Teacher self-efficacy and pupil achievement: much ado about nothing? International evidence from TIMSS

John Jerrim, Sam Sims & Mary Oliver

View supplementary material ⧉

Published online: 31 Jan 2023.

Submit your article to this journal ⧉

Article views: 2136

View related articles ⧉

View Crossmark data ⧉

Teacher self-efficacy and pupil achievement: much ado about nothing? International evidence from TIMSS

John Jerrim 🅸🅳[a], Sam Sims 🅸🅳[b] and Mary Oliver 🅸🅳[c]

[a]UCL Social Research Institute, London, UK; [b]UCL Centre for Education Policy and Equalising Opportunities, London, UK; [c]School of Education, University of Nottingham, London, UK

ABSTRACT

Bandura's influential theory has been used to argue that teachers with high self-efficacy will be more effective at increasing pupil achievement—and a voluminous empirical literature has repeatedly documented associations consistent with this claim. However, few studies have considered whether these correlations reflect an underlying causal relationship. In this paper we utilise across-subject, within-pupil variation in teacher self-efficacy in the Trends in Mathematics and Science Study (TIMSS) 2015 data to provide new evidence on this question. By focusing upon relative differences in teacher self-efficacy and pupil achievement within pupil-teacher pairs, our estimates control for more potential confounders than much of the existing literature. Contrary to that literature, we find no evidence of a relationship. Instead, this paper presents clear and consistent findings of null effects.

Introduction

Across the world, governments are seeking to raise the academic achievement of young people, particularly amongst those from disadvantaged socio-economic backgrounds. Although pupil achievement is the result of a complex interplay between a wide array of factors—including schools, parents and the home learning environment—teachers are widely regarded as one of the most important influences into children's academic development outside of the home (Burgess, 2019). Yet accurately measuring teacher quality is difficult and easily-observable characteristics such as postgraduate qualifications provide little to no indication of quality (Bitler et al., 2019; Hill et al., 2019). It is hence important that we collectively develop a better understanding of the attributes of teachers that are related to stronger levels of academic performance amongst pupils.

Teacher self-efficacy

In light of the failure to identify readily-observable characteristics of teachers that predict effectiveness, researchers have increasingly looked to intangible psychological variables. Perhaps the most widely studied in the literature is teacher self-efficacy (TSE), which

refers to 'a [teacher's] judgement of his or her capabilities to bring about desired outcomes of student engagement and learning' (Tschannen-Moran & Woolfolk Hoy, 2001, p. 783). Educational psychologists have long argued that teachers' self-efficacy will in turn influence pupils' academic outcomes (Klassen et al., 2011). Indeed, recent meta-analyses by Klassen and Tze (2014) and Zee and Koomen (2016) identified around thirty studies looking at the relationship between TSE and pupil achievement. Research in this area has continued at pace since the publication of these influential reviews (Burić & Kim, 2020; Künsting et al., 2016; Perera & John, 2020).

The foundations of the TSE concept can be traced back to Rotter's (1966) theory of locus of control (Zee & Koomen, 2016). Rotter hypothesised that individuals differ in terms of their beliefs about whether outcomes are generally due to luck or fate (external locus) or the result of their own actions (internal locus). Bandura (1977, 1986, 1997) built on this work but argued that an individual's locus of control would also depend on their own personal capabilities. Since individual capabilities are domain-specific, the self-efficacy construct has been since adapted and applied to the specific domain of teaching (Tschannen-Moran & Woolfolk Hoy, 2001; Tschannen-Moran et al., 1998), where it is thought to encompass teachers' beliefs with regard to instructional practice, classroom management and student engagement.

Teacher self-efficacy and pupil outcomes

Self-efficacy beliefs 'influence thought patterns and emotions, which in turn enable or inhibit actions' (Gavora, 2010, p. 2). Thus, it has been argued that teachers with high levels of self-efficacy are more likely to perceive difficulties as something that can be overcome, will feel less fatalistic about initial failure and may have greater confidence to take on new challenges (Gibson & Dembo, 1984). As a result, they are also more likely to persist with practising and successfully acquiring new pedagogical skills (Holzberger et al., 2013, 2014). Pupils in turn are theorised to benefit from this in terms of enhanced learning and, by extension, greater academic self-confidence (Woolfolk Hoy et al., 2009; Zee & Koomen, 2016).

Two broad theoretical pathways have been proposed connecting teacher self-efficacy and pupil achievement. The first is the *indirect path*, which assumes that increased teacher self-efficacy will improve pupil achievement via the mediating variable of tea-chers' behaviour/practice in the classroom (Lauermann & Butler, 2021). More precisely, teachers with higher levels of self-efficacy are more likely to persist in the face of difficulties or to employ a wider range of teaching techniques, which may be better suited to the specific and varied challenges they face in the classroom (Lauermann, ten Hagen, 2021). The second pathway is the *direct path*, which assumes that increased teacher self-efficacy may 'rub off' directly on pupils via a role-modelling process. This increased student self-efficacy may in turn improve pupils persistence with regards to their schoolwork, thus benefiting their achievement (Lauermann & ten Hagen, 2021).

Consistent with this theory, a large body of empirical research has found that TSE is linked to 'a range of instructional outcomes, teacher instructional behaviour, and teacher well-being, including student motivation, student engagement, student achievement, student self-efficacy, teacher work satisfaction, work commitment, teacher effectiveness and instructional behaviour' (Mok & Moore, 2019). With respect to academic

achievement, Klassen and Tze (2014) found a meta-analytic Cohen's D effect size of 0.2. A more recent meta-analysis by Kim and Seo (2018) found a slightly smaller, though still statistically significant, mean effect. This body of theory and evidence has led to the TSE concept becoming highly influential in both academic educational psychology and in the classroom (American Psychological Association, 2020).

The present study

Given the existing evidence, and the consistent findings of a positive association with pupil outcomes, what does this paper contribute to an already sizeable literature? One of the main limitations with the existing evidence base is that there has been relatively little consideration of whether estimates capture cause and effect (Lauermann & Butler, 2021; Pekrun, 2021). Indeed, among the 27 studies identified by Zee and Koomen (2016) focusing on pupil achievment, all but four relied on cross-sectional data (Caprara et al., 2006; Guo et al., 2010, 2012; Midgley et al., 1989). Even with the addition of more recent longitudinal studies (Praetorius et al., 2017), causal evidence is lacking. This is in large part due to the difficulty of convincingly accounting for key confounders in the TSE— pupil achievement relationship. For instance, teacher quality might cause self-efficacy rather than the other way around. Or both variables may be caused by a third hard-to-measure variable, such as the quality of teachers' colleagues (Kirabo & Bruegmann, 2009) or the quality of working conditions in the school (Kraft & Papay, 2014).

The present study addresses this limitation. Whereas most studies in this literature use correlation, OLS regression or structural equation modelling, our headline analysis is based upon a child fixed-effects approach (Jerrim et al., 2020). These models are estimated for a subset of primary school pupils who are taught by the same teacher across two core subjects (science and mathematics), which allows us to focus upon differences across subjects in teacher self-efficacy and how this relates to differences in children's academic achievement and self-confidence *within a given pupil-teacher pairing*. Our paper therefore responds to recent calls for greater use of between-person designs in this literature (Pekrun, 2021). Although we are still unable to rule out the possibility of unobserved confounders affecting our results, this approach does at least allow us to implicitly control for subject-invariant aspects of teacher quality—which, we believe, is a first in this literature. Using the Trends in Mathematics and Science Study (TIMSS) 2015 data,[1] these models are estimated for a large sample (72,637 pupils) across multiple countries. In doing so, we argue that the estimates presented here are at least one step closer to establishing whether teacher self-efficacy is indeed causally related to pupil outcomes than elsewhere in the existing literature. Thus, in summary, we address the following two research questions:

- Research question 1. Are higher levels of teacher self-efficacy associated with higher levels of pupil achievement?
- Research question 2. Are higher levels of teacher self-efficacy associated with higher levels of pupil's academic self-confidence?

To trail our key findings, in contrast to conventional wisdom and existing empirical work, we find no robust positive relationship between teacher self-efficacy and pupil

academic achievement or self-confidence. Interestingly, this holds true within both our child fixed-effects models and when using a more conventional OLS regression approach. This finding holds across several different model specifications and almost all the countries included in our analysis. Hence this paper turns out to be a story of null effects.

Method

Study design and sample

The Trends in Mathematics and Science Study (TIMSS) is an international assessment of school children's achievement in science and mathematics. Although TIMSS covers both fourth grade (≈ age 9/10) and eighth grade (≈ age 13/14) pupils, we focus upon the former, given the tendency for primary school teachers to teach both science and mathematics to the same pupils in many countries (further details as to why this is important are provided in the methodology section that follows). Nationally representative samples are drawn, with schools initially randomly selected with probability proportional to size. Then, from within each school, one or two classes are randomly selected to complete the TIMSS assessment and background questionnaires. All pupils within the selected classes—regardless of their age—are eligible to take part. Response rates of both schools and pupils are high in most countries, with further details provided in Appendix A. As part of TIMSS, pupils complete a mathematics and science assessment, while they (along with their parents and the class teacher) complete background questionnaires. Further details on the key information captured within these questionnaires are provided below.

Within this paper, we restrict our attention to a subset of countries and a subset of pupils from the full TIMSS sample. We focus upon members of the OECD only, leaving a pool of 25 countries available for analysis. This restriction has been made given that most of the existing literature investigating the link between teacher self-efficacy and student outcomes has been conducted in developed countries. It also limits the number of estimates to be presented, helping to facilitate communication and interpretation of our results. The Netherlands and Flemish-Belgium are also excluded due to difficulties in estimating the child fixed-effects approach set out in the following section. This leaves a total of 23 countries in the analysis.

Within these nations, we make some further sample restrictions. Most importantly, we focus upon those pupils who are taught by the same teacher for both science and mathematics.[2] This is to ensure that we estimate the link between teacher self-efficacy and pupil outcomes across different subjects (science and mathematics) for the same teacher. As noted in the following section, this will allow us to rule out any subject-invariant characteristics of teachers within our child fixed-effects approach. Table 1 illustrates how this leads to a reduction in the sample size, with some countries (e.g. Chile, Hungary, Denmark, Japan) affected more than others (e.g. Ireland, Turkey, Slovenia, New Zealand). We also restrict the analytic sample to those teachers/pupils where the key outcomes (pupil achievement and academic self-confidence) and key covariate (teacher self-efficacy) are available. In countries where the sample size has been greatly reduced, this will limit the generalisability (external validity) of the results. This leaves a final analytic sample size of 72,637 pupils.

Table 1. Details about sample selection.

	Full sample	OLS analytic sample	Fixed-effect analytic sample
Canada	12,283	8,862	8,096
United States	10,029	7,172	6,621
Spain	7,764	4,798	4,331
Turkey	6,456	6,431	5,138
New Zealand	6,322	5,191	4,671
Slovak Republic	5,773	3,597	3,357
Czech Republic	5,202	3,637	3,392
Hungary	5,036	3,436	2,678
Finland	5,015	3,600	3,565
France	4,873	3,729	1,945
Chile	4,756	2,387	1,765
Portugal	4,693	4,412	3,932
South Korea	4,669	3,383	3,344
Slovenia	4,445	4,113	4,069
Japan	4,383	1,239	1,129
Italy	4,373	2,935	626
Ireland	4,344	4,325	3,716
Norway	4,164	1,701	1,579
Sweden	4,142	1,701	1,501
England	4,006	2,745	2,210
Germany	3,948	1,811	1,766
Denmark	3,710	1,303	961
Northern Ireland	3,116	2,639	2,252
Total	**123,502**	**85,147**	**72,637**

Notes: OLS and fixed-effect sample sizes differ due to imputation being used to account for missing covariate data in the former but not the latter. Figures refer to the number of pupils. OLS analytic sample size also based upon mathematics data only.

Measures

Teacher Self Efficacy. As part of the teacher questionnaire, respondents were asked the following question in reference to mathematics:

In teaching mathematics to this class, how would you characterise your confidence in doing the following?

(a) Inspiring students to learn mathematics
(b) Showing students a variety of problem solving strategies
(c) Providing challenging tasks for the highest-achieving students
(d) Adapting my teaching to engage students' interest
(e) Helping students appreciate the value of learning mathematics
(f) Assessing student comprehension of mathematics
(g) Improving the understanding of struggling students
(h) Making mathematics relevant to students
(i) Developing students' higher-order thinking skills

Teachers were then asked the same questions about their confidence with respect to teaching science—but with item (b), which is maths-specific replaced with an item about 'teaching science using inquiry methods'. Each statement required a response using a four-point scale (very low confidence to very high confidence).

In line with our definition, these eight items clearly capture teachers' 'judgement of their abilities' (i.e. confidence) to bring about 'desired outcomes of student engagement and learning' (e.g. engaging their interests, improving their understanding). Teacher self-

efficacy is one of several related constructs that capture different aspects of teachers' beliefs about their teaching capability, some of which are liable to be confused (Lauermann & ten Hagen, 2021). In particular, it is important to distinguish teacher self-efficacy from teacher self-concept (teachers' evaluated perceptions of their teaching effectiveness; Yeung et al., 2014). Zhu et al. (2018) outline that self-efficacy differs from self-concept in that the former is context specific, does not rely on social comparisons, focuses on teachers' confidence, and is future oriented rather than past oriented. On the first three of these criteria, we believe that the eight items above clearly capture teacher self-efficacy beliefs, rather than self-concept: they are focused on specific classes (context specific), rely on improvement in certain domains (rather than social comparisons), and focuses on confidence ('how would you characterise your confidence'). On the fourth criteria, the item is worded in the present tense, which is neither past- or future-oriented. In the terms introduce by Lauermann & ten Hagen (2021) these items are capturing a global (across teaching competencies) and subject-specific (maths, science) measure of teacher self-efficacy.

Using responses to the eight items that were asked for both science and mathematics, a continuous 'teacher self-efficacy' scale is constructed. This has been done via estimation of a partial credit item-response model within each country, using the pooled mathematics and science data when our focus is upon relative differences between subjects (i.e. when implementing the child fixed-effects approach). Teacher self-efficacy is thus treated as a latent variable, with our scale created using latent rather than manifest scores. This scale is then divided into thirds within each country, defining our 'low', 'moderate' and 'high' teacher self-efficacy groups. While we recognise that dividing a continuous variable into a discrete categorical variable has some limitations—e.g. some loss of statistical power, somewhat arbitrary cut-points—it also has at least two important advantages. First, it aids communication of results. It is easier to describe to non-specialists differences between teachers with low, average and high levels of self-efficacy than changes in outcomes associated with a given increase in a continuous teacher self-efficacy scale. Second, it also provides a simple way to explore whether the link between teacher self-efficacy and pupil outcomes may be non-linear—e.g. that there may be difference between teachers with low and average levels of self-efficacy, but not between those with average and high-levels.[3] Hence the focus of our analysis is whether pupil outcomes differ depending upon whether their teacher has high or low levels of self-efficacy (relative to 'moderate' as the reference group). However, appreciating the limitations of this approach, in Appendix E we provide a full set of alternative estimates using the underlying continuous teacher self-efficacy scale instead. Both approaches (using a continuous measure of teacher self-efficacy versus using terciles) lead to the same substantive results.

Pupil Achievement. As part of the TIMSS fourth grade study, children sit a 72-minute test, which includes questions covering skills in both science and mathematics. This assesses pupils' knowledge and skills on an internationally defined curricula. One important issue is the 'curricular validity' of this test—i.e. the extent that the test is connected to the curriculum actually taught. For instance, if TIMSS mostly assesses mathematics/science skills not actually taught in schools, then teacher characteristics (including teacher self-efficacy) are unlikely to be related to pupil test scores. Fortunately, as part of participating in TIMSS, education experts from each country provide data that

feeds into the 'test-curriculum matching analysis'. Specifically, for each TIMSS test question, these experts provided information as to whether it was part of the country's curriculum in the fourth grade. According to the TIMSS 2015 technical report: '*4 of the 47 countries that took part in the TCMA analysis judged 100% of the items to be included in their curricula . . . . . . A further 34 countries . . . . judged 75% or more to be appropriate. All of the participants concurred that more than half of the mathematics items were included in their curricula*' (Mullis et al., 2016: Appendix F). This illustrates how there was close alignment between the TIMSS test and national curricula in most participating countries.

A multiple matrix test design has been used, with test questions divided into different test booklets. These booklets were then randomly assigned for pupils to complete. The main implication of this type of test design is that pupils only complete a (random) sample of all test questions for that year, although, all pupils do answer some questions in both science and mathematics.[4] Using pupils' responses to these questions, as well as information gathered in the background questionnaire, the survey organisers create a set of five 'plausible values' for each child in mathematics and science. These can be broadly interpreted as measures of pupils' achievement in the two subjects, though with an element of uncertainty added to reflect that pupils only answer a random subset of all possible test questions. The first plausible value is used throughout the analysis, although the substantive results do not change if any of the other four plausible values are used instead (see Appendix H for further details).[5] The plausible values have been standardised to mean zero and standard deviation one in each country, so that the estimates can be interpreted in terms of effect sizes.

Pupil academic self-confidence. In the background questionnaire, pupils were asked about their self-confidence in mathematics and science. Specifically, they were posed the following question in reference to mathematics:

'*How much do you agree with these statements about mathematics?*'

 (a) I usually do well in mathematics*
 (b) Mathematics is harder for me than for many of my classmates*
 (c) I am just not good at mathematics*
 (d) I learn things quickly in mathematics*
 (e) Mathematics makes me nervous
 (f) I am good at working out difficult mathematics problems
 (g) My teacher tells me I am good at mathematics*
 (h) Mathematics is harder for me than any other subject*
 (i) Mathematics makes me confused*

Answers to these statements were answered by pupils using a four-point scale (agree a lot, agree a little, disagree a little, disagree a lot). The items highlighted using a * were also asked to pupils in reference to science (i.e. all items except e and f). Using the seven common questions pupils were asked about for both science and mathematics, we construct a pupil academic self-confidence scale. This is again done using the pooled science and mathematics data when our focus is upon relative differences between subjects. This is implemented via estimation of a partial credit item-response model within each country included in the analysis. We standardise the scale to mean zero and

standard deviation one within each country, meaning estimates can be interpreted in terms of effect sizes.

Control variables. The pupil, teacher and parent[6] questionnaires also captured a range of other information, which are used as statistical controls within parts of our analysis. This includes:

- Pupil demographics (e.g. age, gender, immigrant background).
- Frequency with which a pupil is absent from school.
- Whether the pupil eats breakfast regularly.
- Whether the pupil is bullied at school.
- Teacher demographics (e.g. gender).
- Teacher experience.
- Whether the teacher has an educational background/specialism in mathematics or science.
- The amount of class time allocated each week to mathematics and science.
- Teachers' views on school resource shortages and whether the school has a safe environment.
- Class size.
- Home educational resources.
- Parental reports of the child's numeracy/literacy skills as they started school.
- Highest level of parental education.
- Parental occupation.
- Parental attitudes towards mathematics and science.

While many of these variables are basic demographic information, some are likely to be impacted by a degree of recall bias and/or measurement error. For instance, parents are only likely to be able to partially remember their child's numeracy and literacy skills when they started school, while some pupils may not provide a truthful answer when asked about their experiences of bullying and whether they eat breakfast before school. Consequently, to test the robustness of our OLS results to such issues, in Appendix B and C we present alternative estimates using different sets of controls—including specifications where those variables most likely affected by measurement error have been excluded

Data analysis

Intuition

We undertake two forms of empirical analysis—a 'between-teacher' approach and a 'within-teacher' approach. These can be conceptualised as two separate empirical paradigms, with some important differences between them. With respect to teacher self-efficacy, the 'between-teacher' approach exploits variation across individual teachers in their self-efficacy levels to explore how this is then limited to variation in pupil outcomes. The main advantage of this approach is that it maximises variation in teacher self-efficacy that can be exploited within the analysis. An important limitation, however, is that individual teachers will differ in many ways which could potentially confound the results. Although one may attempt to control for such differences across teachers—and the

pupils they teach—via regression analysis, this is likely to overcome such issues only partially.

In By contrast a 'within–teacher' approach removes all the between–teacher variation in the data. This means that the analysis will now capture how the self-efficacy of individual teachers vary across different aspects of their job (e.g. the different subjects that they teach). A major advantage of the 'within-teacher' approach is that, as one is now effectively looking at differences within the same teachers and pupils, many of the possible confounders will effectively be controlled. However, in doing so, variation in key measures of interest is likely to be greatly reduced—i.e. there will be less information available in the data left to exploit. Thus, as both approaches have both pros and cons, we undertake both a within–teacher and between-teacher analysis to establish whether they provide consistent evidence of the link between teacher self-efficacy and pupil outcomes.

Between–teacher approach

OLS regression is used to estimate the strength of the cross-sectional relationship between teacher self-efficacy and pupil achievement. This model is specified as:

$$A_{ij} = \alpha + \beta.TSE_j + \gamma.C_i + \delta.P_i + \theta.T_j + \varepsilon_{ij} \tag{1}$$

Where:

A_{ij} = Pupil's achievement in science/mathematics, as measured by TIMSS scores.

TSE_j = A set of dummy variables reflecting teachers' self-efficacy, with the 'moderate' category as the reference group.

C_i = A vector of controls capturing aspects of the child's background. In the primary model specification this includes pupil demographics (age, gender, home language, immigrant status) frequency the pupil is absent from school, whether the pupil regularly has breakfast and whether the pupil is bullied at school.

P_i = A vector of controls capturing information gathered within the TIMSS parental questionnaire. In the primary model specification this includes home parental resources, parental reports of pupils' early numeracy/literacy skills, parental education, parental occupation and parental attitude towards mathematics/science.

T_j = A vector of controls capturing information about the class and the teacher. In the primary model specification this includes gender, years of teaching experience, teacher specialism in maths/science, class time allocated to maths/science, teachers' views on school shortages and whether school is safe and class size.

i = Pupil i.
j = Teacher j.

ε_{ij} = The error term. Standard errors are clustered at the classroom level to account for the hierarchical nature of the data.

The model is estimated twice—once where the focus is mathematics and once where the focus is science—within each country. The parameter of interest from this model is β, which captures the strength of the association between teacher self-efficacy and pupil

outcomes. The focus will be upon whether pupil outcomes differ depending upon whether their teacher has high or low levels of self-efficacy, with the middle category as the reference group. Multiple imputation using chained equations has been used to account for missing covariate data, with standard errors clustered at the class (teacher) level to account for the hierarchical structure of the data. As we have standardised TIMSS scores to mean zero and standard deviation one, the β parameter estimates can be interpreted in terms of effect sizes. Analogous estimates using pupil confidence as the outcome are also presented, with the model including the same set of statistical controls. The robustness of the results to alternate model specifications are presented in Appendix B and Appendix C.

Within–teacher approach

As noted above, there are some limitations with the 'between–teacher' approach if one wishes to know whether teacher self-efficacy and pupil outcomes are causally related. The most obvious problem is potential omitted variable bias—that important confounders have not been controlled for. To take a step towards addressing this issue, we also estimate a set of pupil fixed-effect models. These focus upon *relative* differences in pupils' science and mathematics skills and how these relate to *relative* differences in teacher self-efficacy. This is hence our 'within-teacher' approach and has two main advantages over the OLS approach outlined in Equation (1). First, the pupil fixed-effect models control for all things that are invariant within pupils and do not vary by subject, such as gender, socio-economic status, parents' general attitude towards

Table 2. Estimated intra-cluster correlations (ICC) for teacher self-efficacy, pupil self-confidence and pupil achievement.

	Teacher self-efficacy	Pupil self-confidence	Pupil achievement
Czech Republic	0.80	0.41	0.85
Slovak Republic	0.74	0.47	0.88
Italy	0.74	0.51	0.83
Turkey	0.73	0.57	0.91
South Korea	0.72	0.51	0.83
Spain	0.72	0.46	0.86
Finland	0.68	0.46	0.81
Slovenia	0.67	0.47	0.87
Hungary	0.67	0.52	0.88
Sweden	0.63	0.50	0.82
Germany	0.62	0.45	0.84
Chile	0.61	0.45	0.84
Portugal	0.60	0.38	0.80
France	0.60	0.40	0.85
Japan	0.58	0.45	0.83
Canada	0.57	0.46	0.81
Denmark	0.54	0.49	0.80
USA	0.54	0.43	0.85
England	0.51	0.42	0.82
Norway	0.50	0.48	0.83
Ireland	0.46	0.47	0.84
New Zealand	0.45	0.44	0.87
Northern Ireland	0.37	0.38	0.85
Median	**0.61**	**0.46**	**0.84**
Mean	**0.61**	**0.46**	**0.84**

Notes: Figures refer to the proportion of the variation in the given measure that is explained between teachers/ pupils. Values of 1 would indicate that all the variation occurs between pupil/teachers. Estimates based upon a pupil fixed effects model including no other covariates.

education, pupil's general cognitive ability. Second, as we have restricted the sample to those pupils who are taught by the same teacher for mathematics and science, these models also implicitly control for all factors that are invariant within teachers (i.e. that do not vary by subject). Estimates of the intra-cluster correlation for our key measures of interest are presented in Table 2. These capture the proportion of the variation in teacher self-efficacy, pupil self-confidence and pupil achievement that occurs between teachers/ pupils. While around half of the variation in teacher self-efficacy and pupil self-confidence occurs between teachers/pupils, it reaches around 85% for pupil achievement in mathematics relative to science.

The pupil fixed-effects model we estimate is specified as follows:

$$A_{ijk} = \alpha + \beta.TSE_k + \gamma.Special_k + \delta.Time_k + u_i + \varepsilon_{ijk} \tag{2}$$

Where

A_{ijk} = The academic achievement (as measured by TIMSS scores) of pupil i, taught by teacher j in subject k.

TSE_k = Teacher self-efficacy in subject k. This is captured by a vector of dummy variables (low and high, with moderate as the reference group).

$Special_k$ = Whether the teacher has an educational background or specialisation in subject k.

$Time_k$ = The amount of class time devoted to subject k each week.

u_i = Pupil fixed-effects. Note that, as pupils are nested within teachers—and we have restricted the sample to only those who share the same teacher for mathematics and science—that this encompasses a teacher fixed-effect as well.

ε_{ijk} = A subject-specific error term.

i = Pupil i.
j = Teacher j.
k = Subject k (maths or science).

The parameter of interest from this model is β, capturing the link between teacher self-efficacy and pupil outcomes, and can be interpreted in terms of an effect size. Given the limited number of controls in the model, cases with missing data in the covariates are dropped from the analysis. The model is estimated separately for each country, with an analogous approach (including the same set of controls) used to investigate the link between teacher self-efficacy and pupil confidence. Robustness tests are presented in Appendix D, where the fixed-effects models are re-estimated including a different set of controls.

Although this 'within-teacher' approach has certain advantages, we note that caution is still needed when interpreting these results. Two particular issues stand out. First, there could still be subject-specific confounders that we have been unable to control for,

meaning estimates still do not capture causal effects. For instance, there could be a subject specific element to primary school teacher quality (e.g. a teacher being better —and thus more confident—in teaching mathematics than science, or vice-versa) that could have an impact upon the results. Likewise, differences in prior achievement of pupils across mathematics and science cannot be observed in the TIMSS data, meaning this factor is also not controlled. Consequently, we believe that it is prudent to interpret these pupil fixed-effect estimates as capturing conditional associations only, rather than causal effects.

Second, we note the criticism of the pupil-fixed effects approach used here by Jerrim et al. (2017), who focus upon its application to the PISA data. In particular, *author cite* note how the complex PISA test design means that not all pupils answer test questions in all three of the core PISA subjects. This, in turn, means that pupil fixed-effect approaches applied to the PISA data are potentially problematic due to the imputation methodology the survey organisers use to generate the plausible values. Although the TIMSS data we use have a number of similarities to the PISA data considered by Jerrim et al. (2017), there is at least one key difference. Namely, in TIMSS, all pupils answer test questions in each of the subjects, which is not true of PISA (where test scores for some pupils in some subjects are simply imputed based upon how they performed in other subjects and their background characteristics). This in-turn makes a pupil fixed-effect approach in TIMSS more reasonable than in other settings (e.g. PISA) in which it has previously been applied.

Results

Pupil academic achievement

Table 3 begins by presenting OLS estimates of the link between teacher self-efficacy and pupil achievement. Figures in panel (a) refer to mathematics and panel (b) to science. The 'low self-efficacy' column provides the difference in pupil achievement between those taught by teachers with low (bottom third) and average (middle third) self-efficacy levels. Similarly, the 'high self-efficacy' column captures differences in achievement between teachers with average (middle third) and high (top third) levels of self-efficacy.

Together, these results provide a clear and consistent message—there is little evidence to suggest that teacher self-efficacy is associated with pupil achievement in either subject. Effect sizes are generally small (0.1 standard deviations or less) and are rarely statistically significant at conventional thresholds. This finding is summarised in the final two rows of Table 3, where the cross-country median and mean effect sizes are presented. For both mathematics and science, the difference in achievement between the low, moderate and high teacher self-efficacy groups is essentially zero. Similarly, on the whole, there is little evidence to suggest that the relationship between teacher self-efficacy and pupil achievement varies significantly across countries. The OLS estimates therefore suggest that teacher self-efficacy and pupil achievement are, on the whole, not related.

Table 4 presents analogous results, though now based upon the pupil fixed-effects approach. Recall that these estimates focus upon *relative* differences in pupils' mathematics and science achievement, and how this is related to *relative* differences in teachers' mathematics and science self-efficacy. Similar findings once again emerge. In most

Table 3. OLS estimates of the link between teacher self-efficacy and pupils' achievement.

Country	Low self-efficacy		High self-efficacy		
	Beta	SE	Beta	SE	N
(a) Mathematics					
England	0.00	0.09	0.15	0.11	2,745
Japan	0.01	0.10	0.15	0.11	1,239
Sweden	0.09	0.09	0.09	0.10	1,701
Slovak Republic	−0.05	0.08	0.07	0.08	3,597
South Korea	0.01	0.05	0.04	0.05	3,383
Norway	0.01	0.10	0.03	0.11	1,701
Slovenia	0.04	0.06	0.02	0.07	4,113
France	0.16*	0.07	0.02	0.07	3,729
Finland	0.03	0.06	0.02	0.06	3,600
New Zealand	−0.04	0.06	0.01	0.06	5,191
Northern Ireland	0.01	0.07	−0.01	0.07	2,639
Denmark	0.10	0.12	−0.02	0.12	1,303
Portugal	−0.07	0.06	−0.02	0.06	4,412
Chile	−0.06	0.12	−0.02	0.13	2,387
Italy	−0.04	0.08	−0.02	0.11	2,935
Ireland	0.00	0.05	−0.03	0.05	4,325
Canada	0.05	0.06	−0.04	0.07	8,862
Czech Republic	−0.05	0.08	−0.05	0.08	3,637
Turkey	−0.02	0.05	−0.06	0.07	6,431
Hungary	−0.17*	0.05	−0.12	0.06	3,436
USA	−0.20*	0.07	−0.12	0.08	7,172
Spain	−0.03	0.07	−0.14	0.09	4,798
Germany	0.02	0.09	−0.20	0.11	1,811
Median	**0.00**		**−0.02**		
Mean	**−0.01**		**−0.01**		
(b) Science					
Sweden	0.11	0.09	0.14	0.10	1,701
Turkey	0.09	0.05	0.08	0.05	6,447
Ireland	0.09	0.06	0.06	0.06	4,304
France	0.11	0.06	0.04	0.07	3,692
Northern Ireland	−0.05	0.09	0.04	0.07	2,603
Canada	0.15*	0.06	0.01	0.07	8,711
South Korea	−0.02	0.05	0.00	0.05	3,383
Hungary	−0.06	0.05	0.00	0.07	3,470
USA	−0.02	0.06	0.00	0.07	7,017
Denmark	−0.17	0.12	−0.01	0.12	1,302
Slovenia	0.09	0.06	−0.01	0.06	4,112
Spain	0.10	0.08	−0.02	0.08	4,811
Finland	0.03	0.06	−0.02	0.06	3,596
New Zealand	−0.19*	0.06	−0.02	0.06	5,175
Czech Republic	0.02	0.07	−0.02	0.07	3,620
Japan	−0.09	0.08	−0.03	0.11	1,215
Slovak Republic	−0.04	0.07	−0.03	0.08	3,555
Portugal	0.01	0.05	−0.04	0.05	4,446
England	−0.06	0.11	−0.05	0.12	2,351
Norway	0.06	0.10	−0.08	0.11	1,633
Italy	0.00	0.08	−0.12	0.09	2,918
Chile	0.01	0.10	−0.16	0.12	2,424
Germany	−0.03	0.07	−0.19*	0.09	1,776
Median	**0.01**		**−0.02**		
Mean	**0.01**		**−0.02**		

Notes: Figures refer to differences in achievement for pupils with teachers with low and high levels of self-efficacy (compared to average levels of self-efficacy as the reference group). Figures refer to effect sizes. Median/Mean in final rows refer to cross-country averages. Sample restricted to those pupils who have the same teacher for science and maths classes to facilitate comparisons with the fixed-effects estimated. Controls included for pupil and teacher demographics, frequency pupil absent from school, whether child regularly has breakfast, whether pupil bullied at school, teacher experience, teacher specialism in maths/science, class time allocated to science, teachers views on school shortages and whether school is safe, class size, home parental resources, parental reports of early numeracy/literacy skills, parental education, parental occupation and parental attitude towards mathematics/science. A * indicates statistical significance at the five percent level.

Table 4. Child fixed-effect estimates of the link between teacher self-efficacy and pupils' achievement.

| Country | Low self-efficacy | | High self-efficacy | | |
	Beta	SE	Beta	SE	N
England	0.06*	0.03	0.10*	0.03	2,210
Sweden	−0.03	0.03	0.08*	0.04	1,501
Japan	−0.01	0.04	0.07	0.04	1,129
Slovak Republic	0.00	0.02	0.06*	0.02	3,357
Germany	0.00	0.03	0.06	0.04	1,766
Czech Republic	−0.04	0.02	0.06*	0.02	3,392
Chile	0.02	0.03	0.06	0.03	1,765
Northern Ireland	0.05	0.02	0.05*	0.02	2,252
Finland	−0.04	0.02	0.04	0.02	3,565
New Zealand	0.00	0.02	0.04*	0.02	4,671
Canada	−0.04*	0.02	0.02	0.01	8,096
Spain	−0.01	0.02	0.02	0.02	4,331
Turkey	−0.02	0.02	0.01	0.01	5,138
Denmark	−0.03	0.04	0.01	0.05	961
USA	−0.03	0.01	0.00	0.01	6,621
Norway	0.04	0.03	0.00	0.03	1,579
South Korea	0.05	0.03	−0.01	0.02	3,344
Italy	−0.04	0.06	−0.01	0.04	626
France	−0.06*	0.03	−0.01	0.03	1,945
Slovenia	0.04*	0.02	−0.01	0.02	4,069
Ireland	−0.08*	0.02	−0.02	0.02	3,716
Hungary	−0.05*	0.02	−0.03	0.02	2,678
Portugal	−0.08*	0.02	−0.03	0.02	3,932
Median	**−0.02**		**0.02**		
Mean	**−0.01**		**0.02**		

Notes: Figures refer to differences in achievement for pupils with teachers with low and high levels of self-efficacy (compared to average levels of self-efficacy as the reference group). Figures refer to effect sizes. Median/Mean in final rows refer to cross-country averages. Sample restricted to those pupils who have the same teacher for science and maths classes. Controls included for whether the teacher has post-secondary education in mathematics/science, whether the teacher has a specialisation in mathematics/science and amount of class time devoted to mathematics/science. A * indicates statistical significance at the five percent level.

countries estimated effect sizes are small, standing at less than 0.1 standard deviations for the difference between the low, moderate and high self-efficacy groups. This is reflected in the cross-country averages in the bottom row. The difference in achievement between pupils who have teachers with low and moderate levels of teacher self-efficacy is just 0.02 standard deviations. The same holds true with respect to the difference between the moderate and high teacher self-efficacy groups, with the difference again standing at an effect size of just 0.02. Hence, consistent with the OLS estimates presented in Table 3, results from our pupil fixed-effect models strongly suggest that teacher self-efficacy is not an important determinant of pupil achievement.

Pupil self-confidence

Table 5 turns to the results from our secondary analysis, where we consider the link between teacher self-efficacy and pupil academic self-confidence. These are again OLS estimates, with those for mathematics in panel (a) and those for science in panel (b). A broadly similar finding holds as for pupil achievement above. In the vast majority of countries any apparent relationship between teacher and pupil self-confidence is weak, if not zero. Effect sizes again typically stand at 0.1 standard deviations or less, as highlighted

Table 5. OLS estimates of the link between teacher self-efficacy and pupils' confidence.

| Country | Low self-efficacy | | High self-efficacy | | |
	Beta	SE	Beta	SE	N
(a) Mathematics					
Japan	0.09	0.06	0.14	0.08	1,237
Germany	−0.03	0.07	0.13	0.08	1,776
Slovak Republic	−0.07	0.06	0.09	0.07	3,550
Ireland	0.00	0.04	0.08	0.05	4,300
Canada	−0.04	0.03	0.07	0.04	8,758
Slovenia	0.11	0.06	0.07	0.06	4,069
USA	−0.01	0.04	0.07	0.04	7,052
New Zealand	0.04	0.04	0.03	0.04	5,121
Spain	−0.02	0.06	0.03	0.08	4,769
Hungary	−0.04	0.05	0.03	0.06	3,423
Sweden	0.12	0.09	0.03	0.09	1,689
England	−0.07	0.07	0.01	0.08	2,731
Turkey	−0.04	0.06	0.01	0.07	6,381
Finland	−0.05	0.06	0.00	0.06	3,580
Italy	−0.14*	0.05	0.00	0.06	2,896
France	−0.06	0.06	−0.01	0.05	3,696
Northern Ireland	0.03	0.06	−0.01	0.07	2,622
Norway	−0.05	0.07	−0.01	0.10	1,683
Portugal	0.00	0.05	−0.01	0.06	4,384
South Korea	−0.08	0.04	−0.07	0.06	3,374
Denmark	−0.07	0.10	−0.07	0.10	1,287
Czech Republic	−0.04	0.05	−0.07	0.05	3,623
Chile	−0.05	0.09	−0.08	0.08	2,337
Median	**−0.04**		**0.01**		
Mean	**−0.02**		**0.02**		
(b) Science					
France	0.13	0.07	0.18*	0.07	3,628
Germany	0.09	0.09	0.14	0.10	1,736
Japan	0.13	0.08	0.12	0.12	1,211
Sweden	0.20*	0.08	0.12	0.08	1,671
New Zealand	0.01	0.05	0.11	0.05	5,081
Hungary	0.00	0.06	0.10	0.07	3,446
Turkey	0.01	0.05	0.06	0.05	6,401
USA	0.00	0.04	0.05	0.05	6,914
Canada	−0.04	0.05	0.05	0.05	8,552
Northern Ireland	−0.06	0.07	0.05	0.08	2,570
England	−0.07	0.06	0.05	0.08	2,330
Finland	0.04	0.05	0.04	0.06	3,561
Chile	0.08	0.07	0.04	0.09	2,312
Slovenia	−0.07	0.05	0.00	0.05	4,069
Spain	−0.06	0.06	−0.01	0.07	4,767
South Korea	0.00	0.05	−0.02	0.06	3,373
Czech Republic	−0.08	0.05	−0.02	0.07	3,600
Norway	−0.16	0.10	−0.02	0.12	1,624
Ireland	0.01	0.06	−0.02	0.06	4,254
Portugal	−0.08	0.05	−0.04	0.05	4,419
Italy	−0.15*	0.07	−0.05	0.07	2,837
Slovak Republic	−0.07	0.05	−0.06	0.07	3,525
Denmark	−0.31*	0.11	−0.10	0.11	1,280
Median	**0.00**		**0.04**		
Mean	**−0.02**		**0.03**		

Notes: Figures refer to differences in academic self-confidence for pupils with teachers with low and high levels of self-efficacy (compared to average levels of self-efficacy as the reference group). Figures refer to effect sizes. Median/Mean in final rows refer to cross-country averages. Sample restricted to those pupils who have the same teacher for science and maths classes. Controls included for pupil and teacher demographics, frequency pupil absent from school, whether child regularly has breakfast, whether pupil bullied at school, teacher experience, teacher specialism in maths/science, class time allocated to science, teachers views on school shortages and whether school is safe, class size, home parental resources, parental reports of early numeracy/literacy skills, parental education, parental occupation and parental attitude towards mathematics/science. A * indicates statistical significance at the five percent level.

by the bottom rows displaying the cross-country averages. For instance, in the median country, the difference in pupil confidence in mathematics is just 0.04 standard deviations when comparing the low and moderate teacher self-efficacy groups (and zero in science). The same holds true for the comparison of pupil confidence between teachers with moderate and high levels of self-efficacy (0.01 standard deviations for the median country in mathematics and 0.04 standard deviations in science). Moreover, evidence of any substantial cross-national variation in the effect sizes is very limited. Our overall interpretation of Table 5 is therefore that, based upon OLS estimates alone, there does not appear to be any meaningful relationship between teacher self-efficacy and pupil self-confidence.

A broadly similar finding emerges in Table 6, where we focus upon teachers' *relative* self-efficacy in science and mathematics, and how this is related to pupils' *relative* self-confidence across these two subjects. In almost all countries, the difference in pupil academic self-confidence between the low and moderate teacher self-efficacy groups is very small (less than 0.1 standard deviations with the exception of Portugal) and rarely statistically significant. On average across countries the difference is just 0.02. The same broadly holds true for the comparison between moderate and high teacher self-efficacy groups, though with a handful of notable exceptions. In particular, there are five

Table 6. Child fixed-effect estimates of the link between teacher self-efficacy and pupils' self-confidence.

Country	Low self-efficacy		High self-efficacy		
	Beta	SE	Beta	SE	N
Sweden	0.02	0.06	0.21*	0.07	1,476
Japan	0.00	0.07	0.17*	0.08	1,126
Norway	0.01	0.06	0.14*	0.06	1,561
Czech Republic	−0.03	0.05	0.13*	0.06	3,371
Spain	0.01	0.04	0.13*	0.04	4,280
Northern Ireland	0.06	0.06	0.10	0.06	2,227
Slovak Republic	0.03	0.05	0.10*	0.05	3,320
Turkey	0.05	0.04	0.08*	0.03	5,092
Germany	−0.07	0.07	0.08	0.08	1,498
New Zealand	−0.08*	0.04	0.08	0.04	4,574
Finland	0.04	0.04	0.06	0.05	3,528
USA	−0.04	0.03	0.05	0.03	6,451
Denmark	0.02	0.07	0.03	0.09	918
Canada	−0.06	0.03	0.03	0.03	7,897
South Korea	0.07	0.05	0.02	0.05	3,334
Ireland	−0.08*	0.04	0.02	0.04	3,670
Slovenia	0.04	0.04	0.01	0.04	4,011
Portugal	−0.14*	0.04	0.00	0.05	3,908
Italy	−0.08	0.12	−0.02	0.09	611
Chile	−0.06	0.07	−0.02	0.07	1,699
England	−0.08	0.05	−0.06	0.06	2,185
France	−0.09	0.06	−0.06	0.08	1,911
Hungary	−0.05	0.04	−0.09	0.06	2,659
Median	**−0.03**		**0.05**		
Mean	**−0.02**		**0.05**		

Figures refer to differences in academic self-confidence for pupils with teachers with low and high levels of self-efficacy (compared to average levels of self-efficacy as the reference group). Figures refer to effect sizes. Median/Mean in final rows refer to cross-country averages. Sample restricted to those pupils who have the same teacher for science and maths classes. Controls included for whether the teacher has post-secondary education in mathematics/science, whether the teacher has a specialisation in mathematics/science and amount of class time devoted to mathematics/science. A * indicates statistical significance at the five percent le.

countries (Sweden, Japan, Norway, Czech Republic and Spain) where the estimated effect size is above 0.1 standard deviations and statistically significant at the five percent level. Yet these potential exceptions should not distract from the fact that the average difference across countries (for the comparison between teachers with moderate and high levels of self-efficacy) remains small, standing at just 0.05 standard deviations. Consequently, on the whole, estimates from our OLS and pupil fixed-effects models are consistent; there is little clear evidence of a strong link between teacher self-efficacy and pupils' academic self-confidence.

Robustness checks

The robustness of these results is tested in Appendix B (alternative OLS estimates for mathematics), Appendix C (alternative OLS estimates for science) and Appendix D (alternative fixed-effect estimates). These present a range of alternative estimates based upon different model specifications, which vary the set of statistical controls (further details are provided in the notes to the appendix tables). Overall, the results from these alternative model specifications are consistent with those reported above. In most countries, effect sizes tend to be small and statistically insignificant at conventional thresholds. Relatedly, the average effect size across countries is consistently less than 0.05 standard deviations. We hence reject the conventional wisdom that has emerged in the education literature, and argue is seems unlikely that teacher self-efficacy and pupil achievement are causally related.

Discussion

There are compelling theoretical reasons to believe that teachers with higher levels of self-efficacy might improve pupil outcomes (Lauermann & Butler, 2021). As a result, researchers have spent decades trying to empirically establish the relationship between these two variables. However, empirical findings remained mixed and very little research goes beyond descriptive/correlational research that relies on a 'selection on observables' assumption (Lauermann & ten Hagen, 2021; Pekrun, 2021). This paper sought to address these limitations in the existing literature by presenting new international evidence on the link between teacher self-efficacy and pupil outcomes. Using large-scale data drawn across multiple countries, we apply a pupil fixed-effects approach that has rarely been utilised in this literature before. Against the main thrust of the existing literature, we found no evidence of a link between teacher self-efficacy and pupil outcomes, both in terms of academic achievement and pupils' academic self-confidence. These results hold true across both science and mathematics, can be observed across both our OLS and pupil fixed-effect models, with similar findings emerging across multiple countries. They also persist across several different model specifications. We therefore conclude that there is no evidence of a link between teacher self-efficacy and pupil's outcomes.

It is interesting to consider why our results differ to most of the existing literature. In a recent review of methods and findings in this literature, Lauermann & ten Hagen (2021) suggest three possible theoretical explanations for such conflicting findings. First, we might expect fewer null findings when the self-efficacy measures and the outcome measures are at a similar level of specificity/generality. However, in our study, the self-efficacy measures

and pupil achievement measures are at the same level of generality—the subject—meaning this is unlikely to explain our null findings. Second, teacher self-efficacy may be very stable, thus limiting the amount of variation available to use in longitudinal research designs. However, once again, this seems unlikely to explain our null findings because our within-teacher design uses variation across subjects, rather than across time. Third, Lauermann & ten Hagen observe that relationships tend to be larger when pupils are younger, know the teacher well, and are faced with harder tasks. Our analysis is based on pupils who are 9/10 years old, based in primary/elementary classroom in which a given teacher works with them for the majority of their lessons, studying maths/science. Based on this, we believe our study should have a reasonable chance of finding a relationship, if indeed one exists.

Pekrun (2021) provides an alternative explanation for why changes in teacher self-efficacy may not translate into improved pupil outcomes. In particular, the increased persistence with different teaching approaches may not be sufficient to improve teaching practice. Rather, it is necessary to consider the interplay with other factors such as knowledge and skills related to effective teaching. This appears to be the most elegant theoretical account for the null findings we observed in this research.

Alongside these theoretical considerations, there may be limitations of our data or research design that account for the lack of a finding. First, although our pupil fixed-effects approach means we are likely to control for many more key confounders than almost all the existing literature, estimates may still not capture cause and effect. Specifically, we are unable to account for confounders of the efficacy achievement relationship that vary across subjects within pupil-teacher pairs. Second, although the fixed-effects ('within-person') approach has some important benefits, it does also lead to substantially less variation in the data for our analysis to exploit. This could potentially explain the null effects produced under this approach, particularly for pupil achievement, where only around 15% of the total variation occurs within individuals (recall Table 2). However, our OLS (between-person) estimates are very similar to those using individual fixed-effects (within-person estimates), suggesting our substantive findings are robust to this potential issue. Third, teacher self-efficacy has been measured in TIMSS in one particular way, with previous research suggesting that the precise teacher self-efficacy scale used can have an impact upon the magnitude of the results (Kim & Seo, 2018).

Finally, it may be the case that TIMSS does not capture the academic progress pupils make during the current school year—when they would have been taught by the teachers included in the teacher sample. In other words, the TIMSS test captures the cumulation of pupils' mathematics and science learning over several years, while only the self-efficacy levels of their current teacher are recorded. To resolve this issue, we see great potential in the longitudinal TIMSS study that is currently being planned (https://www.iea.nl/pub lications/timss-longitudinal-study-introductory-package). This will allow researchers to investigate the learning gains pupils make over an academic year and how this relates to the self-efficacy levels of who they were taught by. As such, it offers the opportunity for a third analytic approach to be used (a 'value-added' model) potentially in combination with within-pupil analysis.

Conclusion

Were such a study to replicate our main findings, then the suggestion that teacher self-efficacy are causally related to pupil achievement would be further brought into doubt. While it may be true that there are plausible theoretical reasons to believe that teacher self-efficacy might benefit pupils, this provides no guarantee that it is indeed the case. There are only so many times that well designed empirical studies can find no relationship before the hypothesis that teacher self-efficacy benefits pupils in any measurable way will need to be either very substantially revised or dropped altogether.

Notes

1. Other international studies such as PISA do not include measures of teacher self-efficacy, and do not allowed teachers to be linked to individual pupils.
2. In most countries, the TIMSS variable ITCOURSE was used to identify these teachers (those coded as 7). England is an exception, where we identify pupils as having the same teacher for science and maths if the teacher in these two subjects works in the same school, has the same number of years teaching experience, is the same gender, holds the same educational qualifications and has the same class size and characteristics across the two subjects.
3. While we recognise there are other ways that this can be achieved—e.g. including a quadratic term for the teacher-self-efficacy scale—this again complicates the communication and interpretability of results still further.
4. In the Programme for International Student Assessment (PISA), three subjects (reading, mathematics and science) are covered. Yet, because of the test design, students may only answer question in one or two of these domains.
5. We have chosen to use just one plausible value rather than all five for speed of estimation and convenience. As the technical literature on plausible values notes, '*using one plausible value or five plausible values does not really make a substantial difference on large samples*' (OECD, 2009, p. 46) and that '*working with one plausible value instead of five will provide unbiased estimates of population parameters*' (OECD, 2009, p. 43). In our application, using all five plausible values instead of just one would lead to a very slight increase in the reported standard errors.
6. The parent questionnaire was not conducted in England or the United States. This means that some of the parent reported information is not available for these two countries.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

John Jerrim 🆔 http://orcid.org/0000-0001-5705-7954
Sam Sims 🆔 http://orcid.org/0000-0002-5585-8202
Mary Oliver 🆔 http://orcid.org/0000-0002-4177-0920

References

American Psychological Association. (2020). Teaching Tip Sheet; Self-Efficacy. Accessed 29/06/2020 from https://www.apa.org/pi/aids/resources/education/self-efficacy.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Bitler, M., Corcoran, S., Domina, T., & Penner, E. (2019). Teacher effects on student achievement and height: A cautionary tale. *NBER working paper 26480*.

Burgess, S. (2019). Understanding teacher effectiveness to raise pupil attainment. *IZA World of Labor: 465*. https://doi.org/10.15185/izawol.465

Burić, I., & Kim, L. E. (2020). Teacher self-efficacy, instructional quality, and student motivational beliefs: An analysis using multilevel structural equation modeling. *Learning and Instruction*, *66*, 101302. https://doi.org/10.1016/j.learninstruc.2019.101302

Caprara, G., Baarbaranelli, C., Steca, P., & Malone, P. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study as the school level. *Journal of School Psychology*, *44*(6), 473–490. https://doi.org/10.1016/j.jsp.2006.09.001

Gavora, P. (2010). Slovak pre-service teacher self-efficacy: Theoretical and research considerations. *The New Educational Review*, *21*(2), 17–30.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, *76*(4), 569–582. https://doi.org/10.1037/0022-0663.76.4.569

Guo, Y., Connor, C. M., Yang, Y., Roehrig, A. D., & Morrison, F. J. (2012). The effects of teacher qualification, teacher self-efficacy, and classroom practices on fifth graders' literacy outcomes. *The Elementary School Journal*, *113*(1), 3–24. https://doi.org/10.1086/665816

Guo, Y., Piasta, S. B., Justice, L. M., & Kaderavek, J. N. (2010). Relations among preschool teachers' self-efficacy, classroom quality, and children's language and literacy gains. *Teaching and Teacher Education*, *26*(4), 1094–1103. https://doi.org/10.1016/j.tate.2009.11.005

Hill, H. C., Charalambous, C. Y., & Chin, M. J. (2019). Teacher characteristics and student learning in mathematics: A comprehensive assessment. *Educational Policy*, *33*(7), 1103–1134. https://doi.org/10.1177/0895904818755468

Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: A longitudinal analysis. *Journal of Educational Psychology*, *105*(3), 774–786. https://doi.org/10.1037/a0032198

Holzberger, D., Philipp, A., & Kunter, M. (2014). Predicting teachers' instructional behaviors: The interplay between self-efficacy and intrinsic needs. *Contemporary Educational Psychology*, *39*(2), 100–111. https://doi.org/10.1016/j.cedpsych.2014.02.001

Jerrim, J., Lopez-Agudo, L., & Marcenaro-Gutierrez, O. (2020). The association between homework and primary school children's academic achievement. International evidence from PIRLS and TIMSS. *European Journal of Education*, *55*(2), 248–260. https://doi.org/10.1111/ejed.12374

Jerrim, J., Lopez-Agudo, L. A., Marcenaro-Gutierrez, O. D., & Shure, N. (2017). What happens when econometrics and psychometrics collide? An example using the PISA data. *Economics of Education Review*, *61*, 51–58.

Kim, K., & Seo, E. (2018). The relationship between teacher efficacy and students' academic achievement: A meta-analysis. *Social Behavior and Personality: An International Journal*, *46*(4), 529–540. https://doi.org/10.2224/sbp.6554

Kirabo, J., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, *1*(4), 85–108. https://doi.org/10.1257/app.1.4.85

Klassen, R., & Tze, V. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, *12*, 59–76. https://doi.org/10.1016/j.edurev.2014.06.001

Klassen, R. M., Tze, V. M., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise? *Educational Psychology Review*, *23*(1), 21–43. https://doi.org/10.1007/s10648-010-9141-8

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, *36*(4), 476–500. https://doi.org/10.3102/0162373713519496

Künsting, J., Neuber, V., & Lipowsky, F. (2016). Teacher self-efficacy as a long-term predictor of instructional quality in the classroom. *European Journal of Psychology of Education*, *31*(3), 299–322. https://doi.org/10.1007/s10212-015-0272-7

Lauermann, F., & Butler, R. (2021). The elusive links between teachers' teaching-related emotions, motivations, and self-regulation and students' educational outcomes. *Educational Psychologist*, *56*(4), 243–249. https://doi.org/10.1080/00461520.2021.1991800

Lauermann, F., & ten Hagen, I. (2021). Do teachers' perceived teaching competence and self-efficacy affect students' academic outcomes? A closer look at student-reported classroom processes and outcomes. *Educational Psychologist*, *56*(4), 265–282. https://doi.org/10.1080/00461520.2021.1991355

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self-and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology*, *81*(2), 247–258. https://doi.org/10.1037/0022-0663.81.2.247

Mok, M., & Moore, P. (2019). Teachers and self-efficacy. *Educational Psychology*, *39*(1), 1–3. https://doi.org/10.1080/01443410.2019.1567070

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics*. Retrieved from TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/international-results/

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

Pekrun, R. (2021). Teachers need more than knowledge: Why motivation, emotion, and self-regulation are indispensable. *Educational Psychologist*, *56*(4), 312–322. https://doi.org/10.1080/00461520.2021.1991356

Perera, H. N., & John, J. E. (2020). Teachers' Self-efficacy Beliefs for Teaching Math: Relations with Teacher and Student Outcomes. *Contemporary Educational Psychology*, *61*, 101842. https://doi.org/10.1016/j.cedpsych.2020.101842

Praetorius, A. K., Lauermann, F., Klassen, R. M., Dickhäuser, O., Janke, S., & Dresel, M. (2017). Longitudinal relations between teaching-related motivations and student-reported teaching quality. *Teaching and Teacher Education*, *65*, 241–254. https://doi.org/10.1016/j.tate.2017.03.023

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, *80*(1), 1–28. https://doi.org/10.1037/h0092976

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, *68*(2), 202–248. https://doi.org/10.3102/00346543068002202

Woolfolk Hoy, A., Hoy, W. K., & Davis, H. A. (2009). Teachers' self-efficacy beliefs. In K. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 627–653). Routledge.

Yeung, A. S., Craven, R. G., & Kaur, G. (2014). Teachers' self-concept and valuing of learning: Relations with teaching approaches and beliefs about students. *Asia-Pacific Journal of Teacher Education*, *42*(3), 305–320. https://doi.org/10.1080/1359866X.2014.905670

Zee, M., & Koomen, H. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research*, *86*(4), 981–1015. https://doi.org/10.3102/0034654315626801

Zhu, M., Liu, Q., Fu, Y., Yang, T., Zhang, X., & Shi, J. (2018). The relationship between teacher self-concept, teacher efficacy and burnout. *Teachers and Teaching*, *24*(7), 788–801. https://doi.org/10.1080/13540602.2018.1483913