



## Looping: does keeping the same secondary school mathematics teacher lead to better outcomes?

John Jerrim & Loic Menzies

**To cite this article:** John Jerrim & Loic Menzies (16 Feb 2024): Looping: does keeping the same secondary school mathematics teacher lead to better outcomes?, Oxford Review of Education, DOI: [10.1080/03054985.2024.2305462](https://doi.org/10.1080/03054985.2024.2305462)

**To link to this article:** <https://doi.org/10.1080/03054985.2024.2305462>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 16 Feb 2024.



[Submit your article to this journal](#)



Article views: 140




[View related articles](#)



[View Crossmark data](#)

# Looping: does keeping the same secondary school mathematics teacher lead to better outcomes?

John Jerrim <sup>a</sup> and Loic Menzies<sup>b</sup>

<sup>a</sup>Social Research Institute, University College London, London, UK; <sup>b</sup>Jesus College, University of Cambridge, Cambridge, UK

## ABSTRACT

Previous research has suggested that pupils may benefit from retaining the same teacher for more than one academic year, leading to better relationships, lower levels of absence and higher test scores. This has led some to suggest that leaders should consider so-called ‘looping’ when devising their school timetable. We provide new evidence on how looping relates to an array of pupil, teacher and class outcomes. Using rich international data from the TALIS video study, we typically find small associations, particularly when ‘looping’ is considered at the class-level. We thus conclude that looping is probably not something that school leaders should either purposefully avoid or dogmatically pursue.



## KEYWORDS


Teacher effectiveness; TALIS; mathematics; looping

Countries across the world are constantly attempting to improve young people’s academic achievement. Yet, with ever increasing pressure on budgets, there are limited resources to achieve such goals. This makes initiatives proven to be effective but expensive – such as one-to-one tuition (Education Endowment Foundation, 2023) – difficult to scale. Alternative options are hence increasingly sought where schools’ existing practice can be modified at little additional cost.

One option that has gained some interest – particularly in the United States – is ‘looping’. This is where teachers provide instruction to the same class for more than one year. In theory, keeping the same teacher with the same pupils allows better pupil-staff and pupil-pupil relationships to be formed and a degree of ‘continuity of care’ (Menzies, 2023). Teachers may therefore know their pupils better and can ‘adjust and target their teaching styles’ (Hill & Jones, 2018, p. 8). This interpretation finds support in studies suggesting that teachers’ ‘newness’ can be seen on a continuum linked to their effectiveness (Atteberry et al., 2017). Adaptation between teachers and pupils may also be a two-way process, with pupils adjusting to their teachers in instances of repeat-matches (Wedenoja et al., 2022).

Where looping takes place, classroom rules and expectations may already be established. This might translate into fewer disruptions, better classroom management and

**CONTACT** John Jerrim  [J.Jerrim@ucl.ac.uk](mailto:J.Jerrim@ucl.ac.uk)  Social Research Institute, University College London, 55-59 Gordon Square, Bloomsbury, WC1H 0AL London

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/03054985.2024.2305462>

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

improved behaviour. This may increase the quantity and quality of learning time, and then better pupil outcomes (Tourigny et al., 2020). Because these effects manifest at a classroom level due to changes in classroom climate (rather than solely through individual teacher-pupil relationships), there may be spill-over benefits to pupils not taught by the same teacher previously. These spill-over effects have been found to be greater in classes with a greater proportion of repeat teacher-pupil matches (Hill & Jones, 2018; Wedenoja et al., 2022).

Teachers may also prefer, or benefit from, retaining the same class, though there could also be negative consequences such as increased workload due to teachers having to plan new material (Menzies, 2023). Looping might therefore inadvertently drive teachers out of schools by increasing 'grade reassignment' – a strong predictor of turnover (Ost & Schiman, 2015).

Previous research into looping has nevertheless found some positive benefits. Albornoz et al. (2023) find a positive association between looping of 8<sup>th</sup> graders and attendance, progression, behaviour, test scores and university admission exams in Chile. Using data on the mathematics attainment of pupils in Grades 3–5 in North Carolina, Hill and Jones (2018) find looping to result in 'small but significant test score gains', noting these gains are greatest for pupils taught by less effective teachers. Hwang et al. (2021) apply a similar approach to administrative data on reading and mathematics attainment among pupils in Grades 3–8 in Indiana, finding similar results.

Exploring the attitudes of parents towards looping in the United States, Nichols and Nichols (2003) report it to be associated with pupil motivation and positive attitudes towards the school environment. Based upon a sample of 206 pupils from Mississippi in Grades 6–8, Franz et al. (2010) found looping to be associated with higher test scores. In Tennessee, Wedenoja et al. (2022) report looping to have small positive effects on high school pupils, with high achievers benefitting the most academically and boys of colour to benefit most behaviourally. In a small sample of pupils in Grades 2–5 in Florida, Cistone and Shneyderman (2004) find a difference of around 0.16 standard deviations in reading scores and 0.21 in mathematics. Tourigny et al. (2020) investigate looping amongst 192 French-Canadian pupils in Grade 4 from 12 schools, finding looping to be positively associated with achievement, but no association with pupil-teacher relationships. Thus, overall, the existing literature typically suggests looping to have positive effects upon pupils' outcomes. However, conclusions regarding which groups benefit most from looping are more mixed.

Despite this impressive body of research, there continue to be gaps in the evidence base. While much previous work has considered the link between looping and pupils' attendance levels and test scores, few quantitative studies have considered its relationship with broader outcomes such as subject interest and self-confidence. Likewise, although looping in theory leads to better staff-pupil relationships and a more positive classroom environment, limited empirical evidence has been presented on this issue. One small-scale qualitative study in a childcare setting reports positive findings in this respect (Hegde & Cassidy, 2004). In contrast, a small-scale quantitative study in Canada finds contradictory and inconclusive results regarding the relationship between looping and teacher-pupil relationships (Tourigny et al., 2020). While pupil outcomes have rightly been the focus, how

looping impacts teachers has received less attention. Meanwhile, most existing research into looping stems from the United States, with evidence from other countries more limited.

This paper attempts to resolve such issues using longitudinal data from across eight countries that participated in the Teaching and Learning International Survey (TALIS) Video Study. It does so by addressing five research questions. To begin, we replicate and extend the (mostly US) evidence on looping by considering a broader array of pupil outcomes. We consider not only how looping is related to achievement and absences, but also subject interest and self-confidence, using data from a diverse mix of education systems. This allows us to establish whether similar patterns are observed internationally as in the United States, while also extending the evidence to a broader set of outcomes:

- Research question 1: Is there a positive association between looping and pupil outcomes?

Next, we investigate the suggested mechanisms behind why looping may be effective. In theory, by retaining the same class, better relationships between pupils and teachers (and between pupils) are formed, leading to a more positive learning environment. This includes better classroom management and behaviour, with rules and expectations already clear. There has, however, been little research measuring the strength of the association between looping and such outcomes. Our second and third research questions thus ask:

- Research question 2: Is classroom behaviour and management better in looped classrooms?
- Research question 3: Are pupil-teacher and pupil-pupil relationships better in looped classrooms?

Pupils are not the only group that may be impacted by looping; teachers are as well. They may, for instance, feel more able to reach their learning goals, leading to greater enjoyment of teaching the class and improvements to their self-efficacy. On the other hand, looping could lead to increased workload (e.g. having to create extra material) or reduced enthusiasm for their job – particularly when retaining a class of challenging pupils. Research question 4 thus focuses on teacher outcomes:

- Research question 4: How is looping linked to teacher outcomes? Are teachers ‘happier’ teaching looped classes?

Finally, although most pupils may remain with the same teacher and peer-group, some new pupils may enter the class as well (e.g. due to moving up or down an ability group, changing school, timetable clashes). But how outcomes compare for pupils who are new to a looped class relative to their peers (who have been part of the group for a prolonged period) has been underexplored – particularly non-achievement outcomes. Do they feel their relationship with their teacher is weaker, like they are an outsider in the class, with their achievement suffering as a result? Our final research question asks:

- Research question 5: How do outcomes and perceptions of the class environment differ between pupils retaining the same teacher to those new to the class?

## Data

This study draws upon the 2018 TALIS Video Study (TVS). These data were gathered from eight jurisdictions (Chile, Colombia, England, Germany, Japan, Spain [Madrid], Mexico and China [Shanghai]). The intention was for schools to be randomly sampled and then a randomly selected mathematics teacher from each school to take part. However, Germany, Japan, Madrid and Shanghai did not randomly select teachers, while in England response rates were low. It is hence better to consider these data as a convenience sample. See OECD (2021, Chapter 12) for further details. This issue is considered further in our sensitivity analyses.

As argued by an anonymous reviewer, there is debate about whether one should include standard errors and tests of statistical significance when analysing a convenience sample. Conventional practice is to report such inferential statistics, even if the sample is not randomly selected from the population. However, some argue this is not appropriate, and propose no such inferential statistics should be presented (Gorard, 2015). Irrespective of the approach taken, our substantive conclusions remain unchanged. But, to acknowledge both perspectives, we refrain from discussing ‘statistical significance’ in the main text and, instead, concentrate on the strength of the associations, as indicated by effect sizes. Standard errors and results from significance tests are however documented within the tables for those interested in this information.

Near the start of the academic year, pupils completed a 40-minute mathematics test and completed a background questionnaire. Their mathematics teacher completed a baseline questionnaire as well. A selection of their mathematics lessons – those focused on quadratic equations – were then recorded and judged by a set of expert observers. Once the lessons on quadratic equations were complete, pupils sat another assessment and – along with teachers – answered an end-of-study questionnaire. OECD (2021) provides further information. The average age of participating pupils was 14 years and 9 months (interquartile range 14 years 2 months to 15 years 4 months).

### *Identification of ‘looped’ classes*

In the baseline questionnaire, pupils were asked ‘since when have you had your CURRENT mathematics teacher’: (a) this school year; (b) last school year; (c) before the last school year, with the instruction to ‘just include the time you have had your current teacher continuously up until this school year (don’t include the time you had the current teacher in earlier school years followed by different teachers)’.

To identify looped classrooms, we first convert responses into a binary variable, coded zero if pupils had a new mathematics teacher this year (option a above) and coded one if their mathematics teacher had taught them for at least the previous academic year (options b and c). This information is then aggregated to the class level, providing a continuous variable capturing the proportion of pupils in each class reporting having the same mathematics teacher now as previously. We treat any class where at least 75% of pupils retained the same mathematics teacher as a looped classroom, while those with less than

**Table 1.** Number of looped classrooms in the TALIS video study by country.

	Not looped		Looped	
	Classes	Pupils	Classes	Pupils
Chile	38	1034	40	1101
Colombia	53	1581	20	587
England	59	1386	16	371
Germany	25	591	23	522
Japan	27	755	40	1161
Madrid	42	1013	10	242
Mexico	68	1868	16	435
Shanghai	14	443	64	1972
<b>Total</b>	<b>326</b>	<b>8,671</b>	<b>229</b>	<b>6,391</b>

Note: Looped classes are defined as those where at least 75% of pupils were taught by their mathematics teacher for at least two consecutive years.

25% of pupils identified as a non-looped class.<sup>1</sup> Appendix A (see online supplementary material) tests the sensitivity of results to using a 50% threshold to identify looped classrooms instead. Table 1 provides information on sample sizes, including the number of pupils in looped and non-looped classrooms. While more than 50% of pupils in Shanghai, Japan and Chile are taught in looped classrooms, only around 20% of pupils are in England, Madrid and Mexico.

### **Test scores**

Pupils completed two 40-minute tests: one at the start of the data collection period (30 questions; Cronbach alpha = 0.84) and one at the end (25 questions; Cronbach alpha = 0.71). The baseline test covered a mix of basic mathematics skills, pre-requisite knowledge needed to tackle quadratic equations. The post-test focused on young people's skills in solving problems involving quadratic equations. An advantage of the post-test having this narrow focus is that it assesses pupils' skills on material they have just been taught in their school lessons. It should thus be particularly sensitive to the learning and instruction that occurred during the video-recorded lessons – more so than a general mathematics test assessing the full spectrum of students' abilities. The average duration between the pre- and post-test was 41 days, reflecting the period during which lessons covering quadratic equations occurred. Online Appendix B illustrates how the test scores are distributed.

### **Other pupil outcomes**

Pupils were asked questions capturing their broader attitudes towards mathematics. This included three questions about their interest in the material taught (e.g. 'I was interested in the topic of quadratic equations'); five questions capturing their self-confidence (e.g. 'Quadratic equations were easy for me'); five questions about their self-efficacy (e.g. 'I was confident I could master the skills being taught during the unit on quadratic equations'); and 11 questions regarding their confidence in solving relevant tasks (e.g. 'Solving any quadratic equation – example:  $4x^2 + 6x + 3 = 0$ '). Pupils' responses have been combined by the survey organisers into a set of continuous scales. We draw upon these to consider how

looping is related to broader pupil outcomes. They were also asked how many quadratic equations lessons they were absent from, which we use to consider the relationship between looping and absenteeism.

### ***Classroom behaviour and teachers' classroom management***

Three sources of information are used to measure class behaviour and teachers' classroom management. The first are pupil responses to ten questions (repeated at baseline and endpoint), such as 'there was much disruptive noise in the classroom' and 'in our teacher's class, we were aware of what was allowed and what was not'. The survey organisers have converted responses into an overall classroom environment scale, along with sub-scales capturing (i) frequency of classroom disruptions and (ii) pupils' perceptions of their teacher's classroom management skills. We can thus investigate whether looping is associated with pupil reports of the class environment at two points during the academic year.

The second information source is the class teacher, who were asked very similar questions about the classroom environment (see online Appendix B). The survey organisers have also created an overall classroom environment scale (plus 'disruptions' and 'teacher classroom management' sub-scales) using teacher reports. We thus investigate whether results obtained using pupils' responses are replicated when using teachers' responses instead.

Finally, teachers had two of their quadratic equation lessons recorded and rated along six dimensions of instructional quality by two expert observers. One of these domains was 'classroom management' and comprised three components:

- Routines. How effectively common tasks (such as handing out workbooks) were carried out.
- Disruptions. How frequent disruptions were, how effectively the teacher managed them and how much learning time was lost.
- Monitoring. How frequently and consistently the teacher monitored the activity of the entire class.

Each component was scored by the expert raters using a four-point scale for each lesson. The survey organisers then created overall scores across these three dimensions by taking the average awarded across raters over the two lessons. An overall 'classroom management' domain score was also derived by averaging over the routines, disruptions and monitoring components. Our analysis focuses on the relationship between looping and the overall classroom management score, as well as the routines and disruptions sub-domains. Online Appendix B presents the distribution of these scales.

### ***Teacher-pupil and pupil-pupil relationships***

Pupils were asked about the nature of the relationship they had with the teacher and other pupils in the class at both baseline and endpoint. We focus on three scales derived by the survey organisers:

- (a) The pupil-teacher relationship. Captured by five questions such as ‘my mathematics teacher really listened to what I had to say’ and ‘my mathematics teacher made me feel she/he really cared about me’.
- (b) Relationships with other pupils. Captured by four questions such as ‘I felt like an outsider (or left out of things) in my mathematics class’ and ‘I felt like I belonged in my mathematics class’.
- (c) Teacher support for competence. Captured by four questions such as ‘I felt that our mathematics teacher understood me’ and ‘Our mathematics teacher listened to my view on how to do things’.

Very similar questions were asked to teachers regarding ‘pupil-teacher relationships’ and ‘teacher support for competence’, with the survey organisers also creating scales for these measures using teachers’ reports. We again investigate the consistency of results across pupil and teacher reports.

Expert observers also provided scores regarding intra-classroom relationships from the video recorded lessons. The first is whether respect is shown within the classroom. The expert judges awarded higher ratings when pupils and teachers listened to one another, used respectful language and were generally well-mannered. The second aspect is whether pupils were willing to take ‘risks’ – e.g. being willing to ask the teacher for guidance or share their thoughts publicly (i.e. to the rest of the class). If looping makes pupils feel more comfortable in the classroom – e.g. by knowing and developing a relationship with their teacher and peers – one may expect them to be more willing to take such ‘risks’.

### ***Teacher views of teaching the class***

Teachers were asked numerous questions about how they felt when teaching the relevant class. We consider whether looping is associated with:

- Enthusiasm for teaching the class. Four questions such as ‘I teach mathematics in this class with great enthusiasm’. Recorded at both baseline and endpoint.
- Enjoyment of teaching the class. Four questions such as ‘I enjoyed teaching these students’. Endpoint only.
- Anxiety when teaching the class. Four questions including ‘I feel uneasy when I think about teaching these students’. Endpoint only.
- Anger when teaching the class. Four questions such as ‘I often feel annoyed while teaching these students’. Endpoint only.
- Achievement of learning goals. Teachers’ responses to five statements answering whether they felt they met their learning goals, such as ‘enhancing students’ mathematical knowledge and skills’. Endpoint only.
- Teacher self-efficacy. Teachers were asked ‘to what extent could you do the following in the target class during the quadratic equations unit’ and then to respond to 12 statements such as ‘get students in this class to believe they can do well in work on quadratic equations’. Baseline and endpoint.



- Self-efficacy in classroom management. A sub-scale of the teacher self-efficacy scale described above, focusing on four statements such as ‘make my expectations about student behaviour clear in this class’. Baseline and endpoint.

Further details about these scales are provided in Appendix B online.

## Methodology

### *What might drive selection into looped classes?*

Looping may be either intentional or unintentional and is affected by national norms and historic policy. Nichols and Nichols (2003) note that looping is the norm in some Asian countries and Albornoz et al. (2023) report that it is particularly common in Germany, China, Finland, Israel, Sweden and Japan. Table 1 provides further evidence suggesting that norms vary between countries.

School leaders may also make intentional, class-specific decisions about looping, for example if they observe particularly strong teacher-pupil relationships in the current allocation of classes. They may thus decide to keep a teacher with the same pupils the following year to maintain this positive relationship. Empirically, this would likely lead to an upward bias in the estimated association between looping and pupil outcomes. Schools may reserve this policy for certain teachers; for instance, staff that are new to the school while they settle in to their new working environment. However, there is little evidence on looping as a conscious school policy from the previous literature (particularly outside of the United States).

Class-specific looping may also happen due to teachers’ individual influence over class allocation and their desire to keep preferred classes and move on from those they favour less. The degree of teachers’ influence over looping may be linked to years of service or seniority. For example in certain US states senior teachers’ right to influence class allocation is specified as part of collective bargaining agreements (Grissom et al., 2015). There is also evidence to suggest that in England, teachers’ influence over class allocation is linked to seniority and that teachers’ preference for looping may be linked to the extent to which they consider it to be in their pupils’ interests (Menzies, 2023). In Canada, Tourigny et al. (2020, p. 745) suggest that more ‘competent’ teachers may be more likely to implement looping, though they do not present any empirical evidence for this.

Looping may also be unintentional; timetabling may simply lead to some staff being allocated to the same group of pupils for more than one year. For instance, a teacher currently allocated to Year 9 may next year be allocated to Year 10. Then, given their and their colleagues’ teaching schedules, the most pragmatic choice may be to keep them teaching the same class. Likewise, a teacher may already provide instruction across multiple school year groups, with teaching schedules leading (somewhat unintentionally) to repeat teacher-class matches. This scenario may be particularly prevalent in small schools where mixed age year groups are more prevalent or in secondary schools where only a small number of teachers are qualified to teach each subject.

Unintentional looping may also occur at an individual pupil level. For instance, when moving between school years, a child may move up or down an ‘ability’ group (class set/stream). While most of their former classmates then have a different

teacher the following academic year, they – having moved up/down a set – may be taught by the same member of staff leading to ‘unsystematic repeat student-teacher matches’ (Wedenoja et al., 2022 p. 2).

In the United States, it has been suggested that unintentional looping is most common (Wedenoja et al., 2022), with intentional looping being relatively rare. Little evidence exists, however, on the characteristics of looped and not-looped classes and pupils in other national settings. Table 2 hence provides some evidence on this issue using the TALIS Video Study.

Overall, differences in background characteristics across the looped and not looped groups are generally small. Schools that loop tend to be slightly smaller, both in terms of staff headcount (46 versus 53 teachers) and number of pupils (770 versus 909). Teachers of looped classes are also slightly more experienced (15.2 years versus 13.5 years) and slightly more likely to report that the class contains pupils from a background that make instruction challenging (e.g. disadvantaged pupils). The background of pupils appears very similar across looped and not looped classes, including in terms of gender, socio-economic status, prior academic achievement and mathematics interest/self-efficacy under their previous mathematics teacher. Thus, overall, observable systematic differences across the two groups appear to be small and unlikely to substantially confound the association between looping and our outcomes of interest. This could either be due to unplanned looping within schools occurring unsystematically, or with planned looping not differing across schools in terms of the characteristics considered.

### Statistical models

As part of research questions 1–3 we explore the link between looping and pupil outcomes. A set of pupil-level OLS regression models are hence estimated:

**Table 2.** Background characteristics of schools, teachers and pupils where classes are looped compared to not looped.

		Not looped	Looped
School characteristics	Private school	14%	11%
	Urban school	74%	76%
	Number of teachers	53	46
	Number of pupils	909	770
Teacher characteristics	Teacher years of experience	13.5	15.2
	Male teacher	55%	50%
	Teacher holds postgraduate qualification	41%	36%
	Teacher completed a teacher training course	84%	83%
	Number of students in the class	28.9	28.2
	Teacher reports of whether pupil background hinder instruction (standardised)	−0.02	0.15
Pupil characteristics	Pupil self-efficacy under previous teacher (standardised)	0.00	0.00
	Pupil mathematics interest under previous teacher (standardised)	0.02	−0.05
	Possessions at home scale (standardised)	0.04	0.00
	Father holds degree	34%	32%
	Mother holds degree	33%	32%
	Male pupil	49%	49%
	Baseline test scores (standardised)	−0.04	−0.03

Notes: Data on number of teachers in school only available for Madrid, Mexico and Shanghai. Data on number of pupils in school only available for Chile, Colombia, England and Japan. Figures refer to cross-country average, with each country assigned equal weight.

$$O_{ijk} = \alpha + \beta.Loop_{jk} + \gamma.P_{ijk} + \delta.T_{jk} + \pi.Prev\_T_{ijk} + \varphi.B_{ijk} + u_k + \varepsilon_{ij} \quad (1)$$

Where:

$O_{ijk}$  = One of our pupil-level outcomes (as described in the preceding section).

$Loop_{jk}$  = Binary variable capturing whether the class is looped (1) or not (0).

$P_{ijk}$  = Pupil characteristics, including gender, home possessions and parental education (amongst others).

$T_{jk}$  = Teacher characteristics, including gender and experience (amongst others).

$Prev\_T_{ijk}$  = Pupils' views of mathematics when taught by their *previous* mathematics teacher.<sup>2</sup>

$B_{ijk}$  = Mathematics test scores at baseline.

$u_k$  = Country fixed effects.

$\varepsilon_{ij}$  = Random error term.

i = Pupil i.

j = Teacher j.

k = Country k.

Online Appendix C provides a full list of the covariates included in the model. The model is estimated once per outcome measure. The parameter of interest is  $\beta$ . This reflects the association between looping and the outcome (e.g. endpoint test scores in the context of research question 1, views of classroom behaviour and pupil-teacher relationships in the context of research questions 2 and 3). All continuous outcome measures have been standardised, meaning the estimates capture differences in effect sizes (i.e. standard deviation differences). Standard errors have been clustered at the teacher level. The data from each country makes an equal contribution to the analysis via the application of so-called senate weights.

To address research questions 2, 3 and 4, we also conduct an analysis of teachers' responses to the background questionnaires, along with the observations made by expert observers. As teachers are now the unit of analysis, our OLS regression models are re-specified as:

$$O_{jk} = \alpha + \beta.Loop_{jk} + \gamma.\overline{P}_{jk} + \delta.T_{jk} + \pi.\overline{Prev\_T}_{jk} + \varphi.\overline{B}_{jk} + u_k + \varepsilon_{ij} \quad (2)$$

Where:

$O_{jk}$  = A teacher-level outcome (a scale based upon teacher responses to the background questionnaire or judgements of the expert observers).

$Loop_{jk}$  = Binary variable capturing whether the class is looped (1) or not (0).

$\overline{P}_{jk}$  = Pupil characteristics averaged across the class. (e.g. average number of years of parental education across pupils in the class).

$T_{jk}$  = Teacher characteristics.

$\overline{Prev\_T}_{jk}$  = Pupils' views of mathematics under their *previous* mathematics teacher averaged across the class.

$\overline{B}_{jk}$  = Baseline mathematics test scores averaged across the class.

$u_k$  = Country fixed effects.

$\varepsilon_{ij}$  = Random error term.

This model is also estimated once per teacher-level outcome.  $\beta$  again reflects the difference in each outcomes across looped and non-looped classes. Senate weights have been applied, with standard errors now clustered at the country level.

Our final research question involves comparing outcomes across pupils within the same class – drawing on the fact that some will have had the same teacher the previous academic year, and others not. The model we specify to address this issue is:

$$O_{ijk} = \alpha + \beta.Same\_T_{ijk} + \gamma.P_{ijk} + \pi.Prev\_T_{ijk} + \phi.B_{ijk} + u_{jk} + \varepsilon_i \quad (3)$$

Where:

$Same\_T_{ijk}$  = Binary variable coded 1 if the pupil had the same teacher the previous year and 0 if they had a different teacher.

$u_{jk}$  = Class fixed effects.

$\varepsilon_i$  = Random error term, with standard errors clustered at the class level.

With all other variables defined as under equation (1). The inclusion of class fixed effects ( $u_{jk}$ ) means all between-class variation is removed. The  $\beta$  parameter thus captures how outcomes differ between pupils new to the teacher/class this year in comparison to pupils already in the class the previous year. For instance, do they feel like an outsider in the class, have a weaker relationship with their teacher or have less clarity over the class rules? Note that this model is estimated using both the full sample and after it is restricted only to those in looped classes.

We test the robustness of our results in several ways. Online Appendix C explores how estimates change under different model specifications (as suggested by Table 2, our results do not appreciably change whether controls are included or not). A different threshold for defining looped and not-looped classes is used in online Appendix A. Instead of running Ordinary Least Squares (OLS) models with clustered standard errors, multilevel (random effects) models are estimated in online Appendix D. Meanwhile, in online Appendix E we re-weight the data to adjust for non-random elements of the sampling. Our estimates remain broadly stable across these robustness tests, with little substantive change to the conclusions reached.

## Results

### ***Research question 1: Is there a positive association between looping and pupil outcomes?***

Table 3 illustrates the association between looping and pupil outcomes. All estimates can be interpreted as effect sizes, except absence from lessons which – as a binary outcome – refers to probability differences.

There is little evidence that looping is associated with better pupil outcomes. This includes pupils' test scores – the most widely studied outcome in this literature – where the effect size is essentially zero. On the other hand, absences appear to be slightly lower in looped classes, though the magnitude of the difference is relatively small (around three percentage points). There is a negative relationship between looping and mathematics self-confidence and self-efficacy, though effect sizes are small (0.06 and 0.07 standard deviations). Overall, Table 3 suggests that looping is unlikely to lead to substantial gains in pupils' outcomes, though with signs of a small improvement in absence rates.

**Table 3.** The association between looping and pupil outcomes.

	Effect size	SE
Test scores	0.00	0.02
Absence from lessons	−2.9%**	1.2%
Self-confidence	−0.07**	0.03
Personal interest	0.01	0.04
General self-efficacy	−0.06**	0.03

Notes: N ~ 13,000 pupils and ~ 550 teachers. 'Effect size' captures the standard deviation change in the outcome for those in a looped versus a non-looped class. SE = standard error. \* and \*\* indicates effect size significantly different from zero at the ten and five percent levels. Models include controls for country fixed effects, pupil, teacher and class background characteristics, pupils' views of mathematics when taught by their *previous* mathematics teacher and baseline test scores. See online Appendix C for analogous estimates using alternative model specifications.

### **Research question 2: Is classroom behaviour and management better in looped classrooms?**

One potential explanation for the lack of strong associations reported above is that the outcome measures are too distal. We therefore now explore how looping is linked to some of the mechanisms thought to mediate its relationship with pupil outcomes. Table 4 presents results from one such example, exploring the link with class behaviour.<sup>3</sup>

Overall, evidence of a link between looping, class behaviour and teacher classroom management is mixed. Starting with panel (a) – based upon pupil reports – there are some tentative signs of a positive association. The parameter estimates all point in the anticipated direction, though relatively small in magnitude (~0.1 standard deviation difference). Hence, based upon pupil reports, there is some suggestion looping may be linked to small gains in class behaviour and the teacher's ability to manage the class. Albeit we recognise this evidence is relatively weak.

**Table 4.** The association between looping and behaviour in the classroom.

	Effect size	SE
<b>(a) Pupil reports</b>		
Teacher classroom management (start of year)	0.14**	0.05
Class behaviour environment (start of year)	0.12**	0.06
Class behaviour environment (end of year)	0.10	0.06
Teacher classroom management (end of year)	0.08	0.05
Disruptions (end of year)	0.07	0.05
Disruptions (start of year)	0.04	0.06
<b>(b) Teacher reports</b>		
Class behaviour environment (start of year)	0.02	0.04
Disruptions (start of year)	−0.01	0.07
Teacher classroom management (start of year)	−0.04	0.07
Class behaviour environment (end of year)	−0.05	0.05
Disruptions (end of year)	−0.08	0.05
Teacher classroom management (end of year)	−0.08	0.05
<b>(c) Expert observers</b>		
Classroom routines (video)	0.00	0.02
Classroom disruptions (video)	−0.01	0.01
Classroom management (video)	−0.07**	0.03

See notes to Table 3.

Moreover, the same finding is not replicated in panels (b) and (c), where analogous information has been reported by teachers and expert observers. Regarding the former, most of the effect sizes reported in panel (b) are negative, suggesting behaviour and classroom management could be slightly worse in looped classrooms. The magnitude of these estimates are however small (typically 0.05 standard deviations or less). Turning to the results based upon expert observers in panel (c), no difference between looped and non-looped classes is found with respect to the efficiency that class routines are carried out (e.g. handing out workbooks, taking the register) or the frequency of disruptions and how well they are managed. A small negative effect is, however, observed using the overall classroom management scale. In other words, based upon the ratings of expert observers, teachers leading looped classes seem to manage them slightly less effectively.

Overall, the evidence is inconclusive. Some small positive effects appear when using pupils' reports of classroom behaviour and management, while zero or small negative effects emerge using information reported by teachers and independent observers.

### **Research question 3: Are pupil-teacher and pupil-pupil relationships better in looped classrooms?**

A further mechanism via which looping may operate is by improving classroom relationships. As pupils get to know each other and the teacher over time, stronger bonds are formed, which then lead to improved pupil outcomes. Table 5 thus presents evidence on the association between looping and various intra-class relationships.

There is no clear evidence that looping is associated with better pupil-pupil or pupil-teacher relationships. The estimated effect sizes in panel (a) using pupil reports are small in magnitude. In contrast, teachers rate their relationship with pupils as slightly worse in looped classrooms, particularly towards the end of the study period (−0.15 effect size). Meanwhile, based upon the ratings of expert observers, there is no difference across looped and non-looped classes. In particular, results from the expert observers do not suggest that classroom interactions are any more or less respectful in looped classes, or that pupils are more willing to take 'risks' (e.g. openly share their work and thoughts with others). Hence, also in terms of intra-class relationships, the benefits of looping are not clear.

**Table 5.** The association between looping and pupil/teacher relations.

	Effect size	SE
<b>(a) Pupil reports</b>		
Student-teacher relationship (start of year)	0.05	0.05
Teacher support for competence (start of year)	0.04	0.05
Student-student relationship (start of year)	0.03	0.03
Student-teacher relationship (end of year)	0.03	0.05
Teacher support for competence (end of year)	0.02	0.05
Student-student relationship (end of year)	0.02	0.03
<b>(b) Teacher reports</b>		
Teacher support for competence (start of year)	0.01	0.07
Student-teacher relationship (start of year)	−0.08	0.10
Teacher support for competence (end of year)	−0.14*	0.08
Student-teacher relationship (end of year)	−0.15*	0.09
<b>(c) Expert observers</b>		
Pupils willing to take risks in class (video)	0.00	0.09
Respectful classroom (video)	0.00	0.05

See notes to Table 3.

#### **Research question 4: How is looping linked to teachers' outcomes?**

Next, we turn to the perspectives of teachers. Even if looping is not associated with clear gains in pupils' outcomes or the class environment, it may still be beneficial if teachers prefer it, or benefit in some other way. Table 6 thus explores the association between looping and the perspective of teachers regarding their experience teaching the relevant class.

The estimated associations are all weak, and many have a negative sign. For instance, the point estimates suggest that, if anything, teachers in looped classes report enjoying the teaching less (effect size  $-0.09$ ). On the other hand, a positive association is observed for the 'enthusiasm' scale. Such inconsistencies again point towards no clear evidence that teachers prefer teaching looped classes, nor that they feel more able to achieve their learning goals.

#### **Research question 5: How do outcomes and perceptions of the class environment differ between pupils retaining the same teacher to those new to the class?**

Our analysis has thus far focused on differences between looped and non-looped classes; instances where most pupils in the class retained the same teacher to instances where for most pupils the teacher was new. We now turn to *within-class* variation, investigating whether outcomes differ between classmates for whom the teacher was new to those who had the teacher the previous year. The results based upon the full sample can be found in Table 7.

The estimates in panel (a) again point towards small effect sizes. Consistent with previous studies (e.g. Albornoz et al., 2023; Hill & Jones, 2018), pupils retaining the same teacher achieve marginally higher test scores (effect size = 0.03) and have slightly lower absence rates ( $-2.6\%$  points). There is, however, no evidence of a link between retaining the same teacher and pupils' subject interest, self-confidence and self-efficacy. A similar pattern emerges in panel (b) where estimates are presented for pupils' views of classroom relationships and their teacher's classroom management. Although consistently positive, the effect sizes are small (less than 0.05 standard deviations). Hence, even after changing focus from looping of entire classes to looping of individual pupils, the benefits for those keeping the same teacher are marginal, at best.

Table 8 presents findings from a sub-group analysis where we investigate differences in outcomes where most – but not all – pupils had the same teacher (and classmates)

**Table 6.** The association between looping and teacher outcomes.

	Effect size	SE
Enthusiasm for teaching the class (start of year)	0.12	0.08
Enthusiasm for teaching the class (end of year)	0.06	0.08
Anxious when teaching the class (end of year)	0.00	0.07
Angry when teaching the class (end of year)	$-0.02$	0.04
Self-efficacy in classroom management (end of year)	$-0.02$	0.06
Achievement of learning goals (end of year)	$-0.03$	0.08
Teacher self-efficacy (end of year)	$-0.03$	0.08
Self-efficacy in classroom management (start of year)	$-0.07$	0.07
Enjoyed teaching the class (end of year)	$-0.09$	0.10
Teacher self-efficacy (start of year)	$-0.10$	0.11

See notes to Table 3.

**Table 7.** Comparison of outcomes across pupils who remained with the same teacher to those who were new to their class.

	Effect size	SE
<b>(a) Pupil outcomes</b>		
Self-efficacy mathematics tasks	0.03	0.02
Test scores	0.03*	0.02
Personal interest	0.00	0.02
General self-efficacy	-0.01	0.03
Self-confidence	-0.02	0.03
Absence from lessons	-2.6%**	1.3%
<b>(b) Pupil-teacher relationship and views on class management</b>		
Student-teacher relationship (start of year)	0.04	0.03
Student-teacher relationship (end of year)	0.04	0.03
Teacher support for competence (start of year)	0.04	0.03
Teacher support for competence (end of year)	0.04	0.03
Teacher classroom management (end of year)	0.03	0.03
Teacher classroom management (start of year)	0.03	0.03
Student-student relationship (start of year)	0.02	0.03
Student-student relationship (end of year)	0.00	0.03

Notes: Estimates based on TALIS video study data pooled across all participating countries. N ~15,000 pupils and ~ 650 teachers. The 'effect size' column captures the standard deviation difference in the outcome for pupils who had the same mathematics teacher the previous academic year compared to those who had a different teacher. The SE column records the standard error. Standard errors have been clustered by teacher. \* and \*\* indicates that the effect size is significantly different from zero at the ten and five percent levels respectively. Models include controls for class fixed effects, pupil characteristics, pupils' views of mathematics when taught by their *previous* mathematics teacher and baseline test scores. See online Appendix C for analogous estimates using alternative model specifications.

previously. This restricts the sample to looped classes – where at least 75% of pupils had the same teacher the previous year. We then investigate how outcomes for the minority of new entrants into these classes compare to their peers (who were all in the class and had the same teacher the previous year).

Starting with panel (a), there is some evidence that new pupils entering an otherwise looped class experience slightly worse outcomes in certain dimensions, achieving slightly lower test scores (effect size = -0.07) and having a somewhat higher absence rate (4.5% points). On the other hand, there is no difference in their mathematics self-confidence, while a positive – if small – association emerges for personal interest. These findings are strengthened by our robustness tests in online Appendix A, where we illustrate how similarly sized effects for test scores and absence levels emerge when an alternative definition of a looped classroom is used.

The results in panel (b) are to some extent consistent with this finding. Pupils who are new members of an otherwise looped class report worse relations with their peers at the start of the study period (effect size = -0.11) and a worse relationship with their teacher at the end (effect size = -0.09). The sign of the coefficient for all other outcomes considered in panel (b) are negative as well. Our robustness tests in online Appendix A using a more liberal definition of a looped classroom (and hence sample selection) supports these findings; a negative relationship with effect sizes around -0.10 are found for inter-pupil relationships (start of the study period), teacher support for competence (start of the study period) and pupil-teacher relations (both start and end). Table 8 hence suggests pupils who enter a looped class may struggle to form the same bond with their teacher and classmates as their peers already part of the group.



**Table 8.** Outcomes for pupils who are new entrants into an otherwise looped class.

	Effect size	SE
<b>(a) Pupil outcomes</b>		
Personal interest	0.07	0.04
Absence from lessons	4.5%*	2.7%
Self-confidence	0.03	0.05
General self-efficacy	0.02	0.04
Self-efficacy mathematics tasks	0.01	0.04
Test scores	-0.07**	0.03
<b>(b) Pupil-teacher relationship and views on class management</b>		
Student-student relationship (end of year)	-0.04	0.05
Teacher support for competence (start of year)	-0.04	0.04
Teacher support for competence (end of year)	-0.04	0.05
Student-teacher relationship (start of year)	-0.05	0.04
Teacher classroom management (end of year)	-0.07	0.06
Teacher classroom management (start of year)	-0.08	0.05
Student-teacher relationship (end of year)	-0.09*	0.05
Student-student relationship (start of year)	-0.11**	0.05

Notes: N ~5,500 pupils and ~229 teachers. See notes to [Table 7](#) for further details.

## Conclusions

Policymakers and school leaders are looking for the most effective ways to increase the skills of young people. Although some expensive options such as intensive one-to-one tutoring have proven effective (Education Endowment Foundation, 2023) smaller, more localised changes to a school's organisation could have a positive impact as well. Looping is one such option, where teachers remain with the same class for more than one year. It also has the attraction of not requiring policy or system-level reforms. Although in some contexts such repeated pupil-teacher allocations occur more by chance than design (Wedenoja et al., 2022), previous studies – mainly from the United States – suggest it may have some benefits (Hill & Jones, 2018; Hwang et al., 2021), including improving attendance and test scores (Wedenoja et al., 2022).

The potential benefits of looping have, however, been underexplored in other national settings. Moreover, existing studies focus on a narrow set of outcomes (usually test scores, attendance and exclusions) with little quantitative research into how looping relates to the mechanisms thought to underpin its effects. This paper has hence provided new international evidence on such matters. We also provide evidence about the relationship between looping and teacher outcomes – such as whether they enjoy teaching looped classes more – and how the small subset of pupils who are new to an otherwise looped class progress.

Our results are largely a story of null effects. We find no evidence of a meaningful relationship between looping and pupil outcomes, including test scores, absence rates and self-efficacy. Likewise, evidence of an association between looping and the class environment – including behaviour, classroom management and pupil-teacher relationships – is mixed (at best). There is also no suggestion that teachers enjoy or otherwise prefer teaching looped classes more than non-looped classes, or tangibly benefit in other ways.

At first glance, our findings may seem at odds with the existing literature. There is, however, more consistency than first meets the eye. Take the link between looping and test scores – the most widely studied outcome in this literature. We find an effect size of essentially zero in our analysis of looped classes and +0.03 for pupils who remain with the same teacher. This is very similar to the associations reported by Wedenoja et al. (2022) in

Tennessee (+0.02), Albornoz et al. (2023) in Chile (+0.02), Hwang et al. (2021) in Indiana (+0.015) and Hill and Jones (2018) in North Carolina (+0.02). The common theme – including our study – is the estimated effect sizes are extremely small. Thus, despite headline findings that ‘student-teacher rematches’ increase test scores (Hwang et al., 2021, p. 1), ‘repeating matches has a robust positive effect on test scores’ (Albornoz et al., 2023, p. 2) and ‘having a repeat teacher improves achievement’ (Wedenoja et al., 2022, p. 1), the magnitude of such benefits are marginal.

There are of course limitations of our work. First, the limited sample size has meant we have not explored cross-country variation, with a lack of statistical power to capture very small effects. Second, the sample analysed focuses on secondary mathematics teachers, with teachers within participating schools not randomly selected in some countries. Given that looping has been reported to have a particularly beneficial impact on classes taught by less effective teachers (Hill & Jones, 2018), non-random sampling could be a particularly important limitation if it resulted in less effective teachers being under-represented in the sample. As most existing studies focus on upper primary and lower secondary pupils from a small number of countries, future research should seek to generalise our findings to other settings – including primary schools, other countries and other subject areas. Third, while the narrow focus of the outcome test means it should be particularly sensitive to the instruction pupils had just received in schools, the length of time between the pre- and post-assessments was relatively short (on average, 41 days). This may arguably be too short a timeframe for substantial effects on pupil achievement to occur. Finally, due to the observational nature of the data, estimates refer to conditional associations and may not capture cause and effect. While we recognise that a Randomised Controlled Trial (RCT) of looping would provide stronger evidence with respect to the causal effect, our results – and those from other observational studies – suggest that stronger evidence of more sizeable effects from quasi-experimental studies are perhaps needed first.

What advice do we offer to school leaders and other stakeholders based on our results? It may be argued that education researchers are sometimes guilty of overprescribing solutions when, in reality, the impact is likely marginal. Looping probably falls into this category. The magnitude of the likely effects is not strong enough to justify school leaders going out of their way to enact looping as a school policy with pupils purposefully assigned to the same teacher. Yet school leaders should not actively avoid looping either. Rather they should encourage it to happen when other aspects of timetabling make it the most pragmatic choice, or when localised professional judgement indicates that it is an appropriate course of action.

## Notes

1. Classes where 25–75% of pupils reported they kept the same teacher have been excluded due to ambiguity over whether the class is ‘looped’ or not. These classes are included within our robustness tests in Appendix A.
2. At baseline, pupils were instructed: ‘*please think about the time when you were taught by your PREVIOUS mathematics teacher (the teacher you had before your current mathematics teacher): how did you think about mathematics back then?*’ A set of statements were then presented, such as ‘*I often thought that what we were talking about in my mathematics class was interesting*’.

3. In the raw data file, the endpoint 'Teacher classroom management' and 'teacher disruption' scale scores appear identical. The results using these two scales are hence identical in [Table 4](#). We believe this is likely a data entry error made by the survey organisers, though this does not impact upon our substantive conclusions.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Notes on contributors

*John Jerrim* is Professor of Education and Social Statistics at the UCL Social Research Institute. His research interests include international comparisons of educational achievement and inequalities in education.

*Loic Menzies* is a researcher and policy specialist. He has authored numerous high-profile studies of issues ranging from teacher recruitment and educational assessment to youth homelessness. Loic was previously Chief Executive of the 'think and action-tank' The Centre for Education and Youth, and is a former teacher.

## ORCID

John Jerrim  <http://orcid.org/0000-0001-5705-7954>

## Ethics

The paper is a secondary analysis of data that is publicly available online. The BERA code of ethical practise has been followed.

## References

- Albornoz, F., Contreras, D., & Upward, R. (2023). Let's stay together: The effects of repeat student-teacher matches on academic achievement. *Economics of Education Review*, 94, 102375. <https://doi.org/10.1016/j.econedurev.2023.102375>
- Atteberry, A., Loeb, S., & Wyckoff, J. (2017). Teacher churning: Reassignment rates and implications for student achievement. *Educational Evaluation and Policy Analysis*, 39(1), 3–30. <https://doi.org/10.3102/0162373716659929>
- Cistone, P., & Shneyderman, A. (2004). Looping: An Empirical Evaluation. *International Journal of Educational Policy, Research, and Practice: Reconceptualizing Childhood Studies*, 5(1), 47–61. <https://eric.ed.gov/?redir=http%3A%2F%2Fwww.caddogap.com%2Fperiodicals.shtml>
- Education Endowment Foundation. (2023). One to one tuition. <https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/one-to-one-tuition>
- Franz, D. P., Thompson, N. L., Fuller, B., Hare, R. D., Miller, N. C., & Walker, J. (2010). Evaluating mathematics achievement of middle school students in a looping environment. *School Science and Mathematics*, 110(6), 298–308. <https://doi.org/10.1111/j.1949-8594.2010.00038.x>
- Gorard, S. (2015). Rethinking 'quantitative' methods and the development of new researchers. *Review of Education*, 3(1), 72–96. <https://doi.org/10.1002/rev3.3041>
- Grissom, J. A., Kalogrides, D. & Loeb, S. (2015). The micropolitics of educational inequality: The case of Teacher–Student assignments. *Peabody Journal of Education*, 90(5), 601–614. <https://doi.org/10.1080/0161956X.2015.1087768>

- Hegde, A. V. & Cassidy, D. J. (2004). Teacher and parent perspectives on looping. *Early Childhood Education Journal*, 32, 133–138. <https://doi.org/10.1007/s10643-004-1080-x>
- Hill, A. J., & Jones, D. B. (2018). A teacher who knows me: The academic benefits of repeat student-teacher matches. *Economics of Education Review*, 64, 1–12. <https://doi.org/10.1016/j.econedurev.2018.03.004>
- Hwang, N., Kisida, B., & Koedel, C. (2021). A familiar face: Student-teacher rematches and student achievement. *Economics of Education Review*, 85, 102194. <https://doi.org/10.1016/j.econedurev.2021.102194>
- Menzies, L. (2023). Continuity and churn: Understanding and responding to the impact of teacher turnover. *London Review of Education*, 21(1).
- Nichols, J. D. & Nichols, G. W. (2003). The impact of looping classroom environments on parental attitudes. *Preventing School Failure: Alternative Education for Children & Youth*, 47(1), 18–25. <https://doi.org/10.1080/10459880309604424>
- OECD. (2021). Global teaching InSights technical documents. Retrived from <https://www.oecd.org/education/school/global-teaching-insights-technical-documents.htm>
- Ost, B., & Schiman, J. C. (2015). Grade-specific experience, grade reassignments, and teacher turnover. *Economics of Education Review*, 46, 112–126. <https://doi.org/10.1016/j.econedurev.2015.03.004>
- Tourigny, R., Plante, I., & Raby, C. (2020). Do students in a looping classroom get higher grades and report a better teacher-student relationship than those in a traditional setting? *Educational Studies*, 46(6), 744–759. <https://doi.org/10.1080/03055698.2019.1663152>
- Wedenoja, L., Papay, J., & Kraft, M. A. (2022). Second time's the charm? How sustained relationships from repeat student-teacher matches build academic and behavioral skills. (EdWorkingPaper: 22-590). <https://doi.org/10.26300/sddw-ag22>