Revista de Economía Aplicada E A Número 71 (vol. XXIV), 2016, págs. 93 a 120

MIND THE GAP: ANALYSING THE FACTORS BEHIND THE GAP IN STUDENTS' PERFORMANCE BETWEEN PENCIL AND COMPUTER BASED ASSESSMENT METHODS*

ÓSCAR MARCENARO-GUTIÉRREZ
LUIS ALEJANDRO LÓPEZ-AGUDO
Universidad de Málaga

The implementation of computer-based assessment (CBA) in PISA and the OECD's intention of relying only on this from 2015 suggests the need to evaluate to what extent it provides us with the same information as traditional paper and pencil assessment (PPA) conducted until 2012. Our results show that there is a significant gap between PPA and CBA for Spanish students that can be explained by factors such as availability and access to ICT or students' socioeconomic and cultural status. This leads us to conclude that educational policy reforms based on PISA scores and on any of these factors which affect them, as well as the comparison of these results along time, should be made in light of the specific assessment method used as reference.

Key words: computer-based assessment, paper and pencil assessment, gap.

JEL Classification: I20, I21.

The development of assessment indicators of students' academic performance is a key issue to be able to implement improvements in education systems [Battauz *et al*. (2011), Marcenaro and Vignoles (2015)]. The results from some type of students' assessment are often used as a one-dimensional measurement to characterise the product of the learning process. Paper and pencil (PPA) have been the basic elements for filling out these assessment tests in different subjects, either through exams held regularly as part of the subjects programme or through standardised tests of the kind developed by the Programme for International

(*) We are grateful for very helpful comments from two anonymous referees. This work has been partly supported by the *Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía* (PAI group SEJ-532 and Excellence research group SEJ-2727); by the *Ministerio de Economía y Competitividad* (Research Project ECO2014-56397-P) and scholarship FPU2014 04518 of the *Ministerio de Educación, Cultura y Deporte*. Luis Alejandro López-Agudo also acknowledge the training received from the University of Malaga PhD Programme in Economy and Business [*Programa de Doctorado en Economía y Empresa de la Universidad de Malaga*].

Student Assessment (PISA), among others. But in a digital era, with an unstoppable and fast extension of ICT –particularly among youngsters[1]–, very diverse assessments based on increasingly sophisticated technological media are gaining ground. In fact in PISA 2012 the assessment of students' performance was conducted by the traditional format, through written tests, as well as computer based assessments (CBA) for a restricted sample, what gives us the first opportunity to evaluate the factors behind the potential differences between the students' results in PPA and CBA in the Spanish education system.

The literature on the differences between PPA and CBA relating to other countries is limited and far from conclusive. Perhaps the most cited results are those of Wallace and Clariana (2002), who allude to the existence of a significant gap in favour of CBA for the group of students with higher academic performance –also highlighted by Bennett *et al*. (2008)–. However, in general, the existing evidence seems to opt for the superiority of PPA over CBA [Sim and Horton (2005), Jeong (2014)].

It is worth mentioning other contributions on the potential impact on students' performance of the availability of ICT in schools as, e.g., Cabras and Tena (2013), who found some evidence of its positive effect, while others as Fuchs and Woessman (2004) did not. In the case of the use of computers at home, Harrison *et al*. (2002) conclude that it is strongly correlated with higher academic achievement, while Valentine *et al*. (2005) indicate that it is positive only when used for educational purposes and that out-of-school use of ICT is negative[2].

Bearing in mind the previous literature, this paper is aimed at disentangling the degree to which "traditional" factors –thought to be significant determinants of students' performance–, e.g., socio-economic and cultural status, as well as "non-traditional" factors –which might evolve faster in this digital era–, e.g., availability and access to ICT, contribute to the gap between PPA and CBA. To do that, students' characteristics and their learning environment at home and at school will be examined. The importance of this issue can be seen in the increasing recognition by countries of PISA results for monitoring their performance and equity [Hopkins *et al*. (2008)]. Thus, in the event that the detected differences between PPA and CBA may be explained significantly by differences in the previous stated factors, policy reforms designed for increasing equity and optimizing investments into the education system –in a context of financial constraints– should account for this gap when using PISA as best-practice model [Breakspear (2012)]. It is also important to underline that the analyses in our study will be performed

(1) According to EUROSTAT (2014), Spain achieved a high level of ICT development in 2013, as 70% of the households had access to the Internet (9% under the UE-28). Additionally, according to INE (2014), 74.8% of Spanish households had a computer of any kind and the use of ICT by children (10 to 15 years old) was, in general, very high: 93.8% used a computer and 92% the Internet, and gender differences were not significative.

(2) Mediavilla and Escardíbul (2014) found, for the case of Spain, that ICT had more influence on mathematics than in other competences; they also detected a positive relationship between computer use in an early age and results in CBA; although they did not analyse the effect on the PPA-CBA gap. Likewise the meta-analysis of Sosa *et al*. (2011) proposed a random effects model using 45 studies evaluating the differences between students who use computer-based tools in their statistics courses and those who receive only face-to-face classes. They found that the first one obtained higher achievement than the former.

by splitting the sample into repeaters and non-repeaters[3], as in Jimerson (1999), in order to avoid the repeaters' specific characteristics[4] biasing the results.

Our estimates allow us to conclude, on the one hand, that there is a significant gap between both modes of assessment which can be explained by factors such as access and availability of ICT; thus investments into education should be allocated bearing in mind the relative contribution of ICT resources to the new assessment method. On the other hand, our empirical results highlight that the substitution of PPA by CBA would shorten the distance with respect to the students with lower socio-economic background, what is very important in terms of equity. The relevance of our conclusions in this subject can be found in the emphasis placed in PISA results –in terms of the implementation of education policy reforms–, what has been well documented in countries like, e.g., Norway, whose trend of decreasing scores triggered many policy reforms, mainly based on increasing the annual expenditure per student at higher rate (11% from 2005 to 2010) than enrolment, which increased only 2% along that period [Baird *et al*. (2011), OECD (2013)]. Additionally, to the extent that CBA will be regarded as reference level to evaluate the performance of educational systems from 2015 onwards, the presence of a significant gap –between PPA and CBA– should be taken into account when making comparisons of PISA results for any country before and after 2015.

The rest of this research is structured as follows: in the second section the data contained in PISA 2012 will be described, as a starting point for briefly describing, in the third section, the methodology. In the fourth section the estimate of the established models will be addressed. This paper ends in section five, which presents the main conclusions and provides a reflection by the authors, based on the results.

1. Data

The information contained in PISA 2012, regarding the sample for Spain, makes reference to a total of 25,313 students spread over 902 secondary schools. However, standardised CBA-type tests were only completed by a random subsample (40%) of the selected students per school (10,175 students enrolled in 368 schools). Table 1 (Appendix) provides statistics, on the one hand, of the subsample of students who only performed PPA exams and, on the other hand, of the subsample of students who took both PPA and CBA. From its comparison we infer that sample selection should not introduce any bias in the performed estimates.

(3) García-Pérez *et al*. (2014) pointed out that the inclusion of repeater and non-repeater students together in the same specification –even when controlling by this feature– could cause problems of endogeneity in the estimation. Consequently they suggested the use of instrumental variables (namely, the quarter of birth). However, in our study the dependent variable is the gap (PPA-CBA), thus the presence of endogeneity is not that clear. Despite of this, as robustness check we did perform an instrumental variable approach using the same instrument as in García-Pérez *et al*. (2014). The results showed that the instrument was not significant.

(4) Many authors have recently deepened into the question of grade retention and its negative effect on students' achievement, e.g. Cordero *et al*. (2013), Calero *et al*. (2010) and Dolton *et al*. (2003). Carabaña (2013) established a significant correlation between students' actual scores and previous episodes of repetition. In our dataset we observe a similar pattern –Tables 2.1 and 2.2 (Appendix)–.

The use of information relating to some potential highly relevant variables for the teaching-learning process such as, e.g., time spent by students doing schoolwork with the help of their parents, guided work outside the school, hours of study with the computer, etc., further reduce the sample size, as they are only available for two-thirds of the total sample (due to the structure of the PISA 2012 questionnaire). Furthermore, we have to bear in mind that approximately one-third of the students have repeated. Regarding to the design of the CBA, it gathered reading, mathematics and problem-solving items which were allocated to 24 booklets [MECD (2013)]. Firstly PPA was done in two hours of reading, maths and science and then CBA (for reading, maths and problem solving) in forty minutes[5].

Bivariate descriptive analyses performed at a stage prior to the estimate have revealed possible sources for explaining the variations observed in the gap between PPA and CBA tests. This gap has been calculated for each student as the difference between the "average" plausible value –obtained according to the PISA guidelines– of the results on PPA minus the "average" plausible value on CBA. Broadly speaking, as statistics of Tables 2.1 and 2.2 (Appendix) show, the scores of CBA are lower than those obtained in PPA[6]. However, as can be seen in Figures 1.1 and 1.2 –which show the relationship between PPA and CBA scores, and also PPA-CBA gap with CBA scores–, the profile of this gap shows a decreasing gradient in relation to the scores obtained in CBA; i.e. the gap is reduced as we move towards the "average" student in academic performance, becoming negative –i.e. a higher score in CBA than in PPA– for the most gifted students –those to the right–, who seem to benefit more from the use of CBA.

In Figure 2.1 we show the average value of the PPA-CBA gap according to the degree of technological development of the school (as expressed by their students through the ICT index). Competence in maths shows a clear trend in the reduction of the gap as the technological resources of the school increase, even turning positive towards the CBA scores when the opinions of repeaters are collected.

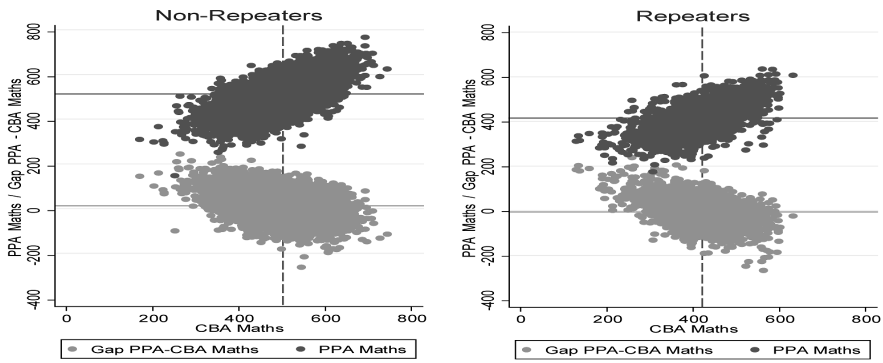
In order to complement the analysis in Figure 2.1 we constructed Figure 2.2, whose results may seem counter-intuitive. Together with the statistics reported in Tables 2.1 and 2.2 (Appendix) it can be inferred that a late start in using a computer is closely correlated with a low score in PPA and CBA, presenting CBA higher scores than PPA in maths. At first the opposite may have been expected, assuming that an early start in the use of computers is associated with a greater skill with them. Consequently, the effect on the gap would be better explained by the lesser ability in the standardised PPA tests of the most disadvantaged students than by a greater skill in the use of computers, as scores in CBA also decrease.

The performance gap between students from favourable socio-economic and cultural context and those from disadvantaged ones (high and low ESCS index, respectively) is consolidated at positive levels –for non-repeaters– (Tables 2.1 and 2.2), just as speaking a different language at home and at school (diglossia) produces the widening of the gap in reading in favour of PPA, and also immigrant status. Students'

(5) There is evidence [Borghans and Schils (2012)] that PISA scores decay when advancing in the position of the question, although to the best of our knowledge there is no specific evidence that this could be the case when students perform PPA right before CBA.

(6) Precisely, 501 in CBA compared to 513.1 in PPA, in maths, and 497.7 compared to 517.8, in reading, for non-repeaters.

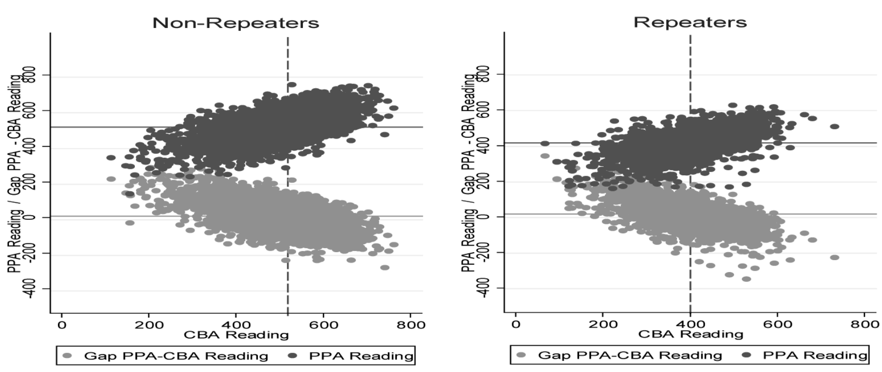
Figure 1.1: Gap between PPA and CBA in maths competence



Note: The horizontal lines drawn from the vertical axis represent, respectively, the average scores in PPA Mathematics (dark grey) and in PPA-CBA Mathematics gap (light grey). The dashed vertical line drawn from the horizontal axis represents the average scores in CBA Mathematics.

Source: Authors' own calculation from PISA 2012.

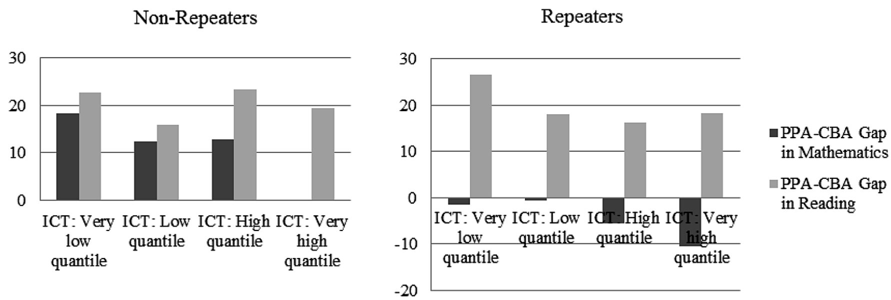
Figure 1.2: Gap between PPA and CBA in reading competence



Note: The horizontal lines drawn from the vertical axis represent, respectively, the average scores in PPA Reading (dark grey) and in PPA-CBA Reading gap (light grey). The dashed vertical line drawn from the horizontal axis represents the average scores in CBA Reading.

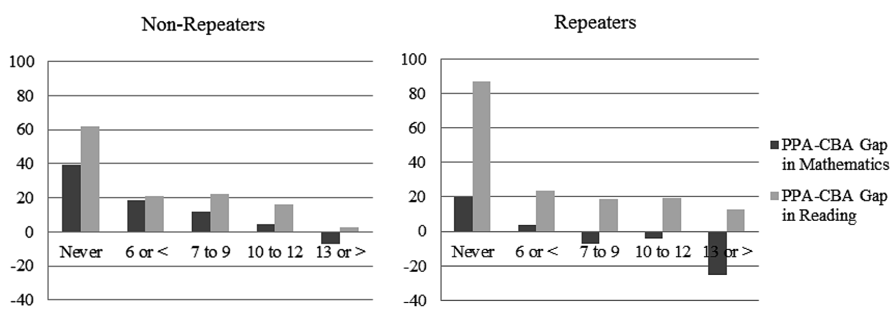
Source: Authors' own calculation from PISA 2012.

Figure 2.1: DISTRIBUTION OF THE GAP BETWEEN PPA-CBA ACCORDING TO THE DEGREE OF ICT RESOURCES REVEALED BY STUDENTS REGARDING THEIR SCHOOL



Source: Authors' own calculation from PISA 2012.

Figure 2.2: DISTRIBUTION OF THE GAP BETWEEN PPA-CBA ACCORDING TO THE AGE AT WHICH STUDENTS BEGAN USING A COMPUTER FOR THE FIRST TIME



Source: Authors' own calculation from PISA 2012.

complete absences[7] go hand in hand with a deterioration of the scores in both assessments, being higher in PPA than in CBA. In addition, repeater students who have a tablet available at school and use it present a high increase in the maths' gap in favour of CBA, while they have a positive gap in reading.

In order to evaluate to what extent the results collected from the bivariate analysis are maintained when determined simultaneously by the group of variables mentioned in the results' section, we present a set of estimates based on the methodology discussed briefly in the next section.

(7) Data for this variable are available for the entire sample, not just for two-thirds as occurred for the variables included in the alternative specifications displayed in Tables 5, 6 and 7 (Appendix) commented in section 4.

2. METHODOLOGY

In PISA a sample of approximately 35 students per school is used, so some "homogeneity" of the characteristics of students can be expected as a result of belonging to the same school. This is true if attendance at one school or another is determined largely by the distance between the home and the school, so we are likely to observe "common" socioeconomic and cultural patterns among students of the same school, and different ones between schools. This "between" and "within" school double heterogeneity cannot be accurately captured using the simple techniques of linear regression (Ordinary Least Squares), since this methodological approach will provide an inaccurate quantification of the standard errors of the parameters if the students belonging to the same school show similar values in the school variables. Because of that the average correlation between variables referred to students in the same school will be higher than that existing between students from different schools [Hox (1998)]; this will lead us to underestimate the standard deviations of the parameters of the estimated regression and, therefore, to consider the correlation between the explained variable and any of the explanatory variables as significant when sometimes it should not.

Consequently, an alternative multilevel analysis is normally used [Raudenbush and Bryk (2002)], in which the contribution to the variance (total heterogeneity) of the characteristics of the students is estimated, as a first level, and that added according to school, as a second level. In algebraic terms we start with the equation [1]:

$$Y_{ij} = \alpha_j + \beta_1 X_{kij} + ... + \beta_p X_{pij} + \varepsilon_{ij} \quad [1]$$

where Y_{ij} denotes the difference between the "average" plausible value of the results of the PPA test minus the "average" plausible value of the CBA test for a student "" who attends school "j"; α_j is a parameter of level that reflects the differential effect on Y_{ij} of each school, i.e., it is a proxy of the "quality" of the school; X_{kij} represents student's k characteristic from a total of k = 1, ... p, and ε_{ij} is the idiosyncratic error term

Equation [1] assumes that the intercept of the regression is the same for all transverse units. The random effects model suggests that each school has a different level parameter (interception term). So that:

$$\alpha_j = \alpha + u_j \quad [2]$$

Substituting [2] in [1] we obtain:

$$Y_{ij} = \alpha + \beta_1 X_{1ij} + ... + \beta_p X_{pij} + u_j + \varepsilon_{ij} \quad [3]$$

This is the expression of the random effects model. If we go a step further and add the effect on the scores (Y_{ij}) of the variables considered at the school level (Z_{sj}) –for s = 1, ..., q– to the expression [3] we come to the following model:

$$Y_{ij} = \alpha + \beta_1 X_{1ij} + ... + \beta_p X_{pij} + \delta_1 Z_{1j} + ... + \delta_{qj} Z_{qj} + u_j + \varepsilon_{ij} \quad [4]$$

which represents the multilevel model using random effects. This methodology has become widespread in the analysis of PISA data, especially in the reports accompanying the latest editions [see, for example, García-Montalvo (2013)].

3. Results

We proceeded to estimate the coefficients of the multilevel analysis for the sample of students with scores in both PPA and CBA; the results are provided in Tables 3.1 and 3.2 (Appendix) for maths and reading, respectively[8]. The ICT variables selected for the base case have been the availability and use –by the students– of a tablet in the school and the age at which the student began to use the computer[9]. In order to test for the potential existence of multicollinearity between these ICT variables, we performed several specifications where they were omitted alternatively. The results showed that the sign and significance of the correlation with the gap remained the same for the alternative specifications, so there is no evidence of high multicollinearity between the ICT variables[10]. Thus, they have been included in the analysis reported.

The second part of the estimates presented in Tables 3.1 to 3.2 (Appendix) shows the effect of second-level variables, i.e. variables whose effect is aggregated at a school level –starting from "type of school"–. In relation to this, we have presented the coefficients of intra-class correlation (ICC) in Table 4 (Appendix). The ICC value depends largely on the variability of the observed values: the more homogeneous the sample, the lower the value of the ICC tends to be. In the case of the gaps in PPA-CBA scores, they show an ICC which is significant and not at all negligible for both non-repeaters and repeaters, indicating that differences between schools affect the gaps in the maths and reading tests, consequently supporting the use of multilevel regressions[11].

However, if we look at the variables aggregated at school level, few of them stand out as significant in explaining the PPA-CBA gap. This is the case, for example, of the student-teacher ratio variable, which contributes to turn the gap slightly more positive. In the case of the average years of schooling of the father, it widens the gap between PPA and CBA in favour of the former for both competences, while average years of schooling of the mother increases the gap in favour of CBA for the reading competence –affecting this only to worse performers (i.e. repeaters)–. The use of ICT in the school also contributes to increase the gap in reading for non-repeaters, turning it more positive; likewise, the effect of the proportion of immigrants at the school in the gap of reading is significant. Thus, the gap in mathematics is not affected by any of these two variables.

Turning to the variables measured at student level, central to this paper are some variables which reflect the level of ICT development of the school and the student. The availability and/or use of a tablet device in the school leads us to assert that students who claim to have a tablet available for their use in school achieve a significantly lower

(8) As an alternative to the variable which represents the number of days that the student missed a whole class, we estimated a different specification where this variable was substituted by hours of study –which is only available for 2/3 of the total sample– due to its relevance in the economics of education literature. Results are not reported following an anonymous referee advice (available upon request).

(9) Many alternative specifications were defined in order to select the most representative ICT variables.

(10) Additionally, a variance inflation factor (VIF) analysis has been performed in order to check for the existence of multicollinearity in the estimations. The results obtained for tablet variables showed a VIF of approximately "1", which denotes the absence of multicollinearity problems caused by these variables.

(11) The results from research on previous editions of PISA show a relatively small contribution of schools (about 20%) to the variation of students' scores.

score in PPA –as compared to CBA– than those who do not, and this difference is more substantial when they use it actively for maths comprehension. Thus, the reduction of the gap between PPA and CBA scores is highly determined by this availability. It is also interesting to note that this effect is diluted if the availability of the device refers to the student's home[12]. One could probably think that the presence of tablets in the school might respond to the type of school, classified by the total amount of public funding received. However, the descriptive statistics showed that this is not the case[13]. Another relevant variable to consider, within this set, is that concerning the starting age of the use of computers; the estimated coefficients imply that the earlier the students begin to use computers, the greater the gap in favour of PPA in maths for non-repeaters, while in reading competence the same effect can be appreciated for repeaters.

A key variable in the analyses of the educational production function is the socioeconomic and cultural status (ESCS) of the students, as it explains a substantial percentage of the variance observed in PPA [see, e.g., Cordero *et al*. (2014)]. At first we run the estimations using different alternative proxies to this status, such as the highest level of education reached by the father and the mother, number of books at home, or the possession of specific books of humanities; none of them showed substantially different effects when included in alternative specifications, being highly correlated with the ESCS index (which PISA provides as a derived variable), so the latter was used to make the estimations more parsimonious. Despite the relevance of the effect of the ESCS index on the PPA scores, when this index is included in the form of quartiles to collect possible non-linearities, its influence is weak in maths competence. However, high levels of ESCS favour PPA results in reading over CBA among non-repeaters.

Missing a whole class contributed to widen the gap in favour of CBA, except for the case of repeaters in mathematics. Many variables were also employed in alternative specifications (concretely, hours of study with the computer, guided work hours outside the school and help time provided by parents or other family member), which were answered only by two thirds of the sample[14]. Firstly, we studied the number of hours –using discrete variables– which the teenager claims to devote each week to study using a computer (Table 5, Appendix); their coefficients were not significant in reading, but the gap is drastically reduced in maths (approximately in 30 points) for non-repeaters who spend 7 or more hours compared to those who do not spend any, in line with what is observed for the availability of tablets at school. Furthermore, when estimating the impact of this last variable on the CBA scores we ob-

(12) Table available upon request.

(13) The proportion of students who stated that they have tablet devices at their schools is similar in private, semi-private and public schools.

(14) We compared whether the subset of students of this subsample –from among those who were selected to be assessed either using PPA as well as CBA– showed any differences in their individual characteristics compared to those who were not in it –also from that who answered both PPA and CBA– and no statistically significant differences were found, what guarantees the randomness of the subsample.

(15) Another argument, which seems less plausible, is that students with lower average productivity, measured by the relationship between effort in terms of time commitment and results, use the computer to try to compensate for their slow learning, what would explain the concentration of students who use the computer for academic tasks among the least advantaged in terms of scores.

(16) http://www2.ed.gov/programs/racetothetop/index.html.

tain coefficients which are negative and increasing with the number of hours[15], so we can conclude that there is a selection bias in the use of computers for studying among "low performance" students, since they get lesser scores in both types of assessments, reducing their PPA results more than their CBA results. To sum up, as far as the main objective of this research is concerned, the help of computers for study does not make students more competent in general, but using CBA could mean that "less competent" students "artificially" cut back the disadvantage that they show in PPA, without implying a truly better performance.

Regarding the "non-independent" learning (i.e. guided academic work performed outside the school), the number of hours devoted to this turns more negative the maths gap for non-repeaters (it is negligible in reading) –Table 6, Appendix–. Finally, time spent with the help of parents or other family member doing schoolwork at home shows an insignificant effect on the PPA-CBA gap –Table 7, Appendix–. In Marcenaro (2013) it is argued that the potential adverse influence on PPA could be explained, on the one hand, by the limited confidence of parents in their children given their previous results. This could be leading parents to help them in academic tasks outside school or, on the other hand, by a "passivity" which would arise in the students receiving such help, contributing to poorer results. This line of argument could also apply to CBA, which would result in the lack of effect on the gap. From what has been said it appears that, in general, "supervised" learning outside the school has enough impact to tip the balance in favour of one or the other mode of assessment.

Focusing on family structure (Tables 3.1 and 3.2, Appendix), non-repeater students who live in a single parent home obtain a more negative gap in maths than those who live with both parents. This could be explained by the more frequent use of the computer and the Internet of these pupils to keep themselves distracted, what can enhance their competences in their use [Notten *et al.* (2009)].

Regarding to gender, the PPA-CBA gap is reduced at a higher rate for women than for men when assessing the achievement in maths competence of the non-repeaters. Interestingly, despite that among repeater students boys show better outcomes in mathematics, the PPA-CBA gap is not significantly different.

The negative effect for immigrant status typically observed in the literature with respect to the results in the standardised PPA-type tests [Calero and Escardíbul (2013)] is maintained when the PPA-CBA gap in reading competence is analysed –for non-repeaters–, so that it may be concluded that it contributes even more to a reduction, compared to the native students, of the scores in PPA than in CBA. The opposite occurs in maths, a competence in which the gap is positively accentuated as a result, at least partially, of the greater proportion of immigrant students among the repeaters.

4. Conclusions

Over 90% of boys and girls (10 to 15 years old) use ICT on a daily base. This extension of ICT has reached the assessment methods, fostering the debate on the potential differences in results when performing the same assessment test in both a written and computer-based form. This is particularly relevant to the degree that assessment results are taken as indicators of students' academic performance. In this context PISA results have been invoked as a major justification for the implementation of education policy reforms (e.g., the ambitious "Race to the Top" program in USA)[16]. In fact, to the extent that CBA will be regarded as a reference level to evaluate the performance of educational systems from 2015 onwards, it is necessary to be aware of the changes in the influence of factors which affect the gap between PPA and CBA. This is particularly relevant in order to inform policy makers about the potential differences in the effects of the educational reforms based on such factors, as budgetary decisions regarding them could affect students' performance in a different way when the reference is PPA than in the case of CBA.

As evidence on this issue is not only limited but also far from conclusive, we have deepen into this subject by examining which students' characteristics and those of their learning environment may increase or reduce the gap between PPA and CBA. Among these factors special attention has been paid to the role of the availability and use of ICT in the school and at home to explain these potential gaps.

Focusing on the results obtained, the descriptive statistics showed feeble evidence since the gap was reduced and even reversed, placing the results in CBA above those of PPA for students with less relative digital literacy (approximated from the age when the use of computers starts). However, in other indicators, such as the perception of students on the available ICT resources at school, it showed the opposite behaviour. Despite the apparent lack of consistency of the results from the bivariate analysis, the multivariate conditional regression analysis has shed light on the matter. In particular, it could be interpreted that conducting the tests in a CBA format "favours" female students over their male classmates in helping to improve their relative position in the maths competence, in which they have historically shown a lower average performance than boys –measured in terms of PPA–. In addition, the advantage they already had in reading competence remained unchanged or even slightly increased.

No less important is that the substitution of PPA by the CBA would reduce the distance with respect to the students found in "more disadvantaged" environments (in terms of ESCS), which leads back to the hypothesis that differences between PPA and CBA go beyond the simple assessment mode effect. This idea also rises up since the more negative gap estimated for students who spend more hours studying with the help of a computer imply the existence of a selection bias among the "less gifted", which are those who get the greatest relative performance in CBA.

In summary, at least two issues can be inferred from the results reported in this investigation. Namely, on the one hand, using an assessment method based on ICT could mean that the "less gifted" students "artificially" reduce the relative disadvantage they show when performing PPA-type tests, without implying a better performance in the teaching-learning process. Consequently, educational policy initiatives taking the PISA assessment framework as a best-practice model should be carefully considered as the size of the effect of interventions on some groups of students might be conditioned by the assessment method. On the other hand, the comparability along time on the education performance of a country based on PISA, before and after 2015, also has to be made with caution as the size of the gap correlates significantly with "non-traditional" factors whose contribution to the education production function may evolve much faster than "traditional factors" in a digital era.

Appendix

Table 1: Descriptive Statistics for the subsamples by assessment mode

	Sample with only PPA			Sample with PPA and CBA		
Variables	N	Average	St. Dev.	N	Average	St. Dev.
Female	15,138	0.50	0.50	10,175	0.50	0.50
Male	15,138	0.50	0.50	10,175	0.50	0.50
Immigrant	14,895	0.10	0.31	9,929	0.09	0.29
Native	14,895	0.90	0.31	9,929	0.91	0.29
Diglossia: Yes	14,973	0.13	0.34	10,027	0.18	0.39
Diglossia: No	14,973	0.87	0.34	10,027	0.82	0.39
ESCS (Very low quartile)	15,074	0.25	0.43	10,047	0.25	0.43
ESCS (Low quartile)	15,074	0.25	0.43	10,047	0.25	0.43
ESCS (High quartile)	15,074	0.25	0.43	10,047	0.25	0.43
ESCS (Very high quartile)	15,074	0.25	0.43	10,047	0.25	0.43
Miss the whole class: Never	15,054	0.77	0.43	10,053	0.79	0.41
Miss the whole class: 1 to 2 times	15,054	0.20	0.40	10,053	0.18	0.38
Miss the whole class: 3 to 4 times	15,054	0.02	0.15	10,053	0.02	0.14
Miss the whole class: 5 or more times	15,054	0.01	0.11	10,053	0.01	0.10
Single parent home: Yes	14,306	0.10	0.30	9,548	0.10	0.30
Single parent home: No	14,306	0.90	0.30	9,548	0.90	0.30
Tablet in the school: Yes and they use it	14,627	0.04	0.19	9,631	0.04	0.20
Tablet in the school: Yes but they do not use it	14,627	0.03	0.18	9,631	0.03	0.18
Tablet in the school: No	14,627	0.93	0.25	9,631	0.93	0.26
Age when started to use the computer: Never	14,696	0.00	0.05	9,644	0.00	0.05
Age when started to use the computer: 6 or <	14,696	0.34	0.47	9,644	0.33	0.47
Age when started to use the computer: 7 – 9	14,696	0.45	0.50	9,644	0.46	0.50
Age when started to use the computer: 10 – 12	14,696	0.18	0.38	9,644	0.19	0.39
Age when started to use the computer: 13 or >	14,696	0.03	0.16	9,644	0.02	0.15
Private school	15,112	0.05	0.21	10,175	0.05	0.22
Semi-private school	15,112	0.29	0.45	10,175	0.40	0.49
Public school	15,112	0.66	0.47	10,175	0.55	0.50
ICT Quartile: Very low	14,720	0.30	0.46	9,691	0.28	0.45
ICT Quartile: Low	14,720	0.29	0.45	9,691	0.29	0.45
ICT Quartile: High	14,720	0.27	0.44	9,691	0.27	0.45
ICT Quartile: Very high	14,720	0.14	0.35	9,691	0.16	0.37

Source: Authors' own calculations from PISA 2012.

Table 2.1: Descriptive statistics for the characteristics of the non-repeater students

	PPA Maths		PPA Reading		CBA Maths		CBA Reading		PPA-CBA Maths		PPA-CBA Reading	
Variables	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.
Female	500.7	2.5	527	2.3	489.5	3.5	505.1	3.4	11.2	3.4	22	3.8
Male	526.4	2.9	507.8	3.2	513.4	3.5	489.8	5	13	3.5	18	4.9
Immigrant	468.8	6.2	480.5	6.7	448.4	6.4	449.4	8.3	20.4	6.1	31.1	7.7
Native	516.6	2.3	520.8	2.4	505.2	3.1	501.3	3.9	11.4	3.3	19.5	4.2
Diglossia: Yes	498.4	3.5	507.6	4.8	487.5	7.2	471.8	9.7	10.9	6.9	35.8	10.8
Diglossia: No	518	2.3	521.3	2.3	505.3	2.8	506.1	3.8	12.7	2.8	15.2	3.5
ESCS (Very low quartile)	484	3	493.5	3.2	476	4.3	474.4	5.5	8	4.2	19.1	5
ESCS (Low quartile)	507.8	2.6	512.7	2.7	498.1	3.9	487.9	5.5	9.7	3.3	24.7	5.6
ESCS (High quartile)	520.5	4.2	524.1	4	508.2	3.5	503.9	4.3	12.3	3.8	20.2	4.5
ESCS (Very high quartile)	547.4	2.9	547.1	4.2	527.9	4.3	530.9	4.6	19.5	4.3	16.2	5.3
Miss the whole class: Never	519.8	2.2	522.3	2.3	508	3.2	501.3	4	11.8	3.3	21.1	4.2
Miss the whole class: 1 to 2 times	492.2	3.3	504.5	3.7	476.6	4.3	485.4	4.6	15.6	3.7	19.1	4.9
Miss the whole class: 3 to 4 times	493.1	8.6	499.7	10.2	511.2	12.8	510.5	11.2	-18.1	13.8	-10.8	10.6
Miss the whole class: 5 or more times	454.4	15.5	455.4	20.7	452	12.4	448.2	19.3	2.5	13	7.2	9.9
Single parent home: Yes	507.8	4.7	514.2	4.6	506.2	5.9	493.6	6.5	1.5	4.7	20.6	5.5
Single parent home: No	514.3	2.2	517.9	2.5	501.3	3.2	497.2	3.9	13	3.3	20.7	4.2
Tablet in the school: Yes and they use it	475.1	11.6	462.5	11.5	478	11.9	440.4	20.5	-2.9	5.8	22.1	12.8
Tablet in the school: Yes but they do not use it	495.1	7.6	493.6	7.8	506.1	9.7	486.2	14	-11	9	7.4	13.5

Source: Authors' own calculations from PISA 2012 (the sample only includes students who performed the PPA and CBA tests).

Table 2.1: DESCRIPTIVE STATISTICS FOR THE CHARACTERISTICS OF THE NON-REPEATER STUDENTS (continuation)

	PPA Maths		PPA Reading		CBA Maths		CBA Reading		PPA-CBA Maths		PPA-CBA Reading	
Variables	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.
Tablet in the school: No	516	2.3	521.9	2.4	501.9	3.2	501.3	3.9	14	3.1	20.6	4.1
Age when started to use the computer: Never	508.6	29.3	471.7	28.8	469.5	45.1	410.1	48.3	39.1	32.1	61.6	50.5
Age when started to use the computer: 6 or <	525.2	2.9	528.5	3.5	506.5	3	507.6	4.5	18.7	3.4	20.9	5
Age when started to use the computer: 7-9	515	2.8	518.7	2.7	502.9	3.9	496.7	4.5	12	3.4	22	4.4
Age when started to use the computer: 10-12	495.1	3.3	507.5	3.6	490.5	4.3	491.6	5	4.6	4.4	15.9	4.5
Age when started to use the computer: 13 or >	458.9	9.5	470.6	7.5	465.8	8	468.1	9.1	-6.8	6.8	2.4	7.2
Private school	541	5	544.3	6.3	520.4	8.9	514.7	9.2	20.6	8.1	29.6	9.3
Semi-private school	525.6	4.7	526.7	5	511.5	5.9	513.3	7.5	14.1	6.2	13.4	9.3
Public school	503.3	3	509.9	3.1	493.3	4.2	487.9	6	10	3.8	22	5.3
ICT Quartile: Very low	512.8	3.9	514.8	3.8	494.5	4.5	492	5.5	18.3	4.4	22.8	5.4
ICT Quartile: Low	519.4	2.4	526.4	2.7	507	3.7	510.5	4.3	12.5	3.6	15.9	4
ICT Quartile: High	517.9	3.5	525.3	3	505	4.3	501.9	4.9	12.9	3.5	23.4	5
ICT Quartile: Very high	492.4	4.4	494.8	5.6	492.3	5.4	475.3	8.6	0.1	5	19.5	7.5
Total	513.1	2.3	517.8	2.5	501	3.2	497.7	4	12.1	3.2	20.1	4.2

Source: Authors' own calculations from PISA 2012 (the sample only includes students who performed the PPA and CBA tests).

Table 2.2: Descriptive statistics for the characteristics of the repeater students

	PPA Maths		PPA Reading		CBA Maths		CBA Reading		PPA-CBA Maths		PPA-CBA Reading	
Variables	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.
Female	402.2	2.7	433.6	3	412.8	4.1	414.3	6.3	-10.6	3.6	19.4	6.8
Male	424.7	2.9	409.8	3.8	426.9	3.4	390.1	5.5	-2.2	3.4	19.7	4.4
Immigrant	404.4	7.9	409.9	6.1	397.4	5	386.6	7.4	7	7.2	23.4	10
Native	417.2	2.1	422.2	3	426.7	3	402.4	5.6	-9.5	3.1	19.8	4.7
Diglossia: Yes	413.5	5.4	418.8	5.4	418.8	4.9	384	7.6	-5.3	5.8	34.8	8
Diglossia: No	415	2.4	420.5	3	420.9	3.2	405.3	6.3	-5.9	3.2	15.2	5.7
ESCS (Very low quartile)	399.2	3.3	408.1	3.9	407.9	4.1	387.5	5.9	-8.7	3.9	20.6	6.4
ESCS (Low quartile)	408.1	4.4	413.8	4.8	412.4	4.2	394.3	8.1	-4.3	4.8	19.5	8.4
ESCS (High quartile)	424.8	3.9	431.4	5.2	429	4.8	409.6	7.3	-4.1	3.5	21.7	6.2
ESCS (Very high quartile)	430.6	4.9	431	4.5	436.5	4.4	414.2	7.7	-5.8	4.7	16.8	5.9
Miss the whole class: Never	423.8	2.8	424.9	3	426.2	3.8	398.7	5.4	-2.4	3.4	26.2	5
Miss the whole class: 1 to 2 times	407.5	4.4	419	5	414.2	2.9	409.2	7.7	-6.7	3.7	9.8	7.9
Miss the whole class: 3 to 4 times	386.5	9.4	403.4	9.4	416.2	11.1	400.4	11.1	-29.7	11.7	3	9.2
Miss the whole class: 5 or more times	379.4	9.6	387.8	11	398.6	15.7	346.6	13.3	-19.2	11	41.2	12.9
Single parent home: Yes	422.5	4.2	431.6	4.8	430.7	5.1	416.2	9.4	-8.3	4.7	15.4	10.9
Single parent home: No	414.8	2.5	418.9	3.2	419.8	2.9	396.7	5.3	-5	3.3	22.2	5
Tablet in the school: Yes and they use it	385.1	6.7	374.8	9.1	415.6	7.6	369.9	12.6	-30.5	4.3	4.9	8.2
Tablet in the school: Yes but they do not use it	391.5	12.9	397.5	11.7	398.6	11.6	373.1	17	-7.1	11.8	24.4	14.9

Source: Authors' own calculations from PISA 2012 (the sample only includes students who performed the PPA and CBA tests).

Table 2.2: DESCRIPTIVE STATISTICS FOR THE CHARACTERISTICS OF THE REPEATER STUDENTS (continuation)

Variables	PPA Maths		PPA Reading		CBA Maths		CBA Reading		PPA-CBA Maths		PPA-CBA Reading	
	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.
Tablet in the school: No	420.9	2.1	428.3	2.9	422.6	2.7	406.7	5.5	-1.7	3.1	21.6	5.3
Age when started to use the computer: Never	338	28.8	294.5	27.5	318	23.3	207.6	56.9	20	23.1	86.9	30.6
Age when started to use the computer: 6 or <	427.3	3.9	427.5	5	423.9	4.2	404.2	8.4	3.5	3.3	23.3	8.7
Age when started to use the computer: 7-9	419.3	3.4	427.3	3.5	426.7	3.8	408.6	6.4	-7.5	4.2	18.7	5.6
Age when started to use the computer: 10-12	403.1	4.7	414.6	4.1	407	4.8	395	6.3	-3.9	4.8	19.6	6.2
Age when started to use the computer: 13 or >	390.6	11.4	393.6	10.8	416.2	12.5	380.9	17.7	-25.6	9.4	12.8	12.1
Private school	448	9.4	452.4	9	433.8	9.9	421	6.8	14.1	7.9	31.3	8.1
Semi-private school	437.7	5.1	435.1	6.5	439.7	7.3	421.3	10.8	-2	6.1	13.8	9
Public school	408.4	2.8	415.6	3.2	416.1	3.1	395.4	6.8	-7.7	3.8	20.1	6.5
ICT Quartile: Very low	413.7	3.4	422.9	4.6	415.3	5.1	396.3	7.8	-1.6	3.8	26.6	8.2
ICT Quartile: Low	426.9	4.6	434.7	5.1	427.4	3.7	416.7	6.8	-0.5	3.1	18	6.6
ICT Quartile: High	420.4	4.1	425.6	4.9	426	5.8	409.3	8.6	-5.5	6.1	16.2	6.4
ICT Quartile: Very high	402.7	4.9	402.8	6.3	413.1	4.9	384.5	6.7	-10.5	5.1	18.3	6.9
Total	414.8	2.3	420.3	2.8	420.7	2.8	400.7	5.3	-5.9	3.1	19.6	5.1

Source: Authors' own calculations from PISA 2012 (the sample only includes students who performed the PPA and CBA tests).

Table 3.1: MULTILEVEL ESTIMATES TO EXPLAIN THE GAP BETWEEN PPA AND CBA IN MATHS COMPETENCE

Variables	Non-Repeaters	Repeaters
FEMALE	-3.343***	-1.910
	(1.044)	(2.353)
IMMIGRANT	15.201***	23.180***
	(2.503)	(3.634)
DIGLOSSIA	-0.632	-0.163
	(1.780)	(3.884)
ESCS. Reference category: (Very low quartile)		
ESCS (Low quartile)	-2.238	0.332
	(1.632)	(2.856)
ESCS (High quartile)	0.353	-5.283
	(1.640)	(3.400)
ESCS (Very high quartile)	4.618***	-4.413
	(1.719)	(4.531)
MISS THE WHOLE CLASS. Reference category: (Never missed a whole class)		
Missed the whole class (1 to 2 times)	2.476	6.677**
	(1.526)	(2.674)
Missed the whole class (3 to 4 times)	-27.896***	-7.644
	(4.664)	(5.601)
Missed the whole class (5 or more times)	-25.019***	-9.670
	(6.905)	(7.114)
SINGLE PARENT HOME	-5.597***	-3.462
	(1.866)	(3.311)
TABLET IN THE SCHOOL. Reference category: (No tablet in the school)		
There is a tablet and they use it	-18.002***	-17.860***
	(3.179)	(4.592)
There is a tablet but they do not use it	-8.522***	-12.072**
	(3.091)	(5.633)
AGE WHEN STARTED TO USE A COMPUTER. Reference category: (thirteen or older)		
Six years or less	13.877***	5.497
	(4.146)	(5.923)
Seven to nine years	11.742***	3.532
	(4.107)	(5.753)
Ten to twelve years	4.293	1.254
	(4.217)	(5.880)

Table 3.1: Multilevel estimates to explain the gap between PPA and CBA in maths competence (continuation)

Variables	Non-Repeaters	Repeaters
TYPE OF SCHOOL. Reference category: (Public school)		
Private school	2.320	14.594
	(15.816)	(19.030)
Semi-private school	-1.815	10.503
	(14.002)	(16.587)
PERCENTAGE OF GIRLS IN THE SCHOOL	-32.821	-34.845
	(23.060)	(30.324)
YEARS OF SCHOOLING OF THE FATHER	4.924*	6.753**
	(2.561)	(2.904)
YEARS OF SCHOOLING OF THE MOTHER	-1.600	-1.174
	(2.503)	(2.887)
PROPORTION OF IMMIGRANTS	-0.038	-0.244
	(0.186)	(0.199)
STUDENTS/TEACHER RATIO	1.848***	1.238
	(0.691)	(0.868)
FLAG STUDENTS/TEACHER RATIO	4.349	4.536
	(14.597)	(18.312)
USE OF ICT IN THE SCHOOL	6.091	9.650
	(5.272)	(6.180)
USE OF ICT IN MATHS LESSONS	0.461	-3.434
	(5.271)	(6.303)
GROUPING FOR THE CLASSES. Reference category: (No grouping)		
Grouping in all classes	8.013	4.307
	(7.896)	(9.622)
Grouping in some classes	-9.855*	-19.193***
	(5.435)	(6.360)
ADMISSION BECAUSE PARENTS SHARE THE PHILOSOPHY OF THE SCHOOL. Reference category: (This criterion is not used)		
This criterion is always used	-4.057	-0.085
	(6.118)	(7.630)
This criterion is sometimes used	5.394	0.814
	(6.759)	(8.191)
INDEPENDENCE OF HEAD/TEACHERS/BOARD TO APPOINT/FIRE TEACHERS	-12.626 (13.223)	-16.663 (15.434)
INDEPENDENCE OF HEAD/TEACHERS/BOARD TO SET SALARIES AND SALARY INCREASES	-0.923 (6.948)	-5.530 (9.184)

Table 3.1: Multilevel estimates to explain the gap between PPA and CBA in maths competence (continuation)

Variables	Non-Repeaters	Repeaters
INDEPENDENCE OF HEAD/TEACHERS/BOARD TO DECIDE BUDGETS AND BUDGET ALLOCATION	11.468 (11.116)	21.630 (13.235)
INDEPENDENCE OF HEAD/TEACHERS/BOARD TO SET ASSESSMENT RULES, DISCIPLINE AND STUDENT SELECTION	28.149 (21.338)	29.801 (27.452)
Constant	-54.778	-55.202
	(34.870)	(44.841)
Observations	6,591	1,796
Number of schools	343	319
Wald test	284.77***	152.65***

*** Denotes variable significant to level 1%; ** to 5%; * to 10%.

(1) A dummy variable for each autonomous region was also included.

Source: Authors' own calculations from PISA 2012.

Table 3.2: Multilevel estimates to explain the gap between PPA and CBA in reading competence

Variables	Non-Repeaters	Repeaters
FEMALE	3.680***	4.895*
	(1.187)	(2.815)
IMMIGRANT	-9.549***	-5.367
	(2.845)	(4.348)
DIGLOSSIA	19.758***	18.059***
	(2.026)	(4.654)
ESCS. Reference category: (Very low quartile)		
ESCS (Low quartile)	3.839**	3.735
	(1.855)	(3.419)
ESCS (High quartile)	8.222***	1.231
	(1.864)	(4.069)
ESCS (Very high quartile)	6.856***	1.858
	(1.954)	(5.426)
MISS THE WHOLE CLASS. Reference category: (Never missed a whole class)		
Missed the whole class (1 to 2 times)	-12.422***	-13.453***
	(1.735)	(3.200)

Table 3.2: MULTILEVEL ESTIMATES TO EXPLAIN THE GAP BETWEEN PPA AND CBA IN READING COMPETENCE (continuation)

Variables	Non-Repeaters	Repeaters
Missed the whole class (3 to 4 times)	-15.911***	-7.147
	(5.301)	(6.702)
Missed the whole class (5 or more times)	-12.048	9.840
	(7.848)	(8.511)
SINGLE PARENT HOME	-0.210	5.728
	(2.121)	(3.962)
TABLET IN THE SCHOOL. Reference category: (No tablet in the school)		
There is a tablet and they use it	-6.563*	-10.514*
	(3.615)	(5.497)
There is a tablet but they do not use it	-6.792*	-2.581
	(3.514)	(6.740)
AGE WHEN STARTED TO USE A COMPUTER. Reference category: (thirteen or older)		
Six years or less	-8.063*	21.387***
	(4.713)	(7.083)
Seven to nine years	-2.133	18.852***
	(4.669)	(6.879)
Ten to twelve years	-3.158	17.860**
	(4.794)	(7.030)
TYPE OF SCHOOL. Reference category: (Public school)		
Private school	13.447	-0.759
	(19.992)	(23.703)
Semi-private school	-2.436	-12.712
	(17.698)	(20.660)
PERCENTAGE OF GIRLS IN THE SCHOOL	-14.544	22.155
	(29.165)	(37.694)
YEARS OF SCHOOLING OF THE FATHER	0.643	6.694*
	(3.224)	(3.612)
YEARS OF SCHOOLING OF THE MOTHER	-0.139	-6.311*
	(3.151)	(3.597)
PROPORTION OF IMMIGRANTS	0.685***	0.578**
	(0.233)	(0.248)
STUDENTS/TEACHER RATIO	1.619*	2.843***
	(0.874)	(1.078)
FLAG STUDENTS/TEACHER RATIO	12.679	48.537**
	(18.459)	(22.776)
USE OF ICT IN THE SCHOOL	12.364*	8.089
	(6.655)	(7.699)

Table 3.2: MULTILEVEL ESTIMATES TO EXPLAIN THE GAP BETWEEN PPA AND CBA IN READING COMPETENCE (continuation)

Variables	Non-Repeaters	Repeaters
USE OF ICT IN MATHS LESSONS	5.360	8.312
	(6.660)	(7.857)
GROUPING FOR THE CLASSES. Reference category: (No grouping)		
Grouping in all classes	-1.160	-11.197
	(9.988)	(11.977)
Grouping in some classes	3.817	7.416
	(6.862)	(7.927)
ADMISSION BECAUSE PARENTS SHARE THE PHILOSOPHY OF THE SCHOOL. Reference category: (This criterion is not used)		
This criterion is always used	1.242	-1.610
	(7.736)	(9.501)
This criterion is sometimes used	-4.667	-5.389
	(8.539)	(10.207)
INDEPENDENCE OF HEAD/TEACHERS/BOARD TO APPOINT/FIRE TEACHERS	-0.975	3.366
	(16.712)	(19.229)
INDEPENDENCE OF HEAD/TEACHERS/BOARD TO SET SALARIES AND SALARY INCREASES	-9.757	-5.003
	(8.790)	(11.413)
INDEPENDENCE OF HEAD/TEACHERS/BOARD TO DECIDE BUDGETS AND BUDGET ALLOCATION	-6.095	-17.041
	(14.049)	(16.507)
INDEPENDENCE OF HEAD/TEACHERS/BOARD TO SET ASSESSMENT RULES, DISCIPLINE AND STUDENT SELECTION	33.014	55.464
	(26.993)	(34.516)
Constant	9.558	-30.422
	(43.649)	(55.592)
Observations	6,591	1,796
Number of schools	343	319
Wald test	261.51***	116.67***

*** Denotes variable significant to level 1%; ** to 5%; * to 10%.

(1) A dummy variable for each autonomous region was also included.

Source: Authors' own calculations from PISA 2012.

Table 4: RANDOM VALUES OF THE MULTILEVEL REGRESSION MODELS FOR THE COMPLETE SAMPLE

		Maths		Reading	
		Non-Repeaters	Non-Repeaters	Non-Repeaters	Repeaters
Coefficient of intra-class correlation	Coeff.	0.419	0.360	0.475	0.388
	St. Dev.	0.022	0.029	0.022	0.029
Standard deviation of the constant across schools	Coeff.	34.564	34.167	44.020	43.324
	St. Dev.	1.557	1.976	1.919	2.454
Standard deviation of the total error term	Coeff.	40.730	45.529	46.280	54.370
	St. Dev.	0.365	0.836	0.415	1.000

Source: Authors' own calculations from PISA 2012.

Table 5: MULTILEVEL ESTIMATES TO EXPLAIN THE GAP BETWEEN PPA AND CBA (ALTERNATIVE SPECIFICATION II)

Variables	Specification II			
	Maths Gap		Reading Gap	
	Non-Repeaters	Repeaters	Non-Repeaters	Repeaters
Additional control variables	✓	✓	✓	✓
HOURS OF STUDY WITH THE COMPUTER. Reference category: (0 hours)				
Hours of study with the computer (1 hour)	-5.630***	0.403	1.592	-7.160
	(1.722)	(4.102)	(1.932)	(4.561)
Hours of study with the computer (2 hours)	-4.791**	3.471	-1.076	2.464
	(2.213)	(5.043)	(2.484)	(5.616)
Hours of study with the computer (3 hours)	-10.205***	-3.877	-2.398	-8.194
	(3.172)	(7.824)	(3.560)	(8.686)
Hours of study with the computer (4 hours)	-20.701***	-7.161	3.164	-5.988
	(4.381)	(10.304)	(4.917)	(11.426)
Hours of study with the computer (5 hours)	-17.081***	-13.951	-0.001	-24.627**
	(5.309)	(11.158)	(5.958)	(12.413)
Hours of study with the computer (6 hours)	-21.072**	-5.624	-15.533	-28.725*
	(8.996)	(14.513)	(10.096)	(16.157)
Hours of study with the computer	-29.615***	-8.555	-16.317**	-1.181
(7 hours or more)	(5.677)	(9.539)	(6.372)	(10.570)
Constant	-64.875*	-76.112*	-14.294	-70.391
	(34.159)	(45.450)	(42.812)	(59.324)
Observations	4,226	1,095	4,226	1,095
Number of schools	342	307	342	307
Wald test	273.88***	105.90***	188.10***	112.69***

*** Denotes variable significant to level 1%; ** to 5%; * to 10%.

(1) A dummy variable for each autonomous region was also included.

(2) The tick (✓) means that additional control variables have been included in the estimates. These are: female, immigrant, diglossia, ESCS, miss the whole class, single parent home, tablet in the school, age at which started to use a computer, type of school, percentage of girls in the school, years of schooling of the father, years of schooling of the mother, proportion of immigrants, students/teacher ratio, flag students/teacher ratio, use of ICT in the school, use of ICT in maths lessons, grouping for the classes, admission because parents share the philosophy of the school, independence of head/teachers/board to appoint/fire teachers, independence of head/teachers/board to set salaries and salary increases, independence of head/teachers/board to decide budgets and budget allocation, independence of head/teachers/board to set assessment rules, discipline and student selection.

Source: Authors' own calculations from PISA 2012.

Table 6: MULTILEVEL ESTIMATES TO EXPLAIN THE GAP BETWEEN PPA AND CBA (ALTERNATIVE SPECIFICATION III)

Variables	Specification III			
	Maths Gap		Reading Gap	
	Non-Repeaters	Repeaters	Non-Repeaters	Repeaters
Additional control variables	✓	✓	✓	✓
GUIDED WORK OUTSIDE THE SCHOOL. Reference category: (0 hours)				
Guided work outside the school (1 hour)	-10.448***	-8.467**	-9.193***	6.865
	(1.904)	(4.268)	(2.124)	(5.214)
Guided work outside the school (2 hours)	10.121***	4.604	-4.312*	0.347
	(2.013)	(4.889)	(2.246)	(5.973)
Guided work outside the school (3 hours)	-7.134***	-2.866	-4.554	6.142
	(2.672)	(6.977)	(2.981)	(8.521)
Guided work outside the school (4 hours)	-10.204***	4.184	-2.975	1.775
	(2.978)	(7.217)	(3.322)	(8.799)
Guided work outside the school (5 hours)	-11.676***	7.120	-3.427	5.378
	(3.765)	(7.716)	(4.201)	(9.441)
Guided work outside the school (6 hours)	-5.193	-4.943	1.189	5.193
	(5.287)	(11.518)	(5.897)	(14.062)
Guided work outside the school	-13.286***	0.755	14.256***	19.646**
(7 hours or more)	(3.695)	(7.741)	(4.124)	(9.437)
Constant	-55.292	-85.812*	-5.770	-73.972
	(34.161)	(46.971)	(42.765)	(60.090)
Observations	4,075	998	4,075	998
Number of schools	341	295	341	295
Wald test	243.94***	111.18***	213.35***	109.82***

*** Denotes variable significant to level 1%; ** to 5%; * to 10%.

(1) A dummy variable for each autonomous region was also included.

(2) The tick (✓) means that additional control variables have been included in the estimates. These are: female, immigrant, diglossia, ESCS, miss the whole class, single parent home, tablet in the school, age at which started to use a computer, type of school, percentage of girls in the school, years of schooling of the father, years of schooling of the mother, proportion of immigrants, students/teacher ratio, flag students/teacher ratio, use of ICT in the school, use of ICT in maths lessons, grouping for the classes, admission because parents share the philosophy of the school, independence of head/teachers/board to appoint/fire teachers, independence of head/teachers/board to set salaries and salary increases, independence of head/teachers/board to decide budgets and budget allocation, independence of head/teachers/board to set assessment rules, discipline and student selection.

Source: Authors' own calculations from PISA 2012.

Table 7: Multilevel estimates to explain the gap between PPA and CBA (alternative specification IV)

Variables	Specification IV			
	Maths Gap		Reading Gap	
	Non-Repeaters	Repeaters	Non-Repeaters	Repeaters
Additional control variables	✓	✓	✓	✓
TIME SPENT WITH THE HELP OF PARENTS OR OTHER FAMILY MEMBER DOING SCHOOLWORK AT HOME. Reference category: (0 hours)				
Time spent with the help of parents or other family member (1 hour)	0.995 (1.938)	4.322 (4.340)	-1.346 (2.169)	-7.156 (4.987)
Time spent with the help of parents or other other family member (2 hours)	-2.575 (2.512)	-9.126* (5.292)	2.774 (2.811)	3.532 (6.086)
Time spent with the help of parents or family member (3 hours)	-2.905 (3.842)	-10.195 (7.939)	-0.969 (4.298)	-3.628 (9.107)
Time spent with the help of parents or other family member (4 hours)	-10.208* (5.641)	7.994 (9.512)	-11.194* (6.312)	2.522 (10.891)
Time spent with the help of parents or other family member (5 hours)	6.113 (6.745)	-6.159 (14.011)	-9.913 (7.545)	6.446 (16.094)
Time spent with the help of parents or other family member (6 hours)	-3.983 (8.603)	22.044 (16.311)	1.554 (9.623)	1.723 (18.797)
Time spent with the help of parents or other family member (7 hours or more)	-17.708*** (5.575)	-19.062* (9.846)	-2.295 (6.239)	13.522 (11.279)
Constant	-66.672* (34.052)	-70.016 (45.283)	-11.713 (42.726)	-79.792 (58.488)
Observations	4,212	1,087	4,212	1,087
Number of schools	343	306	343	306
Wald test	208.55***	119.97***	180.02***	109.19***

*** Denotes variable significant to level 1%; ** to 5%; * to 10%.

(1) A dummy variable for each autonomous region was also included.

(2) The tick (✓) means that additional control variables have been included in the estimates. These are: female, immigrant, diglossia, ESCS, miss the whole class, single parent home, tablet in the school, age at which started to use a computer, type of school, percentage of girls in the school, years of schooling of the father, years of schooling of the mother, proportion of immigrants, students/teacher ratio, flag students/teacher ratio, use of ICT in the school, use of ICT in maths lessons, grouping for the classes, admission because parents share the philosophy of the school, independence of head/teachers/board to appoint/fire teachers, independence of head/teachers/board to set salaries and salary increases, independence of head/teachers/board to decide budgets and budget allocation, independence of head/teachers/board to set assessment rules, discipline and student selection.

Source: Authors' own calculations from PISA 2012.

REFERENCES

Baird, J.A., Isaacs, T., Johnson, S., Stobart, G., Yu, G., Sprague, T. and Daugherty, R. (2011): *Policy effects of PISA*, Oxford, Oxford University Centre for Educational Assessment.

Battauz, M.R., Bellio, R. and Gori, E. (2011): "Covariate error adjustment for multilevel models with application to educational data", *Journal of Educational and Behavioral Statistics*, vol. 36(3), pp. 283-306.

Bennett, R.E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B. and Yan, F. (2008): "Does it Matter if I Take My Mathematics Test on Computer? A Second Empirical Study of Mode Effects in NAEP", *Journal of Technology, Learning, and Assessment*, vol. 6(9), pp. 1-39.

Borghans, L. and Schils, T. (2012): *The leaning tower of Pisa: decomposing achievement test scores into cognitive and noncognitive components*, Netherlands, Maastricht University.

Breakspear, S. (2012): "The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance", *OECD Education Working Papers*, 71, OECD Publishing. *http://dx.doi.org/10.1787/ 5k9fdfqffr28-en*

Cabras, S. and Tena, J.D. (2013): "Estimación del efecto causal del uso de ordenadores en los resultados de los estudiantes en el test PISA 2012", in INEE (Ed.), *PISA 2012: Programa para la evaluación internacional de los alumnos. Informe español. Volumen II: Análisis secundario*, Madrid, Ministerio de Educación, Cultura y Deporte, Instituto Nacional de Evaluación Educativa, pp. 67-88.

Calero, J., Choi, A. and Waisgrais, S. (2010): "Determinantes del riesgo de fracaso escolar en España: una aproximación a través de un análisis logístico multinivel aplicado a PISA-2006", *Revista de Educación*, vol. Extr., pp. 225-256.

Calero, J. and Escardibul, O. (2013): "El rendimiento del alumnado de origen inmigrante en PISA 2012", in INEE (Ed.), *PISA 2012: Programa para la evaluación internacional de los alumnos. Informe español. Volumen II: Análisis secundario*, Madrid, Ministerio de Educación, Cultura y Deporte, Instituto Nacional de Evaluación Educativa, pp. 4-31.

Carabaña, J. (2013): "Repetición de curso y puntuaciones pisa ¿cuál causa cuál?", in INEE (Ed.), *PISA 2012: Programa para la evaluación internacional de los alumnos. Informe español. Volumen II: Análisis secundario*, Madrid, Ministerio de Educación, Cultura y Deporte, Instituto Nacional de Evaluación Educativa, pp. 32-66.

Cordero, J.M., Cebada, E.C. and Pedraja, F. (2013): "Rendimiento educativo y determinantes según PISA: Una revisión de la literatura en España", *Revista de Educación*, vol. 362, pp. 273-297.

Cordero, J.M., Pedraja, F. and Simancas, R. (2014): "Superando las barreras: factores determinantes del rendimiento en escuelas y estudiantes con un entorno desfavorable", in INEE (Ed.), *PISA 2012: Resolución de problemas de la vida real. Resultados de Matemáticas y Lectura por ordenador. Informe español. Análisis secundario*, Madrid, Ministerio de Educación, Cultura y Deporte, Instituto Nacional de Evaluación Educativa, pp. 39-63.

Dolton, P., Marcenaro, O. and Navarro, L. (2003): "The effective use of student time: a stochastic frontier production function case study", *Economics of Education Review*, vol. 22(6), pp. 547-560.

EUROSTAT (2014): "Internet access and use in 2013", *Eurostat News Release*, retrieved from http://ec.europa.eu/eurostat/documents/2995521/5168694/4-18122013-BP-EN.PDF/b92e0257-3dba-4eb1-97ce-0b42a736dee0?version=1.0

Fuchs, T. and Woessman, L. (2004): "Computers and Student Learning: Bivariate and Multivariate Evidence on the Availability and Use of Computers at Home and at Schools", *Brussels Economic Review*, vol. 47, pp. 359-389.

García-Montalvo, J. (2013): "Crisis, igualdad de oportunidades y resultados educativos en España: una visión retrospectiva desde PISA 2012", in INEE (Ed.), *PISA 2012: Programa* *para la evaluación internacional de los alumnos. Informe español. Volumen II: Análisis secundario*, Madrid, Ministerio de Educación, Cultura y Deporte, Instituto Nacional de Evaluación Educativa, pp. 89-117.

García-Pérez, J.I., Hidalgo-Hidalgo, M. and Robles-Zurita, J.A. (2014): "Does grade retention affect achievement? Some evidence from PISA", *Applied Economics*, vol. 46(12), pp. 1373-1392.

Harrison, C., Comber, C., Fisher, T., Haw, K., Lewin, C., Lunzer, E., McFarlane, A., Mavers, D., Scrimshaw, P., Somekh, B. and Watling, R. (2002): *ImpacCT2: The Impact of Information and Communication Technologies on Pupil Learning and Attainment*, London, DfES and Becta.

Hopkins, D., Pennock, D., Ritzen, J., Ahtaridou, E. and Zimmer, K. (2008): *External Evaluation of the Policy Impact of PISA*, Paris, OECD.

Hox, J.J. (1998): "Multilevel modeling: When and why", in I. Balderjahn, R. Mathar and M. Schader (Eds.), *Classification, data analysis and data highways*, New York, Springer, pp. 147-154.

INE (2014): "Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación en los Hogares 2014", retrieved from http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t25/p450/base_2011/a2014/&file=pcaxis

Jeong, H. (2014): "A comparative study of scores on computer-based tests and paper-based tests", *Behaviour & Information Technology*, vol. 33(4), pp. 410-422.

Jimerson S.R. (1999): "On the Failure of Failure: Examining the Association Between Early Grade Retention and Education and Employment Outcomes During Late Adolescence", *Journal of School Psychology*, vol. 37(3), pp. 243-272.

Marcenaro, O. (2013): *El rendimiento del alumnado andaluz a examen*, Sevilla, Junta de Andalucía.

Marcenaro, O. and Vignoles, A. (2015): "A comparison of teacher and test-based assessment for Spanish primary and secondary education students", *Educational Research*, vol. 57(1), pp. 1-21.

MECD (2013): *Marcos y pruebas de evaluación de PISA 2012: Matemáticas, Lectura y Ciencias*, Madrid, Ministerio de Educación, Cultura y Deporte, Instituto Nacional de Evaluación Educativa.

Mediavilla, M. and Escardíbul, J.O. (2014): "El efecto de las TIC en la adquisición de competencias. Un análisis de género y titularidad de centro para las evaluaciones por ordenador", in INEE (Ed.), *PISA 2012: Resolución de problemas de la vida real. Resultados de Matemáticas y Lectura por ordenador. Informe español. Análisis secundario,* Madrid, Ministerio de Educación, Cultura y Deporte, Instituto Nacional de Evaluación Educativa, pp. 158-179.

Notten, N., Peter, J. and Kraaykamp, G. (2009): "Research Note: Digital Divide Across Borders-A Cross-National Study of Adolescents' Use of Digital Technologies", *European Sociological Review*, vol. 25(5), pp. 551-560.

OECD (2013): *Education and Policy outlook: Norway*, Paris, OECD Publishing.

Raudenbush, S.W. and Bryk, A.S. (2002): *Hierarchical linear models: Applications and data analysis methods*, California, 2nd edition, Newbury Park, Sage.

Sim, G. and Horton, M. (2005): "Performance and attitude of children in computer based versus paper based testing", in P. Kommmers and G. Richards (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2005*, Chesapeake, Virginia, AACE, pp. 3610-3614.

Sosa, G.W., Berger, D.E., Saw, A.T. and Mary, J.C. (2011): "Effectiveness of Computer-Assisted Instruction in Statistics A Meta-Analysis", *Review of Educational Research*, vol. 81(1), pp. 97-128.

Valentine, G., Marsh, J. and Pattie, C. (2005): *Children and Young People's Home Use of ICT for Educational Purposes: the impact of attainment at Key Stages 1-4*, Department for Education and Skills, Research Report RR672, London DfES.

Wallace, R. and Clariana, P. (2002): "Paper-based versus computer-based assessment: key factors associated with the test mode effect", *British Journal of Educational Technology*, vol. 33(5), pp. 593-602.

Fecha de recepción del original: julio, 2014
Versión final: junio, 2015

RESUMEN

La implementación de la evaluación por ordenador (CBA) en PISA y la intención de la OCDE de usar este procedimiento de forma única desde 2015 hace que nos planteemos la necesidad de medir si ésta proporciona la misma información que las pruebas con papel y lápiz (PPA) realizadas hasta 2012. Nuestros resultados muestran la existencia de una brecha significativa entre PPA y CBA que puede ser explicada por factores como la disponibilidad y el acceso a las TICs o el estatus socioeconómico y cultural de los estudiantes. Esto nos lleva a concluir que cualquier reforma de política educativa basada en los resultados de PISA y en los factores que les afectan, así como la comparación de estos resultados a lo largo del tiempo, debe hacerse a la luz del método de evaluación específico utilizado como referencia.

Palabras clave: evaluación asistida por ordenador, evaluación con papel y lápiz, brecha.

Clasificación JEL: I20, I21.