Composite Classes, Class Size and Human Capital Accumulation

Daniel Borbely*, Markus Gehrsitz† ‡ , Stuart McIntyre†,
Gennaro Rossi†, Graeme Roy†

UNIVERSITY of STRATHCLYDE
FRASER OF ALLANDER INSTITUTE

Nuffield Foundation

*School of Business, University of Dundee, Dundee, UK.
†Fraser of Allander Institute, Department of Economics, University of Strathclyde, Glasgow, UK.
‡Institute for the Study of Labor (IZA), Bonn, Germany.

Acknowledgements

This report summarises the analysis conducted as part of the Nuffield Foundation grant EDO/43743. For this research grant, Markus Gehrsitz was the principal investigator and Stuart McIntyre and Graeme Roy were co-investigators. Daniel Borbely and Gennaro Rossi provided excellent research assistance. We thank the Nuffield Foundation for providing funding to carry out this work.

The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare, and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org.

We are grateful to Emma Congreve, Susan Ellis, Gordon McKinlay, Ian Walker and Tanya Wilson as well as to Marco Alfano, David A. Jaeger, and Jonathan Norris for their helpful comments. We thank Mick Wilson and his team at Scottish Government for providing the raw data used in this study. We also thank the electronic Data Research and Innovation Service (eDRIS) team, in particular Julian Augley, Fiona James, Suhail Iqbal, David Stobie, Amy Tilbrook and Dionysis Vragkos for their assistance in accessing the data used in this study. All data are solely owned by the Scottish Government. All views expressed in this report are the authors' and do not reflect the view of eDRIS or the Scottish Government.

Executive Summary

In this project, we studied the effects of classroom composition, in particular multi-grade ("composite") classes, and class size on pupils' attainment in primary schools. Both features have received ample attention from academics, practitioners, and policy makers. Yet, due to a lack of data and robust empirical methods, evidence on peer effects and the impact of class structure has been scant. Our project turned to Scotland to overcome these obstacles and fill this knowledge gap. In particular, our study pursued to answer the following questions:

1. Do composite classes in primary school affect pupil attainment?
2. Is exposure to more mature peers by way of composite classes beneficial? If so, do these gains in attainment come at the expense of children who form the older/more mature part of composite classes?
3. Do smaller class sizes in primary school lead to higher attainment as measured by teacher assessment?
4. Does exposure to older/younger peers by way of composite classes and/or smaller classes improve pupils' attitudes towards learning?
5. Do smaller classes and/or exposure to older/younger peers by way of composite classes reduce pupil absences and suspensions?

In addition, an important "legacy goal" of this project was to demonstrate that it is feasible to create a panel data set that allows researchers and stakeholders to track Scottish pupils as they progress through the education system and beyond, thus allowing for an analysis of both short-term and long-term outcomes of education policies and interventions.

Background: Multi-grade classes (also known as "composite classes") combine pupils from adjacent year groups into a single class room. This class room structure is widespread yet understudied. For instance, more than a quarter of US schools use mixed classrooms and in France about one third of primary school pupils attend multi-grade classes. In Scotland, composite classes feature in virtually every primary school in both urban and rural areas, making Scotland the perfect setting to study their effects on attainment. This is all the more important because to date the only robust research on composite classes stems from thinly populated areas in Norway and Italy. By contrast, the Scottish institutional setting and data infrastructure allowed us to assess the effects of composite classes in a more representative setting. Our project, therefore, fills an important knowledge gap. In addition, we study the effects of class size which is an alternative policy lever that has been shown to affect attainment. While there is more, at times contradictory evidence on the role of class size, there has been very little work on class size effects in the UK.

Data Work: We compiled and linked data from several data sources. The Scottish Pupil Census (SPC) provided the ideal basis for our analysis. The SPC is a well-organised, annual collection of detailed individual-level data on the universe of pupils in Scottish public primary and secondary schools. Such rich and well-documented data are rare internationally, yet often necessary for robust quantitative research. Scotland with its well-maintained high-quality data sets is therefore particularly well suited to study educational interventions and policies. In particular, we coded up each pupil's school, grade, and class, and linked this information across waves from 2007/08-2018/19. The data also contain information on pupil demographics. We also linked these data with teacher assessments of pupils' numeracy and literacy abilities in first grade (P1), fourth grade (P4), seventh grade (P7) and ninth grade (second year of secondary school, S2) from 2015/16 to 2018/19.

We also identified pupils who took part in the Scottish Surveys of Literacy and Numeracy (SSLN) between 2011/12 and 2016/17 which contain information on, among other things, learning attitudes and more detailed data on performance in numeracy and literacy for a sub-sample of pupils. Our data set also allows us to observe whether and when pupils get suspended or expelled. We also added data from the leavers' surveys which allow researchers to track each pupil's qualifications as per the Scottish Credit and Qualifications Framework (SCQF) as well as post-highschool destinations. A legacy of this project is that the code to generate this data set will be placed in an online, publicly accessible GitHub repository. We showcase the value of this data source for scientific research that can inform policy, by analysing the effect of composite classes on pupil performance. These are classes in which students from adjacent grades share a classroom. As a result pupils are often exposed to more mature pupils from preceding cohorts who have been in school longer. These peer effects, and the impact of composite classes in general, are of major policy interest.

Methods: In order to isolate causal effects, we exploit the interaction of several institutional features of the Scottish school system. First, school intake in Scotland is mainly governed by fixed school catchment areas. That is, every pupil residing in a primary school's catchment area is entitled to a place in this very school. As a result, school enrolment is largely determined by random fluctuations in catchment area birth cohort size. Second, the Scottish Government centrally sets national class size limits. That is, if the cohort size in a school is above the corresponding class size maximum, the school is forced to divide such a cohort into two smaller classes. Third, Scotland allows for the creation of composite classes in primary school. That is, schools may pool pupils from adjacent years into a single "mixed" classroom with a class size cap of 25 pupils.

These three features work together in generating a useful natural experiment: As fluctuations in cohort size push enrolment counts above class size limits, for some schools enrolment counts across grades just happen to work out such that a classroom can be saved by creating a composite class. In other schools (or the same school in a different year) with slightly different enrolment counts this option does not exist. For instance, the maximum class size in fourth and fifth grade is 33. A school with 46 fourth-graders and 45 fifth-graders would thus be able to create two single-year classes with 33 pupils each while pooling the remaining 25 pupils in a composite class. However, a school with 46 fourth-graders and 46 fifth-graders would not have this option. The presence of an additional pupil would push this school over the cap of 25 for composite classes, thus forcing the school to create four single year classes with 23 pupils each. In other words, small variations in enrolment counts quasi-randomly trigger/prevent composite classes.

The above example is obviously stylised and in practice all seven primary school grades have to be taken into account when deciding on the number and type of classes. This creates a complex allocation task. School administrators, therefore, use an algorithmic class planner to determine in which grades composite classes should be created in order to minimise the number of classrooms in a school. Again, these class planner predictions ultimately depend on enrolment counts, which in turn are determined by random population variation. Our approach exploits precisely this feature. Intuitively, we compare pupils who - due to random population variation - were assigned to composite classes with those who are not placed in composite classes but would have been assigned to one, had the enrolment counts been marginally different. Econometrically, this is modelled using Two-Stage Least Squares (2SLS) instrumental variable regressions. Because head teachers are most likely to comply ("strong first stage") with class planner suggestions for composite classes that are comprised of first and second graders, we focus on these pupils, but also report our results for fourth and seventh graders.

Results: We find that first graders benefit from sharing a composite classroom with second-graders:

- Every additional more mature peer raises the probability that a pupil performs at level in numeracy by 0.8 to 1.1 percentage points. Put differently, composite classes increase first-graders' performance by 0.28 standard derivations.

- The effects for literacy are slightly larger than for numeracy. The presence of an additional more mature peer increases reading and writing ability by 1.3 to 1.5 percentage points.
- These benefits do not seem to occur at the expense of second-graders. We find no evidence for lower teacher assessments for second-graders who had shared a P1/P2 composite class room with first graders.
- Our analysis also reveals neither positive nor negative effects of composite classes beyond attainment. We find no statistically significant effect of exposure to older or younger peers on attendance and suspension rates, nor with respect to learning attitudes.
- Our analysis of fourth and seventh graders, the only other grades for which teacher assessments are made, are less definite but suggestive of similar patterns: Classroom exposure to more mature peers from a preceding cohort boosts performance in literacy and numeracy and there is no conclusive evidence suggesting that these gains accrue at the expense of adverse learning effects on older pupils in a mixed class.
- Even though composite classes tend to be smaller than single-year classes we show that class size is not driving these gains.
- In fact, more broadly our results suggest that further reductions in class size in Scottish schools offer little return in terms of attainment, at least in the range of reductions that we study. This is true for first graders, fourth graders, and seventh graders.

Summary and Conclusion: Overall, our research concludes that exposure to older peers is highly beneficial to primary school pupils in terms of attainment. Composite classes, which are widespread internationally and very common in Scotland, explicitly create these peer effects while simultaneously allowing administrators to save classrooms and thus costs. Class size reductions, by contrast, offer no statistically significant benefits.

Contents

1 Introduction **9**

2 Background **11**

2.1 Previous Research - A Short Overview 11

2.2 Institutional Setting - Schools in Scotland 12

3 Data **14**

3.1 Main Data Source: Scottish Pupil Census (SPC) 14

3.2 Main Outcome Data: CfE Assessments 15

3.3 Additional Outcome Data 17

4 Methods **18**

5 Main Results **24**

6 Mechanisms and Heterogeneity **29**

7 Conclusion and Recommendations **31**

Appendix A **36**

List of Tables

1 Maximum Class Size Rules . . . 13
2 Summary Statistics . . . 16
3 Main Results - First Graders (P1) . . . 25
4 Second Stage Results - Performance of Second Graders (P2) in Fourth Grade (P4) . . . 27
5 Second Stage Results - Fourth (P4) and Seventh (P7) Graders . . . 28
A1 Self-Selection of Composite Class Pupils . . . 36
A2 Second Stage Results (P1) for Literacy Subcategories . . . 37
A3 Second Stage Results (P1): Effect Heterogeneity . . . 38
A4 Behavioral Outcomes - P4 and P7 . . . 40
A5 P1 Behavioural Outcomes . . . 41
A6 Leavers' Destinations - OLS . . . 41

List of Figures

2 Pupils by Grade and Class Type (2018) . . . 14
3 Class Planner Examples . . . 21
4 P1 Class-Types by School and Local Authority in 2018/19 . . . 22
5 P1 Class-Types by School and Local Authority in 2018/19 (Only schools with 150-300 pupils) . . . 23
A1 Data Merging Plan . . . 39

1 Introduction

Classroom composition and peer effects have received increased attention in recent years. Many teachers face mixed-ability classrooms and have to strike a balance between providing sufficient challenge to high achievers while also ensuring that weaker pupils have the support that they need to learn and develop.

A more homogeneous classroom allows teachers to more easily teach at pupils' levels, but it risks exacerbating inequalities, especially if younger or academically weaker pupils benefit from the presence of older or high-achieving peers. While several studies have shed light on the issues of class room composition in general and peer effects in particular, important questions remain.

Our goal in this project was to study the effects of a measure that - while originally conceived to minimize costs - effectively creates a unique set of age and attainment peer effects: multi–grade classes. In Scotland between one in five and one in six primary school pupils attend multi-grade classes which are also referred to as "composite classes", which is the nomenclature we use in the rest of this report. These are classes comprised of pupils from adjacent grades. In the case of first grade pupils, those just starting school, they may be taught alongside second graders. Pupils in second grade meanwhile, may be taught alongside first or third graders, and so forth for other grades. We study whether exposure to younger/older peers by way of composite classes affects educational outcomes.

Furthermore, we also study the effect of class size on pupil attainment and related outcomes. Class size is another important and related policy lever that might affect class room management and pupil attainment. Similar to homogeneous classes, smaller classes are perceived to be easier to teach and allow for teaching techniques that are harder to apply in large class rooms. Class size is also of major policy interest and most governments - including Scotland's - have set both maximum class size rules and objectives to further reduce average class sizes in their school systems.[1] Both issues, composite classes and class size, also interact with one another, not least because composite classes tend to be smaller than single-year classes. In our study we therefore also investigate their interplay by investigating whether class size moderates or even drives any of the composite effects.

Our study is not a mere academic endeavor, but of major interest to policy makers. Class size is an important topic for policy debate. In the past, policymakers have set targets for maximum class sizes believing that the number of pupils in a class has an impact upon attainment. Composite classes are widespread yet understudied. Almost one in five primary school pupils in Scotland attends a composite class. Composite classes are also common

[1] See: https://www.parliament.scot/parliamentarybusiness/16836.aspx

beyond the UK. According to Leuven and Rønning (2014) about 28% of schools in the US use mixed classrooms, in France more than one third of primary school pupils attend multi-grade classrooms. The only studies of the policy to date (Leuven and Rønning (2014); Barbetta et al. (2019); Checchi and De Paola (2018)) exploit the fact that in rural contexts (Norway and Italy, respectively) cohorts are often so small that they need to be pooled into multi-grade classes.[2] In Scotland, on the other hand, composite classes are not a niche phenomenon, but form a staple of the education system. They are used by schools in more affluent and in less affluent areas, as well as in urban and rural schools. What is more, pupils in Scotland typically experience composite classes in some years and single-year classes the year after. In the rural settings that were the focus of previous research, on the other hand, pupils are usually in multi-grade classes for the entirety of their time at school.

Our research, therefore, pushes a nascent literature forward and generates knowledge that is valuable to policy and decision makers at all levels. There is interest in the effects of multi-grade classes and class size among parents, teachers, heads of school, local authorities, and national governments. Our project also fits directly into the Research Strategy for Scottish Education released in April 2017 which is geared towards using data in order to understand what works and has worked and towards providing insights that can impact on practice.[3] Indeed, as part of this project we built a large, individual-level, longitudinal data set that allows for the evaluation of interventions and policies in the education space. We then provide a proof of concept by leveraging these data to investigate the effect of composite classes in particular, and peer effects in general.

This report proceeds as follows. In the next section, we provide a short review of the literature on peer effects and class size, and place our research in this context. We also provide a short overview of the Scottish institutional setting for primary schools. In section 3, we describe our data and data work. Section 4 introduces our methodology and sections 5 and 6 present our results. Finally, in section 7 we discuss our conclusion and policy recommendations arising from this project.

---

[2]The only study that we are aware of that evaluates combination classes in a state or nationwide context is Sims' (2008) school-level analysis of California's class size reduction program.

[3]See: https://www.gov.scot/publications/research-strategy-scottish-education

2 Background

2.1 Previous Research - A Short Overview

At the turn of the millennium, much of cutting-edge research by education economists centered around an important determinant of attainment that is easy to measure: class size. For instance, the famous Tennessee Student/Teacher Achievement Ratio ("STAR") experiment found that students assigned to smaller classes (13-17 pupils) tend to score on average 5 percentage points higher in standardised tests than those in regular classes (22-25 pupils) (Krueger, 1999). However, these positive effects could often not be found in other countries or even other states. For instance, Hoxby (2000b) finds no effect of class size on test scores in Connecticut's schools.

As a result, the focus of research into what makes for successful learning has shifted beyond assessing the effect of the number of students in a class room. In particular, the role of peer characteristics has received substantial attention. One of the main points of interest are spillover effects in students' ability. For example, Lavy et al. (2012a) and Lavy et al. (2012b) document the detrimental effects of low-ability peers on their classmates' performance. Another strand of this literature (see e.g. Hoxby (2000a), Lavy and Schlosser (2011), and Black et al. (2013)) focuses on demographic characteristics. Black et al. (2013), for example, use Norwegian data and find that school outcomes are influenced by the proportion of girls within the grade, whereas demographic and family characteristics such as parental education are not as important. Other work explores whether the presence of immigrants in the classroom has a negative effect on natives' attainment (see Gould et al. 2009; Ballatore et al. 2018). Finally, a more recent strand of the literature has shed some light on peer-effects beyond the cognitive sphere. Carrell et al. (2018) finds that disruptive peers have detrimental effects that even persist after high school graduation and translate into worse labor market outcomes. Norris (2020) documents spillover effects in learning attitudes in adolescence. Another channel of peers' influence is age, i.e. classroom exposure to older/younger peers. Some studies have examined the effect of relative age within classroom (or grade) on cognitive skills (Black et al. 2011; Cascio and Schanzenbach 2016), noncognitive skills (Crawford et al., 2014) as well as the probability of being bullied (Ballatore et al., 2020).

Whilst a key characteristic of previous work is that they exploit random or quasi-random variation of the age composition in class, in our study pupils are consciously grouped together primarily for reasons of cost-minimization, but also with the intention of lifting younger school starters. In other words, our study tests not just whether peer effects exist, but also shows how they can be created by design. Our work bridges the literature on classroom inputs (in particular class size) and peer effects, by contributing to the new literature on multi-grade classes, i.e. classrooms with more than one year group. This is a developing literature which counts few contributions to date: Leuven and Rønning (2014) study multi-grade classrooms in Norway whereas Barbetta et al. (2019) and Checchi and De Paola (2018) investigate the case of Italy. One common finding in the above-mentioned literature is that the lower-graders benefits from this set-up at expenses of their older classmates. Another common factor is that they exploit classroom allocation policies which are typical of rural contexts, where cohorts tend to be particularly small and indeed at times so small that pupils across three stages are pooled into composite classes. In Scotland, on the other hand, composite classes are a widespread phenomenon used in urban as well as rural schools. Our findings therefore substantially push this nascent literature forward and add to its external validity.

2.2 Institutional Setting - Schools in Scotland

The Scottish education system has always been separate from that of the rest of the UK. However for a long time it was controlled by the UK Government department, the 'Scotland Office', which was responsible for Scottish affairs. With the advent of devolution in 1999, responsibility for the Scottish education and skills system transferred to the Scottish Parliament. Since then, the Scottish Government has determined both education funding and institutional conditions, entirely independent from the UK Government, for approximately 700,000 pupils in Scotland. Government-funded public schools in Scotland are free for children aged 5-19. There is only a small private school sector accounting for about 4% of pupils, which is mostly clustered in the "Central Belt" of the country.

Pupils in Scotland typically start school in August of the calendar year in which they turn five. They attend primary school from first grade (P1) to seventh grade (P7) before transferring into secondary school. The minimum secondary school leaving age is 16, ensuring that most pupils attend high school from eighth grade (S1) to eleventh grade (S4). Pupils may stay on for two more years (S5 and S6) in order to study towards the highest of the certifications offered by the Scottish Qualifications Authority (SQA), namely "Highers" and "Advanced Highers". These are subject-specific qualifications, roughly equivalent to English A-levels, and tend to be entry requirements to universities.

School choice is largely contingent on catchment areas which are drawn up by Local Authorities or "Councils" as they are known, and rarely ever change. Each primary school has a catchment area surrounding it and any pupil whose main residence is within this boundary is entitled to a place in that school. Primary school catchment areas do not overlap, and are nested within secondary school catchment areas. In order to register their children in the relevant catchment area, parents need to provide proof of residence to the Local Authority. Parents, of course, may strategically sort into catchment areas of schools that are perceived to be desirable. Rossi (2020) documents that housing prices on two sides of catchment border areas in Scotland differ on average by 4%.

Parents may also ask for their children to attend a school other than their catchment area school via so-called "placing requests". These are applications to the local council to transfer a child to a specific school. Councils are under no obligation to grant these requests and will only do so if the requested school can accommodate additional students without having to create an extra classroom or recruiting additional teaching staff. If there are more placing requests than spaces, places are allocated based on criteria decided by each Local Authority. For instance, typically children with additional support needs and/or with siblings in the specified schools are prioritised.

Table 1: Maximum Class Size Rules

Stage	Max. Size
Primary 1 (P1)	25
Primary 2 and 3 (P2, P3)	30
Primary 4 to 7 (P4-P7)	33
Composites (**all** stages)	25

Note: Rules as of 2019. P1 cutoff was 30 prior to 2011.

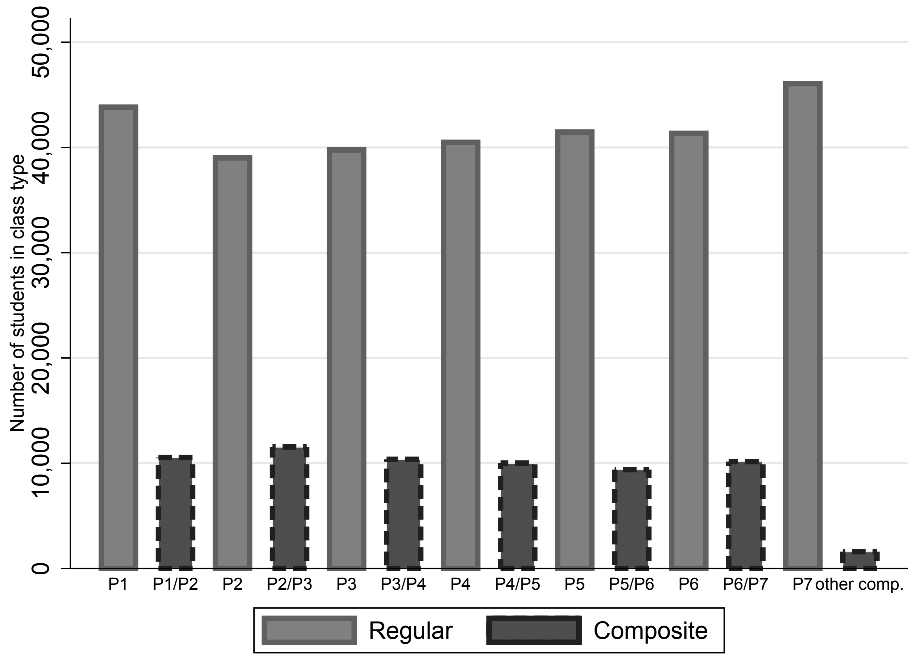
The Scottish Government centrally sets maximum class size rules in primary school which apply to the entire nation. Table 1 shows that class size in P1 must not exceed 25 pupils, the maximum for P2 and P3 is 30, and classes in P4-P7 are formed as multiples of, at most, 33. Prior to 2011, the maximum class size for P1 was 30 students. There are no legally binding grade-based maximum class size rules for secondary school – the only rule relates to maximum class sizes in "practical" versus "non–practical" subjects; the effect of which is that pupils' class size varies throughout the day as they study different subjects.

A widespread feature of Scottish primary education are composite classes. These are classes comprised of pupils from adjacent grades. The maximum class size for composite classes is 25 and each grade needs to contribute a minimum of five pupils. Figure 2 documents that composite classes are very common. In 2018, roughly one in six primary school pupils attended a multi-grade class. Figure 2 also shows that composite classes typically stretch across two grades[4]. In contrast to the examples in the literature to date, composite classes

[4]Most of the small number of cases in which composites stretch across three or more grades are instances

are by no means a rural phenomenon in Scotland. For example, in 2018, 84% of primary schools in the City of Glasgow - the fourth largest city in the UK - featured at least one composite class.

Figure 2: Pupils by Grade and Class Type (2018)



Notes: This bar chart shows the distribution by class type (single-year vs multi-grade) of pupils in Scottish primary schools in 2018

It is also noteworthy that being in a multi-grade class in the previous year is only a weak predictor of composite status in the current year. Indeed, only about 22% of pupils who were in a composite class in the previous school year are again in a composite class the year after. Section 4 will show that this is because small variations in grade enrolment counts can trigger the creation/termination of composite classes in a quasi-random fashion.

3 Data

3.1 Main Data Source: Scottish Pupil Census (SPC)

Our data are drawn from the Scottish Pupil Census (SPC) for school years 2007/08 to 2018/19. The SPC data is collected annually in September and contains information on each

in which parents have elected for a Gaelic education with English taught as the secondary language.

individual pupil and the schools they attend. Enrolment counts and individual-level data are transmitted from each school to the Scottish Government. Upon entering the Scottish school system, each pupil is assigned a unique ID, the so-called Scottish Candidate Number (SCN). We use the SCN to link pupils' records across years and track each individual pupil's progression through the school system.[5]

The SPC data form the main piece of our data linking project. We use these data to create separate 'building block' files that extract and summarise data on key variables from the census. Some of these files summarise information on our composite class and class size measures and instruments, while others contain information on pupils and their demographic characteristics. The information contained in each of these files, and the way we linked them, is summarised in Figure A1. We use the student identifiers to link the data from the building block files to our outcome data.[6]

3.2 Main Outcome Data: CfE Assessments

The primary outcome data we use for our analysis is from the Scottish Government's "Curriculum for Excellence" (CfE) program, which was introduced in 2010. The implementation of CfE involved a pivot towards more topic-based learning that uses knowledge from a range of subjects for projects, rather than purely subject-specific learning. Testing became less frequent, but since 2015/16, each pupil's progress is assessed in both numeracy and literacy as either "Below Early Level", "Early Level", and at "1st/2nd/3rd/4th" level. These assessments are teacher-based but informed by standardised test scores to ensure consistency. Assessments are made at the end of P1 when pupils are expected to perform at early level, and at the end of P4, P7, and S3 when students are expected to perform at the first, second, and third level, respectively. We use the SCN to link each pupil from the pupil census to their CfE assessments and create indicators for whether a pupil performs at the expected level in a given stage.

Table 2 shows that across all grades, between 73% and 85% of pupils perform at least at the expected level in numeracy. The literacy assessment is comprised of assessments of reading, writing, and listening and talking ability. Only if pupils are performing at level in all three subcategories are they considered to perform at level in the aggregate literacy category. This is the case for 69% to 76% of pupils. The average school starting age is 5.2 years. While we have information on free school meal registration, we do not use these as

[5]In practice, the data provider converted the original SNCs into consistent generic IDs for data protection reasons.

[6]Code for both analysis and data cleaning/linkage can be downloaded from GitHub: https://github.com/bodanieli/educationscotland

proxies for family affluence. Scotland introduced universal free school lunches for P1-P3 in 2015, so the variable lacks variation and is no longer properly maintained. We instead use the Scottish Index of Multiple Deprivation (SIMD) as a proxy for background. The SIMD ranks 6,976 'datazones' (small area statistical geographies) from most to least deprived in terms of income, employment, education, health, access to services, crime and housing.[7] Not surprisingly, about 20% of pupils in the sample come from households located in areas ranking in the bottom quintile.

Table 2: Summary Statistics

	First-Graders (P1)		Fourth-Graders (P4)		Seventh-Graders (P7)	
	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Numeracy - Performing at level	0.851	0.356	0.759	0.428	0.731	0.444
Literacy - Performing at level	0.759	0.428	0.690	0.463	0.679	0.467
Reading - Performing at level	0.819	0.385	0.777	0.416	0.775	0.417
Writing - Performing at level	0.791	0.406	0.721	0.449	0.708	0.455
Listening & Talking at level	0.871	0.335	0.844	0.363	0.829	0.377
Class Size	21.813	3.265	26.635	3.955	26.413	4.323
Grade Enrolment	46.168	19.381	46.650	18.788	44.333	17.801
Female	0.491	0.500	0.493	0.500	0.491	0.500
White	0.828	0.377	0.855	0.352	0.878	0.327
Free Meal	0.339	0.473	0.179	0.384	0.167	0.373
Native English Speaker	0.926	0.262	0.924	0.265	0.937	0.243
Bottom 20% SIMD	0.226	0.418	0.217	0.412	0.216	0.411
Age (in Years)	5.210	0.307	8.205	0.308	11.209	0.313
% Female in School	0.490	0.032	0.490	0.032	0.490	0.032
% White British	0.848	0.123	0.852	0.116	0.853	0.119
% Free School Meals	0.247	0.199	0.246	0.197	0.250	0.198
% Native English Speakers	0.922	0.098	0.925	0.092	0.925	0.095
% in Bottom 20% SIMD	0.223	0.265	0.217	0.262	0.217	0.261
No. of Students in School	317.454	126.694	319.516	127.876	317.994	128.655
Observations	190,704		194,804		186,082	
No. of Schools	1,437		1,428		1,435	

Notes: All data stem from Scottish Pupil Census (SPC) 2015/16 - 2018/19, with assessment data added by matching via Scottish Candidate Number (SCN).

The SPC also documents the school and name of the class that each pupil attends as well as each pupil's grade or cohort. Since ours is individual level data, we can simply count the number of pupils in each class and in each grade to construct measures of class size and grade enrolment. Table 2 shows that the average class size experienced by Scottish pupils in P1 is 21.8 whereas it is about 26.5 in fourth and seventh grade. Classes that contain at least five pupils from different grades are flagged up as composite classes. In addition, we cross-

[7] See `https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/` for more information.

checked multi-grade status against the class name which in most local councils specifically flags up composite classes.[8]

3.3 Additional Outcome Data

We also linked the SPC data to several additional data sets which allows for an analysis of outcomes other than same-year attainment. First, we use the SCN to identify pupils who took part in the Scottish Survey of Literacy and Numeracy (SSLN) which run from 2012-2016. The SSLN randomly selected two pupils each from the P4, P7, and S3 grades of every school in Scotland and surveyed these pupils with respect to their attitudes towards learning. For instance, the SSLN obtains information on the degree to which pupils tackle problems independently, whether they enjoy learning and how motivated they are to do well in school. The SSLN also asked participants to take a standardised maths test in 2013 and 2015; and in reading, writing and listening comprehension in 2012, 2014, and 2016.

Second, data on attainment and post-graduation destination are also linked to each pupil's record. Specifically, our data contain information on when pupils leave secondary school along with the numbers and level of Scottish Credit and Qualifications Framework (SCQF) achieved. Pupils are allowed to leave at the age of sixteen, when they have generally completed S4 (eleventh grade) and achieved the relevant qualification for that year, namely National 5, which is equivalent to SCQF level 5. Alternatively, they can progress to S5 and S6 in order to attain Highers and Advanced Highers, which correspond to SCQF level 6 and 7 respectively. These are qualifications equivalent to English "A-Levels". We observe both the number and level of each qualification. Whilst the specific subject for which these qualifications have been attained, is not contained in the original data, our processed data set specifies which SCQF level each student has achieved in the two main SCQF categories of literacy and numeracy. For example, if a student left secondary school with a SCQF level 6 in numeracy, she must have achieved at least three Highers in scientific subjects, such as maths, physics, computer science, or similar subjects.

The Scottish Government also records the destination of pupils about three months after they exit high school, and then again about nine months later. About 95% of pupils are usually tracked and we match their records to our main data set via the Scottish Candidate Number. These "leavers data" run from 2008 to 2019. For every graduate they contain information on whether they are enrolled in higher education, are in employment, receiving further education or training, are unemployed, or pursuing other activities such as volunteer work.

[8]For instance, first-grade year-group classes are typically named P1-A, P1-B, etc, whereas P1/P2 composite classes are typically named P1P2-A, P1P2-B, etc.

It should be noted that the sample we use for our analysis of composite classes differs marginally from the full population data. We excluded special education classes from our sample because class size limits do not apply to them. We dropped pupils who receive a Gaelic Medium education. Because the number of pupils in Gaelic Medium education is very small, primary schools tend to pool them into either a single P1-P7 or two P1-P3 and P4-P7 classes regardless of enrolment counts. We also exclude grades where non-native English speakers were grouped together into large multi-grade classes. These adjustments reduce the sample size by about 1%. Lastly, for our main analysis, we also exclude pupils who attend primary schools in East Renfrewshire because this local authority does not create composite classes in first grade.[9]

4 Methods

A key challenge for our analysis is that pupils are not randomly selected for composite classes. For instance, Table A1 in the Appendix documents that P1/P2 composite classes tend to combine the oldest members of a first-grader (P1) cohort with the youngest and lowest attainment members of the corresponding second-grade (P2) cohort. Failure to account for these patterns will result in a biased estimate of the effect of composite status on attainment. That is because we cannot easily distinguish the effect of age or ability from the effect of exposure to composite class peers. For instance, if only high ability pupils were to be placed in composite classes, then a simple comparison of composite and single-year classes would suggest a positive correlation between composite status and attainment. But, it would be very hard to determine whether this is a spurious correlation that is driven by differences in ability, or a causal effect that is driven by class composition.

We overcome this selection issue by exploiting a "natural experiment" that is created by the institutional setting in Scotland. In Scottish primary schools, an algorithm ("class planner") determines the most efficient number, size, and composition of classes, subject to minimum and maximum class size rules. Specifically there are class size limits for single grade classes which vary by grade, and separate caps for composite classes (see Table 1). The class planner is set up to minimize the number of classrooms a school needs to create. Combined with fluctuations in enrolment counts across years, it generates variation in the composition of classes within a school. For instance, the maximum class size for fourth and fifth grade in Scotland is 33, while composite classes are capped at 25. Therefore, for an enrolment count

[9]All results are robust to the inclusion of East Renfrewshire data, but since an inclusion of these data will lead to a weaker "first stage" (see methods section), our preferred sample excludes observations from these schools.

of 45 fourth-graders and 46 fifth-graders, the class planner would recommend the creation of one 33 pupil-strong fourth and fifth-grade class each, and one 25 pupil strong composite class. Yet with just one additional fourth-grade pupil (i.e. 46 pupils in both grades), class size maxima would force the creation of two fourth-grade and two fifth-grade classes. In effect, small, random variations in enrolment counts trigger the creation of composite classes in some grades, in some schools, in some years, but not in others.

To further illustrate the intuition behind our natural experiment, Figure 3a shows the optimal allocation – as predicted by the class planner – for one of the schools in our sample. Enrolment counts for all seven grades are in the high 40s or low 50s, as is typical for the average school. For illustrative purposes, we zoom in on the bottom three grades. The class planner here determines that the optimal allocation is to create two single-year classes for each grade. Figure 3b, on the other hand, shows the optimal allocation, as calculated by the class planner, for a case which is identical to the one in Figure 3a except that there are now 44 instead of 45 pupils enrolled in first grade. This marginal change triggers several composite classes across different stages, and the suggested reallocation ultimately saves one classroom. This example illustrates that marginal changes in enrolment counts in any grade trigger or prevent composite classes and reshuffle pupils into different class types across all grades. As a result, pupils are quasi-randomly exposed to peers from either the same or older/younger age groups.

Intuitively, our method compares pupils who are placed in composite classes because they happened to be in cohorts for which the class planner determined that they should contribute to a composite class, with pupils who happened to be in cohorts in which the class planner determined that single-year classes should be created. Figure 3 illustrates that these class planner predictions depend on marginal variations in enrolment counts in all grades. Enrolment counts, in turn, are primarily driven by population counts and thus as good as randomly determined. In other words, exploiting this natural experiment allows us to overcome the above selection issue and to isolate a causal effect of composite classes on pupil performance in literacy and numeracy.

On a more technical level, the analysis is then conducted using Two-Stage Least Squares (2SLS) instrumental variable technique in which the class planner prediction is used as an instrument for composite status. This statistical technique also allows us to account for non compliance, i.e. instances in which head teachers deviate from class planner predictions. Figure 4 illustrates the compliance of schools with class planner predictions for P1. Because there is a lot of overlap for local authorities with many medium-sized primary schools, Figure 5 zooms into this segment of schools with 150 to 300 pupils. A class planner suggestion for the school to create a composite class is indicated by a dot. When a school indeed creates a multi-grade class, that is indicated by a circle. So, a dot that is surrounded by a circle refers to a school that - as per planner predictions - should have created a composite class and indeed did so. Circles without dots show schools that created a composite class in P1 even though - given the enrolment count - the class planner will not have suggested one. Correspondingly, dots without circles show instances in which head teachers opted to not create a suggested P1/P2 class. Crosses flag up instances where the class planner did not suggest the creation of a composite class and none was created. Our 2SLS method works as long as there is sufficient compliance, that is empty circles tend to be a rare occurrence relative to the prevalence of filled circles in Figures 4 and 5.

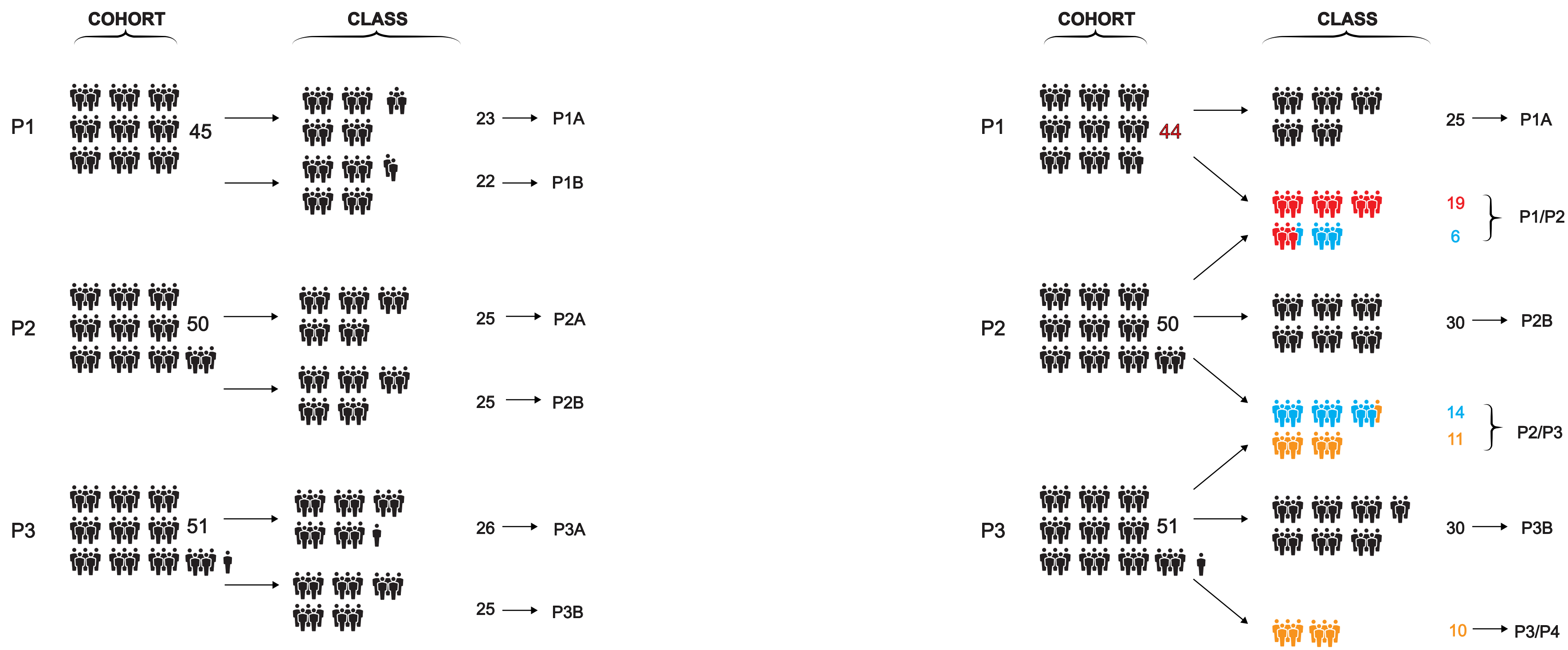
A close inspection of the graphs reveals that indeed filled circles vastly outnumber empty circles. However, compliance is lower in fourth-grade and seventh-grade which mechanically increases the estimated effect range. In our analysis, we therefore focus on the effects of P1/P2 composite classes, although we also report our results for fourth and seventh-graders. It is also important to note that our method does not yield the effect of peer exposure for the average student. That is because we compare pupils who - by virtue of random fluctuations in enrolment counts - end up in a composite class with older peers, against pupils who would have ended up in a composite class, had the enrolment count in their school-year just marginally differed from their actual enrolment count. For instance, P1 pupils in composite classes tend to be older than the average P1 pupil (see Appendix Table A1). In other words, our estimate yields a causal effect that is purged of con-founders such as age and ability, but this effect still pertains to the subset of pupils who typically are selected to attend composite classes.[10]

[10]The econometrics literature refers to our estimates as "Local Average Treatment Effects" (LATEs), see Angrist and Pischke (2008)

Figure 3: Class Planner Examples

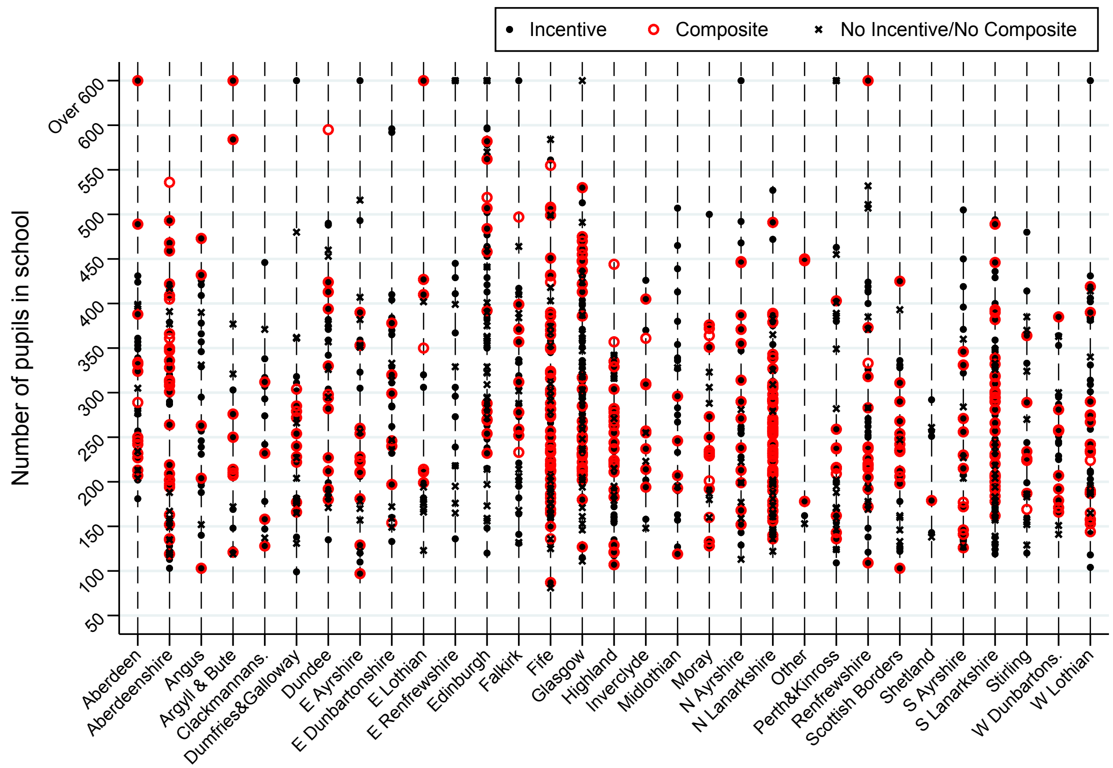
(a) Class Planner Example - Scenario 1

(b) Class Planner Example - Scenario 2



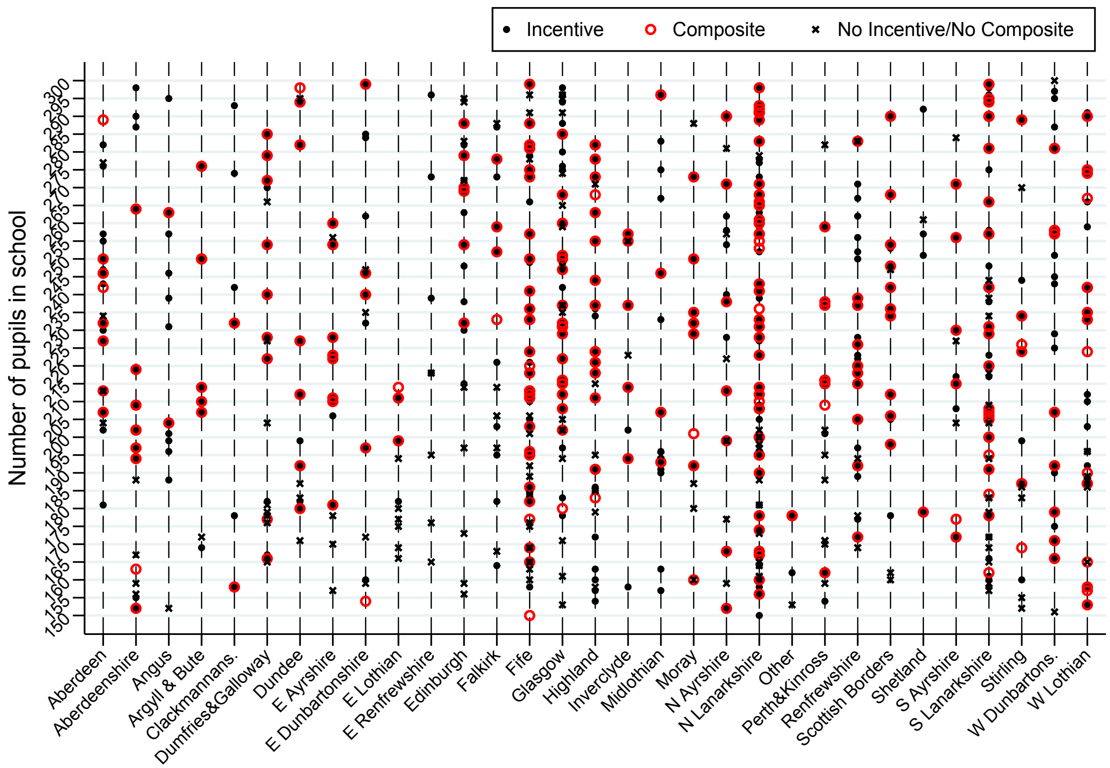
Notes: This is an illustration of the allocations suggested by the class planner. In reality enrolment counts for all seven primary school grades are fed into the class planner, for ease of interpretation we focus here on the bottom three grades of an anonymised primary school. We show two scenarios. The only difference between both scenarios is that in scenario 1 (on the left) this school has an enrolment count of 45 first graders, whereas in scenario 2 (on the right), there are 44 first graders enrolled. As is apparent from the figure, this marginal difference leads to fundamentally different class planner predictions. In scenario 1, none of the pupils is assigned to a composite class, in scenario 2 all grades are assigned to treatment.

Figure 4: P1 Class-Types by School and Local Authority in 2018/19



Notes: This figures shows primary schools' compliance with class planner predictions. Each school is represented by either a dot surrounded by a circle which indicates that a P1/P2 composite class was suggested by the class planner and was created; or by just a dot which means a composite class was created without planner suggestion; or by just a circle which indicates that no composite class was created despite a class planner suggestion: or a cross which suggests that a P1/P2 was neither suggested nor created. The position on the y-axis indicates the size of a school, the position on the x-axis corresponds to the local authority ("Council") a school is located in.

Figure 5: P1 Class-Types by School and Local Authority in 2018/19 (Only schools with 150-300 pupils)



Notes: This figures is identical to the previous figure except that it focuses on schools with enrolment counts between 150 and 300 pupils. It shows primary schools' compliance with class planner predictions. Each school is represented by either a dot surrounded by a circle which indicates that a P1/P2 composite class was suggested by the class planner and was created; or by just a dot which means a composite class was created without planner suggestion; or by just a circle which indicates that no composite class was created despite a class planner suggestion: or a cross which suggests that a P1/P2 was neither suggested nor created. The position on the y-axis indicates the size of a school, the position on the x-axis corresponds to the local authority ("Council") a school is located in.

5 Main Results

In this section we present our estimates for the effect of exposure to older, more (school-) experienced peers by way of composite classes. For comparison, we report two estimates: "OLS" denotes estimates that were obtained using Ordinary Least Squares (OLS) regressions. These reflect potentially spurious correlations. Second, we report Two-Stage Least Squares (2SLS) coefficients which are purged of selection bias and yield causal effects. All coefficients can be interpreted as percentage point changes in the odds of performing at the level appropriate for the corresponding grade. We also report standard errors in parentheses. A useful rule of thumb is that regression coefficients need to be at least twice the size of the corresponding standard errors in order to reflect a statistically significant effect (at the 5% level) that is unlikely to be generated by chance.

Columns (2) and (3) of Panel A in Table 3 show that for first-graders, exposure to an additional older peer (i.e. exposure to an additional second grader) raises the probability of performing at level or better in numeracy by 0.8 to 1.1 percentage points. On average, P1/P2 classes contain about 10 P2 pupils, so this translates into an average increase of 9-11 percentage points (see columns (5) and (6)). These sizable effects stand in contrast to naïve OLS estimates in column (1) which indicate a precisely estimated zero effect. Of course, the OLS coefficients in column (1) and (4) suffer from selection bias. The contrast with our 2SLS coefficients illustrates the value of exploiting a natural experiment rather than relying on mere correlations. Panel B shows that our effects are slightly larger for literacy. Each older peer here increases performance by 1.3 to 1.5 percentage points. The coefficients in both columns (2) and (3) are statistically significant at the 5% level. That translates into a 15-16 percentage point composite class boost in the probability of performing at least at the expected level in literacy.

While our analysis reveals very large benefits of P1/P2 composite classes for P1 pupils, these estimates are in line with the previous literature. For instance, Leuven and Rønning (2014) find that multi-grade classes in Norway increase younger pupils' performance by 0.4 standard deviations. Our point estimates suggest improvements of 0.28 standard deviations for numeracy and 0.35 standard deviations for literacy.

A natural question is whether the benefits for first graders come at the expense of second graders. While second graders are not tested at the end of second grade, they are tested two years later at the end of fourth grade. In other words, our data allow us to test whether there are any medium-run adverse effects on pupils who made up the P2 component of a P1/P2 composite class. Table 4 suggests that this is not the case. All 2SLS coefficients are very small and none of them are statistically significant. We therefore conclude that first-graders

Table 3: Main Results - First Graders (P1)

Panel A: Numeracy - Performing at Least at Level						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Older Peers	0.001*	0.008**	0.011**			
	(0.000)	(0.003)	(0.005)			
Composite				−0.002	0.091**	0.108**
				(0.004)	(0.037)	(0.054)
Class Size	0.002***	0.001	0.006*	0.002***	0.001**	0.005*
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.003)
Observations	190,704	190,704	190,704	190,704	190,704	190,704
No. of Schools	1,437	1,437	1,437	1,437	1,437	1,437
Class-Size Instrumented	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes
F-Stat		556.5	212.6		368.2	190.3
Panel B: Literacy - Performing at Least at Level						
	(7)	(8)	(9)	(10)	(11)	(12)
	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Older Peers	0.001**	0.013***	0.015**			
	(0.000)	(0.004)	(0.007)			
Composite				0.003	0.159***	0.153**
				(0.004)	(0.046)	(0.067)
Class Size	0.002***	0.001	0.004	0.002***	0.002**	0.002
	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.004)
Observations	190,704	190,704	190,704	190,704	190,704	190,704
No. of Schools	1,437	1,437	1,437	1,437	1,437	1,437
Class-Size Instrumented	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes
F-Stat		556.5	212.6		368.2	190.3

Notes: ∗∗∗/∗∗/∗ indicate significance at the 1%/5%/10%-level. Heteroscedasticity-robust standard errors adjusted for clustering at the school and year level are reported in parentheses.

This table shows the results for our estimation by Ordinary Least Squares (OLS) and 2-Stage-Least-Squares (2SLS) regression. Our outcomes of interest are dummy (0/1) indicators for whether a pupil performs at least at the expected level in numeracy or literacy, respectively. All results refer to our sample of first graders (P1).

Covariates include pupil age, sex, and ethnicity an indicator for whether pupil is from a neighborhood in bottom 20% of deprivation (SIMD), grade enrolment counts and its square, the size of the school, and the percentage of pupils in a school that are female, white British, native English speakers, and in the bottom 20% of deprivation respectively. All specifications contain a set of school and year fixed effects.

The reported first-stage F-statistic is heteroscedasticity and autocorrelation consistent (HAC) and was calculated using the method developed by Kleibergen and Paap (2006).

are likely to benefit from exposure to second graders by way of composite classes without there being evidence of a detrimental impact on the second graders who are in the same class. However, it should be noted that we also fail to find persistent benefits for composite P1 pupils when tested in P4. This suggests that costs and benefits are either short-lived and wash out over time, or that we lack the statistical precision to detect these long-run effects.

Additional evidence for our findings is provided by our analysis of fourth- and seventh-graders. Remember that our natural experiment suffered from lower compliance for these grades, so our results are generally less precise and more suggestive in nature. Nonetheless, Table 5 shows results that are consistent with our main findings so far. Exposure to older peers is associated with improvements in numeracy (see columns (2) and (4) of Panel A) and there is no evidence for a statistically significant negative impact on pupils who make up the older component of a P3/P4 or a P6/P7 composite class.

Table 4: Second Stage Results - Performance of Second Graders (P2) in Fourth Grade (P4)

Panel A: Numeracy - Performing at Least at Level

	(1) OLS	(2) 2SLS	(3) 2SLS	(4) OLS	(5) 2SLS	(6) 2SLS
Young Peers	−0.006***	0.004	0.005			
	(0.001)	(0.004)	(0.005)			
Composite				−0.055***	0.032	0.048
				(0.008)	(0.039)	(0.042)
ClassSize	0.002***	0.006***	−0.002	0.002***	0.006***	−0.002
	(0.001)	(0.002)	(0.006)	(0.001)	(0.002)	(0.006)
Observations	97,568	97,568	97,568	97,568	97,568	97,568
No. of Schools	1380	1380	1380	1380	1380	1380
Class-Size Instrumented	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes
F-Stat		195.9	224.9		155.9	165.9

Panel B: Literacy - Performing at Least at Level

	(7) OLS	(8) 2SLS	(9) 2SLS	(10) OLS	(11) 2SLS	(12) 2SLS
Young Peers	−0.006***	−0.005	−0.004			
	(0.001)	(0.005)	(0.005)			
Composite				−0.049***	−0.045	−0.040
				(0.009)	(0.043)	(0.047)
ClassSize	0.002**	0.003	0.005	0.003***	0.003	0.005
	(0.001)	(0.002)	(0.007)	(0.001)	(0.002)	(0.007)
Observations	97,568	97,568	97,568	97,568	97,568	97,568
No. of Schools	1380	1380	1380	1380	1380	1380
Class-Size Instrumented	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes
F-Stat		195.9	224.9		155.9	165.9

Notes: ∗∗∗/∗∗/∗ indicate significance at the 1%/5%/10%-level. Heteroscedasticity-robust standard errors adjusted for clustering at the school and year level are reported in parentheses.

This table shows the results for our estimation by Ordinary Least Squares (OLS) and 2-Stage-Least-Squares (2SLS) regression. Our outcomes of interest are dummy (0/1) indicators for whether a second grader (P2) performs at least at the expected level in numeracy or literacy two years later in fourth grade (P4). The explanatory variable measures whether a second grader was exposed to younger P1 pupils by way of a P1/P2 composite class.

Covariates include pupil age, sex, and ethnicity an indicator for whether pupil is from a neighborhood in bottom 20% of deprivation (SIMD), grade enrolment counts and its square, the size of the school, and the percentage of pupils in a school that are female, white British, native English speakers, and in the bottom 20% of deprivation respectively. All specifications contain a set of school and year fixed effects.

The reported first-stage F-statistic is heteroscedasticity and autocorrelation consistent (HAC) and was calculated using the method developed by Kleibergen and Paap (2006).

Table 5: Second Stage Results - Fourth (P4) and Seventh (P7) Graders

Panel A: Second Stage Results for P4

	Numeracy				Literacy			
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS
Older Peers	0.002***	0.044***			0.002***	0.024		
	(0.001)	(0.017)			(0.001)	(0.017)		
Younger Peers	−0.004***	−0.009			−0.004***	−0.008		
	(0.001)	(0.015)			(0.001)	(0.015)		
Bottom Comp.			0.033***	0.509**			0.027***	0.291
			(0.006)	(0.215)			(0.007)	(0.206)
Top Composite			−0.032***	−0.250			−0.037***	−0.180
			(0.007)	(0.257)			(0.007)	(0.247)
Class Size	0.004***	0.004	0.004***	0.001	0.004***	0.002	0.004***	0.001
	(0.001)	(0.003)	(0.001)	(0.004)	(0.001)	(0.003)	(0.001)	(0.004)
Observations	194,804	194,803	194,804	194,803	194,804	194,803	194,804	194,803
No. of Schools	1428	1428	1428	1428	1428	1428	1428	1428
Class-Size Instr.	No	Yes	No	Yes	No	Yes	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-Stat		12.82		4.918		12.82		4.918

Panel B: Second Stage Results for P7

	Numeracy				Literacy			
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS
Younger Peers	−0.005***	−0.051			−0.005***	0.008		
	(0.001)	(0.035)			(0.001)	(0.031)		
Top Composite			−0.054***	−0.704			−0.051***	0.114
			(0.007)	(0.527)			(0.007)	(0.425)
Class Size	0.004***	−0.017	0.004***	−0.021	0.004***	0.005	0.004***	0.005
	(0.001)	(0.012)	(0.001)	(0.017)	(0.001)	(0.011)	(0.001)	(0.014)
Observations	186,082	186,078	186,082	186,078	186,082	186,078	186,082	186,078
No. of Schools	1,435	1,435	1435	1435	1,435	1,435	1,435	1,435
Class-Size Instr.	No	Yes	No	Yes	No	Yes	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-Stat		17.99		12.91		17.99		12.91

Notes: ∗∗∗/∗∗/∗ indicate significance at the 1%/5%/10%-level. Heteroscedasticity-robust standard errors adjusted for clustering at the school and year level are reported in parentheses.

This table shows the results for our estimation by Ordinary Least Squares (OLS) and 2-Stage-Least-Squares (2SLS) regression. Our outcomes of interest are dummy (0/1) indicators for whether a pupil performs at least at the expected level in numeracy or literacy, respectively. Results in Panel A refer to our sample of fourth graders (P4), results in Panel B refer to seventh graders (P7).

Covariates include pupil age, sex, and ethnicity an indicator for whether pupil is from a neighborhood in bottom 20% of deprivation (SIMD), grade enrolment counts and its square, the size of the school, and the percentage of pupils in a school that are female, white British, native English speakers, and in the bottom 20% of deprivation respectively. All specifications contain a set of school and year fixed effects.

The reported first-stage F-statistic is heteroscedasticity and autocorrelation consistent (HAC) and was calculated using the method developed by Kleibergen and Paap (2006).

6 Mechanisms and Heterogeneity

Overall, our results show that first graders who are selected for P1/P2 classes benefit from exposure to peers who are already in their second year of primary school. By contrast, we find no evidence for adverse effects on pupils who are exposed to younger peers because they make up the upper part of a composite class. We now turn to mechanisms and explore whether some subgroups benefit more (or less) from the peer effects we identified above.

By virtue of a lower cap, composite classes tend to be smaller than single-year classes (see Table 1 for maximum class size rules). As a result, composite classes may affect achievement not just through peer effects but also due to a lower pupil to teacher ratio. In order to disentangle these two competing mechanisms, we have included a control variable for class size in all specifications. We reported the corresponding regression output in Tables 3, 4, and 5. It is noticeable that the effect of class size tends to be both statistically and economically insignificant in virtually all specifications. For instance, the class size coefficients in Table 3 are positive, range from 0.001 to 0.006 depending on the specification, and none of them are statistically significant at the 5% level. This is in stark contrast to our large and statistically significant results for exposure to older peers. We thus conclude that multi-grade classes do not improve first graders' attainment through smaller classroom sizes, but indeed through peer effects. In fact, class size seems to have no measurable effect on attainment.

It should also be noted that while our null result for class size is inconsistent with Fredriksson et al.'s (2013) study of Sweden and Angrist and Lavy's (1999) original findings for Israel, it is in line with, among others, results by Hoxby (2000b) for the US, Leuven et al. (2008) for Norway, as well as Angrist et al.'s (2019) follow-up study in Israel.

We also find larger and more pronounced peer effects for literacy than for numeracy that warrant further investigation. Table A2 in the Appendix breaks down our literacy assessment into its three components: reading, writing, and listening & talking. These subcategories may offer pointers on the channel through which exposure to more mature peers improves literacy. While listening and talking are – by definition – interactive activities, reading and writing can be improved by working on one's own. We find the gains appear to be concentrated in improvements in reading and writing ability respectively, but find no effect on listening and talking. While this breakdown does not allow us to fully disentangle these mechanisms, it suggests that it is not the direct interaction with older peers that is driving these improvements. Instead younger pupils may be motivated and spurred on by observing peers who have already acquired reading and writing proficiency.

There are other mechanisms through which exposure to more mature classroom peers may lead to better performance. It is conceivable that parents may invest more effort into supporting their children if they are placed in a composite class. More generally, there is growing evidence suggesting that parents from higher socio-economic strata offer more support to their offspring (Francesconi and Heckman 2016; Fredriksson et al. 2016). Composite classes may exacerbate these inequalities if gains for first graders are driven by higher time and resource investments by affluent parents.

However, Panel A of Appendix Table A3 indicates few differences in terms of socio-economic background. In fact, our point estimates suggest that pupils from postcodes which are ranked in the two bottom quintiles in terms of deprivation (as measured by the above-mentioned SMID) tend to benefit slightly more from exposure to older peers than pupils in the top three quintiles. However, these differences are not significant at any reasonable level of statistical significance.

It may also be the case that teacher training or teacher effort is affected by class type and that this may feed through into better pupil outcomes. Unfortunately information on each class' main teacher is not included in the data. However, our interviews with decision makers revealed that there is neither special training nor additional support (e.g. teaching assistants) for composite classes. That is confirmed by our data which shows that composite classes receive no additional staffing. Furthermore, urban schools tend to find teacher recruitment easier and are on average larger which may make them more likely to develop teachers who specialise in the instruction of composite classes. But again, Panel B of Appendix Table A3 reveals little in the way of effect differences between urban and rural schools.

We also stratify our sample by pupil gender. There is an extensive literature that documents gender differences. Lavy et al. (2012b), for instance, show that girls benefit more from exposure to high-ability peers than boys. However, consistent with Leuven and Rønning (2014), Panel C of Table A3 shows that while at first glance boys appear to benefit more from composite classes than girls, these gender differences are too small to be considered statistically meaningful.

It is plausible that the experience of a composite class setting may affect behaviour in general and attitudes towards learning in particular. SSLN survey data and the data on exclusions and attendance allow us a glimpse at this channel. Panel B of Table A4 and Table A5 show no evidence of an effect of either class size or exposure to more experienced peers on exclusions, absences, and attendance rates. It should be noted that for primary school pupils, exclusions are very rare events to begin with. This makes it challenging to statistically isolate an effect. But the fact that even our large sample size (e.g. $N > 150{,}000$ in Table A5) analysis fails to pick up an effect, suggests that large effects can be all but ruled out. Information on learning attitudes, on the other hand, are only available for a small subsample and only for P4 and P7 pupils. Panel A of Table A4 again reveals no statistically significant effect of exposure to either older or younger peers on indicators for whether pupils enjoy learning, aspire to do well in school, and work independently on assignments. However, it should be noted that the standard errors for all estimates are very large indicating that our sample is too small to warrant strong conclusions.

Finally, we also assessed whether composite classes influence attainment beyond primary or secondary school. Note that because we only have 12 years of data, we cannot track pupils from P1 until after their high school graduation. Appendix Table A6 thus shows the correlation between the number of years spent in a composite class between P4 and P7 and the probabilities of graduating into a positive destination in general and of attending higher education in particular. We fail to find an economically significant link, nor does exposure to composite classes between fourth and seventh grade seem to be associated with the decision to drop out of highschool at age 16, or the odds of obtaining "Highers" in literacy or numeracy.

7 Conclusion and Recommendations

This project leveraged novel data on the universe of the Scottish pupil population from 2007/08 until 2018/19. We linked several data sources to generate an individual-level panel that, in principle, allows researchers to follow each Scottish pupils educational career and trajectory from first grade until 1 year after high school graduation. The code to generate this novel data set, which will be made publicly available, is a key legacy from this project. We hope that as such we have created a useful resource and public good that can be used by policy makers, practitioners and other researchers alike.

We encourage the Scottish Government and third party organizations to utilise this new resource to evaluate existing or future policies and interventions, and to add new waves of existing data sets as they become available. We also encourage adding new data building blocks that allow for even further-reaching research. For instance, many Scandinavian countries have started to link data on educational trajectories with labor market outcomes. Other data may be held by local authorities and schools themselves. Our project was, in one sense, a proof of concept. Not least thanks to the high data quality of administrative records, we showed that it is feasible to link data from different sources, and to use these data to gain insights into what works in education.

Our project showcased the value of these new data when they are combined with quasi-experimental methods by investigating the importance of peer effects in general and the effect of multi-grade ("composite") classes in particular. We find that the presence of second graders by way of a composite class, improves first-graders' reading, writing, and maths performance, as measured by teacher assessments that are informed by standardised test scores. We find no evidence that achievement of older pupils is adversely affected by the classroom presence of younger peers. We also found that class size is not driving these effects and found no effect of class size in general.

Our results have several important implications. First, we show that peer effects matter and can be consciously generated by way of composite classes. This may help alleviate parental concerns about composite classes, at least as far as attainment is concerned. It will also be reassuring for Scottish local authorities that the cost savings that composite classes can provide, do not appear to come at the expense of lower instructional quality. We also encourage policy makers in other countries to experiment with multi-grade classes. An important insight of our study is that the benefits that have been documented in rural settings where composite classes were born out of necessity, also accrue in urban settings where multi-grade classes are created by design. While further research in this area is certainly warranted, the overall body of evidence suggests that composite classes, especially in the early years of primary education, have the potential to be a useful tool to stimulate the learning of mature, high-ability pupils by exposing them to older peers. While our research did not find evidence for adverse effects of composite classes on older peers, we also cannot rule out these effects. In other words, that there is a chance that adverse effects of composite classes on older peers - which had been identified by previous studies - have simply remained undetected in our study. Composite classes should thus not be considered a panacea, but rather appear to be a useful addition to school administrators' and head teachers' tool boxes.

Furthermore, our results suggest that class structure may be more important for attainment than class size. We document very low returns to further class size reductions. This is not to say that smaller class rooms cannot foster effective learning or are ineffective in all dimensions. For instance, smaller class sizes almost certainly make for easier classroom management and allow for teaching techniques that are not feasible in large class rooms. They might also lower teaching workload thus helping with teacher recruitment and retention. It may just be that most attainment gains from class size reductions have already been realised or that further gains are only attainable if average class size drops into the single digits, which may well be prohibitively expensive and infeasible from a staffing perspective. Therefore, other inputs, such as peer or teacher quality, offer more potential to boost pupil performance. An important recommendation is, therefore, that practitioners and policy makers are aware of these other channels when allocating resources.

Lastly, future research should attempt to better understand the mechanisms through which composite classes and peer effects generate benefits. Our study suggests that composite classes may feature particular models of learning, class room management techniques, or pupil interactions that are conductive to learning. How exactly these benefits accrue is important and shedding more light on mechanisms is likely to offer lessons that generalise beyond the role of composite classes. Despite the apparent cost savings of composite classes, anecdotal evidence suggests that there is also a hidden cost to teachers in the form of longer preparation times for composite classes relative to single year classes. Uncovering all costs and benefits and the channels through which they operate will require more or better data, qualitative approaches, or - ideally - policy experimentation that allows for robust evaluation.

References

Angrist, J. D. and V. Lavy (1999): “Using Maimonides’ rule to estimate the effect of class size on scholastic achievement,” *The Quarterly Journal of Economics*, 114, 533–575.

Angrist, J. D., V. Lavy, J. Leder-Luis, and A. Shany (2019): “Maimonides’ Rule Redux,” *American Economic Review: Insights*, 1, 309–24.

Angrist, J. D. and J.-S. Pischke (2008): *Mostly Harmless Econometrics: An Empiricist’s Companion*, Princeton University Press.

Ballatore, R. M., M. Fort, and A. Ichino (2018): “Tower of Babel in the classroom: immigrants and natives in Italian schools,” *Journal of Labor Economics*, 36, 885–921.

Ballatore, R. M., M. Paccagnella, and M. Tonello (2020): “Bullied because younger than my mates? The effect of age rank on victimisation at school,” *Labour Economics*, 62, 101772.

Barbetta, G. P., G. Sorrenti, and G. Turati (2019): “Multigrading and child achievement,” *Journal of Human Resources*, 0118–9310R4.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2011): “Too young to leave the nest? The effects of school starting age,” *The review of economics and statistics*, 93, 455–467.

——— (2013): “Under pressure? The effect of peers on outcomes of young adults,” *Journal of Labor Economics*, 31, 119–153.

Carrell, S. E., M. Hoekstra, and E. Kuka (2018): “The long-run effects of disruptive peers,” *American Economic Review*, 108, 3377–3415.

Cascio, E. U. and D. W. Schanzenbach (2016): “First in the class? Age and the education production function,” *Education Finance and Policy*, 11, 225–250.

Checchi, D. and M. De Paola (2018): “The effect of multigrade classes on cognitive and non-cognitive skills. Causal evidence exploiting minimum class size rules in Italy,” *Economics of Education Review*, 67, 235–253.

Crawford, C., L. Dearden, and E. Greaves (2014): “The drivers of month-of-birth differences in children’s cognitive and non-cognitive skills,” *Journal of the Royal Statistical Society. Series A,(Statistics in Society)*, 177, 829.

Francesconi, M. and J. J. Heckman (2016): “Child development and parental investment: Introduction,” *The Economic Journal*, 126, F1–F27.

Fredriksson, P., B. Öckert, and H. Oosterbeek (2013): “Long-term effects of class size,” *The Quarterly Journal of Economics*, 128, 249–285.

——— (2016): “Parental responses to public investments in children: Evidence from a maximum class size rule,” *Journal of Human Resources*, 51, 832–868.

Gould, E. D., V. Lavy, and M. D. Paserman (2009): “Long-term Classroom Peer Effects: Evidence from Random Variation in Enrollment of Disadvantaged Immigrants,” *Economic Journal*, 119, 1243–1269.

Hoxby, C. (2000a): “Peer effects in the classroom: Learning from gender and race variation,” Tech. rep., National Bureau of Economic Research.

Hoxby, C. M. (2000b): “The effects of class size on student achievement: New evidence from population variation,” *The Quarterly Journal of Economics*, 115, 1239–1285.

Kleibergen, F. and R. Paap (2006): “Generalized reduced rank tests using the singular value decomposition,” *Journal of Econometrics*, 133, 97–126.

Krueger, A. B. (1999): "Experimental estimates of education production functions," *The Quarterly Journal of Economics*, 114, 497–532.

Lavy, V., M. D. Paserman, and A. Schlosser (2012a): "Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom," *The Economic Journal*, 122, 208–237.

Lavy, V. and A. Schlosser (2011): "Mechanisms and impacts of gender peer effects at school," *American Economic Journal: Applied Economics*, 3, 1–33.

Lavy, V., O. Silva, and F. Weinhardt (2012b): "The good, the bad, and the average: Evidence on ability peer effects in schools," *Journal of Labor Economics*, 30, 367–414.

Leuven, E., H. Oosterbeek, and M. Rønning (2008): "Quasi-experimental estimates of the effect of class size on achievement in Norway," *Scandinavian Journal of Economics*, 110, 663–693.

Leuven, E. and M. Rønning (2014): "Classroom grade composition and pupil achievement," *The Economic Journal*, 126, 1164–1192.

Norris, J. (2020): "Peers, parents, and attitudes about school," *Journal of Human Capital*, 14, 290–342.

Rossi, G. (2020): "School Performance, non-cognitive skills and house prices," *Working Paper*.

Sims, D. (2008): "A strategic response to class size reduction: Combination classes and student achievement in California," *Journal of Policy Analysis and Management*, 457–478.

Appendix A

Table A1: Self-Selection of Composite Class Pupils

	Prob(CompP1/P2) - First Graders			Prob(CompP1/P2) -Second Graders			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	0.004***	0.004***	0.003**	−0.004**	−0.004***	−0.003**	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
White	0.006*	−0.004	−0.004	0.010***	−0.002	−0.002	0.000
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Native English Speaker	0.015***	0.016***	0.015***	−0.010*	−0.009**	−0.008**	−0.002
	(0.005)	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.004)
Bottom 20% SIMD	−0.001	−0.003	−0.003	0.004	0.006***	0.006***	0.002
	(0.004)	(0.002)	(0.002)	(0.005)	(0.002)	(0.002)	(0.003)
Age (in Years)	0.132***	0.135***		−0.103***	−0.105***		−0.104***
	(0.006)	(0.006)		(0.006)	(0.006)		(0.007)
1st Age Quartile			−0.013***			0.051***	
			(0.002)			(0.004)	
3rd Age Quartile			0.027***			−0.020***	
			(0.003)			(0.003)	
4th Age Quartile			0.098***			−0.028***	
			(0.005)			(0.003)	
Low Literacy							0.029***
							(0.004)
Low Numeracy							0.036***
							(0.005)
Observations	190,704	190,704	190,704	203,139	203,139	203,139	139,198
R-squared	0.018	0.179	0.181	0.010	0.162	0.163	0.175
School FE	No	Yes	Yes	No	Yes	Yes	Yes

Notes: ∗∗∗/∗∗/∗ indicate significance at the 1%/5%/10%-level. Heteroscedasticity-robust standard errors adjusted for clustering at the school and year level are reported in parentheses.
This table regresses a dummy indicator for whether a pupil is part of a P1/P2 composite class on pupil characteristics. The first three columns show the results for first-graders who form the bottom component of a P1/P2 composite class. Columns (4) through (7) show the our results for second graders who form the top component of a P1/P2 composite class.
Note that only P1 pupils from our main sample (with valid assessment data) are used. In column (7) only P2 pupils for whom P1 assessments (from previous year) were available, are part of the sample.
Low Literacy and Low Numeracy, respectively, indicate that pupils scored below early level when in first grade.

Table A2: Second Stage Results (P1) for Literacy Subcategories

	Reading		Writing		Listening & Talking	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	2SLS	OLS	2SLS	OLS	2SLS
Older Peers	0.001*	0.008**	0.001	0.012***	0.000	0.004
	(0.000)	(0.003)	(0.000)	(0.004)	(0.000)	(0.003)
ClassSize	0.002***	0.001	0.002***	0.001	0.002***	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	190,704	190,704	190,704	190,704	190,704	190,704
No. of Schools	1,437	1,437	1,437	1,437	1,437	1,437
School FE	Yes	Yes	Yes	Yes	Yes	Yes
F-Stat		556.5		556.5		556.5

Notes: ∗∗∗/∗∗/∗ indicate significance at the 1%/5%/10%-level. Heteroscedasticity-robust standard errors adjusted for clustering at the school and year level are reported in parentheses.
This table shows the results for our estimation of equation **??** by Ordinary Least Squares (OLS) and 2-Stage-Least-Squares (2SLS) regression. Our outcomes of interest are dummy indicators for whether a pupil performs at least at the expected level in three subcategories of literacy. All results refer to our sample of first graders (P1)
Covariates include pupil age, sex, and ethnicity an indicator for whether pupil is from a neighborhood in bottom 20% of deprivation (SIMD), grade enrolment counts and its square, the size of the school, and the percentage of pupils in a school that are female, white British, native English speakers, and in the bottom 20% of deprivation respectively.
All specifications contain a set of school and year fixed effects.
The reported first-stage F-statistic is heteroscedasticity and autocorrelation consistent (HAC) and was calculated using the method developed by Kleibergen and Paap (2006).

Table A3: Second Stage Results (P1): Effect Heterogeneity

	Numeracy				Literacy			
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS
Panel A: Heterogeneous Effects by Level of Deprivation								
	Top 60% SIMD		Bottom 40% SIMD		Top 60% SIMD		Bottom 40% SIMD	
Older Peers	0.001**	0.006*	0.000	0.009*	0.001**	0.011**	0.001	0.016***
	(0.000)	(0.003)	(0.000)	(0.005)	(0.000)	(0.004)	(0.001)	(0.006)
Class Size	0.002***	0.001*	0.001*	0.000	0.002**	0.001	0.002**	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	106,653	106,653	84,051	84,051	106,653	106,653	84,051	84,051
No. of Schools	1411	1411	1269	1269	1411	1411	1269	1269
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-Stat	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
376.9 338.5 376.9 338.5								
Panel B: Heterogeneous Effects by School Size								
	Urban		Rural		Urban		Rural	
Older Peers	0.001**	0.008**	−0.000	0.005	0.001**	0.012***	0.001	0.017**
	(0.000)	(0.004)	(0.001)	(0.006)	(0.000)	(0.005)	(0.001)	(0.008)
Class Size	0.002***	0.001*	0.001	0.000	0.002***	0.001	0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Observations	143,834	143,834	46,870	46,870	143,834	143,834	46,870	46,870
No. of Schools	972	972	486	486	972	972	486	486
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-Stat		424.8		131.2		424.8		131.2
Panel C: Heterogeneous Effects by Pupil Sex								
	Boys		Girls		Boys		Girls	
Older Peers	0.000	0.009**	0.001**	0.006	0.001	0.016***	0.001***	0.011**
	(0.000)	(0.004)	(0.000)	(0.004)	(0.001)	(0.005)	(0.000)	(0.004)
Class Size	0.003***	0.002**	0.001	0.000	0.003***	0.001	0.001	−0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	97,125	97,125	93,579	93,579	97,125	97,125	93,579	93,575
No. of Schools	1435	1435	1435	1435	1435	1435	1435	1435
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-Stat		479.7		489		479.7		489

Notes: ∗∗∗/∗∗/∗ indicate significance at the 1%/5%/10%-level. Heteroscedasticity-robust standard errors adjusted for clustering at the school and year level are reported in parentheses.

This table shows the results for our estimation of equation **??** by Ordinary Least Squares (OLS) and 2-Stage-Least-Squares (2SLS) regression. Our outcomes of interest are dummy indicators for whether a pupil performs at least at the expected level in three subcategories of literacy. All results refer to our sample of first graders (P1).

Unless they are the category of interest, covariates include pupil age, sex, and ethnicity an indicator for whether pupil is from a neighborhood in bottom 20% of deprivation (SIMD), grade enrolment counts and its square, the size of the school, and the percentage of pupils in a school that are female, white British, native English speakers, and in the bottom 20% of deprivation respectively. All specifications contain a set of school and year fixed effects.

The reported first-stage F-statistic is heteroscedasticity and autocorrelation consistent (HAC) and was calculated using the method developed by Kleibergen and Paap (2006).

Figure A1: Data Merging Plan

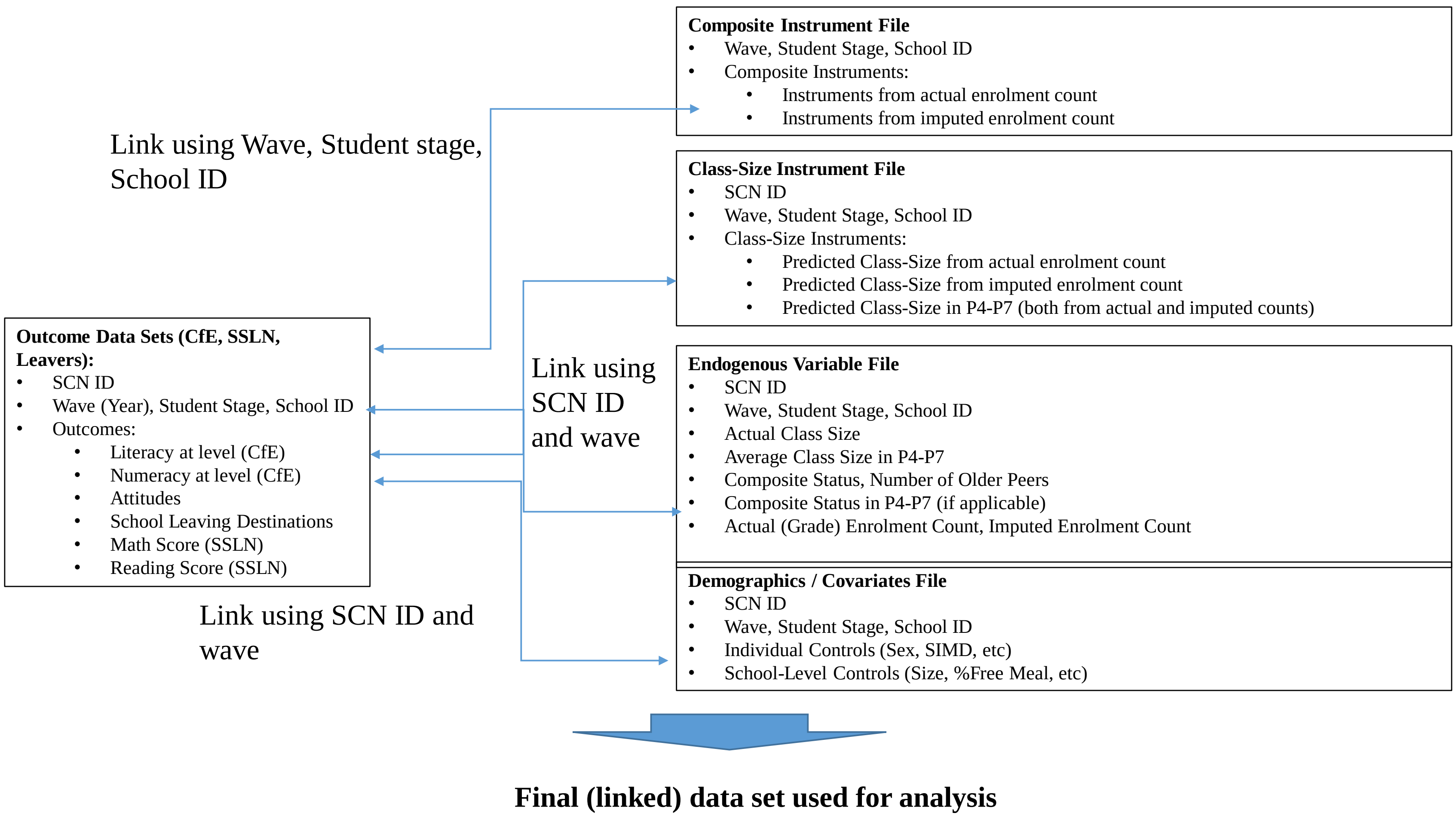


Table A4: Behavioral Outcomes - P4 and P7

Panel A: Attitudes Toward Learning

	P4						P7					
	Enjoys learning		Wants to do well		Answer on own		Enjoys learning		Wants to do well		Answer on own	
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS	(9) OLS	(10) 2SLS	(11) OLS	(12) 2SLS
Older Peers	-0.001	0.045	-0.001	0.060	-0.000	-0.027						
	(0.002)	(0.081)	(0.001)	(0.063)	(0.002)	(0.084)						
Younger Peers	0.003**	0.024	-0.000	0.025	-0.002	-0.041	0.003**	0.077	-0.000	-0.012	0.000	-0.039
	(0.002)	(0.032)	(0.001)	(0.025)	(0.002)	(0.033)	(0.002)	(0.047)	(0.001)	(0.025)	(0.002)	(0.043)
Class Size	-0.002	0.011	-0.001	0.015	0.000	-0.013	0.004**	0.011**	0.000	-0.001	0.003*	-0.001
	(0.002)	(0.019)	(0.001)	(0.015)	(0.002)	(0.020)	(0.001)	(0.005)	(0.001)	(0.003)	(0.001)	(0.005)
Observations	9,905	9,890	9,880	9,865	9,874	9,859	10,194	10,180	10,162	10,148	10,176	10,162
R-squared	0.186	0.098	0.162	-0.171	0.163	0.090	0.181	-0.002	0.153	0.138	0.166	0.111
No. of Schools	1457	1457	1457	1457	1457	1457	1457	1457	1457	1457	1457	1457
F-Stat		1.426		1.534		1.406		11.16		10.64		10.79

Panel B: Behavioural Outcomes

	P4						P7					
	Attendance Rate		Absence Rate		Ever Excluded		Attendance Rate		Absence Rate		Ever Excluded	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Older Peers	0.017*	-0.736	0.000	0.188	0.000	-0.006						
	(0.010)	(0.601)	(0.003)	(0.175)	(0.000)	(0.004)						
Younger Peers	-0.035***	-0.471	0.010***	0.191	0.000	-0.001	-0.034***	0.500	0.013***	0.031	0.000	-0.007*
	(0.011)	(0.433)	(0.003)	(0.126)	(0.000)	(0.003)	(0.010)	(0.370)	(0.004)	(0.105)	(0.000)	(0.004)
Class Size	0.008	-0.322	0.000	0.103	-0.000*	-0.002	-0.004	0.120	0.003	0.007	-0.000**	-0.002**
	(0.011)	(0.266)	(0.003)	(0.078)	(0.000)	(0.002)	(0.009)	(0.086)	(0.003)	(0.025)	(0.000)	(0.001)
Observations	146,133	146,133	146,133	146,133	146,133	146,133	137,579	137,576	137,579	137,576	137,579	137,576
R-squared	0.071	0.026	0.127	0.074	0.031	-0.026	0.075	0.049	0.128	0.128	0.043	0.005
No. of Schools	1444	1444	1444	1444	1444	1444	1443	1443	1443	1443	1443	1443
F-Stat		2.271		2.271		2.271		12.57		12.57		12.57

Notes: $***/**/*$ indicate significance at the 1%/5%/10%-level. Heteroscedasticity-robust standard errors adjusted for clustering at the school and year level are reported in parentheses.
All specifications contain the usual set of covariates alongside school fixed-effects. In panel A, the dependent variables are binary indicators with the following extended names: Enjoys learning a lot, Really wants to do well (a lot), Tries to find answer on own (a lot).
In panel B, Absence rate refers to Unauthorised Absences.
The reported first-stage F-statistic is heteroscedasticity and autocorrelation consistent (HAC) and was calculated using the method developed by Kleibergen and Paap (2006).

Table A5: P1 Behavioural Outcomes

Panel: Behavioural Outcomes - P1	Attendance Rate		Absence Rate		Ever Excluded	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	2SLS	OLS	2SLS	OLS	2SLS
Older Peers	0.023**	-0.002	-0.004*	-0.007	-0.000	-0.000
	(0.009)	(0.113)	(0.002)	(0.025)	(0.000)	(0.000)
Class Size	-0.012	-0.011	0.003	0.003	0.000	0.000
	(0.013)	(0.014)	(0.003)	(0.003)	(0.000)	(0.000)
Observations	152,095	152,095	152,095	152,095	152,095	152,095
R-squared	0.095	0.095	0.116	0.116	0.014	0.014
No. of Schools	1449	1449	1449	1449	1449	1449
F-Stat		277.2		277.2		277.2

Notes: ∗∗∗/∗∗/∗ indicate significance at the 1%/5%/10%-level. Heteroscedasticity-robust standard errors adjusted for clustering at the school and year level are reported in parentheses.
All specifications contain the usual set of covariates alongside school fixed-effects.
Absence rate refers to Unauthorised Absences.
The reported first-stage F-statistic is heteroscedasticity and autocorrelation consistent (HAC) and was calculated using the method developed by Kleibergen and Paap (2006).

Table A6: Leavers' Destinations - OLS

Panel: P1 Behavioural Outcomes	Positive Destination	Higher Education	Dropout	Years of Schooling	Literacy	Numeracy
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
Composite P4-P7	-0.002**	0.001	-0.001	-0.004	-0.007***	-0.004
	(0.001)	(0.002)	(0.002)	(0.005)	(0.002)	(0.003)
Class Size P4-P7	0.001**	0.005***	-0.002**	0.007***	0.006***	0.007***
	(0.000)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Observations	110,779	110,779	110,779	110,779	110,779	110,779
R-squared	0.077	0.233	0.464	0.570	0.321	0.289
No. of Schools	1327	1327	1327	1327	1327	1327

Notes: ∗∗∗/∗∗/∗ indicate significance at the 1%/5%/10%-level. Heteroscedasticity-robust standard errors adjusted for clustering at the school and year level are reported in parentheses.
The two regressors refer to the number of years spent in a composite class between P4-P7 and average class size experiences within the same period.
Literacy and Numeracy are binary variables indicating whether SCQF level 5 (or above) has been achieved.
All specifications contain the usual set of covariates alongside school fixed-effects.