Research Papers in Education

ISSN: 0267-1522 (Print) 1470-1146 (Online) Journal homepage: http://www.tandfonline.com/loi/rred20

How teachers' practices and students' attitudes towards technology affect mathematics achievement: results and insights from PISA 2012

Javier Tourón, Enrique Navarro-Asencio, Luis Lizasoain, Emelina López-González & María José García-San Pedro

To cite this article: Javier Tourón, Enrique Navarro-Asencio, Luis Lizasoain, Emelina López-González & María José García-San Pedro (2018): How teachers' practices and students' attitudes towards technology affect mathematics achievement: results and insights from PISA 2012, Research Papers in Education

To link to this article: https://doi.org/10.1080/02671522.2018.1424927

Published online: 11 Jan 2018.

RESEARCH PAPERS IN EDUCATION, 2018
https://doi.org/10.1080/02671522.2018.1424927

How teachers' practices and students' attitudes towards technology affect mathematics achievement: results and insights from PISA 2012

Javier Tourón[a], Enrique Navarro-Asencio[b], Luis Lizasoain[c], Emelina López-González[d] and María José García-San Pedro[a]

[a]Department of Research Methods and Diagnosis in Education, Universidad Internacional de La Rioja (UNIR), Logroño, Spain; [b]Department of Research Methods and Diagnosis in Education, Universidad Complutense de Madrid, Madrid, Spain; [c]Department of Research Methods and Diagnosis in Education, Universidad del País Vasco, San Sebastián, Spain; [d]Department of Research Methods and Diagnosis in Education, Universidad de Valencia, Valencia, Spain

ABSTRACT

The present work seeks to deepen the impact of factors linked to the characteristics of teaching practices and students' attitudes towards the use of technology on their performance in mathematics in the process of teaching-learning in the Spanish context. In this sense, this study is a secondary analysis of the PISA 2012 data. Therefore, it is an ex post facto design. Regarding the attitudes and the contextual variables, the results do coincide with the accumulated evidence. However, once these contextual effects have been controlled for, the negative relationship found between the pedagogic strategies used by the teachers and the mathematics score cannot but convey perplexity, since the results relative to student-oriented, formative assessment and teacher-directed instruction are clearly contradictory to the solid previous evidence. The data do not allow us to explain this paradoxical result. We dare to point to a conjecture that we find plausible. All these complex variables are informed through questionnaires responded to by students and require a great degree of inference in the answers. Future studies must consider the complexity of the measured variables as well as the students' perception and understanding of them.

ARTICLE HISTORY
Received 28 April 2017
Accepted 7 December 2017

KEYWORDS
Mathematics achievement; Programme of International Student Assessment (PISA); teaching practice; students' attitudes; technology

Teachers' practices and students' attitudes towards technology in mathematics achievement in PISA 2012

The Programme for International Student Assessment (PISA) is a study developed by the Organisation for Economic Cooperation and Development (OECD) that evaluates the level of knowledge acquired by 15-year-old students undergoing compulsory education regarding the information and skills needed to participate actively in modern societies. The 2012 edition of PISA centred on mathematics.

CONTACT Javier Tourón javier.touron@unir.net

The present work seeks to deepen the impact of factors linked to the characteristics of teaching practices and students' attitudes towards the use of technology on their performance in mathematics in the process of teaching-learning in the Spanish context.

Literature overview

The concept of good teaching is based on the supposition that a faculty organises a series of resources and activities in order to develop the competences and meet the objectives of the students' current educational stage. The Analytical Framework of PISA 2012 (OECD 2013) establishes that good teaching promotes self-regulated learning and metacognition and develops the cognitive processes that underpin problem solving. It prepares students to reason effectively in unfamiliar situations and to fill gaps in their knowledge by observation, exploration and interaction with unknown systems. Good teaching also includes the development of metacognitive strategies and effective feedback (Hattie and Timperley 2007). Effective feedback creates the possibility of enhancing students' experiences by starting from the information offered by an agent about their performance or a fulfilled task.

The role teachers and their educational practice play has been emphasised in a number of studies as a decisive element in the results of a learning process. For instance, Chetty, Friedman, and Rockoff (2014) found relevant differences in the later skills of students who had had good teachers. Bietenbeck (2014) analysed the effect of traditional and modern teaching practices on cognitive abilities such as accumulated formal knowledge, routine problem-solving skills and reasoning skills, finding that traditional teaching practises are efficient in improving the first two while more modern practises significantly increase reasoning and problem-solving skills.

Hattie's interesting meta-analytical studies on performance (2009, 2012, 2015) identified six main areas that contribute to learning: student, home, school, curricula, teacher and teaching and learning approaches. Considering that 50% of the learning variance was discovered to depend on each student's personal background (previous knowledge, motivation, commitment with the learning process, study habits, preferences and, more significantly, the student's attitude towards the particular study subject), the second most relevant variance source was a successful teacher (20–25%, Hattie 2015). This concept is related to visible learning, which promotes a more relevant role for teachers in terms of assessing their own teaching (Hattie 2012). Teaching and learning are visible inasmuch as teachers can see learning through their students' eyes. In this respect, good teaching should also promote self-regulated learning and metacognition, i.e. a development of metacognitive strategies and effective feedback (Hattie and Timperley 2007) both teacher-oriented (Clayson 2008) and student-oriented, as well as feedback among teachers (Hattie 2015). A successful teacher receives feedback of their teaching impact and is able to ponder it, being the whole process of significant benefit to the students (Clinton and Hattie 2014). Consequently, those data obtained from original performance or learning results studies are not as relevant as the underlying history, which leads to the need of feedback processes.

However, considering that measurements obtained from the questionnaires answered by PISA students on their teachers' performance are analysed in this study (in other words, feedback measurements on teachers' performance), it is not unusual to believe that this type of measurements and the actual performance of teachers are highly connected (Sullivan and Skanes 1974; Frey, Leonard, and Beatty 1975) but, as Clayson (2008) remarks in his meta-analysis, there has been no study since 1990 that has shown a positive significant connection between learning measurements and teachers' performance assessment by their students.

Regarding the relation between ICT and students, Biagi and Loi (2013) state that the number and hence the diversification of activities, irrespective of the intensity of computer use, is positively correlated with students' proficiency in all three PISA domains in the vast majority of countries. Moreover, Delen and Bulut (2011) note two interesting contributions. On the one hand, the way in which students use computers in schools to attain learning outcomes may have a stronger effect on scientific literacy than how often computers are accessed. In contrast, students with prior experience with basic ICT tasks earned higher scientific literacy scores.

Some critical points that arise from these coincidences are the lack of integration and normalisation of new technologies in the teaching process and in the curriculum in general (Biagi and Loi 2013). In relation to gender issues, the results of the analysis of Thomson and De Bortoli (2007) show that female students are perceived as being less confident than males in their skills when it comes to higher level tasks and that they see computers more as a tool than males do. Hence, the need to deepen students' attitudes towards information and communications technology (ICT) to promote didactic designs that positively impact their learning is justified.

Given that the results of the research do not coincide, the aim of this study is to deepen into these matters by analysing the relation between teaching practises and students' attitudes towards technology in their mathematics achievement, by means of PISA database.

Methodology

This study is a secondary analysis of the PISA 2012 data. Therefore, it is an *ex post facto* design whose main purpose is to study the impact of factors linked to educational practices, strategies used by teachers in classrooms and students' attitudes towards technology use in the process of teaching-learning.

As usual in large-scale assessments, together with the measure of the achievement level, a group of context questionnaires is administered to obtain greater information regarding the students' educational and family environments. In this case, information about the educational methodology developed in the classrooms, or, more precisely, the perceptions or opinions the students have of that process was used. In addition, PISA has some specific characteristics that affect the sampling selection process and the treatment of the resulting data.

Sample

The sampling used by PISA matches a stratified design of two stages, with the schools as primary sampling units (PSUs) in the first stage and a sampling of 15-year-old students in the second stage. A weight variable was included to appropriately calculate the punctual estimators of each variable.

In this work, the sample of Spanish students is composed of 373,691 individuals (considering the weight variable) distributed in a total of 901 schools, 550 public and 351 private. The student body has an average age of 15.86 (SD = 0.29) and is 49.2% female and 50.8% male.

In the PISA evaluation, not all the evaluated subjects answer the same questions on the performance tests; therefore, the error variance is estimated by means of the resampling procedure known as JRR or JK2, a variant of the jackknife method. Therefore, for the performance variables, 80 replicas are used, which are generating 80 weight variables that allow us to obtain a better estimate of the sampling variance of each estimator.

Variables

Dependent variable

Mathematics performance was selected as the dependent variable for this work. Each variable of the achievement of the students in the sample is composed of five plausible values. Thus, instead of obtaining a punctual estimator of the measured competence for each student, an *a posteriori* distribution is obtained for each, and five values are then randomly extracted and denominated as plausible values. This procedure has the advantage of allowing a better estimate of the measure error variance.

Independent variables

Factors and indexes. PISA Technical Report lists and describes the scales and indexes produced that are based on the items included in its questionnaires. In this study, some of the mentioned indexes are considered as independent variables. In accordance with our established aims, we have specifically focused on those rates specified on Table 1, being the items that compose them also highlighted there.

The first 13 items relate to teacher-directed instruction, student orientation, formative assessment. Each item is measured with a four-point Likert scale linked to the frequency of use of that methodology (every lesson, most lessons, some lessons and never or hardly ever). Additionally, this work uses the indexes that PISA provides for each of these three dimensions, which may be considered interval variables, usually distributed with a mean of 0 and a standard deviation of 1.

The second group of nine items is related to teaching methodology and, more specifically, to the support given by the teacher to the student through cognitive activation (teacher support) processes. Here, the Likert scale is linked to the frequency of use of each strategy (always or almost always, often, sometimes and never or rarely).

Finally, the last six items are related to the attitudes of the students towards the technology used in the classrooms. In this case, the Likert scale identifies four agreement points (Strongly Agree, Agree, Disagree and Strongly Disagree). In addition, PISA provides two indexes, the first compounded of questions that note a positive attitude towards the use of ICT (for example, regarding its function as a tool in the school) and the second for those who look at ICT as a limiting factor for learning in the school.

According to what was previously mentioned, these indexes are defined and designed by PISA (see technical report, 2012, 312 et seq.). The values for those indexes were estimated with Rasch models and, in the particular case of ordinal items, the Masters & Wring adaptation was used (partial credit model). PISA calculates each rate separately and the

Table 1. Dimensions and items.

Dimension (PISA index)	Item
Teacher-directed instruction	Sets clear goals
	Encourages thinking and reasoning
	Checks understanding
	Summarises previous lessons
	Informs about learning goals
	Differentiates between students when giving tasks
Student orientation	Assigns complex projects
	Plans classroom activities
	Has students work in small groups
Formative assessment	Gives feedback on strengths and weaknesses
	Informs about expectations
	Gives feedback
	Tells how to get better
Cognitive activation (teacher support)	Encourages reflection on problems
	Gives problems that require thought
	Asks to use own procedures
	Presents problems with no obvious solutions
	Presents problems in different contexts
	Helps learn from mistakes
	Asks for explanations
	Has students apply what they learned
	Problems with multiple solutions
Attitudes towards technology	Useful for schoolwork (+)
	Makes homework more fun (+)
	A source of information (+)
	Troublesome (–)
	Not suitable for schoolwork (–)
	Unreliable (–)

connections made among countries are analysed to validate the construct. The indexes used in this study are as reliable as those in the Technical Report.

Socio-economic and cultural status. Socio-economic and cultural status (ESCS) is used in this study as a control variable for its proven relationship with academic achievement. PISA suggests including it in the studies that analyse its results (OECD 2009). The scores of this index are normally distributed with a mean of 0 and a standard deviation of 1 in the complete sample; in Spain, it has the values shown in Table 2, where the ESCS and the previously mentioned index statistics are shown.

Data analysis

The sample design used by PISA requires the use of appropriate resampling procedures if comparisons of groups will be carried out in the data analyses. Additionally, the plausible

Table 2. Teacher dimensions, technology use and ESCS values for Spain.

	N	Min.	Max.	Mean	SD
Teacher behaviour: teacher-directed instruction	245,466	−3.653	2.563	−0.127	0.986
Teacher behaviour: student orientation	245,366	−1.600	3.311	−0.143	1.045
Teacher behaviour: formative assessment	244,748	−2.392	2.630	−0.061	1.071
Teacher support	245,627	−2.920	1.680	0.117	1.036
Computer as a tool for school learning	343,226	−2.900	1.305	0.210	0.909
Limitations of computer as a tool for school learning	342,387	−2.158	2.408	0.120	0.992
ESCS	370,423	−5.300	2.730	−0.190	.0260

values obtained from the *a posteriori* distribution for each subject must be used suitably. For this reason, the SPSS macros offered in the OECD PISA Data Analysis Manual (2009, 203, 217) have been used to estimate multilevel models with PISA's evaluation data.

First, the work starts from a study of the effects of the different afore-mentioned indexes: the teacher methodology indicators and attitudes towards ICT together with ESCS. Additionally, the schools' sectors and the ESCS school mean were included in this first two-level (students, schools) lineal-hierarchical model (Gaviria and Castro 2005; Snijders and Bosker 2009):

$$y_{ij} = \beta_0 + \beta_1 \text{ESCS} + \beta_2 \text{ESCS}_{\text{mean}} + \beta_3 \text{Sector} + \beta_X \text{Indexes} + \left(\mu_{0j} + \varepsilon_{ij}\right)$$

On one hand, the pattern has as fixed coefficients the intercept of the regression model (β_0); the effect of students' ESCS and its mean for schools (β_1 and β_2), respectively; the sector effect (β_3); and the effects of each of the described indexes (β_X). On the other hand, the model consists of random effects associated with the schools and the students ($\mu_{0j} + \varepsilon_{ij}$).

Second, a separate multilevel regression model has been developed for each item of the teaching methodology dimensions and attitudes towards ICT. As these variables are ordinals with four categories, contrast coding has been used for their inclusion in the analyses, thus creating a model with four predictors in which the intercept (β_0) is the yield average of the subjects that notes the first category of each item, and the β_1, β_2 and β_3 predictors are the differences of the subjects that mark the 2, 3 or 4 categories, respectively.

$$y_{ij} = \beta_0 + \beta_1 C2 + \beta_2 C3 + \beta_3 C4 + \left(\mu_{0j} + \varepsilon_{ij}\right)$$

Therefore, the predictors can be considered to be average increments (or decreases) in performance regarding subjects who select the first category in the Likert scale for each item.

Third, to characterise the type of opposing effect in the estimated analyses of multilevel regression for each of the factors, the effect sizes have been calculated through the difference of the standardised averages between the two categories with better and worse mathematics performance results.

Due to the specificity of the analyses and the information provided by the results of the multilevel model, the calculation uses the average of the two afore-mentioned groups, their sample sizes and the standard deviation of the whole sample, as Lipsey and Wilson proposed (2001):

$$S_{\text{pooled}} = \sqrt{\frac{S^2(N-1) - \frac{(\bar{X}_{G1}^2 + \bar{X}_{G2}^2 - 2\bar{X}_{G1}\bar{X}_{G2})(n_{G1}n_{G2})}{(n_{G1}+n_{G2})}}{N-1}}$$

The interpretation of the effects can be considered, following Cohen (1988), as a small effect if the value is less than or equal to 0.2, a moderate effect if it is equal to 0.5 and a large effect if it is greater than or equal to 0.80.

Results

Table 3 shows the results of the whole multilevel model. The average mathematics score is 487.31. The contextual effects are as follows: for each increment unit in the ESCS at the

Table 3. Multilevel model parameters.

Parameter	Statistic	SE
Intraclass correlation	0.105	0.002
Level 2 residual variance (schools)	569.836	30.636
Level 1 residual variance (students)	4836.901	145.936
Intercept	487.310	1.854
ESCS	23.639	1.095
ESCS school mean	21.043	1.645
Sector (public/private)	6.419	1.192
Teacher-directed instruction	1.066	1.500*
Student orientation	−17.169	1.238
Formative assessment	−4.350	1.172
Teacher support	8.283	1.639
ICT attitudes (negative)	−11.448	1.211
ICT attitudes (positive)	3.140	1.213

*Non-significant coefficient.

individual level, 23.64 points are added; if that increment is given in the school mean ESCS, another 21.04 points are added. Last, the effect of the school's sector is 6.41 points in favour of the private schools.

The effects of the compound indexes are as follows: regarding the educational methodology, we did not find a significant relation between the teacher-directed instruction index and the mathematics results. In addition, in the other two dimensions (student orientation and formative assessment), the effect is negative. This finding is especially noteworthy in regard to student guidance, where each incremental point supposes a descent of 17 points in the dependent variable. In contrast, the relative effects regarding teacher support are direct and add 8.28 incremental points in mathematics. Lastly, the effects associated with attitudes towards technology are direct and are especially relevant in the cases of negative attitude, so that 1 incremental point in negative attitude supposes a drop of 11.45 points in mathematics performance. In the cases of positive attitude, the increment is only 3.14.

To examine these effects in more detail, the item-level results are presented next.

Table 4 shows the questionnaire items regarding the dimensions teacher-directed instruction, student orientation and the uses of formative assessment. For each item, the average score is shown as a function of the use frequency. The last column shows the value of the effect size (Cohen's *d*) among the extreme categories (which appear in bold).

Of the 13 items, only in 3 can the effect size be considered important (superior at 0.7); all three belong to the student orientation dimension. In the remnant, the effects are moderate or very weak. In five items, the highest score in mathematics is associated with the smallest frequency of employment of the instructive practice under consideration (never or hardly ever).

The biggest difference (506-445, Cohen's $d = 0.947$) appears in regard to classroom activity planning; the students who state that their teachers never or hardly ever plan their activities are those who achieve the highest scores in mathematics. Conversely, the students who report that their teachers do plan classroom activities for every lesson turn out to be those obtaining lower scores.

The same inverse trend is shown by the other four items; we found that 'assigns complex projects', 'has students work in small groups', 'differentiates between students when giving tasks' and 'gives feedback on strengths and weaknesses' are practices whose use frequency is inversely related to mathematics performance.

Table 4. Mathematics achievement of Spanish students on the PISA 2012 and effect size (Cohen's *d*) based on the frequency of teacher-directed instruction, student orientation and formative assessment practices.

	Every lesson	Most lessons	Some lessons	Never or hardly ever	Cohen's *d*
Teacher-directed instruction					
Sets clear goals	492.081	**497.201**	493.449	**487.679**	0.147
Encourages thinking and reasoning	**480.396**	489.036	500.577	**502.008**	0.329
Checks understanding	494.171	**496.753**	492.739	**491.298**	0.074
Summarises previous lessons	**484.475**	495.610	**498.289**	494.194	0.207
Informs about learning goals	**487.250**	**499.550**	496.189	495.870	0.177
Differentiates between students when giving tasks	**471.691**	474.467	488.876	**504.213**	0.500
Student orientation					
Assigns complex projects	**455.646**	468.295	487.341	**507.099**	0.800
Plans classroom activities	**445.041**	473.129	491.09	**506.137**	0.947
Has students work in small groups	**459.978**	462.831	490.719	**505.787**	0.703
Formative assessment					
Gives feedback on strengths and weaknesses	**473.095**	484.750	495.476	**503.007**	0.448
Informs about expectations	**483.578**	492.456	496.864	**499.743**	0.237
Gives feedback	**480.378**	494.068	**498.696**	497.932	0.267
Tells how to get better	**478.262**	492.241	500.343	**501.602**	0.343

For the rest of the items, the effects are very weak, and the maximum difference usually appears among the frequencies of never or hardly ever and some of the central values (most lessons or some lessons).

The second studied dimension (cognitive activation/teacher support) consists of nine items. Table 5 shows the results with the same format used in the previous case.

Table 5. Mathematics achievement of Spanish students on the PISA 2012 and effect size (Cohen's *d*) based on the frequency of teachers' cognitive activation.

Cognitive activation	Always or almost always	Often	Sometimes	Never or rarely	Cohen's *d*
Encourages reflection on problems	**488.620**	495.755	**498.201**	491.143	0.147
Gives problems that require thought	**486.520**	495.923	**497.537**	493.582	0.162
Asks to use own procedures	**491.239**	492.263	**496.947**	496.086	0.074
Presents problems with no obvious solutions	495.837	**499.504**	**495.786**	483.471	0.237
Presents problems in different contexts	498.126	**498.779**	**487.159**	491.110	0.162
Helps learn from mistakes	493.770	**497.769**	494.715	**486.229**	0.162
Asks for explanations	**495.67**	494.286	493.645	**493.371**	0.029
Has students apply what they learned	**501.979**	493.581	488.980	**485.074**	0.236
Problems with multiple solutions	494.139	**497.050**	495.601	**478.132**	0.281

The first distinctive feature of this block is that all the effects are very weak or negligible. The maximum value of Cohen's *d* is 0.28 in the item referring to the use of 'problems with multiple solutions'.

The second distinctive feature is that of the nine items, in the first three ('encourages reflection on problems'; 'gives problems that require thought'; 'asks to use own procedures') the trend is inverse so that, as occurred in the previous dimension, the greater the use frequency, the lower the scores.

Next, in this block, the biggest differences are not between the use frequency of two extreme values (always vs. never) but between one of them and a central value (often, sometimes). This pattern could indicate that in this case, an optimum choice is made that corresponds not with the extreme values but with a moderate or relatively frequent use of the employed didactic strategy.

An example is the item 'problems with multiple solutions', which is also the item showing the greatest effect ($d = 0.28$). In this case, when the teacher never outlines problems of this type, the yield in mathematics is 478 points. In contrast, if such problems are often presented, the score increases to 497 points.

An exception to be noted is the item 'has students apply what they learned', in which the highest value in mathematics (502) is achieved when the students always or more often than not employ what they have learned, while the lowest score (485) is given when they never do so.

Despite the fact that this pattern is not completely clear, it is also true that, in certain cases, the level of achievement increases as the item frequency does but, at a certain point, achievement shrinks. For instance, in the case of the item 'sets clear goals', the achievement rate increases in 'most lessons' and decreases when it becomes 'Every lesson'; that is, it changes at the last situation. The same happens in the cases of 'checks understanding', 'Helps learn from mistakes', 'asks for explanations', 'presents problems with no obvious solutions', 'helps learn from mistakes', 'problems with multiple solutions', 'useful for schoolwork'. In other cases, change happens at the third item category ('summarises previous lessons', 'gives feedback', 'encourages reflection on problems', 'gives problems that require thought', 'asks to use own procedures'). This is a descriptive interpretation obtained when analysing the average results for each item category and that would require a more detailed analysis in future studies. A similar trend has been noticed in the meta-analytical review by Castro et al. (2015) on parental involvement and academical performance.

Table 6 shows the values of the third and last studied factor, which refers to the relationship between attitudes towards ICT and performance in mathematics.

In this case, except for item 2 ('makes homework more fun'), all the effects are appreciable, and the tendency is positive, so that a more favourable attitude towards technology is

Table 6. Mathematics achievement of Spanish students on the PISA 2012 and effect size (Cohen's *d*) based on students' attitudes towards technology.

Technology is	Strongly agree	Agree	Disagree	Strongly disagree	Cohen's *d*
Useful for schoolwork	495.894	**498.739**	489.082	**460.673**	0.562
Makes homework more fun	491.625	499.097	**501.267**	**485.324**	0.236
A source of information	**499.030**	494.330	469.580	**457.084**	0.620
Troublesome	**475.431**	486.523	**504.054**	500.440	0.434
Not suitable for schoolwork	**470.370**	490.717	**506.004**	499.223	0.543
Unreliable	**463.033**	481.547	**506.620**	502.479	0.651

associated with better performance in mathematics. In this case, the Likert scale refers not to use frequency but to the degree of agreement or disagreement with the statement. It is important to note that the last three items are of negative polarity, so that the greater the disagreement, the greater the performance.

Thus, the students who agree or strongly agree that technology is useful for schoolwork, makes homework more fun, or is a good source of information or those who strongly disagree or disagree that it is troublesome, not suitable for schoolwork, or unreliable have mathematics results significantly superior to those of their peers who manifest opposing attitudes, with a difference that reaches over 40 points.

To conclude this section, we would like to remark that every analysis has been made from a multi-level perspective, that a school-level random variance has been taken into consideration, that is, that every category average has been represented but there is a significant variance among students and among schools. However, it was decided that the global results of the multi-level analysis would be submitted without going into detail on each group, as it would have implied that 900 schools should be represented. In either case, a more detailed analysis of level two remnants would allow a characterisation of the variety of schools.

Discussion

The results presented in the previous section are in large part contradictory to the available evidence in the literature of educational investigation.

Regarding the attitudes and the contextual variables, such as ESCS and sector, the results do coincide with the accumulated evidence that students belonging to families with a higher ESCS tend to obtain better academic results. Additionally, if these students attend schools where the average ESCS is also high, their score increases again. Finally, if those schools belong to the private sector, this leads to another performance improvement. These findings have been repeatedly evidenced since the Coleman report. The issue is noted in the reports of the OECD itself (2009, 2010, 2013, 2015), and the contextual effects have been studied by Willms (2010).

However, once these contextual effects have been controlled for, the negative relationship found between the pedagogic strategies used by the teachers and the mathematics score cannot but convey perplexity, since the results relative to student-oriented, formative assessment and teacher-directed instruction are contradictory to the previous evidence, what consequently leads us to Hattie's thesis, based on his meta-analytical research (Hattie 2009, 2012, 2015): the need to deepen in the data-underlying history.

It is important to remark that the research conducted by the OCDE itself points in the same direction (see Echazarra et al. 2016) and affirms that students with lower mathematics scores are those who are more frequently exposed to student-oriented, teacher-directed instruction that includes formative assessments. A possible explanation is that the teachers may use these types of strategies more frequently with students who show greater learning difficulties (Echazarra et al. 2016, 62).

With regard to the dimension of cognitive activation, or teacher support, the results are less clear. At a general level, the effect is positive but has little relevance. However, when the items are examined separately, the results show the coexistence of some items with a negative relation and others with a positive relation. In the latter, it is possible that an optimum local value refers not to the extreme values but to a moderate or relatively frequent employment of the didactic strategy under consideration.

Nonetheless, the submitted results raise some problems because of their lack of compatibility with the investigation seemingly should show. What does it mean that the items most frequently used in the dimension of cognitive activation, *teacher-directed instruction* or the use of *feedback*, have a detrimental effect on performance? How is it possible that 'encourages reflection on problems', 'has students apply what they learned' or 'gives feedback on strengths and weaknesses', to name a few (see Tables 4 and 5), have a negative effect on performance?

This work does not contribute direct answers regarding this paradoxical result, since the data do not allow us to answer these questions.

However, these results are connected to problems detected at the teachers' performance assessment by students and its relation with performance measurements (Clayson 2008). Although it is widely accepted that there is a positive relation between student views and performance (Sullivan and Skanes 1974; Frey, Leonard, and Beatty 1975) the more objective the measurement process becomes for both the learning process and the students' views on the teachers' performance, the more the relation between these two construct diminishes (Johnson 2003; Weinberg, Fleisher, and Hashimoto 2007), being particularly relevant the way performance is operationalised. Among those relevant studies with performance measurements of an objective nature to a certain extent there are those of McKeachie (1987) and Rodin and Rodin (1972). The latter is a turning point on PISA performance measurements and also on those procedures related to variables and constructors on teaching practices.

We dare to point to a conjecture that we find plausible. All these variables are informed through questionnaires responded to by students. They are complex variables requiring a great degree of inference in the answers. Thus, we think the data cannot be obtained in this manner.

Hattie (2015) remarks that a common criticism to performance meta-analysis is how survey tools are used with students. According to that argument, it is not which correlates are significant in performance justification what matters, but which correlates are better, consequently implying the need to deepen into the data-underlying history. In accordance with that, the PISA contextual questionnaires need to be redesigned if we intend them to have utility in finding valid data to analyse the relationships among the measurement variables and performance. Future studies must consider the complexity of the measured variables as well as the students' perception and understanding of them.

The same issues do not occur in the case of attitudes towards technology, where we find the expected results in the presumably correct direction. The difference could be precisely that the degree of inference is minimal, and therefore, the students are able to give answers that are more reflective of reality.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by UNIR Research (http://research.unir.net), Universidad Internacional de La Rioja (http://www.unir.net), within Research Plan 3 [2015-2017].

Notes on contributors

Javier Tourón is a vice-chancellor of Innovation and Educational Development, Universidad Internacional de La Rioja (UNIR). His research focuses on evaluation and assessment in education systems, talent development and educational technology. He is flipped learning international ambassador and research fellow of the Flipped Learning Global Initiative (flglobal.org). http://www.javiertouron.es/.

Enrique Navarro-Asencio is a lecturer at Universidad Complutense de Madrid. His research focuses on psychometrics and assessment in education.

Luis Lizasoain is a senior lecturer at Universidad del País Vasco. His research focuses on large-scale educational assessments and the application of hierarchical linear models.

Emelina López-González is a senior lecturer at Universidad de Valencia. Her research focuses on multivariate statistical analysis and data visualisation, linear predictive models, generalised linear models and non-linear models.

María José García-San Pedro is a lecturer at Universidad Internacional de La Rioja (UNIR). Her research focuses on teacher training and student assessment, and she is a member of the EVASSE research group.

ORCID

María José García-San Pedro http://orcid.org/0000-0002-7361-3778

References

Biagi, F., and M. Loi. 2013. "Measuring ICT Use and Learning Outcomes: Evidence from Recent Econometric Studies." *European Journal of Education* 48: 28–42. doi:10.1111/ejed.12016.
Bietenbeck, J. 2014. "Teaching Practice and Cognitive Skills." *Labour Economics* 30: 143–153.
Castro, M., E. Expósito-Casas, E. López-Martín, L. Lizasoain, E. Navarro, and J. L. Gaviria. 2015. "Parental Involvement on Student Academic Achievement: A Meta-analysis." *Educational Research Review* 14: 33–46. doi:10.1016/j.edurev.2015.01.002.
Chetty, R., J. N. Friedman, and J. Rockoff. 2014. "Measuring the Impact of Teachers I: Evaluating, and Perseverance: Critical Factors for Success in Teacher Value-added Estimates." *American Economic Review* 104 (9): 2593–2632. doi:10.1257/aer.104.9.2593.
Clayson, D. E. 2008. "Student Evaluations of Teaching: Are They Related to What Student Learn? A Meta-analysis and Review of the Literature." *Journal of Marketing Education* 31 (1): 16–30.
Clinton, J. M., and J. A. C. Hattie. 2015. "Teachers as Evaluators: An Empowerment Evaluation Approach." In *Empowerment Evaluation: Knowledge and Tool for Self-Assessment, Evaluation Capacity Building, and Accountability*, edited by D. Fetterman, S. Kaftarian, and A. Wandersman, 86–111. Thousand Oaks, CA: Sage.
Cohen, J. 1988. *Statistical Power Analysis for the Behavioural Sciences*. New York: Routledge Academic.
Delen, E., and O. Bulut. 2011. "The Relationship between Students' Exposure to Technology and Their Achievement in Science and Math." *Turkish Online Journal of Educational Technology (TOJET)* 10: 311–317.
Echazarra, A., D. Salinas, I. Méndez, V. Denis, and G. Rech. 2016. "How Teachers Teach and Students Learn: Successful Strategies for School." *OECD Education Working Papers*, 130. Paris: OECD Publishing. doi:10.1787/19939019.
Frey, P. W., D. W. Leonard, and W. W. Beatty. 1975. "Student Ratings of Instruction: Validation Research." *American Educational Research Journal* 12: 435–444.
Gaviria, J. L., and M. Castro. 2005. *Los modelos jerárquicos lineales* [Hierarchical Linear Models]. Madrid: La Muralla.

Hattie, J. 2009. *Visible Learning: A Synthesis of over 800 Meta-analyses Relating to Achievement*. London: Routledge.
Hattie, J. 2012. *Visible Learning for Teachers: Maximizing Impact on Learning*. London: Routledge.
Hattie, J. 2015. "The Applicability of Visible Learning to Higher Education." *Scholarship of Teaching and Learning in Psychology* 1: 79–91. doi:10.1037/stl0000021.
Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77: 81–112. doi:10.3102/003465430298487.
Johnson, V. E. 2003. *Grade Inflation: A Crisis in College Education*. New York: Springer.
Lipsey, M. W., and D. B. Wilson. 2001. *Practical Meta-analysis*. Thousand Oaks, CA: Sage.
McKeachie, W. J. 1987. "Can evaluating instruction improve teaching?" *New Directions for Teaching and Learning* 1987: 3–7. doi:10.1002/tl.37219873103.
OECD. 2009. *PISA 2006 Technical Report*. Paris: OECD Publishing. https://www.oecd.org/pisa/data/42025182.pdf.
OECD. 2010. *Mathematics Teaching and Learning in PISA*. Paris: OECD Publishing. doi:10.1787/9789264039520-en.
OECD. 2013. *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. Paris: OECD Publishing. doi:10.1787/9789264190511-en.
OECD. 2015. *PISA Technical Report 2012*. Paris: OECD Publishing. https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf.
Rodin, M., and B. Rodin. 1972. "Student evaluations of teachers." *Science* 177: 1164–1166.
Snijders, T., and R. Bosker. 2009. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*. London: Sage.
Sullivan, A. M., and G. R. Skanes. 1974. "Validity of Student Evaluation of Teaching and the Characteristics of Successful Instructors." *Journal of Educational Psychology* 66: 584–590.
Thomson, S., and L. De Bortoli. 2007. *PISA 2003 Australia: ICT Use Familiarity at School and Home*. Australian Council for Educational Research (ACER) Research Monograph, 62. https://www.acer.edu.au/files/Mono_62_PISA-ICT_Report.pdf.
Weinberg, B. A., B. M. Fleisher, and M. Hashimoto. 2007. *Evaluating Methods for Evaluating Instruction: The Case of Higher Education*. (NBER Working Paper No. 12844). http://www.nber.org/papers/w12844.
Willms, J. D. 2010. "School Composition and Contextual Effects on Student Outcomes." *Teachers College Record* 112: 1008–1037.