ABSTRACT

Title: The influence on academic performance of test language in Wales

Authors and e-mails:
[1] John Jerrim (j.jerrim@ucl.ac.uk)
[2] Luis Alejandro Lopez-Agudo (lopezagudo@uma.es)
[2] Oscar David Marcenaro-Gutierrez (odmarcenaro@uma.es)

Department:
[1] UCL Institute of Education
[2] Departamento de Economía Aplicada (Estadística y Econometría). Facultad de Ciencias Económicas y Empresariales.

University:
[1] University College London.
[2] Universidad de Málaga.

Subject area: *(please, indicate the subject area which corresponds to the paper)*13. Turismo y cultura.

Abstract: *(minimum 300 words)*
This research study analyses the relationship between the language in which Welsh pupils took the PISA test and their scores in this assessment. In order to approach this research, we employ five cycles of PISA data and an instrumental variables approach. Our results show that Welsh students who took the test in English performed around 0.3 standard deviations (30 PISA test points) higher in reading, mathematics and science than their peers who took the test in Welsh. This result is kept for many model specifications and statistical approaches. We discuss that this may indicate that plans aimed at further extending Welsh language education may not help students improve key academic skills.

Keywords: *(maximum 6 words)* test language; home language; Wales; PISA; instrumental variables.
JEL codes: I20, I21, I28, C10.

1. Introduction

The Programme for International Assessment (PISA) is an instrument designed to measure 15-year-olds' ability in reading, mathematics and science. According to the PISA technical reports (PISA 2.1 technical standard[1]), students should take the test in the language which they are most comfortable. This is so that their score in the test reflects their actual skill in the subject(s) being assessed, and is not unduly influenced by pupils having limited language skills[2] (OECD, 2012, pp. 369-370). In doing so, the OECD implicitly acknowledges how a difference between pupil's "home language" and their "test language" may have an impact upon the results.

Authors such as Kennedy and Park (1994) have studied the link between test language, home language and academic achievement in the context of middle-school Asian-American and Mexican pupils. They found that those students who did not speak English at home obtained lower reading test scores. Similarly, in an analysis of PISA 2000 data for Australia, De Bortoli and Cresswell (2004) found that pupils who took the test in a language they did not regularly speak at home achieved lower scores overall. Similarly, the work by Mancilla-Martinez and Lesaux (2011) in the United States found that students whose home language was Spanish – but who took a test in English – tended to achieve lower test scores than native English speakers.

This issue is also prominent in international public policy and political debates. In many countries the existence of multilingualism has been used as a "symbol" of nationalism by nationalist political parties. That is the case of French in Canada, Catalan in Spain or Welsh in Wales. In the present study, we focus on the latter. According to the Ministerial Foreword in the Welsh in Education Action Plan 2017-2021 (Welsh Government, 2017, p. 2):

"(…) *we also know that the teaching and learning of Welsh in English-medium settings is inconsistent and too often leads to low attainment*".

[1] Following this standard, students with insufficient experience in the language of assessment are excluded from PISA. In particular, these students are those who: (a) are not native speakers of the assessment language, (b) have limited proficiency in the assessment language and (c) have received less than one year of instruction in the assessment language. Furthermore, students are also excluded from PISA when there are no available materials in the language in which the student is taught.

[2] This is also indicated in the technical reports of all the PISA cycles under analysis in OECD (2009; 2014; 2017; 2020).

Thus, it seems that Welsh education authorities have shown some concern about the balance between preserving national identity and culture through the Welsh language, and the academic scores achieved by their students. This may be the consequence of results found in the literature (e.g. Jerrim & Shure, 2016) which found average reading and science scores to be lower amongst pupils attending Welsh-medium schools than for their pupils attending English-medium schools. In addition, the aforementioned action plan promotes the relevance of Welsh language as "at the heart" of ongoing educational reforms.

In this context, a distinctive characteristic of the Welsh education system is the existence of schools that use Welsh as the primary teaching language (Johnes, 2020). This has led most existing work in this area to focus upon differences in academic performance between Welsh-medium and English-medium schools. For instance, Gorard (1998) shows that, once differences in local-area characteristics are taken into account, there is no significant difference between the performance of Welsh-medium and English-medium schools in Wales. However, to our knowledge, there is no evidence as to whether taking the PISA test in English might be detrimental for the scores obtained by Welsh pupils (compared to the alternative of taking the test in Welsh). The aim of this paper is to hence fill this gap in the literature, quantifying this potential gap in achievement.

To do so, we use five cycles (2006-2018) of PISA data for Wales. These data allow us to identify both the language spoken by pupils at home as well as the language in which they took the PISA test (students in Wales could take the test in either English or Welsh). One novel aspect of this research is that it attempts to move a step closer towards estimating a causal effect by implementing an instrumental variable approach. Specifically, we address the following research question:

Do pupils who took the PISA test in English achieve better results than those who took the test Welsh?

The paper now proceeds as follows: section 2 provides a description of the data, followed by an overview of the methodological approach in section 3. Results are reported in section 4, followed by discussion and conclusions in section 5.

2. Data

2.1. Data description

PISA is a test taken by 15-year-old pupils from 80 countries across the world. It aims to assess their skills in reading, mathematics and science. It has been conducted every three years since 2000. Participating students also complete a background questionnaire, while headteachers complete a school questionnaire. In this paper we analyse those PISA cycles in which both test and home languages are known for Wales: 2006, 2009, 2012, 2015 and 2018. These cycles have been chosen because (a) they are the most recent ones; (b) an oversample was drawn for Wales to facilitate national reporting and (c) information on test and home languages has been collected.

The test language variable has two options: "English" and "Welsh". The home language variable indicates the language spoken most often at home by the pupil, with 5 options: "English", "Welsh", "Irish", "Ulster Scots" and "Other languages". Other relevant variables for our analysis are pupil, father and mother regions of birth, which include "Germany", "India", "Ireland", "Pakistan", "Poland", "United Kingdom (England)", and "Other countries".

2.2. Descriptive statistical analyses

In Table 1 the percentage of students who took the PISA test in each language (English or Welsh) is presented, along with the percentage of students who spoke each language at home. It can be seen that the percentage of students who took the test in English is stable across PISA cycles (87%), along with the percentage who speak each language at home (91% for English, 6% for Welsh, 0.1% for Irish, 0.1% for Ulster Scots and 3% for other languages). These figures are broadly similar to those reported by StatsWales (2021).

-Insert Table 1 here-

Table 2 illustrates how almost all pupils who took the PISA test in English also speak English regularly at home. The situation is rather different for those pupils who took the PISA test in Welsh, more than half of them reporting that English is the language they most regularly speak at home. In other words, many pupils who usually

speak English at home end up taking the test in Welsh, most likely due to the directions provided by their school. This would be consistent with standard policy in Welsh-medium schools, where the Welsh language is preferred.

Finally, Table 3 compares the background characteristics of pupils who took the PISA test in England and Welsh. On the whole, the distribution of common observable characteristics appears broadly comparable across the two groups.

-Insert Table 2 here-

-Insert Table 3 here-

In order to get an initial understanding of the implications of this difference for academic performance, Table 4 presents raw differences in PISA scores between those who took the test in English and Welsh (using PISA plausible values, which have a mean across all participating countries of approximately 500). When no other factors are controlled, pupils who took the test in English seem to perform 41 points higher in reading than those who took it in Welsh. The difference is smaller, though still non-trivial, in mathematics (14 points) and science (27 points).

-Insert Table 4 here-

3. Methodology

3.1. Ordinary Least Squares

To begin, we analyse the influence of taking the test in English using ordinary least squares (OLS). The model is specified as:

$$C_{ijc} = \alpha + \beta ET_{ijc} + \gamma X_{ijc} + \delta F_{ijc} + \vartheta SCH_{jc} + \rho PISA_{c} + \varepsilon_{ijc} \quad (1)$$

where i is the individual, j the school and c the PISA cycle; C_{ijc} are students' standardised scores in reading, mathematics and science (alternatively); ET_{ijc} is a dummy variable which indicates whether the pupil took the PISA test in English (1) or Welsh (0); X_{ijc} are students' background characteristics (i.e. sex, grade retention, student's region of birth, if the student has lived in the United Kingdom since age 6 or

older or not[3] and month of birth); F_{ijc} are family characteristics (socio-economic status, father's region of birth and mother's region of birth)[4]; SCH_{jc} are school characteristics (school funding); $PISA_c$ controls for PISA cycle; ε_{ijc} is the idiosyncratic error term.

The estimated β coefficient from this model will illustrate whether taking the PISA test in English continues to be associated with higher PISA test scores, controlling for a wide array of observable characteristics. Note that PISA scores have been standardised within Wales for each cycle in each subject. Results are hence presented in terms of effect sizes.

However, this β coefficient may be biased due to potential unobservables included in ε_{ijc} which have been omitted from our model (for instance, students' prior achievement before they entered secondary school). This omitted variable problem has been highlighted by many authors when dealing with observational cross-sectional data such as PISA (Hanchane & Mostafa, 2010; Micklewright, Schnepf, & Silva, 2012; Lounkaew, 2013; Cordero and Pedraja, 2018). For instance, as an illustrative example of this problem, one could imagine that some schools might force pupils to take the test only in Welsh, regardless of the language they most regularly speak at home. Focusing on those students who speak English at home, there may be: (a) those who have stronger skills in Welsh (which might be correlated with the omitted variable of student ability) and thus perform higher and (b) those who do not have strong Welsh skills and perform lower. The OLS estimate of β might hence be biased due to the proportion of pupils with strong and weak Welsh skills. We hence employ an instrumental variable approach to try and overcome this problem, implemented via two-stage least squares (2SLS).

3.2. Two-stage least squares

Our instrumental variable approach needs the identification of an instrument (Z_{ijc}), and also the use of control variables (X_{ijc}, F_{ijc}, SCH_{jc}) to try and eliminate any

[3] This variable controls the potential differences in language skills between those students who arrived at the United Kingdom and started compulsory education at age 6 or before, and those who arrived and started after that age.

[4] The combination of the student's, father's and mother's region of birth variables also controls for student's immigrant status.

potential confounder. The instrument used in this paper is language spoken at home, denoted as Z_{ijc}. This is a categorical variable which can be decomposed into a set of binary variables, each representing a language spoken at home.

To be a credible estimation approach, our instrument has to fulfil the following assumptions:

(a) The *relevance condition* or *first stage*. This means that the instrument should be strongly associated with the treatment variable. This is clearly the case in this study, as language at home is strongly linked to the language of the test (EF_{ijc}). We have already illustrated this point descriptively in Table 2, with more formal results from the Stock and Yogo (2005) test of weak instruments presented in the results section below.

(b) The *independence/exogeneity assumption*. This means that the instrument (language spoken at home) should not be independently associated with the outcome (PISA scores), other than through the "treatment" (taking the test in English). In our study, the language at home instrument might be considered as good as randomly assigned after controlling by X_{ijc}, F_{ijc} and SCH_{jc} – specifically after controlling for pupil's socio-economic status, their country of birth, their parents' country of birth and school fixed-effects. This exogeneity of the language at home is due to: (a) there being no home language choice in monolingual households; (b) in plurilingual households, once all the control variables have been included, the choice between one language and another is assumed to be as good as random.

(c) The *exclusion restriction*. This means that there is a sole channel for the effect of the instrument (Z_{ijc}, language at home) on the outcome (C_{ijc}, students' competences). This single channel requires the previous independence assumption, to the extent that the other potential channels of influence have been controlled (X_{ijc}, F_{ijc} and SCH_{jc}).

(d) The *monotonicity property* (Fiorini & Stevens, 2014; Barua & Lang, 2016; Dhuey, Figlio, Karbownik, & Roth, 2019). As defined by Barua and Lang (2016) "while the instrument may have no effect on some individuals, all of

those who are affected should be affected unidirectionally" (p. 348). This is also known as the *no defiers assumption*, i.e. there are no students who, if they are assigned to take the test in their home language, always choose to take it in another language. Likewise, if they are assigned to take the test in a different language than the one spoken at home, they always choose to take the test in their home language. In schools where all pupils were made to do the test in the same language, the monotonicity property is fulfilled. In schools in which pupils could choose the test language, it is assumed they always chose the language that they were most comfortable with. In particular, from a total of 607 school observations in our dataset, all pupils took the test in English in 503, in 44 schools all pupils took the test in Welsh and in 60 schools there was a mix of English and Welsh test takers.

Model (1) can then be estimated using two-stage least squares (2SLS). The *first stage* is specified as:

$$ET_{ijc} = \pi_0 + \pi_1 Z_{ijc} + \pi_2 X_{ijc} + \pi_3 F_{ijc} + \pi_4 SCH_{jc} + \omega_{ijc} \quad (2)$$

where ω_{ijc} is the idiosyncratic error term. After estimating this equation we obtain a prediction of the English test language variable ($\widehat{ET}_{ijc}$), which can be included in model (1) to define the following *reduced form* (hereinafter the baseline model):

$$C_{ijc} = \alpha + \beta \widehat{ET}_{ijc} + \gamma X_{ijc} + \delta F_{ijc} + \vartheta SCH_{jc} + \varepsilon_{ijc} \quad (3)$$

where β indicates the influence of taking the test in English on academic performance in reading, mathematics and science, respectively. The results obtained have been checked using the Stock and Yogo (2005) test of weak instruments and the Wooldridge (1995) endogeneity test (reported below).The recommended practices for analysing PISA data (final student weights, balanced repeated replication weights[5] and plausible values) have been used throughout our analysis (OECD, 2020).

4. Results

[5] Balanced repeated replication (BRR) weights control the multi-level structure of the data, producing unbiased standard errors, also clustering at school level (OECD, 2020).

First, in the OLS column in table 5 we present the results from our baseline model, estimated via OLS (the full set of parameter estimates can be found in Table A1 in the Appendix). Taking the PISA test in English seems positively associated with pupils' test scores: a difference of 0.42 standard deviations (SDs) is observed in reading, 0.18 SDs in mathematics and 0.36 SDs in science, compared to pupils who took the test in Welsh.

The model underpinning these results may, however, omit certain (unobservable) variables, which may confound the relationship between students' academic achievement and test language. We hence move on to results from our instrumental variable approach. The first stage estimations from equation (2) are reported in Table A1 (Appendix). The results for the instrument (i.e. language at home) are significant in explaining the treatment variable (English test language), which supports the *relevance condition*. Related to this, the null hypothesis of the Stock and Yogo (2005) test (that the instrument is weak) can clearly and decisively be rejected.

The second stage of our instrumental variable estimates produce very similar results to those produced under OLS. Specifically, there continues to be a positive influence of taking the test in English (relative to taking the test in Welsh) upon pupils' reading (0.39 SDs), mathematics (0.26 SDs) and science (0.33 SDs) PISA scores. Furthermore, the Wooldridge (1995) endogeneity test cannot be rejected (in which the null hypothesis is that the endogenous variable is now exogenous), providing reassurance that these estimates have at least moved us a step closer to obtaining a causal effect.

-Insert Table 5-

5. Conclusion

This paper has investigated the influence of PISA test language on students' academic performance in Wales, an issue that is at the core of this country's education debate. It is the first time that a study has attempted to produce quasi-experimental evidence on this issue, using an instrumental variables approach applied to 5 cycles of PISA data. Our results show that students who took the test in English performed

around 0.39 SDs (39 points on the PISA scale[6]) higher in reading, 0.26 SDs (26 PISA points) higher in mathematics and 0.33 SDs (33 PISA points) higher in science, compared to their peers who took the test in Welsh. Taking into account that 25-30 points in the PISA scale is equivalent to one year of schooling (OECD, 2019), this is clearly a sizable effect. As one would anticipate, reading is the subject most affected by this problem, though with non-trivial differences in achievement between English and Welsh test-takers also observed in science and mathematics.

One possible interpretation of this finding is that there may be issues with the conduct of the PISA test in Wales, particularly with respect to how the test language is being chosen. Specifically, some pupils who are taking the test in Welsh may be better suited to taking the test in English, due to their comparatively limited Welsh language skills. This is important, as PISA is meant to capture pupils' skills in each subject area and not their level of understanding of the test language per se. Thus, it seems that PISA 2.1 technical standard (i.e. pupils should take the test in the language they are most comfortable with) might not be properly applied in Wales. We hence advise the OECD – and those conducting the PISA test in Wales – to ensure that this technical standard is properly applied in the future. This means giving all pupils in Wales a free and open choice to decide in which language they wish to take the PISA test, even if that means doing it in English within a Welsh-medium school. This policy might not be difficult to implement, to the extent that PISA is now a computer based assessment. An alternative could be testing whether students have enough skill in the test language or not before taking it. This issue is also relevant to the extent that PISA results are used to take policy decisions in Wales (Froese-Germain, 2010; Baird, Isaacs, Johnson, Stobart, & Yu, 2011; Breakspear, 2012; Rees & Taylor, 2015). Therefore, due to the impact that PISA has on education policies in Wales, the potential bias generated by test language issues may trigger unwise education policies and, thus, higher public expenses.

The present research study has some limitations. First, the use of observational and cross-sectional data means that, in spite of using an instrumental variable approach, it may be prudent to continue to interpret our estimates as conditional associations.

[6] This scale presents a mean of 500 and standard deviation of 100.

Second, this research has internal validity for Wales; results may or may not generalise to other national settings.

References

Baird, J.-A., Isaacs, T., Johnson, S., Stobart, G., & Yu, G. (2011). *Policy Effects of PISA*. Oxford: Oxford University Centre for Educational Assessment, 1–38.

Barua, R., & Lang, K. (2016). School Entry, Educational Attainment and Quarter of Birth: A Cautionary Tale of a Local Average Treatment Effect. *Journal of Human Capital, 10*(3), 347–376. doi: 10.1086/687599

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, No. 71, 1-31. doi: 10.1787/5k9fdfqffr28-en

Cordero, J. M., & Pedraja, F. (2018). The effect of financial education training on the financial literacy of Spanish students in PISA. *Applied Economics, 51*(16), 1679–1693. doi: 10.1080/00036846.2018.1528336

De Bortoli, L., & Cresswell, J. (2004). *Australia's Indigenous Students in PISA 2000: Results from an International Study*. ACER Research Monograph No 59, 1–42. Australia: Australian Council for Education Research.

Dhuey, E., Figlio, D., Karbownik, K., & Roth, J. (2019). School Starting Age and Cognitive Development. *Journal of Policy Analysis and Management, 38*(9), 538–578. doi: 10.1002/pam.22135

Fiorini, M., & Stevens, K. (2014). *Assessing the Monoticity Assumption in IV and fuzzy RD designs*. The University of Sydney Economics Working Paper Series, 13, 1–52.

Froese-Germain, B. (2010). *The OECD, PISA and the Impacts on Educational Policy*. Canadian Teacher's Federation: Virtual Research Centre, 1–31.

Gorard, S. (1998). Four errors ... and a conspiracy? The effectiveness of schools in Wales. *Oxford Review of Education, 24*(4), 459–472. doi: 10.1080/0305498980240403

Hanchane, S., & Mostafa, T. (2010). *Endogeneity Problems in Multilevel Estimation of Education Production Functions: an Analysis Using PISA Data*. LLAKES Research Paper 14, 1–45. London: Centre for Learning and Life Chances in Knowledge Economies and Societies.

Jerrim, J., & Shure, N. (2016). *Achievement of 15-year- olds in Wales: PISA 2015 national report*. London: UCL Institute of Education.

Johnes, G. (2020). Medium Efficiency: Comparing Inputs and Outputs by Language of Instruction in Secondary Schools in Wales. *Wales Journal of Education, 22*(2), 52–66.

Kennedy, E., & Park, H.-S. (1994). Home language as a predictor of academic achievement: A comparative study of Mexican- and Asian-American youth. *Journal of Research & Development in Education*, *27*(3), 188–194.

Lounkaew, K. (2013). Explaining urban–rural differences in educational achievement in Thailand: Evidence from PISA literacy data. *Economics of Education Review*, *37*, 213–225. doi: 10.1016/j.econedurev.2013.09.003

Mancilla-Martinez, J., & Lesaux, N. K. (2011). Early home language use and later vocabulary development. *Journal of Educational Psychology*, *103*(3), 535–546. doi: 10.1037/a0023655

Micklewright, J., Schnepf, S. V., & Silva, P. N. (2012). Peer effects and measurement error: The impact of sampling variation in school survey data (evidence from PISA). *Economics of Education Review, 31*(6), 1136–1142. doi: 10.1016/j.econedurev.2012.07.015

OECD (2009). *PISA 2006 Technical Report*. Paris: OECD Publishing.

OECD (2012). *PISA 2009 Technical Report*. Paris: OECD Publishing. doi: 10.1787/9789264167872-en

OECD (2014). *PISA 2012 Technical Report*. Paris: OECD Publishing.

OECD (2017). *PISA 2015 Technical Report*. Paris: OECD Publishing.

OECD (2019). *PISA 2018 Results (Volume I): What Students Know and Can Do*. Paris: OECD Publishing.

OECD (2020). *PISA 2018 Technical Report*. Paris: OECD Publishing.

Rees, G., & Taylor, C. M. (2015). *Is there a 'crisis' in Welsh education?* Transactions of the Honourable Society of Cymmrodorion, 97–113.

StatsWales (2021). Speaking Welsh at home, as assessed by parents, of pupils aged 5 and over in primary, middle and secondary schools by year, sector and category. Retrieved from https://statswales.gov.wales/Catalogue/Education-and-Skills/Schools-and-Teachers/Schools-Census/Pupil-Level-Annual-School-Census/Welsh-Language/speakingwelshhomepupils5andover-by-year-sector-category. Last time accessed: February 2021.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (pp. 80–108). New York: Cambridge University Press.

Welsh Government (2017). *Welsh in education: Action plan 2017–21*. Cardiff: Welsh Government.

Wooldridge, J. M. (1995). Score diagnostics for linear models estimated by two stage least squares. In G. S. Maddala, T. N. Srinivasan, and P. C. B. Phillips (Eds.), *Advances in Econometrics and Quantitative Economics: Essays in Honor of Profesor C. R. Rao* (pp. 66–87). Oxford: Blackwell.

Appendix

-Insert Table A1-

Table 1. Percentage of students who took the test in each language and who speak each language at home in Wales

	Language of the test						Language at home					
Language	**2006**	**2009**	**2012**	**2015**	**2018**	**All cycles**	**2006**	**2009**	**2012**	**2015**	**2018**	**All cycles**
English	87%	87%	87%	90%	86%	87%	92%	91%	92%	90%	89%	91%
Welsh	13%	13%	13%	10%	15%	13%	7%	6%	6%	6%	6%	6%
Irish	-	-	-	-	-	-	0.1%<	0.1%<	0.1%<	0.2%	0.1%<	0.1%
Ulster Scots	-	-	-	-	-	-	-	0.1%<	0.1%<	0.2%	0.1%	0.1%<
Other languages	-	-	-	-	-	-	1.0%	2%	2%	4%	5%	3%

Notes: Notes: All OECD recommended practices (final student weights and BRR weights) have been employed. The "-" indicates that there are not data for that region in the PISA cycle.

Source: Authors' own calculations.

Table 2. Percentage of students who took the test in English or Welsh by language spoken at home in Wales

	Language of the test											
	2006		2009		2012		2015		2018		All cycles	
Language spoken at home	**English**	**Welsh**	**English**	**Welsh**	**English**	**Welsh**	**English**	**Welsh**	**English**	**Welsh**	**English**	**Welsh**
English	97%	58%	94%	67%	95%	67%	93%	59%	92%	69%	94%	64%
Welsh	2%	42%	3%	32%	2%	33%	2%	40%	2%	29%	2%	35%
Irish	0.1%	0%	0.1%<	0%	0.1%<	0%	0.2%	0.3%	0.1%<	0%	0.1%	0.05
Ulster Scots	-	-	0.1%<	0%	0.1%<	0%	0.2%	0%	0.1%	0.3%	0.1%<	0.1%<
Other languages	1%	0%	3%	0.3%	3%	0%	4%	1%	6%	1%	3%	0.4%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Notes: All OECD recommended practices (final student weights and BRR weights) have been employed.
Source: Authors' own calculations.

Table 3. Comparison of the demographic characteristics of pupils who took the PISA test in English and Welsh

Variables		Test language: English			Test language: Welsh		
		Obs.	Mean	Standard error	Obs.	Mean	Standard error
Sex of the student	Male	13,467	0.50	0.00	1,942	0.50	0.01
	Female	13,467	0.50	0.00	1,942	0.50	0.01
Socio-economic status quartile	Fourth quartile (Top)	12,936	0.21	0.01	1,824	0.31	0.02
	Third quartile	12,936	0.24	0.00	1,824	0.27	0.01
	Second quartile	12,936	0.27	0.00	1,824	0.24	0.01
	First quartile (bottom)	12,936	0.28	0.01	1,824	0.18	0.01
Father level of education	None	11,391	0.01	0.00	1,547	0.02	0.00
	ISCED 1	11,391	0.01	0.00	1,547	0.01	0.00
	ISCED 2	11,391	0.09	0.00	1,547	0.06	0.01
	ISCED 3b, c	11,391	0.28	0.00	1,547	0.23	0.01
	ISCED 3a, 4	11,391	0.19	0.00	1,547	0.13	0.01
	ISCED 5b	11,391	0.18	0.00	1,547	0.18	0.01
	ISCED 5a, 6	11,391	0.24	0.00	1,547	0.37	0.02
Mother level of education	None	12,164	0.01	0.00	1,650	0.01	0.00
	ISCED 1	12,164	0.01	0.00	1,650	0.01	0.00
	ISCED 2	12,164	0.03	0.00	1,650	0.02	0.00
	ISCED 3b, c	12,164	0.27	0.00	1,650	0.21	0.01
	ISCED 3a, 4	12,164	0.21	0.00	1,650	0.13	0.01
	ISCED 5b	12,164	0.23	0.00	1,650	0.22	0.01
	ISCED 5a, 6	12,164	0.24	0.01	1,650	0.40	0.02
Number of books at home	0 to 10 books	13,081	0.17	0.00	1,849	0.13	0.01
	11 to 25 books	13,081	0.18	0.00	1,849	0.15	0.01
	26 to 100 books	13,081	0.30	0.00	1,849	0.31	0.01
	101 to 200 books	13,081	0.16	0.00	1,849	0.18	0.01
	201 to 500 books	13,081	0.12	0.00	1,849	0.15	0.01
	More than 500 books	13,081	0.07	0.00	1,849	0.08	0.01
Term of birth	First term	13,467	0.25	0.00	1,942	0.24	0.01
	Second term	13,467	0.25	0.00	1,942	0.25	0.01
	Third term	13,467	0.25	0.00	1,942	0.24	0.01
	Fourth term	13,467	0.25	0.00	1,942	0.26	0.01
School funding	Public	12,448	0.98	0.00	1,854	1.00	-
	Private	12,448	0.02	0.00	1,854	0.00	-

Notes: All OECD recommended practices (final student weights and BRR weights) have been employed. "Obs." stands for "Observations".
Source: Authors' own calculations.

Table 4. Average scores for pupils taking the PISA test in English and in Welsh

	English	Welsh
Reading	485***	444***
Mathematics	482***	468***
Science	493***	466***

Notes: All OECD recommended practices (final student weights, BRR weights and plausible values) have been employed. The asterisks indicate if there are significant differences between those who took the PISA test in English (the "English" column) and those who took it in Welsh (the "Welsh" column): *** significant at 1%, ** significant at 5%, * significant at 10%.
Source: Authors' own calculations based upon PISA data for Wales pooled between 2006 and 2018.

Table 5. Influence of taking the test in English on students' competences in Wales

	Reading		Maths		Science	
	OLS	**2SLS**	**OLS**	**2SLS**	**OLS**	**2SLS**
Effect of taking test in England (compared to Welsh)	0.421***	0.392***	0.183***	0.262***	0.363***	0.328***
Standard error	0.036	0.074	0.034	0.069	0.034	0.063
Stock and Yogo (2005) test of weak instruments	-	63.774***	-	63.774***	-	63.774***
Wooldridge (1995) endogeneity test	-	0.226	-	1.880	-	0.380

Notes: PISA recommended practices (final student weights, balanced repeated replication weights and plausible values) have been employed (OECD, 2020). The null hypothesis of the Stock and Yogo (2005) test of weak instruments is that the instrument is weak and the null hypothesis of the Wooldridge (1995) endogeneity test is that the endogenous variable is now exogenous. Complete estimations are presented in Table A1 (Appendix).

Estimation method: Ordinary Least Squares (OLS) and Two-Stage Least Squares (2SLS). The instrument is students' language at home and the rest of variables in the estimation.

Dependent variable: Students' standardised scores in reading, mathematics and science, using Welsh mean and standard deviations in each PISA cycle.

Coefficient: *** significant at 1%, ** significant at 5%, * significant at 10%.

Source: Authors' own calculations.

Table A1. Influence of taking the test in English on students' competences in Wales. Full parameter estimates

	OLS				2SLS		
					Second stage		
Variables	Reading	Mathematics	Science	First stage	Reading	Mathematics	Science
Test language: English (Ref.: Welsh)	0.421***	0.183***	0.363***	-	0.392***	0.262***	0.328***
	(0.036)	(0.034)	(0.034)		(0.074)	(0.069)	(0.063)
Female: Yes (Ref.: no)	0.244***	-0.165***	-0.088***	-0.004	0.244***	-0.165***	-0.088***
	(0.015)	(0.015)	(0.016)	(0.004)	(0.015)	(0.015)	(0.016)
Socio-economic status quartile (Ref.: first quartile)							
Fourth quartile	0.754***	0.807***	0.798***	-0.063***	0.751***	0.815***	0.794***
	(0.025)	(0.024)	(0.026)	(0.011)	(0.026)	(0.025)	(0.027)
Third quartile	0.398***	0.432***	0.434***	-0.042***	0.397***	0.435***	0.432***
	(0.022)	(0.021)	(0.022)	(0.008)	(0.022)	(0.021)	(0.022)
Second quartile	0.242***	0.245***	0.243***	-0.023***	0.241***	0.247***	0.242***
	(0.020)	(0.020)	(0.020)	(0.007)	(0.020)	(0.020)	(0.020)
Socio-economic status quartile. Missing flag	-0.389***	-0.334***	-0.349***	-0.051***	-0.391***	-0.329***	-0.351***
	(0.062)	(0.059)	(0.062)	(0.018)	(0.062)	(0.059)	(0.062)
Grade retention (Ref.: no)							
Repeater	-0.674***	-0.706***	-0.627***	0.041***	-0.673***	-0.710***	-0.626***
	(0.056)	(0.051)	(0.054)	(0.015)	(0.056)	(0.051)	(0.054)
Repeater. Missing flag	-0.555***	-0.451***	-0.576***	0.036	-0.555***	-0.452***	-0.575***
	(0.125)	(0.121)	(0.128)	(0.033)	(0.124)	(0.122)	(0.127)
Country of birth (Ref.: Other countries)							
United Kingdom	0.089	0.089	0.030	-0.010	0.089	0.090	0.029
	(0.058)	(0.057)	(0.056)	(0.013)	(0.057)	(0.057)	(0.056)
Country of birth. Missing flag	-0.302***	-0.312***	-0.380***	0.005	-0.302***	-0.312***	-0.379***
	(0.083)	(0.081)	(0.082)	(0.023)	(0.083)	(0.081)	(0.082)
Father's country of birth (Ref.: Other countries)							
United Kingdom	0.016	-0.014	0.025	-0.026	0.014	-0.010	0.023
	(0.036)	(0.038)	(0.038)	(0.017)	(0.036)	(0.039)	(0.038)
Father's country of birth. Missing flag	-0.172**	-0.218***	-0.159**	-0.036	-0.174**	-0.212***	-0.161**
	(0.069)	(0.067)	(0.069)	(0.025)	(0.069)	(0.067)	(0.069)
Mother's country of birth (Ref.: Other countries)							
United Kingdom	-0.113***	-0.133***	-0.109***	-0.035**	-0.114***	-0.129***	-0.111***

	(0.037)	(0.039)	(0.039)	(0.015)	(0.037)	(0.039)	(0.039)
Mother's country of birth. Missing flag	-0.256***	-0.330***	-0.284***	-0.023	-0.258***	-0.324***	-0.286***
	(0.093)	(0.093)	(0.092)	(0.033)	(0.093)	(0.093)	(0.092)
Lives in the United Kingdom since age 6 or older: yes (Ref.: no)	-0.216***	-0.171**	-0.229***	0.018	-0.216***	-0.172**	-0.228***
	(0.081)	(0.074)	(0.081)	(0.016)	(0.081)	(0.074)	(0.080)
Term of birth (Ref.: Fourth term)							
First term	-0.058***	-0.035*	-0.040**	0.009	-0.058***	-0.036*	-0.040**
	(0.020)	(0.020)	(0.020)	(0.006)	(0.020)	(0.020)	(0.020)
Second term	-0.142***	-0.109***	-0.111***	0.007	-0.142***	-0.109***	-0.111***
	(0.021)	(0.020)	(0.021)	(0.006)	(0.021)	(0.020)	(0.021)
Third term	-0.051**	-0.024	-0.035	0.010	-0.051**	-0.025	-0.034
	(0.021)	(0.021)	(0.022)	(0.006)	(0.021)	(0.021)	(0.022)
School funding (Ref.: public)							
Private	0.543***	0.614***	0.557***	0.110***	0.547***	0.602***	0.562***
	(0.099)	(0.096)	(0.085)	(0.016)	(0.099)	(0.096)	(0.086)
School funding. Missing flag	-0.133**	-0.097**	-0.074	0.013	-0.133**	-0.098**	-0.073
	(0.060)	(0.048)	(0.053)	(0.035)	(0.060)	(0.048)	(0.053)
PISA cycle (Ref.: 2006)							
2018	0.043	0.015	0.025	-0.025	0.043	0.017	0.024
	(0.046)	(0.040)	(0.041)	(0.036)	(0.046)	(0.040)	(0.041)
2015	0.050	0.042	0.029	0.010	0.050	0.041	0.030
	(0.046)	(0.038)	(0.041)	(0.030)	(0.046)	(0.038)	(0.041)
2012	0.024	0.024	0.018	-0.009	0.024	0.024	0.018
	(0.048)	(0.039)	(0.044)	(0.032)	(0.048)	(0.040)	(0.044)
2009	0.027	0.022	0.021	-0.030	0.027	0.024	0.020
	(0.045)	(0.041)	(0.043)	(0.039)	(0.045)	(0.042)	(0.043)
Language at home (Ref.: English)							
Welsh	-	-	-	-0.591***	-	-	-
				(0.037)			
Irish	-	-	-	0.010	-	-	-
				(0.047)			
Ulster Scots	-	-	-	-0.033	-	-	-
				(0.083)			
Other languages	-	-	-	-0.008	-	-	-
				(0.015)			
Constant	-0.716***	-0.299***	-0.488***	1.013***	-0.687***	-0.380***	-0.453***
	(0.078)	(0.078)	(0.075)	(0.028)	(0.102)	(0.101)	(0.092)

Observations	14,951	14,951	14,951	14,951	14,951	14,951	14,951
R-squared	0.168	0.160	0.152	0.208	0.168	0.159	0.152
Stock and Yogo (2005) test of weak instruments	-	-	-	-	63.774***	63.774***	63.774***
Wooldridge (1995) endogeneity test	-	-	-	-	0.226	1.880	0.380

Notes: Standard errors in parenthesis. PISA recommended practices (final student weights, balanced repeated replication weights and plausible values) have been employed (OECD, 2020). The null hypothesis of the Stock and Yogo (2005) test of weak instruments is that the instrument is weak and the null hypothesis of the Wooldridge (1995) endogeneity test is that the endogenous variable is now exogenous.

Estimation method: Ordinary Least Squares (OLS) and Two-Stage Least Squares (2SLS). The instrument is students' language at home and the rest of variables in the estimation.

Dependent variable: Students' standardised scores in reading, mathematics and science, using Welsh mean and standard deviations in each PISA cycle.

Coefficient: *** significant at 1%, ** significant at 5%, * significant at 10%.

Source: Authors' own calculations.