet, it is interesting to know if the effect we identified in the previous sections varies or not across countries, and in the case it does, whether this variation is systematic and can be explained by some country characteristic. Last but not least, if we find little variation of the estimates across countries, it may come as a confirmation of the robustness of our previous results.

In practice, we run each model separately for each country and store the results. We classify the coefficients estimated into three categories: significantly positive, insignificant, and significantly negative. Table 10 in Appendix presents a summary of the findings. It is worth noting that the coefficient associated to PCG, when it is significantly different from zero, is positive in many countries. This is a constant across models and samples. But the table does not tell anything about the magnitude of the effect. Figure 11 in Appendix is here for this purpose. For each country, the figure graphically represents the coefficients' point estimates as well as their 95% confidence interval, the y-scale being in score units. For the sake of concision, we only included the graph regarding the impact of POPCG on the scores in science measured by PISA. A vast majority of point estimates are positive, which is consistent with our previous findings. A lot of them are insignificant too, which means that, in many countries, we cannot reject the hypothesis that video games have no effect on educational achievement. This is probably due to the reduction in sample size, from about 200,000 observations to around 3,000 or 4,000 individuals per country. The other specifications of the model (other datasets and other dependent variables), when estimated separately for each country, are also consistent with their "worldwide" counterpart. This is reassuring: it is unlikely that our results are driven by only a small bunch of countries.

Within-school estimates (Chile only[7]) An excellent way of getting rid of many doubts about potential omitted variable biases is to conduct the analyses at the school-level, and to compare only children within the same school by introducing a school fixed effect in the model. Indeed, students of the same school are likely to share a lot of unobservable characteristics (socio-economic background, neighbourhood quality...). Therefore, we check the robustness of our results by estimating our two-step model with school fixed effects on the sample covering Chile. We remove schools with less than 5 observations to allow the convergence of the maximum likelihood estimation. The results are comforting: within-school estimates are slightly inferior to within-country estimates, but still very significant, and overall absolutely comparable.

6 Conclusion

In this section we will try to draw the lessons from analysis and answer our research questions. What is the expected increase or decrease in scores caused by gaming? How much playing is too much? Which pupils should turn off the video game? From our multiple-faceted analysis, we conclude that:

- There is indeed an inverted-U relationship between game playing and scores, everything else being equal (theoretical hypothesis 1). Playing regularly can raise the scores by a couple of tens of points, and playing every day can bring an around +50 points, or even more depending on the specification of the model. Playing a little each day (1 or 2 hours) can be beneficial too. However, playing several hours each day (typically more than 3 to 5 hours) diminishes the scores by up to 80 points.

[7]For technical reasons, we were not able to run estimations with school fixed effects over the entire sample of countries. Windows O.S. allows a maximum of 2 Go of memory to any software running, including Stata. Hence, the creation of the school dummies (around 8,000 in PISA, for example) over a sample of more than 250,000 observations was not allowed. We chose Chile to run this robustness check because it is the country for which we found one of the highest and most significant impacts of video games on science scores (see figure 11).

- Girls face potentially more harmful consequences from playing a lot, but they also draw the most benefit from playing a little.
- Collaborative online computer games are worse than one-player games in terms of their impact on scores.
- Gaming seems to have a better impact on scores in mathematics and science than in reading.
- Console games (which are included in PIRLS' question) appear to draw our estimations of the impact of playing on scores to the bottom, suggesting that they might hold less educational interest than computer games.
- The evidence is mixed about the gender gap in scores (theoretical hypothesis 2): in some cases (the most reliable ones: PIRLS and PISA), it is entirely explained by the gender gap in gaming, and in other cases, it is not.
- We do not have enough evidence to reject the hypothesis that the effect of playing does not discernably vary with age.

In the end, should the children turn off the video game? Probably yes, if they already have played more than 3 hours today. But they might actually be unintentionally practicing their academic skills while playing: socialising, experiencing diplomacy and cooperation, learning to persevere when challenged, reading, writing, moving in a 3D environment, solving enigmas, exercising their speed and memory... And, in any case, the answer depends on how they would otherwise spend their time. If they would be watching television or chatting on the Internet instead, then it might not be worth turning off the video game. As Blumberg and Altschuler [2011] put it, "playing video games for fun may present children and adolescents with a necessary respite from academic life much the same way that playing on the playground during recess may serve to enhance academic performance".

The major limit of our study is that we cannot track individuals over time. Therefore, all the effects that we have estimated are inter-individual effects, not intra-individual ones: we cannot really say whether one student raising or diminishing her time playing will get the reward or penalty we predicted. To this end, we need panel data. We hope that future international assessment tests will include a longitudinal part so that future researchers will be able to tackle this crucial issue.

References

Megan Glover Adams. Engaging 21st-century adolescents: Video games in the reading classroom. *English Journal*, pages 56–59, 2009.

Yann Algan, Pierre Cahuc, and Andrei Shleifer. Teaching practices and social capital. Technical report, National Bureau of Economic Research, 2011.

Craig Alan Anderson, Douglas A Gentile, and Katherine E Buckley. *Violent video game effects on children and adolescents*. Oxford University Press New York, 2007.

Robert J Barro and Jong-Wha Lee. International data on educational attainment: updates and implications. *Oxford Economic Papers*, 53(3):541–563, 2001.

Fran C Blumberg and Elizabeth Altschuler. From the playroom to the classroom: Children's views of video game play and academic learning. *Child Development Perspectives*, 5(2):99–103, 2011.

Walter R Boot, Daniel P Blakely, and Daniel J Simons. Do action video games improve perception and cognition? *Frontiers in Psychology*, 2, 2011.

Steve M Dorman. Video and computer games: Effect on children and implications for health education. *Journal of School Health*, 67(4):133–138, 1997.

Matthew WG Dye, C Shawn Green, and Daphne Bavelier. Increasing speed of processing with action video games. *Current Directions in Psychological Science*, 18(6):321–326, 2009.

Jing Feng, Ian Spence, and Jay Pratt. Playing an action video game reduces gender differences in spatial cognition. *Psychological science*, 18(10):850–855, 2007.

Nicole M Fortin, Philip Oreopoulos, and Shelley Phipps. Leaving boys behind: Gender disparities in high academic achievement. *UBC, University of Toronto, Dalhousie University, and CIFAR*, 2012.

Douglas A Gentile. The multiple dimensions of video game effects. *Child development perspectives*, 5 (2):75–81, 2011.

Christian Gourieroux, Alain Monfort, Éric Renault, and Alain Trognon. Résidus généralisés résidus simulés et leur utilisation dans les modéles non linéaires. In *Annales de l'inséé*, pages 71–96. JSTOR, 1985.

C Shawn Green and Daphne Bavelier. Action-video-game experience alters the spatial resolution of vision. *Psychological science*, 18(1):88–94, 2007.

Bradley S Greenberg, John Sherry, Kenneth Lachlan, Kristen Lucas, and Amanda Holmstrom. Orientations to video games among gender and age groups. *Simulation & Gaming*, 41(2):238–259, 2010.

Eric A Hanushek and Ludger Woessmann. The economics of international differences in educational achievement. Technical report, National Bureau of Economic Research, 2010.

Jong-Wha Lee and Robert J Barro. Schooling quality in a cross–section of countries. *Economica*, 68 (272):465–488, 2001.

Michael O Martin, Ina VS Mullis, and Ann M Kennedy. *Progress in International Reading Literacy Study (PIRLS): PIRLS 2006 Technical Report.* ERIC, 2007.

OECD. *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I).* OECD Publishing, 2010.

John F Olson, Michael O Martin, Ina VS Mullis, and Alka Arora. *TIMSS 2007 technical report.* IEA TIMSS & PIRLS, 2008.

Mikki H Phan, Jo R Jardina, Sloane Hoyle, and Barbara S Chaparro. Examining the role of gender in video game usage, preference, and behavior. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 56, pages 1496–1500. SAGE Publications, 2012.

Sig J Prais. Cautions on oecd's recent educational survey (pisa). *Oxford Review of Education*, 29(2): 139–163, 2003.

Iris C Rotberg. Myths about test score comparisons. *Science*, 270(5241):1446–1448, 1995.

Vincenzo Spiezia. Does computer use increase educational achievements? student-level evidence from pisa. *OECD Journal: Economic Studies*, 2010:22, 2010.

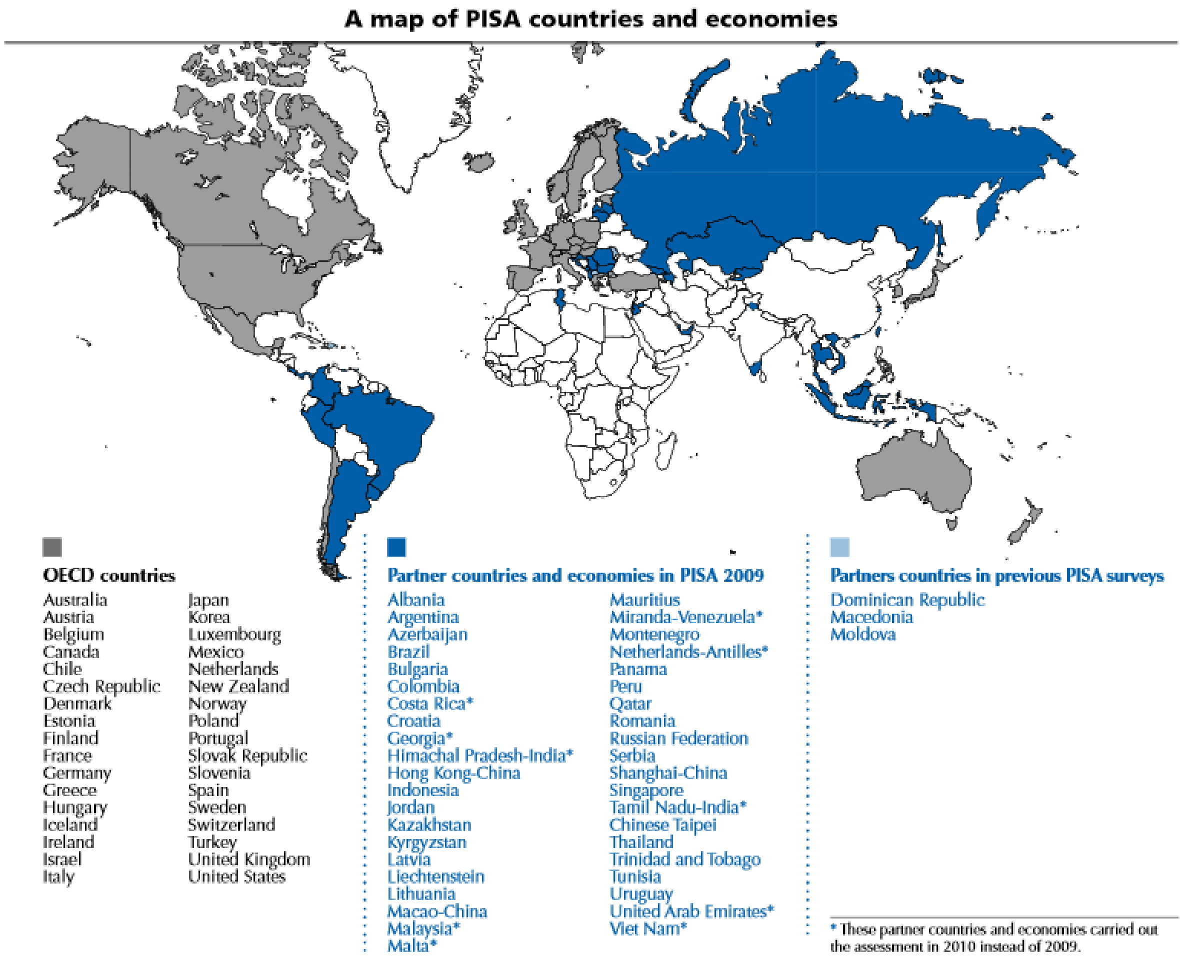
Aysegul Yolga Tahiroglu, Gonca Gul Celik, Ayse Avci, Gulsah Seydaoglu, Mehtap Uzel, and Handan Altunbas. Short-term effects of playing computer games on attention. *Journal of Attention Disorders*, 13(6):668–676, 2010.

Elizabeth A Vandewater, Mi-suk Shim, and Allison G Caplovitz. Linking obesity and activity level with children's television and video game use. *Journal of adolescence*, 27(1):71–85, 2004.

Francis Vella. Estimating models with sample selection bias: a survey. *Journal of Human Resources*, 33:1:127–169, 1998.

Appendix

Figure 5: Countries and economies participating in PISA 2009



Source: OECD [2010, page 19]

Table 7: Country names and abbreviations

	Country or entity	ISO		Country or entity	ISO
1	Alabama, US	UAL	55	Japan	JPN
2	Andalusia, Spain	EAN	56	Jordan	JOR
3	Abu Dhabi, UAE	AAD	57	Kazakhstan	KAZ
4	Algeria	DZA	58	Korea,Rep.of	KOR
5	Argentina	ARG	59	Kuwait	KWT
6	Armenia	ARM	60	Latvia	LVA
7	Australia	AUS	61	Lebanon	LBN
8	Austria	AUT	62	Lithuania	LTU

	Country or entity	ISO		Country or entity	ISO
9	Azerbaijan	AZE	63	Luxembourg	LUX
10	Bahrain	BHR	64	Macedonia, Rep. of	MKD
11	Belgium (Flemish)	BFL	65	Malaysia	MYS
12	Belgium (French)	BFR	66	Malta	MLT
13	Belize	BLZ	67	Maltese-Malta	MLN
14	Bosnia and Herzegovina	BIH	68	Massachusetts, US	UMA
15	Botswana	BWA	69	Minnesota, US	UMN
16	Bulgaria	BGR	70	Moldova, Rep. of	MDA
17	California, US	UCA	71	Mongolia	MNG
18	Canada	CAN	72	Morocco	MAR
19	Canada, Alberta	CAB	73	Morocco 6	MA6
20	Canada, British Columbia	CBC	74	Netherlands	NLD
21	Canada, Nova Scotia	CNS	75	New Zealand	NZL
22	Canada, Ontario	COT	76	North Carolina, US	UNC
23	Canada, Quebec	CQU	77	Northern Ireland	NIR
24	Chile	CHL	78	Norway	NOR
25	Chinese Taipei	TWN	79	Norway (5th grade)	NO5
26	Colombia	COL	80	Oman	OMN
27	Colorado, US	UCO	81	Palestinian	PSE
28	Connecticut, US	UCT	82	Poland	POL
29	Croatia	HRV	83	Portugal	PRT
30	Cyprus	CYP	84	Qatar	QAT
31	Czech Republic	CZE	85	Romania	ROM
32	Denmark	DNK	86	Russian Federation	RUS
33	Dubai,UAE	ADU	87	Saudi Arabia	SAU
34	Egypt	EGY	88	Scotland	SCO
35	El Salavador	SLV	89	Serbia	SRB
36	England	ENG	90	Singapore	SGP
37	Finland	FIN	91	Slovak Republic	SVK
38	Florida, US	UFL	92	Slovenia	SVN
39	France	FRA	93	South Africa	ZAF
40	Georgia	GEO	94	Spain	ESP
41	Germany	DEU	95	Syrian Arab Rep.	SYR
42	Ghana	GHA	96	Sweden	SWE
43	Greece	GRC	97	Thailand	THA
44	Honduras	HND	98	Trinidad and Tobago	TTO
45	Hong Kong SAR	HKG	99	Tunisia	TUN
46	Hungary	HUN	100	Turkey	TUR
47	Iceland	ISL	101	Ukraine	UKR
48	Iceland (5th grade)	IS5	102	United Arab Emirates	ARE
49	Indiana, US	UIN	103	United States	USA
50	Indonesia	IDN	104	Yemen	YEM

	Country or entity	ISO		Country or entity	ISO
51	Iran, Islamic Rep. of	IRN	105	Yemen 6	YE6
52	Ireland	IRL	106	Basque Country, Spain	BSQ
53	Israel	ISR	107	Massachusetts, US	UMA
54	Italy	ITA	108	Minnesota, US	UMN

Table 8: List of control variables in the models

Variable name	Variable description	PISA	TIMSS	PIRLS
AGE	Pupil's age	x	x	x
GIRL	Pupil's gender	x	x	x
IMMIG12	Immigrant 1st or 2nd generation	x	x	x
OWNCOMP	Own computer at home	x	x	x
OWNINT	Own Internet connection at home	x	x	(x)[8]
CONSOLE	Own a videogame console at home	x		
WEALTH	Index of wealth	x	(x)	x
HEDRES	Index of home educational resources	x	(x)	x
HPEL	Highest parents' education level	x	x	x
DISCLIMA	Index of pupil's perception of disciplinary climate in class	x		
SAFE	Index of pupil's perception of safety at school		x	x
STRATIO	Student/teacher ratio at school	x		
SCMATEDU	Availability of educational resources at school	x	x	x
ECONDIS25	More than 25% of pupils at school are economically disadvantaged		x	x
PUBLIC	School is public	x		
TOWN	School is located in town with more than 15,000 inhabitants	x	x	x

[8](x) means that the variable was not directly available in the dataset, but was constructed manually with the help of several variables.

Figure 6: Countries participating in TIMSS 2007

Source: http://timss.bc.edu/

Table 9: Results from step 1 model, PISA 2009 data. Ordered probit explaining the variable "Playing one-player computer games"

	Dep. var.: POPCG	
	point estimate	*s.e.*
AGE	-0.0798***	(-5.04)
WEALTH	0.0149**	(2.71)
HEDRES	0.0812***	(15.23)
HPEL	-0.00664	(-1.92)
DISCLIMA	-0.0828***	(-15.72)
STRATIO	-0.00255*	(-2.56)
SCMATEDU	0.00791	(1.15)
GIRL	-0.626***	(-65.92)
IMMIG12	0.0238	(1.33)
OWNCOMP	0.989***	(49.77)
OWNINT	-0.0899***	(-5.38)
CONSOLE	0.301***	(29.90)

	point estimate	*s.e.*
PUBLIC	0.0493^{**}	(2.65)
TOWN	-0.0109	(-0.81)
Country fixed effects	Yes	
Cut 1	-1.144^{***}	(-4.41)
Cut 2	-0.463	(-1.78)
Cut 3	0.346	(1.33)
N	266,075	
R^2	0.08	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Summary of the results: number of countries where the coefficient associated with each level of the variable PCG is significantly positive, not significantly different from zero, and significantly negative

PISA

		One-player games			Collaborative online games		
	Coefficient	1 or 2/month	1 or 2/week	Every day	1 or 2/month	1 or 2/week	Every day
Reading	Positive	11	9	7	9	11	17
	Insignificant	32	34	36	35	33	27
	Negative	1	1	1	0	0	0
Math	Positive	9	9	5	14	11	20
	Insignificant	35	35	39	30	33	24
	Negative	0	0	0	0	0	0
Science	Positive	10	5	3	7	12	16
	Insignificant	34	39	40	37	32	28
	Negative	0	0	1	0	0	0

TIMSS (8th grade) **PIRLS**

		TIMSS (8th grade): All computer games				PIRLS: All computer games			
	Coefficient	-1h/day	1-2h/day	2-4h/day	+4h/day	-1/day	1-3h/day	3-5h/day	+5h/day
Reading	Positive					6	6	5	5
	Insignificant					34	34	35	35
	Negative					1	1	1	1
Math	Positive	11	11	11	7				
	Insignificant	37	37	37	41				
	Negative	0	0	0	0				
Science	Positive	12	12	11	11				
	Insignificant	36	36	37	37				
	Negative	0	0	0	0				

Figure 7: Countries participating in PIRLS 2006

Source: http://timss.bc.edu/

Figure 8: Scores and computer games (PCCG)

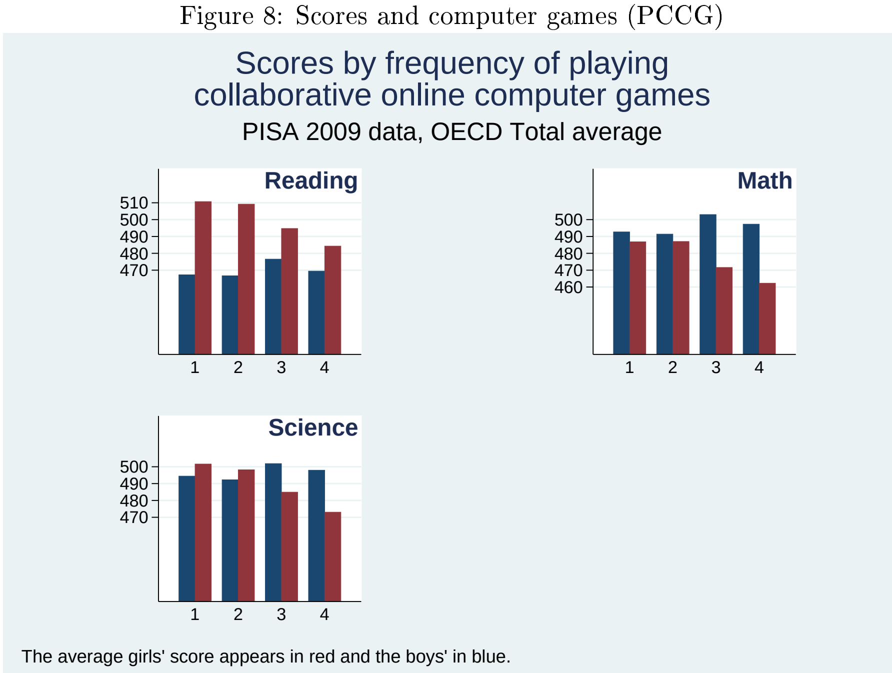


Figure 9: Reading scores and computer games (PCCG) by country

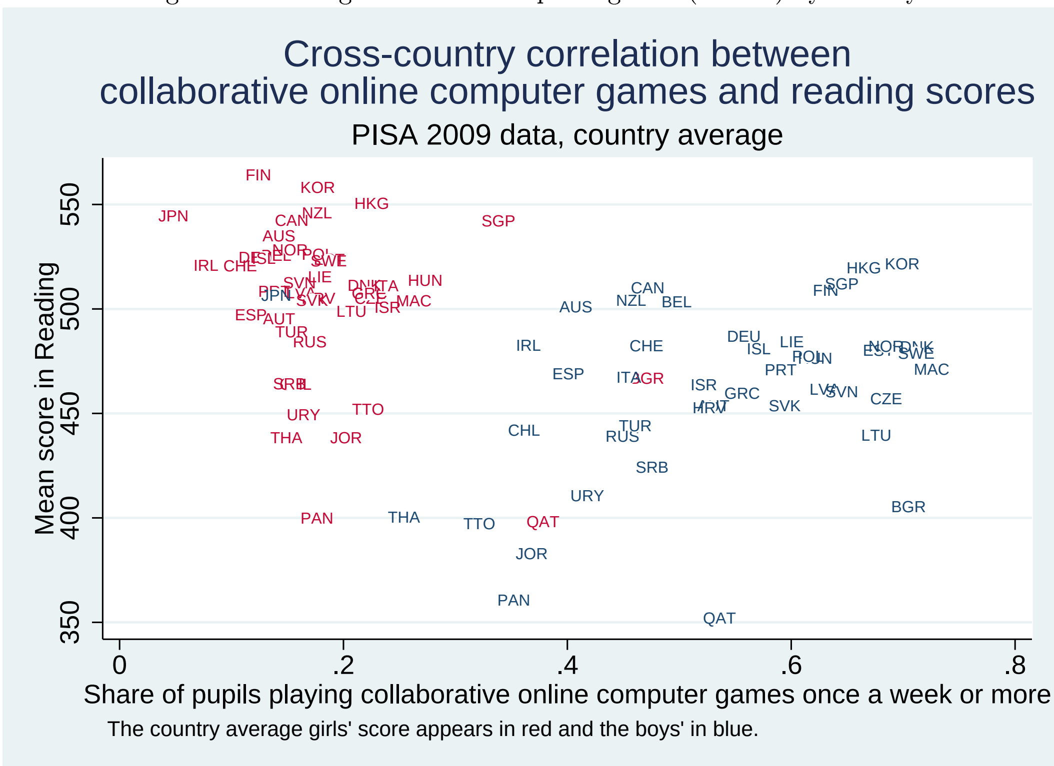


Figure 10: Comparison of gender differences in reading between 2000 and 2009

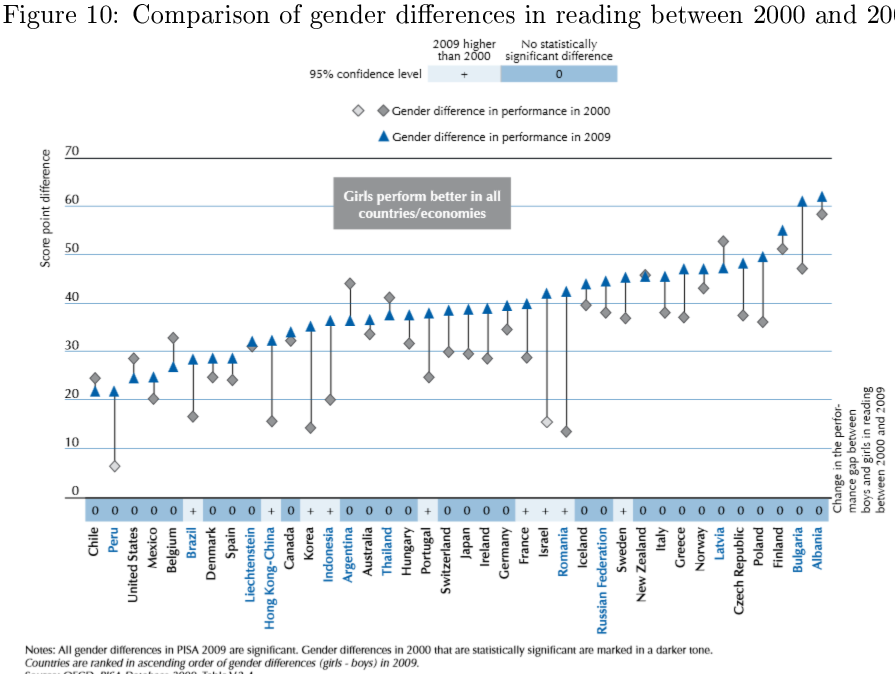


Figure 11: PISA - Playing one-player computer games: impact on science score, by country