Guidelines for data fusion with international large scale assessments: Insights from the TALIS-PISA link

Gil-Izquierdo, María and Cordero, José Manuel

Universidad Autónoma de Madrid, Universidad de Extremadura

June 2017

Guidelines for data fusion with international large scale assessments: Insights from the TALIS-PISA link

María Gil Izquierdo[1], José Manuel Cordero[2]

1 Universidad Autónoma de Madrid, 28049, Madrid, Spain
[2] Universidad de Extremadura, 06006, Badajoz, Spain

Abstract

The educational effectiveness research has experienced a substantial improvement in the last decades thanks to the refinement of large-scale international assessments. Those surveys provide researchers and policy makers with comparative micro data that can be exploited in cross-national studies in order to evaluate educational policies or determinants of educational achievement. This paper focuses on the potential uses and misuses that can be made with the so-called TALIS-PISA link created by the OECD. This is a recently developed instrument that allows for connecting data about teacher characteristics and practices collected in TALIS with students´ academic performance measured in PISA. However, the statistical and technical aspects regarding this link between both surveys are far from straightforward. In this paper we explore the main problematic issues of the data fusion process and provide some guidelines for researchers interested in performing empirical analyses using the resulting dataset.

Keywords: Education, Teachers, International datasets, Large-scale assessments, PISA.

JEL codes: I21; H52; C13.

1. Introduction

It is widely accepted within the educational research community that teachers play a pivotal role in the education sector (Creemers, 1994; Hanushek, 2011). For several decades, researchers have examined the associations between student achievement and a wide variety of teacher variables, including background characteristics, their beliefs and attitudes and the instructional practices applied in the classroom (Palardy and Rumberger, 2008; Boonen et al., 2014). However, the relationships have often been difficult to quantify and understand empirically because there are many factors that might have influence on this relationship (Rockhoff, 2004). As a result, there is still a lack of consensus about which aspects of teachers matter most (Nye et al., 2004, Rivkin et al., 2005; Hattie, 2009).

Until relatively recently, the majority of the available empirical evidence on this topic was referred to the specific context of the United States, since data about teachers were only available in those countries. However, the remarkable development of international large-scale assessments (ILSA) over the past two decades offer researchers new opportunities to explore relationships between teachers´ characteristics and their instructional practices and learning outcomes (Rowan et al., 2002; Chapman et al. 2012) in other countries or even using a cross-country approach. Perhaps the best known ILSAs are the International Association for the Evaluation of Educational Achievement (IEA) Trends in Mathematics and Science Study (TIMSS) and the Organisation for Economic Cooperation and Development (OECD) Programme for International Student Assessment (PISA) and Teaching and Learning International Survey (TALIS).

Most part of recent research on teacher effects with international datasets uses data from TIMSS (Mullis et al., 2012)[1], since this is the only ILSA that provides data on students, teachers and schools. For instance, Schwerdt and Wuppermann (2011) and Van Klaveren (2011) use TIMSS 2003 data for US and Netherlands, respectively, to examine the influence of teaching practices on student achievement. House (2009) and Bietenbek (2014) analyze the effect of different types of instruction using data from TIMSS 2007 for fourth-grade students in Japan and US eight-grade students, respectively. Zuzovsky (2013) and O´Dwyer et al. (2015) also explore the relationship

[1] See Drent et al. (2013) or Cordero et al. (2017) for detailed reviews of this literature

between instructional practices and eighth grade students´ performance using data TIMSS 2007 in a cross-country approach. Finally, the recent book edited by Nilsen and Gustafsson (2016) is a valuable contribution to this growing body of research, since it contains several empirical studies analysing TIMSS data across different countries and grades (four and eight) and taking account of multiple background variables.

In contrast, empirical studies about this topic using the OECD PISA and TALIS surveys are extremely scarce. This can be explained by the fact that data about teachers has been traditionally missing in the PISA dataset[2] and data about students is missing in the TALIS dataset. As pointed out by Kaplan and McCarty (2013), an ideal approach to linking the PISA survey to the TALIS survey would be to sample schools and administer questionnaires from both PISA and TALIS in the same school. This possibility may not be feasible for many countries, but the last wave of the TALIS survey released in 2013 included the possibility of linking the available data to the PISA 2012 dataset through the so-called TALIS-PISA link. Although only eight out of all the countries participating in both surveys chose this option (Australia, Finland, Latvia, Mexico, Portugal, Romania, Singapore and Spain), at least now it is possible to analyse teacher effectiveness using data from those OECD surveys.

Since the statistical and technical issues regarding this link between both surveys are far from straightforward, the aim of this paper is to explore the main characteristics of this data fusion process and provide some guidelines for researchers interested in performing empirical analyses using the resulting dataset. Additionally, we illustrate the alternative fusion process that can be adopted with an empirical analysis of the relationship between teaching practices and student characteristics and outcomes in the specific context of Spain, since this country presents the largest sample of observations among participating countries in this novel process.

The remainder of the paper is organized as follows. In Section 2, we describe the main characteristics of the TALIS and PISA databases, as well as the fusion process. In Section 3, we explore the main strengths and weaknesses of the resulting merged

[2] In 2012 PISA introduced a number of questions in the student questionnaire related to teaching strategies and the instructional context in the mathematics classroom that made it possible to conduct some empirical studies about the effectiveness of teacher strategies (e.g. Caro et al., 2016). Subsequently, in PISA 2015, a teacher questionnaire was offered to PISA-participating countries for the first time.

dataset and provide some recommendations for researchers interested in exploiting this data source. Section 4 presents the results of the empirical analysis performed to show some of the potential utilities of the TALIS-PISA link dataset. Finally, section 5 outlines the main conclusions.

2. Datasets

In this section we explain some basic aspects regarding the structure of the two international surveys analysed in this study as well as some basic methodological aspects related to their fusion via the so-called TALIS-PISA link.

TALIS is an international large-scale survey that focuses on the working conditions of teachers and the learning environment in schools. TALIS links institutional characteristics to aspects of school and classroom climate from the perspective of teachers and school administrators. The study provides insights into the beliefs and attitudes about teaching that teachers bring to the classroom and the pedagogical practices that they adopt. TALIS also the extent to which certain factors may relate to teachers' feelings of job satisfaction and self-efficacy. The first round of TALIS was conducted in 2008 and surveyed lower secondary education teachers and school leaders in 24 countries. The second round was carried out in 2013, including 34 countries. For reliable estimation and modelling, 200 schools and 20 teachers per school were surveyed in each participating country. Therefore, the nominal international sample was around 4,000 teachers. As a result, the dataset includes information from more than 10,000 schools and more than 170,000 teachers. The variables included in the database can be classified into different categories: teachers' opinions and feedback, teachers' background and professional development, school management and mobility indicators.

PISA is an international survey that assesses the extent to which 15-year-old students around the world have acquired competences and skills in three key subjects (mathematics, reading and sciences). The study was first developed in 2000 and it has been carried out periodically every three years with a regular increase in the number of participating countries (65 in 2012). The dataset includes a wide variety of background information on the students collected using individual questionnaires. Most of this information refers to students' family background and personal information, but it also includes their views on the school climate and learning environment, all of which are important aspects of teachers' working environment. In addition, school principals also complete a questionnaire providing information on school resources, the total number of teachers in the school or the school's responsibility for taking decisions.

During the enactment of the first round of TALIS, several countries expressed a desire to have the survey linked to PISA outcome measures, but this option was not fully implemented in the end[3]. During the second round, countries that had taken part in PISA 2012 also had the option of implementing TALIS at the same schools. This option, commonly known as the TALIS-PISA link, made it possible to merge information gathered by teachers and principals in TALIS and by students in PISA into a single dataset. Only eight out of all the countries participating in both surveys chose this option, although this limitation is partly offset by the fact that the sample includes countries with diverse educational systems and cultural contexts. This offers an interesting variability with respect to student achievement, family background, school characteristics and type of instructional practices.

The sample of schools invited to participate in the TALIS-PISA link had to be selected from the existing sample of schools participating in PISA 2012. In order to respect most of the structure of the original sample of schools, a systematic equiprobable random sample of schools was drawn from the PISA 2012 sample, within the original explicit strata and original frame order. Subject to PISA requirements, the nominal sample size for the TALIS-PISA link was set at 150 schools, although the final number of participant schools was lower in some countries. The average number of teachers interviewed was around 3,000 for each country, although the Spanish sample doubles this number (see Table 1).

The target population included a representative sample of 20 teachers of 15-year-olds in the schools that took part in PISA and the principals of the respective schools. In addition, all mathematics teachers available at the schools included in the TALIS-PISA link sample were surveyed. They received an additional questionnaire, the mathematics teacher module, whose main aim was to gather more detailed information on teaching

[3] An experimental link to PISA 2006 was developed for interested countries, but no country took up this option.

practices at classroom level. The teacher questionnaires required teachers to identify a "target class" that would serve as their baseline for responding to questions about their practices and beliefs. This requirement was brought in so as to avoid bias potentially resulting from teachers being free to select a specific or favourite class.

Table 1. Overview of the TALIS-PISA link samples

	Number of schools participating in PISA 2012	Number of participating schools in TALIS-PISA link	Respondent teachers in schools	Weighted estimated size of teacher population
Australia	773	122	2,719	85,750
Finland	298	147	3,326	18,254
Latvia	221	118	2,123	10,228
Mexico	1,602	152	2,167	378,222
Portugal	199	141	3,152	52,101
Romania	201	147	3,275	86,051
Singapore	166	166	4,130	12,052
Spain	910	310	6,130	173,216

Source: OECD, TALIS 2013 Database

Finally, it should be noted that the smaller sample obtained after merging the information provided by schools participating in both the TALIS and PISA surveys is a mere statistical artifice combining two different sources of data. It is not, however, a specific survey created with the aim of conducting a combined analysis of factors related to teachers and student achievement. This is a key point that should be taken into account by analysts when interpreting the results of secondary analyses using these data.

3. Some guidelines for practitioners using data from TALIS-PISA link

The ideal situation for any researcher interested in analysing the relationship between the characteristics and practices of teachers and the academic achievement of students would be to have access to information related to the students of a class and the teachers that teach that class. Unfortunately, the database created by means of the TALIS-PISA link does not fit this ideal scenario. On this ground, a key issue is to identify the main design-dependent limitations caused by the use of the information that this database contains before undertaking any empirical analysis using this data source. Likewise, we also offer some useful guidelines on how to make proper use of this dataset.

3.1. Limitations of the TALIS-PISA link database

First of all, we know that these surveys were not implemented at the same time (PISA was conducted between March and May 2012 for countries in the northern hemisphere and May-August 2012 for countries in the southern hemisphere, while TALIS took place from September to December 2012 for countries in the southern hemisphere and February-June 2013 for countries in the northern hemisphere). Therefore, any empirical research aiming to use the synthetic file output by linking the two databases must take into account that the teachers surveyed in TALIS may not be the same respondents as taught the students evaluated in PISA. Within this framework, we have to assume that teacher mobility is low so that we can be confident about the accuracy of results. Besides, schools might have changed in the interim due to the implementation of some educational policy, thus we also have to assume that the time difference between the implementation of PISA and TALIS did not result in important exogenous changes across schools within a country.

Second, we should again stress that the unit of analysis must always be the school, as this is the common unit of analysis in both surveys. This involves summarizing the original information from the databases by aggregating responses into percentages or building composite indexes before data fusion can proceed. As a result, the multilevel structure of each survey is notably reduced, which might lead into a problem of overestimation when analyzing the influence of those factors on student attainment (Hanushek *et al.*, 1996). Moreover, this limitation makes it far more complicated to conduct empirical analyses aiming to identify causal effects, as pointed out by the TALIS user guide (OECD, 2014b). For instance, the baseline used to analyse the determinants of student achievement in mathematics measured by PISA are the mean characteristics of all the school's mathematics teachers instead of the specific traits of the teacher that actually taught mathematics lessons to the respective students. This is a major weakness with respect to other databases like TIMSS or PIRLS (Mullis et al., 2012a, 2012b), since their design do allows for matching student-teacher data and apply an estimation strategy based on fixed effects to control for unobserved student traits by exploiting between-subject variation (e.g., Schwerdt & Wupperman, 2011; Bietenbeck, 2014).

Another relevant problem is that the number of schools participating in PISA 2012 and TALIS 2013 far outweighs the number of schools included in the TALIS-PISA link. Therefore, the sampling weights provided by the original samples cannot be used for the purpose of estimations that should be representative of the total population[4]. In these cases, the weights provided by the merged database have to be used.

If a variable has been measured on a different scale across the two surveys, previous literature suggests that they need to be converted to *z*-scores (Rässler, 2012), even if the variables are categorical as is usually the case in TALIS. In this case, differences in the scales of categorical variables can also be handled by collapsing one, or both, to a common set of categories.

Finally, as pointed out in the document drafted by TALIS-PISA link working group experts (OCDE, 2014a), the TALIS results should not be used to explain students' PISA results. On the contrary, the results of schools and students should be used to contextualize the responses of principals and teachers. Actually, in the literature we can find some previous studies adopting this approach (e.g. Austin et al., 2015). In section 4 we will illustrate some potential utilities of this approach.

3.2. Recommendations for using data from the TALIS-PISA link

Considering the design-dependent limitations of the TALIS-PISA link described above, we now focus on possible ways of acceptably and consistently addressing the combined analysis of these data. Our starting point for this purpose will be the last point mentioned in the previous section and researchers' data needs. The OECD recommendation is not to avoid contextualizing the PISA results based on TALIS responses, whereas most researchers will be looking precisely to measure the effects of teaching on student outcomes. In statistical terms, both approaches would be equivalent to merging PISA to TALIS (Figure 1, direction a) or TALIS to PISA (Figure 1, direction b). In terms of statistical matching, the above decisions require a distinction to be made between a "donor" dataset and a "recipient" dataset. Direction (a) would mean that TALIS would

[4] Details on the construction of these weights are available within the technical reports (e.g. OECD, 2014a, 2014c).

be the recipient dataset and PISA the donor dataset, whereas PISA would be the recipient and TALIS the donor dataset in direction (b).

As noted by D'Orazio et al. (2006), there are several factors that need to be considered when designating which are the donor and recipient datasets. The two most important concerns are the phenomenon under study and the accuracy of the information that the two surveys contain. With respect to the phenomenon under study, matching PISA and TALIS should yield a synthetic dataset that retains the ability to draw valid and reliable inferences of policy relevance. Regarding accuracy, it would be senseless to match two datasets that contain inaccurate information from either or both surveys.

Figure 1. PISA-TALIS matching

PISA 2012	
Student ID	**School (Principal) ID**
101	1
102	1
103	1
104	1
105	1
106	1
…	1
135	1

(a)

(b)

TALIS 2013				
School (Principal) ID	**Teacher ID**	**Area**	**General Questionnaire**	**Specific Questionnaire**
1	101	Maths	X	X
1	102	Maths	X	X
1	103	Maths	X	X
1	104	Science	X	
1	105	Foreign Language	X	
1	106	Language	X	
1	…	Science	X	
1	120	Science	X	

Source: Own elaboration

If we intended to assess student achievement depending on teaching practices and/or teachers´ characteristics, we would use PISA as a recipient dataset, that is, as the subject of the main analysis. In this case, the TALIS variables should be provided at school level. The ultimate aim is to attribute to PISA a series of indicators that are capable of summarizing issues like the leadership by principals, the teaching practices at the school, level of teacher training, the level of cooperation between teachers or the level of ICT use. The principal variables are provided at school level, therefore they are directly attributable to the respective PISA schools. This would be a way of summarizing school-specific characteristics in terms of teaching activity from the viewpoint of the principals. This would be a complementary perspective to the picture painted by mostly the same school principals in TALIS.

The teacher variables (referred to both mathematics and other subjects) provide the separate opinions of each teacher, which need be summarized at school level.

Additionally, many of the teacher questionnaire items refer to their opinions on numerous aspects that they are asked to rate on a Likert scale. Therefore, it would be wise to first summarize the information as variables referring to a specific aspect of teaching and synthesize the information. The TALIS database does in fact perform this procedure for a set of both principal and teacher variables (measuring classroom management effectiveness, teaching effectiveness or student motivation, etc.). With respect to both issues (aggregation first at teacher and then at school level), several procedures can be applied to create indicators to summarize the information, ranging from the construction of simple indexes to cluster analysis (Sans-Martín et al., 2015) or factorial analysis (TALIS, 2014b). In any case, the idea is to comprise or combine the information according to a procedure that provides a reliable and valid scale and avoid problems of multicollinearity associated with the use of separate variables. The aim in all cases is to measure not only position but also dispersion, which denotes variability (standard deviation, differences between percentiles, maximums and minimums, etc.).

If PISA is the recipient dataset, we have to bear in mind that the number of sampled schools is greater in PISA than in TALIS. Note, therefore, and this is one of the trickiest matters about using PISA as a recipient base, the resulting database will be appreciably smaller than its original size, since only data about units with a common school identifier in both datasets can be linked.

Regarding this issue, it is worth mentioning that there other potential ways of enabling a linkage between both datasets using alternative statistical matching methods, thus we can have a dataset with a higher number of observations. In a previous study, Kaplan and McCarty (2013) provide a systematic evaluation of various alternative data fusion methods that can be applied when there is no link between the original PISA and TALIS data. Actually, this was the situation when the previous waves of those datasets (TALIS 2008 and PISA 2009) were released, thus the only possibility consisted in creating a synthetic cohort of data combining information from both surveys. Those procedures intend to address the issue of missing data, i.e., TALIS is missing student-level data available in PISA, while PISA is missing teacher-level data available in TALIS. It is beyond the scope of this paper to evaluate those matching methods, but in section 4 we provide an example to illustrate its potential usefulness in empirical analysis.

One more issue to be taken into account is that, as mentioned above, because of both the sampling procedure and the time difference between the surveys, the students that completed the PISA tests in 2012 were not necessarily taught by the same teachers that participated in the TALIS survey in 2013. The first problem can be addressed by building indicators, as pointed out above. The second can be tempered by previously filtering the TALIS dataset for teachers with more than two years' service at a school. This procedure will assure that the teachers are very likely to have taught the students that participated in PISA.

On the other hand, if the aim of the analysis is to characterize the opinions of teachers and school principals based on the characteristics of the schools and their students, we would use TALIS as the recipient dataset and PISA as the donor dataset. In this case, again it would be necessary to aggregate data at school level, i.e. the individual responses of each student cannot be linked to their teachers. Therefore, the information from PISA (except information provided by the principal, which is already at school level) would first have to be summarized. In the case of the test results, measures of position, dispersion and order will have to be calculated for the five plausible values. Likewise, for continuous variables representing student characteristics such as the socioeconomic status (ESCS)[5], average values should be calculated as well as for categorical variables, which need to be converted into percentages. In this case, sample weights do not pose a problem as the weights provided by the TALIS-PISA link can be easily applied.

Another option for the combined use of data from PISA to TALIS would be to rank schools according to their PISA mathematics score (considering the five plausible values) and investigate which teaching practices employed by the maths teachers at the respective schools might be considered as a benchmark for the others, i.e., characterize the teaching practices and teaching style of the highest-scoring schools in PISA. As reported in the TALIS-PISA Link International Report (Austin et al., 2015), this approach would make it possible to answer to the following questions:

- In schools with high student outcomes on the PISA assessment, do teachers report higher or lower needs for professional development?

[5] This index provides a measure of family background that includes the highest levels of parents´ occupation, educational resources and cultural possessions at home.

- Do teachers' beliefs about teaching and learning and their teaching practices vary in high-performing or low-performing schools?
- Do teachers' beliefs and practices vary based on the percentage of students from lower socio-economic backgrounds in the school?

Finally, shall we recall that PISA and TALIS offer a common support area, as it provides different views of the classroom climate and the teaching practices (both from the student and the teacher's perspective). The detection and analysis of these variables would allow us to seek for possible divergences or inconsistencies in these two agents' perception of classroom practices (e.g. the possible associations between different aspects of teacher self-efficacy and job satisfaction and characteristics of their schools' student population). This possibility has been explored by Eveleigh and Freeman (2012) in an exploratory analysis of the data using ANOVA and MANOVA models as well as multi-level modeling techniques to identify plausible relationships and explained variation that may be uncovered within the data. Likewise, the recent study conducted by Echazarra et al. (2016) focused on teaching strategies and their association with students' achievement using students´ responses is also complimentary to other about the same topic carried out by Le Donné et al. (2016), which is based on teachers' responses.

4. Some insights about teaching practices using data from Spain

In this section we illustrate some of the utilities of the TALIS-PISA link dataset by exploring the existing relationships between backgrounds and cognitive outcomes of students and backgrounds and teaching practices using data about Spanish schools. We selected this country because its number of available observations is significantly higher than other countries (more than double in some cases) as shown in Table 1. Specifically, we present the results in three different scenarios. First, we explore how the characteristics of students relate to teacher activities, i.e., considering TALIS as the recipient dataset and PISA as the donor dataset. Second, we investigate how teaching practices affect the performance of students, i.e., using PISA as the recipient and TALIS as the donor dataset. Finally, we again use PISA as the recipient data, but instead of using the link provided by the TALIS-PISA link, we use a matching method to construct an artificial dataset with the same size that the original PISA dataset incorporating data about teachers´ beliefs and instructional practices.

4.1. What school factors affect teaching practices?

Our strategy implies aggregating some relevant student variables at school level and linking them to data about teachers from the same school. Specifically, in our empirical analysis we consider the socio-economic status of students in the school (ESCS), the disciplinary climate (DISCLIMA)[6] and the type of school (PRIVATE[7]) as potential school factors that may affect the teaching style of teachers. Regarding data about teachers, we have selected some control variables represented by background characteristics of teachers identified as relevant factors in previous literature (e.g., Ehrenberg and Brewer, 1994; Greenwald et al., 1996; Wayne and Youngs, 2003) such as gender, age, years of experience or qualification (higher than required).

As dependent variables representing teaching practices, we have many different alternatives to be tested, since TALIS dataset provides information about a variety of activities conducted by teachers in the classroom. In this sense, it is important to know that responses about these activities are provided in a Likert scale format, with four possible answers: *(a) never or almost never*; *(b) occasionally*; *(c) frequently* or *(d) in all or nearly all lessons*. In order to construct our core variables we can follow different criteria. One possibility would be creating dummy variables coding answers *(a)* and *(b)* as zeros and *(c)* and *(d)* as ones. According to this criterion, we have built six potential dependent variables from specific questions including in the teachers questionnaire: *(i) students work in small groups to come up with a joint solution to a problem or task; (ii) students use ICT (information and communication technology) for projects or class work; (iii) students work on projects that require at least one week to complete; (iv) teacher presents a summary of recently learned content; (v) teacher let students practice similar tasks until every student has understood the subject matter and (vi) teacher checks students' exercise books or homework.*

Another possibility consists of creating composite indices combining responses about different questions to define an underlying teaching strategy as proposed by Echazarra et al. (2016) or Le Donné et al. (2016). Thus, we can define an index representing active learning activities, i.e., those promoting the engagement of students in their own

[6] This index is derived from the responses to five questions about problems with classroom organization (See OECD, 2014c for details).
[7] This is a dummy variable for which 1 denotes a private and semi-private school (*concertadas*) and 0 represents a public school.

learning, by combining the answers provided to the first three questions *(i, ii* and *iii*) and other index representing teacher-directed instruction, which is mainly based on lecturing, memorization and repetition, by combining answers provided to the other three questions (*iv, v* and *vi*)[8]. Higher values of these indices should be interpreted as a more regular use of this teaching strategy by a specific teacher.

The existence of potential interaction effects between variables representing teaching activities and explanatory variables are examined through the estimation of different hierarchal of multilevel regression models (Raudenbush and Bryk, 2002). The use of this approach allows us to avoid potential problems of bias in the estimations derived from classic methods, such as OLS regression, due to the existence of correlation between the values of student variables aggregated at school level for teachers from the same school (Hox, 2002). This method accounts for this statistical dependence by the complex residual structure thereby producing correct estimates of the standard errors associated with the regression coefficients. For dependent dummy variables, we assume a binomial logistic model structure for the regressions, while for indices we use a simple multilevel regression model. Table 2 reports the estimation results for the six alternative logit multilevel models (one for each dependent variable) and Table 3 for the linear regression models of both composite indices.

Table 2. Relationship between teacher activities and teacher and student variables

	(i)	(ii)	(iii)	(iv)	(v)	(vi)
VARIABLES	smallgroup	usetic	projects	summary	practice	homework
Female	0.128**	-0.119*	-0.0601	0.515***	0.546***	0.621***
	(0.0616)	(0.0612)	(0.0649)	(0.0625)	(0.0634)	(0.0675)
Age	-0.0239***	-0.00266	0.00470	0.00755	-0.0116*	0.00270
	(0.00685)	(0.00669)	(0.00702)	(0.00693)	(0.00703)	(0.00754)
Qualification	0.415***	0.509***	0.471***	0.0364	-0.150	0.363**
	(0.134)	(0.136)	(0.137)	(0.145)	(0.142)	(0.170)
Experience	0.00108	-0.00532	-0.0159**	0.00213	-0.00363	-0.00411
	(0.00633)	(0.00618)	(0.00649)	(0.00637)	(0.00638)	(0.00690)
ESCS	-0.0182	0.0457	0.0384	-0.130*	-0.0856	-0.315***
	(0.0743)	(0.0867)	(0.0685)	(0.0759)	(0.0716)	(0.0737)
DISCLIMA	-0.202**	-0.0888	-0.122	-0.0388	0.135	-0.0215
	(0.100)	(0.117)	(0.0931)	(0.0890)	(0.0845)	(0.101)
PRIVATE	0.0632*	-0.0648	-0.0197	-0.00492	-0.0971	0.0370
	(0.0351)	(0.0415)	(0.0332)	(0.0893)	(0.0843)	(0.0352)
Constant	0.0773	-0.0956	-0.952***	0.248	1.374***	0.701***

[8] See Orlich et al. (2013) for a detailed description of different teaching strategies.

	(0.243)	(0.249)	(0.248)	(0.240)	(0.244)	(0.264)
Observations	6,130	6,130	6,130	6,130	6,130	6,130
Groups	310	310	310	310	310	310

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 3. Relationship between teaching strategies and teachers' and students' variables

	Active learning	Teacher-directed instruction
Female	0.0232	0.696***
	(0.0503)	(0.0418)
Age	-0.00674	-0.000250
	(0.00553)	(0.00460)
Qualification	0.628***	0.123
	(0.115)	(0.0956)
Experience	-0.00801	-0.00460
	(0.00509)	(0.00421)
ESCS	-0.0772	-0.277***
	(0.0809)	(0.0539)
DISCLIMA	-0.174*	0.0119
	(0.0951)	(0.0635)
PRIVATE	0.346***	0.0261
	(0.0943)	(0.0635)
Constant	6.822***	8.401***
	(0.196)	(0.160)
Observations	6,130	6,130
Groups	310	310

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

In general terms, the results indicate that variables at school level do not seem to affect the type of teaching activities carried out by teachers to a significant extent. With respect to the socioeconomic characteristics of students, we only observe a significant (and negative) relationship with the probability of presenting summaries of contents and checking students' homework as well as using teacher-directed instruction, while the disciplinary climate and being a private school only are significantly (and negatively) associated with the implementation of active learning strategies and the probability of working in small groups. Likewise, we also notice that some control variables are more relevant than others. Specifically, being female teacher is positively (and significantly) associated with most part of dependent variables with the exception of the probability of using new technologies (negative association). Similarly, having qualification higher than required is a significant (and positive) factor in the majority of cases. In contrast, experience and age are only found to be significant in one model.

4.2. How teaching strategies affect student performance?

In this case, we aggregate variables representing teaching practices at school level, as well as background characteristics of teachers to control for heterogeneity among schools, and link them to individual data about students from each school. The new dataset includes 8,896 observations about students belonging to 310 schools.

When adopting this approach, the aim is to identify how teaching strategies contribute to student skill acquisition. Therefore, students' mathematics score is our dependent variable now. Regarding this issue, although PISA dataset provides five plausible values for each discipline[9], in our analysis we only consider a single plausible value (the first one), since on large samples using one plausible value or five plausible values does not really make a substantial difference (OECD 2009, p. 44). As explanatory variables, we include as our key variables those representing different teaching strategies (active learning and teacher-directed instruction) aggregated at school level. Moreover, we also include several control variables about the characteristics of teachers from the same school (percentage of female teachers, mean age and mean experience of teachers and proportion of teachers with qualification higher than required), the same school variables included in previous models (ESCS, DISCLIMA and PRIVATE) and, finally, a set of student background variables that have been most frequently identified as influential factors in previous literature, such as gender, attending pre-school, being a repeater, being an immigrant, belonging to a monoparental family, parents´ level of education or different indicators representing possessions at home (own desk and number of books)[10].

Since values of all the variables at school level are highly correlated for students attending the same school, we use again hierarchical or multilevel regression models. Table 4 shows the estimation results for two alternative models (one for each teaching strategy). According to these results, both active learning and teacher-directed practices are significantly related to students´ outcomes, but the relationship is negative. This means that in schools where teachers devote more time to implement many different teaching activities students´ have worse results, independently of which specific activities they conduct.

---

[9] See Wu (2005) for a detailed discussion about the role of plausible values in large-scale surveys.
[10] Todd & Wolpin (2003) survey the educational production function literature.

With regard to control variables, almost all individual factors are significantly associated with mathematic achievement in the expected direction, with the exception of father´s education level. Among variables at school level, the proportion of highly qualified teachers, the mean age of teachers and being a private school are found to be significant factors, showing a positive relationship with achievement, whereas the gender and experience of teachers as well as the average socioeconomic status of schoolmates and the school climate do not seem to affect the performance of students significantly.

Table 4. Student performance and teaching strategies (TALIS-PISA link dataset)

VARIABLES	PV1MATH	PV1MATH
Active learning strategy	-5.382**	
	(2.172)	
Teacher-directed instruction		-8.587**
		(4.140)
% Female teachers at school	4.336	3.577
	(10.36)	(10.45)
Mean age of teachers at the school	1.558***	1.594***
	(0.434)	(0.439)
% Teachers high qualification	62.14**	58.75*
	(30.63)	(31.27)
Mean experience of teachers	-1.040	-1.581
	(2.369)	(2.379)
Gender	-25.17***	-25.17***
	(1.508)	(1.514)
Preprimary	18.63***	18.86***
	(2.434)	(2.458)
Immigrant	-8.260**	-8.055**
	(3.258)	(3.271)
Repeater	-83.01***	-82.38***
	(1.835)	(1.843)
Mono-parental family	-5.724**	-5.590**
	(2.584)	(2.596)
Mother´s highest education level	4.608**	4.887***
	(1.814)	(1.823)
Father´s highest education level	0.428	0.0987
	(1.812)	(1.820)
Owndesk	9.152*	11.57**
	(5.431)	(5.461)
Book25	-31.89***	-31.98***
	(2.053)	(2.062)
Book200	23.75***	24.04***
	(1.896)	(1.902)
Private	10.05*	12.21**
	(5.808)	(5.888)
ESCS	0.00862	0.00902
	(0.00720)	(0.00726)

DISCLIMA	-0.0122***	-0.0114***
	(0.00359)	(0.00362)
Constant	532.6***	534.8***
	(29.08)	(35.65)
Observations	8,896	8,896
Number of groups	310	310

4.3. Robustness check with an alternative fusion process

One of the main drawbacks of databases built from the TALIS-PISA link is that the number of observations available is significantly lower than the original datasets. From an analytical viewpoint, this can be a serious limitation, especially when we work with data about a single country, since statistical power to detect relevant factors affecting the variable of interest might be lower due to the fact that variation among schools within a country is more reduced. In order to check whether our estimates based on the Spanish sample participating in TALIS-PISA link (8.896 students and 310 schools) are reliable, we have replicated the estimation presented in sub-section 4.2 using an artificial dataset composed of 25,313 students and 902 schools (the original number of observations available for Spain in PISA 2012) constructed by applying a multiple imputation method. Specifically, we rely on common information available in both surveys derived from school principal questionnaires to generate the matched datasets.

Table 5 reports the parameters of the estimation made using multilevel regression models for two alternative specifications, one for each teaching strategy (active learning and teacher-directed). These results confirms most part of the evidence presented above, since both types of instructional practices are negatively and significantly related to student performance in mathematics. Concerning control variables, the results are also similar, although for this larger dataset we found that father´s level of education is significantly associated with better results, while most part of teachers´ background characteristics, nor that the type of ownership of the school.

Table 5. Student performance and teaching strategies (matched dataset)

VARIABLES	PV1MATH	PV1MATH
Active learning strategy	-6.575*	
	(3.705)	
Teacher-directed instruction		-18.36*
		(9.994)
% Female teachers at school	-16.58	-36.29**
	(12.16)	(17.69)
Mean age of teachers at the school	0.179	-0.748
	(0.823)	(0.642)
% Teachers high qualification	40.31	53.94
	(53.76)	(51.98)
Mean experience of teachers	-1.758	-1.4271
	(2.072)	(1.918)
Gender	-16.89***	-16.90***
	(1.636)	(1.634)
Preprimary	19.09***	19.12***
	(2.368)	(2.377)
Immigrant	-23.85***	-23.79***
	(2.917)	(2.927)
Repeater	-22.87***	-22.87***
	(2.021)	(2.021)
Mono-parental family	1.392	1.365
	(2.513)	(2.513)
Mother´s highest education level	15.91***	15.89***
	(1.569)	(1.569)
Father´s highest education level	11.74***	11.66***
	(1.750)	(1.749)
Owndesk	5.580	5.597
	(5.776)	(5.781)
Book25	-47.44***	-47.44***
	(2.423)	(2.431)
Book200	30.88***	30.90***
	(1.850)	(1.847)
Private	11.06	4.952
	(9.149)	(7.555)
ESCS	0.000564	0.000590
	(0.00757)	(0.00755)
DISCLIMA	-0.00979**	-0.00975**
	(0.00489)	(0.00490)
Constant	575.5***	686.6***
	(33.49)	(67.37)
Observations	25,313	25,313
Number of groups	902	902

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

5. Concluding remarks

The existing evidence about the impact of teacher quality as a determinant of student achievement in cross-national empirical studies is still scarce because traditionally there have been a shortage of adequate sources of data about this relevant factor of the educational production function. To overcome this limitation, the OECD created the TALIS survey, which provides an extensive dataset about teachers' beliefs about and attitudes towards teaching. Until recently, the usefulness of this information was limited because it could not be linked to student-level data. Nevertheless, the last wave of this survey released in 2013 included the possibility of linking this survey to the OECD PISA outcome measures through the so-called TALIS-PISA link. This resulting combined survey is a mere statistical artifice and not a specific survey created with the aim of conducting a combined analysis of factors related to teachers and student achievement, thus it presents some weaknesses that need to be born in mind by researchers before implementing empirical analysis using this instrument.

In this paper, we have provided a detailed description of those limitations and some guidelines for practitioners using these data in empirical analyses. Among them, maybe the most relevant issue is that data is reported in a format that makes it unfeasible to blindly match teachers from a school to their respective students (or vice versa). Since the school is the school is the only common unit in both surveys, information about specific aspects of teaching or characteristics of students needs to be aggregated at school level. As a result, it is worth mentioning that the possibility of drawing conclusions in terms of causality is very limited.

Likewise, it is important to determine whether the purpose of the analysis is contextualizing the PISA results based on TALIS responses or exploring different aspects related to teachers´ characteristics or activities on student achievement. This decision implies to take TALIS or PISA as the donor or recipient database, which entails different statistical and conceptual implications. In order to illustrate how to deal with all those issues, we have estimated different multilevel regression models with the aim of exploring the existing relationship between teaching practices and students' background characteristics and their performance adopting alternative approaches to establish the link between both datasets. The results of our empirical analysis do not allow us to identify common factors associated with different types of teaching activities and strategies implemented by teachers. However, we found that the more different teaching activities report to implement the worse the results of their students.

Acknowledgments

We are grateful to participants at the XXV Meeting of the Economics of Education Association for their helpful comments and suggestions that contributed notably in improving the quality of the present paper. Likewise, the authors would like to express gratitude to Carmen Tovar and all the personnel working at the Spanish National Institute for Education Evaluation for their technical support. Additionally, the authors also acknowledge the Ramón Areces Foundation for funding this research and the Spanish Ministry of Economy and Competitiveness through grants EDU2016-76414-R and ECO2014-53702-P.

References

Austin, B., Adesope, O. O., French, B. F., Gotch, C., Bélanger, J. and Kubacka, K. (2015). Examining school context and its influence on teachers. Linking TALIS 2013 with PISA 2012 student data. OECD Education Working Papers, No. 115, OECD Publishing, Paris. doi: http://dx.doi.org/10.1787/5js3f5fgkns4-en

Boonen, T., Van Damme, J. and Onghena, P. (2014). Teacher effects on student achievement in first grade: which aspects matter most?, *School Effectiveness and School Improvement*, 25(1), 126-152.

Caro, D. H., Lenkeit, J. and Kyriakides, L. (2016). Teaching strategies and differential effectiveness across learning contexts: Evidence from PISA 2012, *Studies in Educational Evaluation*, 49, 30-41.

Chapman, C., Armstrong, P., Harris, A., Muijs, D., Reynolds, D. and Sammons, P. (eds.). (2012). *School effectiveness and improvement research, policy and practice: Challenging the orthodoxy?* Abingdon, Oxon: Routledge.

Cordero, J.M., Cristóbal, V. and Santín, D. (2017). Causal Inference on Education Policies: A Survey of Empirical Studies Using PISA, TIMSS and PIRLS, *Journal of Economic Surveys, in press*. doi: 10.1111/joes.12217.

Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.

Creemers, B. P. M. and Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London: Routledge.

D'Orazio, M., Di Zio, M. and Scanu, M. (2006). *Statistical matching: Theory and practice*. John Wiley & Sons.

Drent, M., Meelissen, M. R. and van der Kleij, F. M. (2013). The contribution of TIMSS to the link between school and classroom factors and student achievement. *Journal of Curriculum Studies*, 45(2), 198-224.

Echazarra, A., Salinas, D., Méndez, I., Denis, V. and Rech, G. (2016). How teachers teach and students learn: Successful strategies for school, *OECD Education Working Papers, No. 130, Paris: OECD Publishing*. http://dx.doi.org/10.1787/5jm29kpt0xxx-en

Ehrenberg, R. G. and Brewer, D. J. (1994). Do school and teacher characteristics matter? Evidence from high school and beyond, *Economics of Education Review*, 13, 1–17.

Eveleigh, F. and Freeman, C. (2012). An exploratory analysis of the TALIS and PISA link data: an investigation of the possible relationships. Contribution for the European Conference on Educational Research (ECER) (2012) Conference.

Greenwald, R., Hedges, L. V. and Laine, R. D. (1996). The effect of school resources on student achievement, *Review of Educational Research*, 66(3), 361-396.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*(3), 466-479.

Hanushek, E. A., Rivkin, S. G. and Taylor, L. L. (1996). Aggregation and the Estimated Effects of School Resources. *The Review of Economics and Statistics*, 78(4), 611-627.

Hattie, J. A. C. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. Oxon: Routledge.

House, J. D. (2009). Elementary-school mathematics instruction and achievement of fourth-grade students in Japan: Findings from the TIMSS 2007 assessment. *Education*, 130(2), 301.

Hox, J. (2002). *Multilevel Analysis. Techniques and Applications*, Mahwah: Lawrence Erlbaum Associates.

Kaplan, D. and McCarty, A. T. (2013). Data fusion with international large scale assessments: a case study using the OECD PISA and TALIS surveys. *Large-scale Assessments in Education*, *1*(1), 1-26.

Le Donné, N., Fraser, P. and Bousquet, G. (2016). Teaching Strategies for Instructional Quality: Insights from the TALIS-PISA Link Data, *OECD Education Working Papers, No. 148, Paris: OECD Publishing*. http://dx.doi.org/10.1787/5jln1hlsr0lr-en

Mullis, I. V., Martin, M. O., Foy, P. and Arora, A. (2012). *TIMSS 2011 international results in mathematics*. International Association for the Evaluation of Educational Achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

Nilsen, T. and Gustafsson, J. E. (2016). *Teacher Quality, Instructional Quality and Student Outcomes*. Springer International Pu.

Palardy, G. J. and Rumberger, R. W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning, *Educational Evaluation and Policy Analysis*, *30*(2), 111-140.

Nye, B., Konstantopoulos, S. and Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26, 237–257.

O'Dwyer, L. M., Wang, Y., & Shields, K. A. (2015). Teaching for conceptual understanding: A cross-national comparison of the relationship between teachers' instructional practices and student achievement in mathematics. *Large-scale Assessments in Education*, 3(1), 1-30. doi: 10.1186/s40536-014-0011-6

OECD (2009). *PISA Data Analysis Manual, SPSS Second Edition*, OECD Publishing, Paris.

OECD (2014a). TALIS 2013 *Results An International Perspective on Teaching and Learning*, OECD, Paris.

OECD (2014b). *TALIS 2013 User Guide to the International Dataset*, OECD, Paris.

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

Orlich, D. C., Harder, R. J., Callahan, R. C., Trevisan, M. S. and Brown, A. H. (2013). *Teaching Strategies: A Guide to Effective Instruction*, 10th edition, Wadsworth, Cengage Learning, Boston, MA.

Rässler, S. (2012). *Statistical matching: A frequentist theory, practical applications, and alternative Bayesian approaches* (Vol. 168). Springer Science & Business Media.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods,* Sage Publications.

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C. and Stringfield, S. (2014). Educational effectiveness research (EER): a state-of-the-art review. *School Effectiveness and School Improvement*, *25*(2), 197-230.

Rivkin, S.G., Hanushek, E.A. and Kain, J.F. (2005). Teachers, schools, and academic achievement, *Econometrica* 73, no. 2: 417–58.

Rockoff, J.E. (2004) The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review* 94, no. 2: 247–52.

Rowan, B., Correnti, R. and Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the prospects study of elementary schools. *Teacher College Record*, *104*, 1525–1567.

Sans-Martín, A., Guardia, J. and Triadó, X.M. (2016). Educational leadership in Europe:A transcultural approach, *Revista de Educación*, 371, 83-106.

Schwerdt, G. and Wuppermann, A. C. (2011). Is traditional teaching really all that bad? A within-student between-subject approach. *Economics of Education Review*, *30*(2), 365-379.

Todd, P. E. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement, *The Economic Journal*, 113(485).

Van Klaveren, C. (2011). Lecturing style teaching and student performance, *Economics of Education Review*, 30(4), 729-739.

Wayne, A. J. and Youngs, P. (2003). Teacher characteristics and student achievement gains: A review, *Review of Educational Research*, 73, 89–122.

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2-3), 114-128.

Zuzovsky, R. (2013). What works where? The relationship between instructional variables and schools' mean scores in mathematics and science in low-, medium-, and high-achieving countries. *Large-scale Assessments in Education*, 1(1), 2.