European Expert Network on
Economics of Education (EENEE)

Class size and student outcomes in Europe

EENEE Analytical Report No. 33
Prepared for the European Commission

Edwin Leuven and Hessel Oosterbeek
January 2018

33

Erasmus+

Class size and student outcomes in Europe

Edwin Leuven and Hessel Oosterbeek

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*)**:**

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on the Internet (http://europa.eu).

Luxembourg: Publications Office of the European Union, 2018

ISBN: 978-92-79-70085-9
doi: 10.2766/28340

Executive Summary (English)

Class size and student outcomes in Europe

Class size reductions are popular with parents, teachers and politicians. In recent years various contributors to the public debate have increasingly expressed the opinion that class size does not matter. For example, the popular science author Gladwell (2013) uses class size reduction as an example of a "thing we are convinced is such a big advantage [but] might not be such an advantage at all" (as cited in Schanzenbach, 2014). Likewise, in a recent report the OECD concludes that "overall, evidence of the effect of differences in class size on student performance is weak" (OECD, 2016, p.349). These strong statements are made without (OECD) or with very selective (Gladwell) reference to evidence, let alone a careful assessment of the evidence. The same can often be said of advocates of class size reductions.

Our systematic review of the empirical studies on the effects of class size on student outcomes in Europe shows mixed evidence. Some credible studies report substantial beneficial effects of smaller class sizes, whereas other equally credible studies find effects that are rather small or not significantly different from zero. While some studies show that smaller classes did indeed offer sustained benefits, using this evidence for prescriptive policy purposes typically requires an extrapolation to a different population. Such extrapolations require a solid understanding of how population characteristics, incentives and constraints enter the production function and mediate class-size effects. While there are some studies that investigate such mediating factors, the evidence falls short in providing definitive answers. Based on the current evidence it is therefore hard to give an unqualified recommendation about how and when to use class size policies to improve student outcomes.

Additional research is therefore necessary to provide better answers to the questions whether, why, when and for whom class size matters. These questions must be answered through research design, and the evidence must be causal. Field experiments are typically considered the "gold standard". While the single large-scale randomised experiment regarding class size (Project STAR) has produced many insights, they have their own threats to internal and external validity (Heckman and Smith, 1995). We therefore believe it is important to collect evidence from a variety of causal designs. These can be nonexperiment designs derived from maximum class-size rules, or alternative experimental designs (Rockoff, 2009).

Finally it is worth pointing out that the notion of a class-size effect depends on traditional modes of teaching and learning where we can think of class size as a well-defined and policy-relevant proxy for inputs in schools' production function. While today's schools are remarkably similar to the schools of our grandparents, technology and pedagogical innovations may change the nature of classrooms and thereby the relationship between class size and inputs. Estimates of class-size effects are likely to have expiration dates when the production function is changing.

References

Gladwell, M. (2013). *David and Goliath: Underdogs, Misfits, and the Art of Battling Giants*. Little, Brown & Company., New York.

Heckman, J. J. and Smith, J. A. (1995). Assessing the case for social experiments. *The Journal of Economic Perspectives*, 9(2):85–110.

OECD (2016). *Education at a Glance 2016: OECD Indicators*. OECD Publishing, Paris.

Rockoff, J. (2009). Field experiments in class size from the early twentieth century. *Journal of Economic Perspectives*, 23:211–230.

Schanzenbach, D. W. (2014). Does class size matter? National Education Policy Center.

Executive Summary (French)

Taille des classes et résultats des élèves en Europe

Les réductions du nombre d'élèves par classe sont populaires auprès des parents, des enseignants et des politiciens. Au cours des dernières années, divers participants au débat public ont de plus en plus exprimé l'opinion selon laquelle la taille des classes n'a pas d'importance. Par exemple, l'auteur de vulgarisation scientifique Gladwell (2013) utilise la réduction de la taille des classes comme illustration du principe selon lequel une « chose dont nous sommes convaincus qu'elle constitue un grand avantage [mais] pourrait ne pas du tout être si avantageuse » (tel que cité dans Schanzenbach, 2014). De même, dans un récent rapport, l'OCDE conclut que « pourtant, les effets de la variation de la taille des classes sur la performance des élèves ne sont pas étayés par des éléments probants » (OCDE, 2016, p.426). Ces déclarations fortes sont faites sans aucune référence (OCDE) à des données factuelles ou en faisant une référence très sélective (Gladwell) à des données factuelles, sans parler d'une quelconque évaluation minutieuse des données factuelles. On peut souvent en dire autant des partisans de la réduction de la taille des classes.

Notre examen systématique des études empiriques concernant les effets de la taille des classes sur les résultats des élèves en Europe fait apparaître des données factuelles mitigées. Certaines études crédibles relèvent des effets bénéfiques substantiels de classes de plus petite taille, alors que d'autres études tout aussi crédibles constatent des effets plutôt réduits ou pas significativement différents d'un effet nul. Bien que certaines études démontrent que les classes plus petites offrent effectivement des avantages durables, l'utilisation de ces données à des fins de politique normative nécessite généralement une extrapolation à une population différente. De telles extrapolations nécessitent une solide compréhension de la façon dont les caractéristiques des populations concernées, les incitations et les contraintes entrent dans la fonction de production et modulent les effets de la taille des classes. Bien qu'il existe des études consacrées à ces facteurs intermédiaires, les données factuelles ne permettent pas d'obtenir des réponses définitives. Sur la base des données actuelles, il est donc difficile de formuler une recommandation sans réserve sur la façon et le moment d'utiliser des politiques en matière de taille des classes afin d'améliorer les résultats des élèves.

Des recherches supplémentaires sont donc nécessaires pour fournir de meilleures réponses aux questions de savoir si la taille des classes est importante, pourquoi, quand et pour qui. Ces questions doivent être résolues par un plan de recherche et les données doivent être causales. Les expériences sur le terrain sont généralement considérées comme la « méthode de référence ». Bien que la seule expérience randomisée à grande échelle concernant la taille des classes (Projet STAR) ait produit de nombreux éclairages, ils comportent leurs propres risques quant à leur validité interne et externe (Heckman et Smith, 1995). Nous croyons donc qu'il est important de recueillir des données à partir de différents modèles de causalité. Il peut s'agir de modèles non-expérimentaux dérivés de règles de taille de classe maximale ou de modèles expérimentaux alternatifs (Rockoff, 2009).

Enfin, il convient de souligner que la notion d'effet de taille de classe dépend des modes traditionnels d'enseignement et d'apprentissage dans lesquels on peut considérer la taille des classes comme un substitut bien défini et politiquement pertinent des intrants de la fonction de production des écoles. Bien que les écoles d'aujourd'hui soient remarquablement similaires à celles de nos grands-parents, les innovations technologiques et pédagogiques peuvent changer la nature des salles de classe et, de ce fait, la relation entre la taille des classes et les intrants. Les estimations

des effets de la taille des classes sont susceptibles d'être assorties de dates d'expiration lorsque la fonction de production vient à changer.

Références

Gladwell, M. (2013). *David and Goliath: Underdogs, Misfits, and the Art of Battling Giants.* Little, Brown & Company., New York.

Heckman, J. J. et Smith, J. A. (1995). Assessing the case for social experiments. *The Journal of Economic Perspectives*, 9(2):85–110.

OCDE (2016). *Regards sur l'éducation 2016 : Les indicateurs de l'OCDE.* Editions OCDE, Paris.

Rockoff, J. (2009). Field experiments in class size from the early twentieth century. *Journal of Economic Perspectives*, 23:211–230.

Schanzenbach, D. W. (2014). Does class size matter? National Education Policy Center.

Executive Summary (German)

Klassengröße und Schülerleistung in Europa

Verringerungen der Klassengröße sind bei Eltern, Lehrern und Politikern gern gesehen. In den letzten Jahren wurde in der öffentlichen Debatte vermehrt die Meinung zum Ausdruck gebracht, dass die Klassengröße nicht von Bedeutung sei. Der populärwissenschaftliche Autor Gladwell (2013) bemüht beispielsweise die Verringerung der Klassengröße als Beispiel für „etwas, das wir als großen Vorteil sehen, das jedoch eventuell gar keiner ist“ (zit. n. Schanzenbach, 2014). In einem neueren Bericht kommt auch die OECD zu dem Schluss, dass es „insgesamt jedoch [...] kaum eindeutige Belege für die Auswirkungen unterschiedlicher Klassengrößen auf die Leistungen der Schüler“ gibt (OECD, 2016, S. 506). Diese doch sehr deutlichen Aussagen werden bei der OECD gar nicht und bei Gladwell nur durch sehr selektive Verweise auf Belege gestützt, ganz zu schweigen von einer sorgfältigen Bewertung der Fakten. Gleiches gilt in vielen Fällen für die Befürworter von kleineren Klassen.

Unsere systematische Übersicht empirischer Studien zu den Auswirkungen der Klassengröße auf die Schülerleistungen in Europa zeigt ein gemischtes Bild. Einige ernstzunehmende Studien berichten von erheblichen positiven Effekten kleiner Klassengrößen, wohingegen andere, genauso glaubhafte Studien Effekte feststellen, die eher gering oder gar nicht signifikant sind. Auch wenn manche Untersuchungen zeigen, dass kleinere Klassen tatsächlich nachhaltige Vorteile bieten, erfordert die Nutzung dieser Erkenntnisse für die Zwecke einer präskriptiven Politik in der Regel die Übertragbarkeit auf eine andere Population. Solche Extrapolationen erfordern ein solides Verständnis davon, wie bestimmte Merkmale der Population, Anreize und Hindernisse in die Produktionsfunktion hineinwirken und die der Klassengröße zugeschriebenen Effekte beeinflussen. Obwohl einige Studien solche Einflussfaktoren untersuchen, lassen die Ergebnisse keine eindeutigen Schlüsse zu. Basierend auf dem aktuellen Wissensstand ist es deshalb schwierig, klare Empfehlungen auszusprechen, wie und wann Klassengröße als Mittel eingesetzt werden sollte, um den Lernerfolg zu verbessern.

Um besser beantworten zu können ob, warum, wann und für wen Klassengröße von Bedeutung ist, ist daher zusätzliche Forschungsarbeit notwendig. Diese Fragen müssen mithilfe eines Studienaufbaus beantwortet werden, der einen kausalen Zusammenhang darlegt. In der Regel gilt die Feldforschung hier als Goldstandard. „Project STAR“ war ein groß angelegtes randomisiertes Einzelexperiment zur Untersuchung von Klassengrößeneffekten. Obwohl es zu vielen Erkenntnissen führte, sind diese hinsichtlich ihrer internen und externen Validität mit Schwächen behaftet (Heckman und Smith, 1995). Wir sind deshalb der Meinung, dass es wichtig ist, unterschiedliche kausale Forschungsdesigns zu nutzen. Dabei kann es sich um nicht-experimentelle Designs handeln, die sich auf Vorgaben für maximale Klassengröße stützen, oder um alternative Versuchsdesigns (Rockoff, 2009).

Abschließend sei darauf hingewiesen, dass der Begriff des Klassengrößeneffekts in engem Zusammenhang mit traditionellen Lehr- und Lernformen steht, bei denen die Klassengröße eine klar definierte und politikrelevante Variable in der schulischen Produktionsfunktion darstellt. Auch wenn die heutigen Schulen denen unserer Großeltern auffallend ähnlich sind, können technische und pädagogische Innovationen die Art des Klassenunterrichts und damit den Zusammenhang zwischen Klassengröße und Input verändern. Schätzungen von Klassengrößeneffekten sind deshalb wahrscheinlich nur so lange gültig, bis sich die Produktionsfunktion ändert.

Literatur

Gladwell, M. (2013). *David and Goliath: Underdogs, Misfits, and the Art of Battling Giants.* Little, Brown & Company, New York.

Heckman, J. J. und Smith, J. A. (1995). Assessing the case for social experiments. *The Journal of Economic Perspectives*, 9(2):85–110.

OECD (2016). *Bildung auf einen Blick 2016: OECD-Indikatoren*. OECD Publishing, Paris.

Rockoff, J. (2009). Field experiments in class size from the early twentieth century. *Journal of Economic Perspectives*, 23:211–230.

Schanzenbach, D. W. (2014). Does class size matter? National Education Policy Center.

Class size and student outcomes in Europe*

Edwin Leuven Hessel Oosterbeek

Abstract

This paper reviews the empirical evidence about the effects of class size on student outcomes with an emphasis on results for European countries. The key finding is that the results are mixed. Some studies show credible evidence on substantial beneficial effects of smaller class sizes, whereas other equally credible studies find effects that are rather small or not significantly different from zero. Evidence on the factors that may explain these differences is limited.

1 Introduction

Various vocal contributors to the public debate have expressed the opinion that class size does not matter. For example, the popular science author Gladwell (2013) uses class size reduction as an example of a “thing we are convinced is such a big advantage [but] might not be such an advantage at all” (as cited in Schanzenbach, 2014). Likewise, in a recent report the OECD concludes that "overall, evidence of the effect of differences in class size on student performance is weak" (OECD, 2016, p.349). These strong statements are made without (OECD) or with very selective (Gladwell) reference to evidence, let alone a careful assessment of the evidence. But advocates of class size reductions also seem a bit selective in the studies they cite.

The present paper reviews the empirical evidence about the effects of class size on student outcomes with an emphasis on results for European countries, and tries to reconcile apparently contradictory results.

Class size varies considerably across countries and between education levels. Figure 1 shows average class size in 2014 by level of education for countries included in OECD’s

*This version: November 2017. Edwin Leuven is affiliated with the University of Oslo, Oosterbeek is affiliated with the University of Amsterdam. This paper was written as part of the programme of the European Expert Network on Economics of Education (EENEE). Constructive comments from Sandra McNally and Bjorn Öckert are gratefully acknowledged. We thank Marc Gurgand, Ludger Woessmann, Elena del Rey Canteli, Giorgio Brunello, Sandra McNally, Kristof de Witte, Bjorn Öckert, Tuomos Pekkarinen, Pawel Bukowski, Daniel Munich, Rudolf Winter-Ebmer Daniel Horn, Lara Lebedinski, Martina Viarengo, and Ioana Schiopu for their help in our literature search. All omissions are our own.

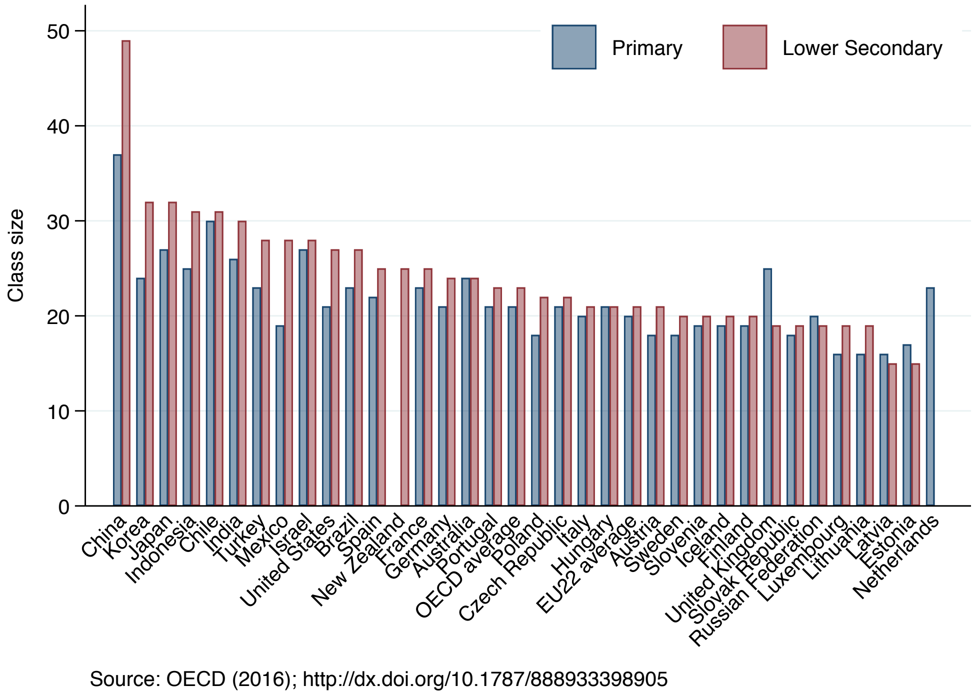


Figure 1: Average class size, by level of education and country (2014)

Education at a Glance 2016 report.[1] This shows average class sizes in lower secondary education as high as 49 in China and as low as 15 in Estonia and Latvia. The range of average class sizes is somewhat smaller in primary education; from 37 in China to 16 in Luxembourg and Lithuania.

Figure 2 shows changes in average class size between 2005 and 2014 by level of education and country. This shows that the average class size not only differs between countries but also within countries over time. In a period of just 10 years, Estonia reduced average class size in primary schools by more than 30 percent, while average class size in lower secondary schools in the Russian Federation went up by more than 30 percent.

With teachers' salaries making up a large share of the costs of education, substantial changes in average class size have a large impact on the education budget. It is therefore not surprising that when public budgets are under pressure, it is sometimes proposed to increase class size (e.g. Chingos, 2012). Whether this is a wise policy in the longer run depends on the effects that increases in class size have on student outcomes.

This paper is structured as follows. Section 2 provides a theoretical framework that

[1] The OECD defines class size as the number of students who are following a common course of study, based on the highest number of common courses (usually compulsory studies), and excluding teaching in subgroups. The calculation is done by dividing the number of students by the number of classes. In contrast, the student-teacher ratio is calculated by dividing the number of full-time equivalent students by the number of full-time equivalent teachers at a given level of education and type of institution.

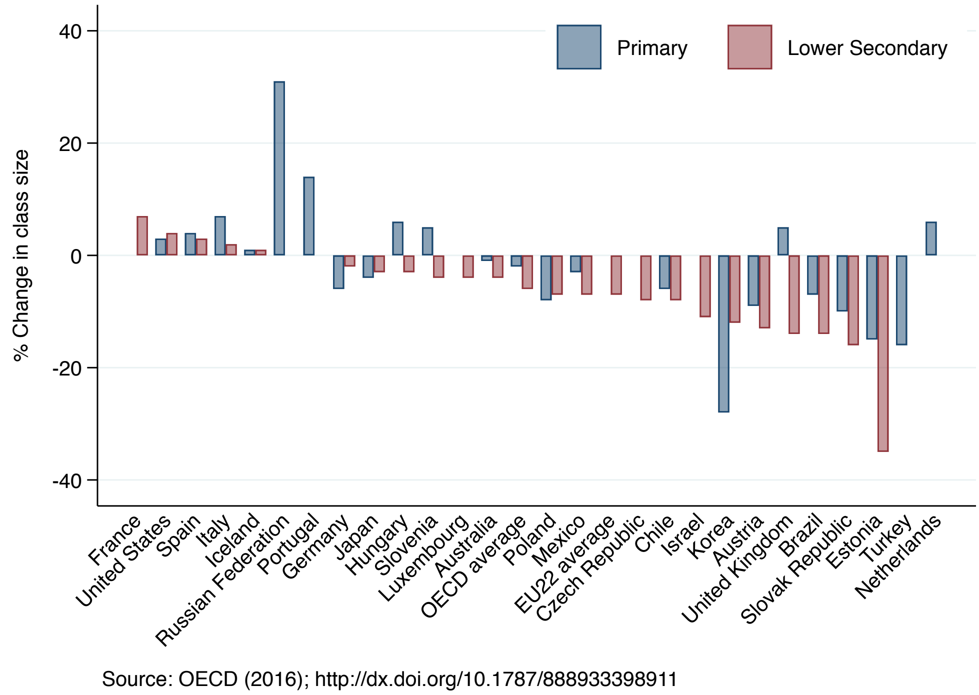


Figure 2: Changes in average class size, by level of education and country (2005-2014)

helps to organise the thinking about the effects of class size. Section 3 discusses the empirical approaches that researchers have used to estimate the causal effects of class size. Section 4 presents and discusses the evidence, where we focus on results for European countries. Section 5 attempts to understand the findings, and section 6 summarises and concludes.

2 Theoretical framework

This section provides a theoretical framework that can be used to organise different insights about the process through which class size may affect student outcomes.

In an important study, Todd and Wolpin (2003) make the distinction between direct and indirect effects of class size. Direct effects are the effects that occur when all other factors that influence student outcomes (i.e. inputs in the education production function) are held constant. This is the effect of a change in class size with the same teacher with the same effort and the same qualifications, the same classroom, the same other resources, the same involvement of parents and the same composition of the class in terms of means and variance of peer characteristics. While other inputs are fixed, this form of variation in class size can still affect how much students learn through various channels (Ehrenberg et al., 2001, p.1): It can change the interaction among students; it can change the amount of disruption in the classroom; it can also change the amount of time the teacher can focus on each individual student (even without a change in the total amounts of time the teacher spends on individual versus group instruction).

Variation in class size can, however, also cause changes in other factors that influence student outcomes. Students can respond by changing their study efforts. Such changes can in principle go in either direction. For example, increased individual attention from the teacher in smaller classes can increase a student's motivation to learn. In that case, the student's response reinforces the positive direct impact that smaller classes may have on student outcomes. It can, however, also be the case that a student responds to the increased individual attention by reducing her own effort because it now requires less student effort to achieve the same learning outcome (e.g. Correa and Gruver, 1987). In that case, the student's response undoes part of the positive direct impact that smaller classes may have on student outcomes.

Parents' responses to changes in class size can also either reinforce or (partially) undo the direct effects of smaller classes. When the student is learning more in school due to the direct effect of smaller classes, the parents may believe that their child is a good student and that further investment in her education are worthwhile. In contrast, when the student is learning a lot in school, the parents may cut back on helping them with homework or other forms of support.[2]

Finally, the reactions of teachers can also go either way. Teachers in small classes can undo part of the direct benefits of small classes by taking it easier, or by spending more time on non-academic activities that students enjoy. On the other hand, if smaller classes are more attractive to teach, schools with small classes may be able to attract better teachers which may reinforce the direct effect of small classes.

A second important theoretical study is Lazear (2001), in which the author develops a disruption model of education production. One implication of this model is that optimal class sizes may be larger for well-behaved students. Lazear's contribution indicates that direct effects of class size may vary between different types of students. This insight naturally extends to indirect effects; not all students, parents and teachers will respond in the same way to variations in class size. The implication of this is that the effects of class size are potentially heterogeneous across students, families and schools.[3]

Todd and Wolpin (2003) refer to the sum of direct and indirect effects as the policy effect, while the direct effect captures the production function parameter. In many cases it is only possible to estimate the total effect of changes in class size (see below). For

[2]Ideas along these lines are formalised in Fredriksson et al. (2016), who pay special attention to differences between low and high income parents.

[3]Albornoz et al. (2017) provide a further formalisation of the interactions between students, teachers, parents and policymakers in the case of heterogenous agents.

an evaluation of a policy it is important to be aware that class size was probably not the only input that changed. Ignoring this may easily lead to the wrong policy conclusions. Assume, for example, that one finds rather large positive effects of smaller classes on student outcomes. It is then important to know whether these effects can be attributed to the change in class size alone, or part of these effects are the result of smaller classes attracting better teachers. (These better teachers may come at the cost of having worse employees in other sectors of the economy.) Or, assume that one finds that smaller classes have no effect on student outcomes. Again, it is then important to know whether this is solely due to the direct effects being equal to zero, or whether positive direct effects are undone by parents no longer spending money on private tutoring. In short, it is not only important to ask whether class size affects student outcomes, but to also ask through which channels.

3 Empirical approaches

Estimation of the effects of class size on student outcomes is a textbook example (Stock and Watson, 2015; Imai, 2017) of the challenges in disentangling causal effects from mere correlations. While in many applications selection bias is likely to go in one specific direction, this is not the case for the effects of class size. Selection bias can lead to an underestimation of the true effect but can also lead to an overestimation of the true effect. The true effect will be overestimated when parents who care a lot about school performance unroll their children in another school with small classes. The finding that students in small classes perform better than students in large classes may in that case be partially attributed to the fact that students in small classes have parents who care a lot about the school performance, and would probably have done better in any event. Alternatively, schools with multiple divisions (classes) per grade level can decide to form one small class with students who need extra attention and larger classes in which easier students are put together. In that scenario it is likely to find that students in small classes do worse than students in larger classes, but this is not caused by the difference in class size but by the difference in characteristics between students assigned to small and large classes. The true effect would then be underestimated.[4]

This section discusses the different approaches that researchers have used to estimate the causal effects of class size. These are: randomised control trials, regression discontinuity design and instrumental variable approaches.

[4]Note that the difference between causal effects and correlations is not the same as the difference between direct effects and indirect effects. Direct and indirect effects are both causal and induced by a change in class size.

Early approaches

Before the end of the 1990s, the standard method to estimate the effects of class size (and other education interventions) was the education production function approach, introduced in the Coleman report (Coleman et al., 1966). In this approach student achievement or the change in student achievement is regressed on a large set of observable variables related to school resources (including class size), teacher characteristics, family characteristics and characteristics of the school community and characteristics of the community where the student lives (Ehrenberg et al., 2001). The conditioning serves two purposes. One is that different conditioning sets will identify different – i.e. more direct or indirect – class-size effects parameters; the second is that conditioning is necessary to address potential omitted variable bias. These two objectives may be mutually incompatible. The key concern with this approach is that there will always be variables omitted from the regression that are potentially correlated with the inputs of interest (class size) and have an impact on student achievement. Such omitted variables will bias the estimate of the class-size effect. In other words: the conditional independence assumption that this approach requires is typically questionable.

Randomised control trials

Randomised control trials (RCTs) are sometimes regarded as the gold standard in evaluation research. In an RCT units are randomly assigned to different treatments (e.g. small class versus regular class). If the randomisation is properly conducted and the number of units is large enough, it can be assumed that the units assigned to the treatment status (small classes) are on average the same as the units assigned to the control status (regular classes). This implies that the treated units would without treatment have had the same average outcome as the average outcome observed for the control units. Likewise, the control units would with treatment have had the same average outcome as the average outcome observed for the treated units. As a consequence, simply subtracting the average outcome of the control units from the average outcome of the treated units gives an unbiased estimate of the average effect of the treatment.

RCTs can differ in the unit that is assigned to treatment. In the case of class sizes, they can vary within schools, and students and teachers are randomly assigned to classes of varying sizes. Alternatively, entire schools can be assigned to different treatments with some schools having only small classes and other schools only having regular classes. The second design requires a much larger sample of schools than the first design (unless students and teachers can be randomly assigned to schools). A disadvantage of the first design, however, is that control units are potentially affected by the treatment status of others. One channel through which this may occur is that parents of students in regular classes are dissatisfied with the assignment of their children and leave the school or demand that their children are reassigned to a small class.

The mere awareness of being studied could lead to different behaviour than would have been observed if subjects would not have been aware of the experiment. When the treated individuals behave differently, this is referred to as Hawthorne effects, while we speak of John Henry effects if the control subjects change their behaviour. Within-school randomisation makes it very salient that an experiment is being conducted, and teachers in experiment schools may change their behaviour in response to their assumption that the results of the experiment may have an impact on future policies (Hoxby, 1998).

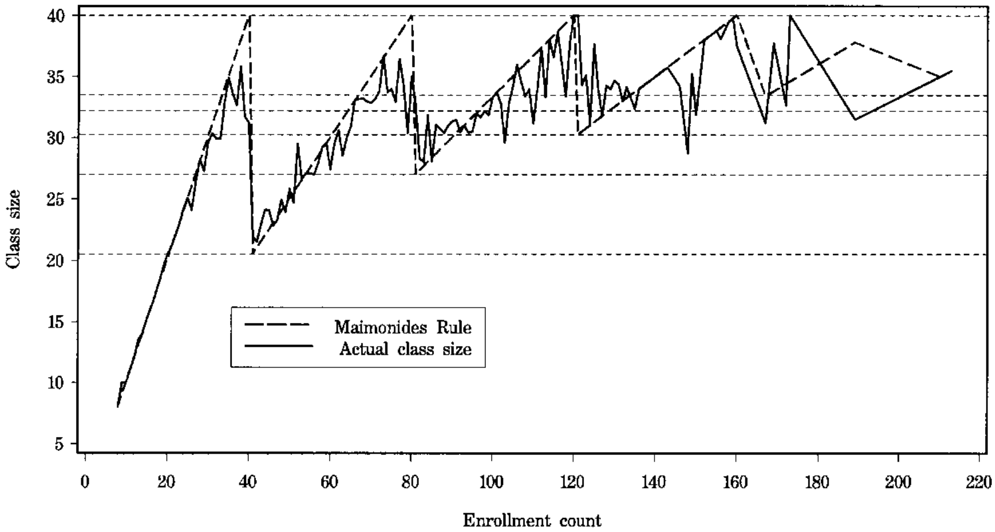
The STAR experiment conducted in 79 primary schools in the 1980s in Tennessee is the only RCT that has taken place to study the effects of class size. Students in grades K-3 (kindergarten through 3rd grade) were – within schools – randomly assigned to small classes, regular classes or regular classes with a teaching aide. The data from this experiment were thoroughly analysed by Krueger (1999); Krueger and Whitmore (2001) and Chetty et al. (2011) have looked at the long-term effects. We describe Krueger's study in more detail in subsection 4.1. This will reveal some of the complications that occur, even in a well-designed large-scale high-cost RCT.

Regression discontinuity (RD) designs

Various countries have rules that stipulate that class size should not exceed a certain number of students. Historically such rules go back at least to the 12th century rabbinic scholar Maimonides who interpreted the Talmud's discussion of class size as implying a maximum class size of 40. These rules were used in primary schools in Israel, and Angrist and Lavy (1999) were the first to use the fact that – to the extent that schools comply with such maximum class-size rules – they are a source of exogenous variation that can be used to estimate the causal effect of class size.

Figure 3 illustrates the working of Maimonides' rule. This graph is taken from Angrist and Lavy (1999).[5] The dashed line shows the relationship between school enrolment in a grade level (5th grade) and predicted average class size if all schools complied with the rule. Whenever grade enrolment reaches a multiple of 40, an extra class is opened and predicted average class size drops. The solid line shows the relationship between school enrolment in 5th grade and the actual average class size in 5th grade. Although the solid line does not fully coincide with the dashed line, it tracks it quite closely; actual average class size also drops discontinuously at multiples of 40 (especially at 40 and 80).

[5]The results of this study are further discussed in subsection 4.1.



Source: Angrist and Lavy (1999, p.541)

Figure 3: Illustration of maximum class-size rule

The RD design compares outcomes of students in schools where grade enrolment is just above (a multiple of) 40 with outcomes of students in schools where grade enrolment is just below (a multiple of) 40. This assumes that potential outcomes are continuous around the thresholds. While it cannot be proven that this assumption holds, it is possible to assess its plausibility. Two tests are commonly reported in studies that use this method. The first test is that there are not many more observations at one side of a kink than at the other side. Urquiola and Verhoogen (2009) use data from Chile showing that in that country the assumption fails; there are many more schools just below (having large classes) than just above the kinks. The second test is to check whether students/schools just below the kinks and students/schools just above the kinks are balanced in their observable characteristics. Also on this test, the data in Chile show signs for non-random assignment to small and large classes with students from poor families being more likely to end up in large classes than students from high-income families. Another attractive feature of the RD design is that the analysis can still allow for a direct relationship between enrolment and student achievement (for example, because better schools attract more students).

The seminal study of Angrist and Lavy (1999) has inspired others to apply the same method to data from other countries, including various European countries. The studies for European countries are summarised in subsection 4.2.

Instrumental variable approaches

The maximum class-size rule discussed in the previous subsection is an example of an instrumental variable (IV) approach. More generally, to identify the causal effect of variable X (class size) on Y (student outcomes) the IV approach uses a third variable Z (the instrumental variable) which i) should be correlated with X but which ii) is unrelated to Y apart from its relationship through X. The first condition is the relevance of the instrument, which can be assessed by regressing X on Z (and the other predetermined variables included in the model). The second condition is the exogeneity of the instrument. Exogeneity has two parts: independence and exclusion. Independence means that the instrument is unrelated to the values of the potential outcomes. Exclusion means that the instrument has no other influence on the outcomes than through X. In the setting of the maximum class-size rule: relevance requires that the maximum class-size rule has an impact on the actual class size, independence requires that the observations around kinks are as good as randomly assigned to the regions just above and just below the kinks, and exclusion requires that the maximum class-size rule only affects student outcomes through its impact on actual class size.

In the class-size literature, other instrumental variables than the kinks resulting from the maximum class-size rule, have been proposed as well. An early example is Akerhielm (1995), who uses the average class size in a grade in a school as instrument for actual class size. This instrument takes account of selection bias that occurs within schools but it does not address biases caused by the school choices of students and their parents. This shortcoming is addressed in the study of Boozer and Rouse (2001) who use average pupil-teacher ratio of schools in a state as the instrumental variable.

A relatively convincing contribution applying the IV approach is Hoxby (2000), who uses year-to-year variation in schools' cohort size. The idea is that "biology causes random variation in the timing and number of births" (p.1246), which then gives rise to random variation in enrolment. With data from a sequence of years, it is possible to correct variation in the size of cohorts for systematic upward or downward trends in grade enrolment in a school. In her study, Hoxby is careful to distinguish between cases where the population variation triggers the opening or closing of a class (through a maximum class-size rule), and where it only causes variation in class size without opening or closing a class. A potential limitation of Hoxby's method is that it induces only limited variation in class size.

4 Findings

This section gives an overview of empirical findings. It has two parts. The first part presents and discusses findings from two influential studies that were conducted outside of Europe. The second part presents and discusses findings from credible studies for European countries.

4.1 Two influential studies

Project STAR (US)

The only large-scale randomised experiment of class size in a developed country is the Tennessee Student/Teacher Achievement Ratio experiment, known as Project STAR.[6] This experiment took place in the 1980s and involved almost 12,000 students in grades K-3 in 79 schools. Around 6,000 students entering kindergarten were within schools randomly assigned to one of three treatments: a small class (13-17 students per teacher), a regular class (22-25 students per teacher) or a regular class with a teaching assistant (22-25 students). Another close to 6,000 students entered one of the 79 participating schools during the experiment in a later grade, and were also randomly assigned to one of the three treatments. Not only students, but also their teachers were randomly assigned. The within-school randomisation requires that schools are large enough to have at least one class in each treatment.[7]

Krueger (1999) provides a careful analysis of the data from this project in which he addresses various threats to the validity of the results. These threats include a high level of attrition from the sample, imperfect compliance, the reassignment of some students after one year between regular classes and regular classes with a teaching assistant, and Hawthorne effects.

The bottom line of Krueger's study is that addressing these issues does not change the main conclusion from previous evaluations of Project STAR: small classes have a positive effect on student outcomes. Students assigned to small classes perform five to seven percentile points (0.20-0.28 SD units) higher than students assigned to regular classes, which had on average about seven more students than small classes. Students assigned to regular classes with a teaching assistant do not perform significantly differently than students assigned to regular classes. The pattern of the findings per grade indicate that the main benefit of being assigned to a small class occurs at the end of the first year. Subsequent years add little to the initial effect (see also Ding and Lehrer, 2010). Krueger suggests that this surprising result can be explained if being in a small class improves school socialisation which then leads to a permanent increase in students' achievement without changing the trajectory (p.529). Furthermore, Krueger reports that the initial effect is larger for boys than for girls, while the opposite holds for the cumulative effect. Initial effects and cumulative effects tend to be larger for students on free lunch programmes and for black students than for students not on free lunch and for white students (p.525). Ding and Lehrer (2011) find that students with higher test scores benefitted more from small

[6] Rockoff (2009) describes smaller scale experiments with class size that have been conducted in the US before WWII.

[7] Of the schools that met this requirement, only one in five participated (Rockoff, 2009, p.227).

classes.

Krueger and Whitmore (2001) and Finn et al. (2005) have used the data of Project STAR to examine the effects of being in a small class on high school graduation and participation in college entry exams. Krueger and Whitmore report that students who were assigned to small classes are 4 percentage points (10%) more likely to take college entry exams than students who were assigned to regular classes. This difference is largest for black students. Chetty et al. (2011) use the data from Project STAR to estimate the effects of small classes on labour market earnings. Their estimates are imprecise and have a negative sign (smaller classes reduce earnings). They attribute the absence of a positive effect of small classes on earnings to the age of the cohort (28) at the moment the analysis was carried out. Class size also affects college attendance so that earnings effects may only materialise when people are in their 30s, especially since college graduates have much steeper earnings profiles than non-college graduates (Chetty et al., 2011, p.1624).

Maimonides (Israel)

Angrist and Lavy (1999) show that the saw-tooth pattern of class size shown in Figure 3 is mirrored by an inverse pattern for student outcomes. When class size drops, achievement goes up and vice versa. This is suggestive of a negative effect of class size on student outcomes. Indeed, Angrist and Lavy report significantly negative effects of class size on students' math and reading scores. A one-student increase in class size in 5th grade, reduces reading scores by 0.024 SD units and math scores by 0.02 SD units. Estimated effects are smaller for students in the 4th and 3rd grades and not always statistically significant.

Angrist and Lavy were not only the first applying the maximum class-size rule to estimate the effects of class size, but were also among the first who applied the RD design. Since the first applications of this method, researchers have improved the method. As was mentioned in section 3, it is now common practice to check for discontinuities in the density of the running variable (enrolment) and to check whether observations at either side of the thresholds are balanced in terms of their observable characteristics. In a recent paper Angrist et al. (2017b) redo the analysis of Angrist and Lavy (1999) and extend it with more recent data. For the more recent cohorts they no longer find significant effects of class size on student outcomes. The re-analysis of the original data confirms that for older cohorts there is a negative class-size effect, but that the evidence is weaker than presented in the original study. In the more recent data they also find that too many schools just pass the thresholds that produces an extra class.

4.2 Findings for Europe

In this subsection we present findings from studies for European countries. For a study to be included in our review it has to satisfy a number of criteria. First, it must use data from a European country. Second, it must at least be under revision for an international peer-reviewed journal. Third, it must explicitly address omitted variable bias through a research design. The last two criteria we use to define credibility/quality. To identify eligible studies we relied on Google Scholar searches on "class size AND achievement" in combination with additional keywords like "experiment", "maximum class size", "Maimonides", "regression discontinuity" and "difference-in-differences". We furthermore tracked down studies from the reference lists in matching articles.

For each country we describe the main features of the credible studies that have been conducted. These descriptions contain: i) the empirical approach and possible issues with it; ii) the grade level; iii) the size of class-size reduction and the length of exposure; and iv) the main findings.

Denmark

Browning and Heinesen (2007) use a maximum class-size rule to estimate the effect of class size in 8th grade (in the period 1985-1992) on years of completed post-secondary education (measured in 2003).[8] The official maximum class size in Danish secondary school was in the relevant period equal to 28. In practice class size was determined by municipalities and they applied smaller maxima than 28. The authors therefore use thresholds at multiples of 24. Results are presented for the full sample (including controls for enrolment and enrolment squared) and for ±5, ±4 and ±3 discontinuity samples, and for specifications with and without a rich set of control variables. The authors' preferred specification indicates that a reduction of class size of 5% (about one student) increases the average length of education by 0.02 years (and the probability to complete secondary education by 0.4 percentage points).[9]

Bingley et al. (2010) extend the study of Browning and Heinesen (2007). They are also interested in the effect of class size in 8th grade (in the period 1981-1990) on the length of post-compulsory education. The authors, however, are worried about the integrity of the maximum class-size rule design. They suspect that parents are able to predict class sizes and that some parents may react to that by opting for a private school or delay school entry by a year. To address this concern, the authors augment the maximum class size rule

[8] In addition to a maximum class-size rule where class size varies with enrolment, school funding in Denmark also has a compensatory component where extra teacher hours are awarded to larger classes. As a result, the authors not only present estimates of the effect of class size on years of completed education but also of the effect of the number of students per teacher hour.

[9] Results of all studies are summarised in Table 1.

approach with a sibling-pair fixed-effects approach. That is, they only use variation in class size generated by the maximum class-size rule which leads two siblings who attend the same school to be exposed to different class sizes. Bingley et al. find a coefficient on log class size of about -0.3 (p.32). With an average class size of around 20, this implies that a one-student increase in class size reduces the length of post-compulsory schooling by 0.015 years, which is rather similar to the estimate of Browning and Heinesen.

A concern with the studies of Browning and Heinesen (2007) and Bingley et al. (2010) is that the estimates of the (log) class-size effects in both studies are sensitive to the inclusion of control variables. For example, the preferred estimate of -0.2 of Browning and Heinesen is based on a specification including control variables. Without control variables, the equivalent estimate is 14% smaller. Furthermore, the outcome "years of post-compulsory education" might not capture the full benefit of smaller classes. The outcome variable also does not differentiate between more years of schooling resulting in a diploma at a higher level and more years of schooling resulting from delays.

Students in Danish lower secondary schools (grades 7-9) start learning a second foreign language besides English. While all schools offer German, only a minority of schools also offer French. In these schools students can choose between German and French. Heinesen (2010) identifies class-size effects by year-to-year variation in the same school in the number of students that unroll in French classes. A one-student increase in class size reduces the average examination mark for French by 0.029 of a standard deviation unit. This estimate changes only little when controls for enrolment and individual characteristics are not included. Various sensitivity tests (e.g. examining the impact of class size for French on other marks) support a causal interpretation of the results. When looking at heterogeneity, Heinesen finds that boys and weaker-performing students are the ones that benefit the most from a reduction in class size for French.

Krassel and Heinesen (2014) estimate the effect of class size on the GPA (grade point average) of students enrolled in the 10th grade of lower secondary education in Denmark. Tenth grade is an optional year offered after compulsory schooling ends. It is chosen by about half of the students, typically the weaker ones and those who do not know yet what type of upper secondary education to pursue. For identification, the authors combine the maximum class-size rule (now abiding to the official maximum of 28) with school fixed effects and a rich set of student characteristics, including GPA in grade 9. The preferred estimate implies that a one-student increase in class size reduces GPA by 0.008 standard deviation units in the distribution of individual GPAs. This estimate is robust to different specifications of the controls for enrolment. The authors also report that students at the two sides of the kinks are not significantly different in terms of their observed characteristics. They further argue that the reported estimate is probably a lower bound on the true effect since the sample is restricted to students who take six advanced exit exams. Other results show that the probability of doing so decreases with class size. Hence, smaller classes may have an overrepresentation of weaker students.

Nandrup (2016) adds to the previous studies by estimating class-size effects for primary education in Denmark. She uses the variation in class size generated by the official maximum class-size rule of 28 to estimate class-size effects in grades 2, 3, 6 and 8 on reading and math scores. Data are for the cohorts starting these grades in 2009, 2010 and 2011. The author's preferred results are based on a specification that restricts the samples to students in classes that are at most four students away from the kinks. All points estimates are negative and larger classes have a negative impact on test scores, but it is small and not always significantly different from zero. The largest effect is found for 6th grade math scores (-0.0149) and the smallest for 8th grade physics/chemistry. There is no clear evidence for heterogeneity of effects by gender, immigrant status or parental earnings.[10]

To summarise: all five studies for Denmark report in the preferred specification negative effects of class size on the outcome of interest.

France

Using data from students in lower secondary school in France, Gary-Bobo and Mahjoub (2013) estimate class-size effects on grade progression. Data cover the period 1989 to 1994. In their preferred specification the authors combine a maximum class-size rule of 30 students with enrolment at the grade level (linear and squared) as instrumental variables, while controlling for school enrolment (linear and squared). This assumes that grade enrolment is random given school enrolment and other controls. This is similar in spirit to the population variation exploited by Hoxby (2000). They find that a one-student increase in class size reduces the probability to be promoted to the next grade by 1.4 percentage points for 6th graders and by 2 percentage points for 7th graders. For students in the 8th and 9th grades, the estimated effects are not significantly different from zero. The distribution of enrolment reveals no indication of bunching at either side of the kinks, and characteristics and test scores at entry are balanced.

Bressoux et al. (2009) analyse data from students in 3rd grade in France to estimate effects of class size on reading and math scores. Data are from a relatively small sample from the year 1991 and are restricted to classes taught by a novice teacher. The data include scores on tests taken before and at the end of the school year. Using the before data, the authors show that class size does not "predict" test scores at baseline. From this they conclude that students are as good as randomly assigned to classes of different sizes.

[10]The author does not report how exposure to smaller class sizes for multiple years affects students outcomes.

The authors find that a one-student increase in class size in 3rd grade reduces reading scores by 0.026 SD units and math scores by 0.027 SD units. The causal interpretation of these results is supported by results from analyses based on a maximum class-size rule (at 30). Estimated effects are rather similar for students of different achievement levels. Effects are larger for students in disadvantaged schools.

Italy

Angrist et al. (2017a) analyse data from students in the 2nd and the 5th grade in Italy to estimate class-size effects on language and math outcomes. Data are for the years 2009 to 2011. When they use a maximum class-size rule at 25 (27), they find estimates equal to -0.0061 (s.e. 0.0020) for math and to -0.0041 (s.e. 0.0016) for language. The authors argue, however, that these estimates do not reflect a causal effect of class size on students' learning. Instead, they argue that these effects are due to the effect of class size on score manipulation (by teachers). In smaller classes, teachers are less likely to share the grading of tests with others. Moreover, in small classes it requires less effort to increase the average score of a class through manipulation than in large classes. To investigate the importance of manipulation of test scores, Angrist et al. extend their analysis using information about class-level measures of manipulation and the random assignment of external monitors (see also Bertoni et al., 2013). When they estimate a model in which test scores depend on class size and manipulation, the estimated class-size effects are small and not significantly different from zero. When the analysis is split by different regions in Italy, it turns out that the initial class size, but also the importance of manipulation is much larger in the south of Italy than in the rest of the country.

Norway

Bonesrønning (2003) uses a maximum class-size rule at 30 students to estimate the effects of class size on math outcomes of students in lower secondary education in Norway. Data were collected from 1998 to 2000. Instead of including controls for grade enrolment, Bonesrønning controls for school size. A one-student increase in class size reduces math score by 0.01 SD unit. This estimate is significant at the 10%-level. Further results show that the class-size effect is larger for boys than for girls, and also larger for students whose parents are little education than for students whose parents are highly educated. Bonesrønning also calls attention to the fact that smaller classes may induce students to reduce their efforts (see also Correa and Gruver, 1987).

Bonesrønning (2003) analyses data that he collected himself with the consequence that the dataset is relatively small (the main result is based on 931 students). Moreover, to achieve sufficient variation in the treatment variable, he collected data that are not representative for Norway. Leuven et al. (2008) use register data from all Norwegian students participating in the final math or language exams of lower secondary education in 2001 or 2002. To estimate the effects of class size they employ the maximum class size rule of 30 students. In this study the treatment is defined as the average class size to which a student is exposed during her three years in lower secondary school.[11] The instrumental variable, however, is based on enrolment in grade 8. The analysis controls for the running variable by including a third-order polynomial of enrolment in grade 8. The results of Leuven et al. show effects for both outcomes variables that are close to zero. The estimates are quite precise so that even modest effect sizes are highly unlikely. In addition to results from a maximum class-size rule, Leuven et al. also present results from analyses exploiting population variation. The findings from these approaches confirm the findings based on the maximum class-size rule.

For a long time there were no comparable data on student performance in Norwegian primary schools. National tests in math and reading were first introduced for 4th graders in the spring of 2004. Iversen and Bonesrønning (2013) have taken advantage of this by relating students' math scores on this test to the class size they have been exposed to in grades 1 to 3. For identification they use the maximum class-size rule of 28 students which is rather strictly applied for primary schools in Norway. Information is available for more than 50,000 students. The results show statistically significant but rather small, negative class size effects on math scores. A one-student increase in class size during grades 1-3 reduces performance on the math test by close to 0.004 SD unit. Effects are larger for students with less-educated parents and for children with divorced parents. The authors find no evidence for bunching around the kinks, nor for differences in parental education or earnings between students in large and small classes. Furthermore, they find no indications that principals favour large classes with extra resources (assistants, special resources).

Poland

Jakubowski and Sakowski (2006) use two approaches to estimate the effect of (log) class size on students' scores on the final exam in primary school (6th grade) in Poland. The first approach instruments actual class size with average class size in the school. This addresses selective assignment within schools. The second approach uses a maximum class0size rule at 29 students. This is not a formal rule in Poland but according to the authors it seems that most schools in Poland add new classes when the average class size

[11] This approach makes the intensity of the treatment very explicit. Studies that estimate the effect of class size during a specific grade can generally not rule out that class size in that grade is correlated with class sizes in earlier grades and that the estimated class-size effect picks up the effects of class sizes in earlier grades.

is about 28 or 29 (p.209).[12] Data are for 2002-2004. Both methods produce fairly similar estimates. Based on the maximum class-size rule approach, doubling the class size (on average a 21-student increase) reduces test scores by 0.25 of a standard deviation. This translates into a 0.012 SD reduction per one-student increase in class size.

Sweden

Lindahl (2005) uses a novel difference-in-differences approach to estimate class-size effects. He collected data on math scores at three moments: at the end of 5th grade, at the beginning of 6th grade and at the end of 6th grade. With the latter two measures he can estimate a traditional value added equation. This gives an estimate of the effect of class size on learning gains in 6th grade. To correct for the selection bias in this estimate, Lindahl uses the difference between the math scores at the beginning of 6th grade and at the end of 5th grade. This difference measures what would have happened to math scores in the absence of exposure to different class sizes. A crucial assumption is that unobserved non-school inputs affect math scores in the same way during the school year and during the summer holiday. Applying this method to a relatively small sample of 556 students in 38 classes in 16 schools in the Stockholm area, Lindahl finds that a one-student increase in class size reduces math scores by 0.03 SD units.[13]

Fredriksson et al. (2013) use data from samples of the birth cohorts from the years 1967, 1972, 1977 and 1982 in Sweden to estimate effects of class size during the three years in upper primary school (grades 4 to 6). For identification they exploit a maximum class-size rule at 30 students. The treatment variable is the average class size to which a student is exposed during her years in grades 4 to 6. The source of exogenous variation is grade 4 enrolment in the school.[14] In the preferred specification, the authors find that a-one student increase in class size during three years reduces cognitive ability by 0.033 SD units.[15] In addition to examining the effects of class size on cognitive scores immediately after exposure to the treatment, this study also investigates the effects of class size on

[12] Jakubowski and Sakowski (2006) do not present much evidence in support of the maximum class-size rule at 29. They do not show the distribution of enrolment (to detect bunching) nor do they show that students/schools just above and just below thresholds are balanced in observed characteristics. They also do not report the first stage estimates of the impact of the maximum class-size rule on actual class size.

[13] Using another specification Lindahl finds an estimated effect that equals around 0.4 times this estimate. He argues that this smaller estimate can be regarded as a lower bound (p.391).

[14] Fredriksson et al. (2013) have to restrict their analysis to schools located in districts that only have a single school. The reason is that for schools from other districts they find evidence of bunching: there are more schools just below the thresholds than just above the thresholds. This seems to be caused by districts redrawing the borders of schools' catchment areas in order to avoid opening an extra class.

[15] Using the same data in a study that focuses on parental responses, Fredriksson et al. (2016) report an impact of 0.044 SD units on academic achievement. Academic achievement measures performance on achievement tests in Swedish and mathematics whereas cognitive ability is based on traditional "IQ-type" tests with logical and verbal elements.

non-cognitive outcomes and on long-term outcomes. The results show that an increase in class size also reduces non-cognitive outcomes and earnings later in life. Another feature of this study is that it carefully considers whether passing a maximum class-size threshold influences other factors that may affect student outcomes. For example, passing the first threshold not only reduces class size, but also implies that the number of classes goes from one to two. With two classes instead of one, it is in principle possible to group students into homogenous groups (tracking), which may have an independent effect on student outcomes. The authors find no evidence for the formation of homogenous groups (or for effects on other factors that may influence student outcomes).

UK

Dearden et al. (2002) and Dustmann et al. (2003) use a conditional independence assumption to estimate the relationship between school quality and outcomes using UK data. The association between the pupil/teacher ratio and outcomes (attainment/exam success) is typically small and insignificant.

Denny and Oppedisano (2013) use the PISA data for the UK to estimate class-size effects on math scores. Their identification method is based on the within-school variation in cohort sizes in grades 10 and 11. Their results indicate that a one-student increase in class size increases students' math scores by 0.06 SD units. This sizeable effect estimate is statistically significant.

Table 1 summarises the results from the European studies. Of the 30 estimates in this table, 15 are significantly negative, seven are insignificantly negative, seven are insignificantly positive and one significantly positive. On balance the results from separate country studies thus lend support the hypothesis that smaller classes have a positive effect on student outcomes. Although it should be noted that the size of the estimated effects are often small.[16]

Multiple countries

A few studies have used international comparative datasets to inquire into the effects of class size for multiple countries. The advantages of this approach are that outcomes are measured in the same way for different countries and that the same identifying method can be applied to the data from different countries. The disadvantages are that the number of observations per country are sometimes small and that the identifying method is not necessarily the most convincing one that can be applied for specific countries or that the chosen method can not be applied to the data for certain countries.

[16]We are currently working on a meta-analysis in which we will relate effect sizes to i) characteristics of the students; ii) the variation in class size; and iii) to their precision.

Table 1: Summary of European class size studies

Study	Cnt	Grade	Exposure	Outcome	Effect	SE	Method
Browning & Heinesen (2007)	DNK	8	1 yr	yrs of schooling	-0.020	0.0094	MCSR
Bingley et al (2010)	DNK	8	1 yr	yrs of schooling	-0.015	0.005***	MCSR
Heinesen (2010)	DNK	7-9	3 yr	exam score French	-0.029	0.0074***	PV+SFE
Krassel & Heinesen (2014)	DNK	10	1 yr	GPA	-0.008	0.0035**	MCSR+SFE
Nandrup (2016)	DNK	2	1 yr	reading score	-0.0101	0.0036***	MCSR
Nandrup (2016)	DNK	3	1 yr	math score	-0.0059	0.0046	MCSR
Nandrup (2016)	DNK	6	1 yr	reading score	-0.0125	0.0037***	MCSR
Nandrup (2016)	DNK	6	1 yr	math score	-0.0149	0.0041***	MCSR
Nandrup (2016)	DNK	8	1 yr	reading score	-0.0078	0.0067	MCSR
Nandrup (2016)	DNK	8	1 yr	physics/chemistry score	-0.0014	0.0048	MCSR
Gary-Bobo & Mahjoub (2013)	FRA	6	1 yr	grade progression	-0.0137	0.0044***	MCSR+PV
Gary-Bobo & Mahjoub (2013)	FRA	7	1 yr	grade progression	-0.0203	0.0060***	MCSR+PV
Gary-Bobo & Mahjoub (2013)	FRA	8	1 yr	grade progression	0.0059	0.0041	MCSR+PV
Gary-Bobo & Mahjoub (2013)	FRA	9	1 yr	grade progression	0.0041	0.0052	MCSR+PV
Bressoux et al (2009)	FRA	3	1 yr	reading score	-0.026	0.009***	QE
Bressoux et al (2009)	FRA	3	1 yr	math score	-0.027	0.010***	QE
Angrist et al (2017)	ITA	2/5	1 yr	math score	0.00075	0.0023	MCSR
Angrist et al (2017)	ITA	2/5	1 yr	language score	0.00121	0.00173	MCSR
Bonesrønning (2003)	NOR	7-9	1 yr	math score	-0.0094	0.0063	MCSR
Leuven et al (2008)	NOR	7-9	3 yr	math score	0.002	0.006	MCSR
Leuven et al (2008)	NOR	7-9	3 yr	language score	0.008	0.004	MCSR
Iversen and Bonesrønning (2003)	NOR	1-3	3 yr	math score	-0.00374	0.00212*	MCSR
Jakubowski & Sakowski (2006)	POL	6	1 yr	test score	-0.012	0.0036***	MCSR
Lindahl (2005)	SWE	6	1 yr	math score	-0.03	0.011***	DD
Fredriksson et al (2013)	SWE	4-6	3 yr	cognitive score	-0.033	0.0146**	MCSR
Fredriksson et al (2016)	SWE	4-6	3 yr	academic achievement	-0.044	0.0122***	MCSR
Dearden et al (2002)	GBR (m)	1-6	6 yr	attainment	-0.002	0.0025	CIA
Dearden et al (2002)	GBR (f)	1-6	6 yr	attainment	0.002	0.0024	CIA
Dustmann et al (2003)	GBR	1-6	6 yr	exam success	-0.0099	0.033	CIA
Denny and Oppedisano (2013)	GBR	10/11	1 yr	math score	0.0655	0.0331**	PV

Note: Column Cnt (country): DNK=Denmark; FRA=France; ITA=Italy; NOR=Norway; POL=Poland; GBR=United Kingdom. Column Effect: effect sizes are scaled as effects of a one-student increase in class size. Hence, negative (positive) effects mean that a reduction in class size increases (decreases) student outcomes. When the outcome variable is a score, effects are expressed in standard deviation units. Otherwise in years or percentage points. Column Method MCSR=maximum class size rule; PV=population variation; SFE=school fixed effects; QE=quasi-experiment; DD=difference-in-differences; CIA=conditional independence assumption. * $p < 0.10$. ** $p < 0.05$. *** $p < 0.01$.

Wößmann (2005) uses data from the first method Third International Mathematics and Science Study (TIMSS) to estimate the effect of class size for 13-year-old students on math and science scores. He uses two different methods. The first method makes use of the fact that 13-year-old students are divided over two different grades (7th and 8th grade typically).[17] This feature allows him to use variation in class size between grades in the same schools. To address selection within grade levels, he furthermore instruments actual class size by the school's average class size at the grade level. This approach is similar in spirit to the population variation approach of Hoxby (2000). But because Wößmann has only used data from two consecutive cohorts, he cannot correct his estimates for trends in school enrolment. Wößmann and West (2006) inquire whether there are systematic within-school differences in observable characteristics between students in grades with larger and smaller average classes, and find no evidence for such differences.[18] In addition to the population variation method, Wößmann (2005) also presents results from analyses where identification is based on maximum class-size rules. This method can be applied to 10 countries where the relationship between actual class size and some maximum class size rule is detectable in the data. He uses maximum class sizes that are sometimes different from the official maximum class sizes pertaining to countries.

Results from the studies by Wößmann (2005) and Wößmann and West (2006) are reported in the first three columns of Table 2. All estimates have been converted to the effects of one-student increases in class size expressed in standard deviation units. The estimates in column (1) show little support for negative effects of class size on student outcomes. Seven out of 15 estimates have a negative sign, and only one of these is significantly different from zero. Some of the coefficients, however, come with relatively large standard errors meaning that reasonably sized negative effects cannot be ruled out. Seven out of the 10 estimates reported in column (2) have a negative sign, with those for Iceland and Spain being significantly different from zero. And out of the nine effects reported in column (3), five are negative, with those for France and (again) Iceland being significantly different from zero.

Shen and Konstantopoulos (2017) use data from the Progress in International Reading Literacy Study (PIRLS) to estimate class-size effects on reading outcomes of 4th graders

[17] See also Wößmann and West (2006), who use the same method and data but restrict the analysis to six countries that have sufficiently strong first stages, and drop any observations of schools with missing data on actual or average class size.

[18] Of the 38 countries included in TIMSS, Wößmann and West (2006) can only present estimates of class-size effects for 11 of these countries. Twenty countries are excluded because they had fewer than 50 schools with the required data, that is: test results from both 7th and 8th graders, information on actual class sizes and average class size per grade for both 7th and 8th grade, and no imputed data for the variables of interest. Another seven countries were excluded because the first-stage regression of actual class size on grade average class size returned an F-statistic smaller than 10. Wößmann (2005) is less strict in the selection of countries and also reports results for which the first stage F-statistics are below 10.

Table 2: Results from multi-country studies

	W (2005) (1)		W (2005) (2)		WW (2006) (3)		SK (2017) (4)		LK (2016) (5)		AK (2012) (6)	
Country	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE
Austria	0.1268	(0.1425)							-0.0290	(0.0203)		
Belgium, D	0.0365	(0.0451)									0.0090	(0.0200)
Belgium, F	0.0126	(0.0270)			0.0097	(0.0119)						
Bulgaria							0.0039	(0.0135)			-0.0131***	(0.0048)
Cyprus											0.0010	(0.0161)
Czech Rep					0.0313	(0.0264)			0.0034	(0.0165)		
Denmark	-0.0161	(0.0164)	-0.0053	(0.0092)								
England											0.0162	(0.0103)
Estonia											0.0002	(0.0020)
France	-0.0409	(0.0272)	-0.0030	(0.0087)	-0.0347**	(0.0174)						
Germany	-0.0018	(0.0154)	0.0075	(0.0110)			0.0225**	(0.0107)	0.0214	(0.0203)		
Greece	-0.0021	(0.0194)	0.0023	(0.0030)	-0.0171	(0.0111)						
Hungary							-0.0003	(0.0182)	0.0050	(0.0166)	-0.0096***	(0.0025)
Iceland	-0.0334***	(0.0111)	-0.0124***	(0.0034)	-0.0363***	(0.0119)						
Ireland	-0.0221	(0.0794)	-0.0113	(0.0110)								
Italy							-0.0507	(0.0635)				
Latvia											0.0017	0.0010
Lithuania							-0.0083	(0.0135)	-0.0169	(0.0173)	-0.0056*	0.0031
Macedonia											0.0066	0.0059
Moldavia											0.0120	0.0086
Netherlands	0.0414***	(0.0142)									-0.0076**	(0.0038)
Norway	0.0944	(0.1333)	-0.0021	(0.0032)								
Portugal	0.0195	(0.0175)			0.0242**	(0.0111)			-0.0553	(0.0389)		
Romania					-0.0034	(0.0191)	-0.0560***	(0.0173)	-0.0459**	(0.0216)	-0.0021	(0.0029)
Scotland	0.0031	(0.0136)									0.0253***	(0.0097)
Serbia											-0.0029**	(0.0013)
Slovakia							-0.0131	(0.0084)	-0.0553***	(0.0198)	0.0177***	(0.0056)
Slovenia					0.0141	(0.0164)	-0.0015	(0.0118)	-0.0273	(0.0193)	0.0055	(0.0083)
Spain	-0.0058	(0.0423)	-0.0096*	(0.0053)	-0.0042	(0.0116)						
Sweden			-0.0059	(0.0037)							-0.0003	(0.0060)
Switzerland	0.1810	(1.2478)	0.0055**	(0.0024)								
Data	TIMSS		TIMSS		TIMSS		PIRLS		TIMSS		TIMSS	
Grade	7/8		7/8		7/8		4		4		4	
Outcome	math		math				reading		math		math/science	
Method	PV		MCSR		PV		MCSR		MCSR		PFE	

Note: W (2005)=Wößmann (2005); WW (2006)=Wößmann and West (2006); SK (2017)=Shen and Konstantopoulos (2017); LK (2016)=Li and Konstantopoulos (2016); AK=Altinok and Kingdon (2012). * $p < 0.10$. ** $p < 0.05$. *** $p < 0.01$. MCSR=maximum class size rule; PV=population variation; PFE=pupil fixed effects.

in various countries. Following Wößmann (2005), they also use information about maximum class-size rules. Their study uses data from eight European countries that participated in PIRLS in all three cycles (2001, 2006 and 2011) and had rules about maximum class size for 4th grade. Results are reported in column (4). Six out of eight estimated effects have a negative sign, but only one of these (and one of the two positive estimates) is significantly different from zero. Li and Konstantopoulos (2016) also use maximum class-size rules to estimate class-size effects on math outcomes of 4th graders. Their analysis is based on the TIMSS data and includes nine countries for which the first stage is strong enough. The results in column (5) show that six out of nine estimated effects have a negative sign, but only for Romania and Slovakia are these significantly different from zero.

Altinok and Kingdon (2012) use the TIMSS data to estimate class-size effects on math and science scores of 8th graders. For identification they use within-student variation in class size across subjects. This approach assumes that students are not selected into classes of different sizes across subjects on the basis of their potential outcomes for these subjects. It also requires that teacher quality (and other class unobservables) does not systematically vary between classes of different sizes. Results for European countries are reported in column (6) of Table 2. Out of the 17 estimates that are reported, only seven have a negative sign. Of the seven coefficients that are significantly different from zero, five have a negative sign.

Long-term effects

Some recent studies have looked beyond the short-term effects of class size on student outcomes, and have estimated effects of class size on labour market outcomes. Fredriksson et al. (2013) merged the Swedish data from their cohort studies to administrative data that contain information about earnings observed between ages 27 and 42. They find that a one-student increase in class size decreases earnings by 0.6%, hence a reduction in class size from 25 to 20 increases earnings on average by 3%.[19] Leuven and Løken (2017) analyse much larger Norwegian samples and report precisely estimated zero effects of class size on earnings (see also Falch et al., 2017). The differences in long-term effects of class size in the two neighbouring countries Sweden and Norway are consistent with the differences in short-term effects discussed above.

[19] Fredriksson et al. also present results showing that class size has a negative effect on short-term cognitive outcomes.

5 Understanding the findings

The presentation of results so far relates class size to student outcomes without examining how class size affects outcomes. In section 2 we described various ways in which students, teachers and parents can respond to class size. These responses translate into indirect effects and may either go in the same direction or in the opposite direction as the direct effects. In this section we summarise what we know about responses of students, teachers and parents.

Students

Using the same sample and identification as in their previous study, Fredriksson et al. (2014) and Fredriksson et al. (2016) analyse the effects of variation in class size on a number of variables that shed light on responses of students, teachers and parents, as well as on changes in classroom environment. Based on information from surveys, they construct variables measuring i) whether students give up when s/he is assigned a difficult task, ii) absenteeism and iii) hours spent on doing homework. Their results indicate that there are no differences in these three variables between students in large classes and students in small classes.

Furthermore they find that students in large classes are less likely to think it's easy to understand the lessons when the teacher presents them in front of the class. This suggests that (in their particular setting) group instruction possibly becomes less effective when class size increases. On the other hand, Fredriksson et al. (2014) find no significant effects of class size on the tendency of students to think about other things while in school (distraction) or to ask the teacher for help if they do not understand the intruction. All these estimates are based on the students' perceptions.

Teachers

Betts and Shkolnik (1999) use survey data from more than 2,000 math classes in grades 7 to 11 in the US to inquire whether math teachers change their teaching behaviour depending on class size. The data include multiple observations per teacher so that the authors can include teacher fixed effects in their analysis. They find that when class size is reduced teachers spend less of their time on group instruction and more on individual instruction. The magnitudes of the changes in teachers' time allocation, however, are modest. Teachers in smaller classes do not spend more time on new material nor do they finish more of the assigned textbook. But teachers in smaller classes spend less time on student discipline and on routine administration.

Fredriksson et al. (2014) and Fredriksson et al. (2016) find that teachers in large classes attach less importance to students taking responsibility for their learning than teachers in small classes. Teachers in large and small classes do not differ in the importance they attach to exams and to homework for students' learning.

Schools with smaller classes may attract different teachers than schools with large classes. Smaller classes may be attractive for all teachers enabling such schools to choose from a larger pool of applicants and to pick the best teachers. It can also be that smaller classes are more attractive for specific types of teachers. Finally, smaller classes imply – other things being equal – that a school has to hire more teachers, which may have an impact on their average quality. Fredriksson et al. (2013) present estimates of the impact of being above a maximum class size kink on teacher qualifications and find that teachers in schools above a kink have somewhat lower educational qualifications than teachers in schools below a kink. To the extent that teacher qualifications improve student outcomes, this means that students in schools above a kink (smaller classes) perform somewhat worse.[20] Using data from primary school teachers in the Netherlands, Bonhomme et al. (2016) show that these teachers have a preference for schools where the pupil-teacher ratio is lower. For a 4-pupil increase in the pupil-teacher ratio, teachers need to be compensated by a reduction in the share of disadvantaged minority students by 10 percentage points.

Parents

Bonesrønning (2004) uses a maximum class-size rule in Norwegian lower secondary education to estimate parental responses to changes in class size. He finds that parental effort (measured as a composite of homework assistance, monitoring and subject-related discussions) declines when class size increases.

Fredriksson et al. (2016) also investigate the effect of class size on parental effort (help with homework). Their results are the opposite of Bonesrønning's findings. According to their findings, a one-student increase in class size increases the probability that a child receives parental help with homework by almost one percentage point. As mentioned before, this effect is not driven by students in larger classes spending more time on homework. Not surprisingly, this effect on help with homework is entirely due to parents with an above-median income since they are more capable of helping (although they may also face higher opportunity costs). Furthermore, Fredriksson et al. examined whether students in large classes are more likely to move to another school district, and find that they do. Every one-student increase in class size increases the probability to move to another

[20]Using year-to-year variation in class size due to California's class-size reduction program, Jepsen and Rivkin (2009) find that the negative effect of class size on math and reading outcomes was partially and initially offset by increases in the share of inexperienced and not fully certified teachers.

district by slightly more than one percentage point. Moving to another school district leads on average to a reduction in class size of seven students. The effect on moving to another district (and the resulting reduction in class size) is larger among parents with below-median incomes, although responses of parents with above-median incomes are in the same direction.

Iversen and Bonesrønning (2013) examine whether interactions of the dummy for being above the kink and parents' education and income influence student outcomes and conclude that they cannot reject this hypothesis. This suggets that the influence of parents' education on student outcomes is the same in small classes and in large classes.

6 Summary and conclusion

Our review of the empirical studies on the effects of class size on student outcomes shows mixed evidence. Some studies show credible evidence on substantial beneficial effects of smaller class sizes, whereas other equally credible studies find effects that are rather small or not significantly different from zero. Hence, the answer to the question whether class size matters is the typical economics answers to many questions: it depends.

So while the existing studies in some cases show that smaller classes offered sustained benefits, it is not clear how to use this evidence for prescriptive policy purposes. The latter requires extrapolation and therefore a solid understanding of how population characteristics, incentives and constraints enter the production function and mediate class-size effects. While there are some studies that investigate such mediating factors, the evidence falls short in providing definitive answers. Moreover, knowing that the effects of class size depend for example on parents' response, is not sufficient for the purposes formulating effective policy. The next question to ask then is which factors determine how parents respond. Is it solely their level of education or does it also depend on other constraints and incentives they face?

The desirability of class-size reduction does not only depend on its effects on (long-term) student outcomes, but also on its costs. The costs of class-size reduction are largely determined by the costs of teachers. When the supply of teachers is low, for example due to a high demand for skilled workers in other sectors, it may not be a good moment to introduce a reduction of class size. At the same time, the results of Bonhomme et al. (2016) suggest that class-size reduction makes the teaching profession more attractive and may thereby lead to an increase in the supply of teachers. A related issue is that the potential benefits of small classes may be attenuated by the lower quality of incoming teachers (Jepsen and Rivkin, 2009; Dieterle, 2015).

Additional research is necessary to provide better answers to the questions whether, why, when and for whom class size matters. While the single large-scale randomised experiment regarding class size (Project STAR) has produced many insights, we also recognise the shortcomings of this study. Non-compliance and attrition were high and Hawthorne effects cannot be excluded. It is doubtful whether such complications can be avoided in new large-scale experiments.

An example in place is Norway, where in 2015 the government decided to spend NOK 450 million (ca. 0.12% of GDP) to fund research on how an extra teacher in grades 1 to 4 affects student achievement. Two different projects will allocate extra resources for teachers to study how this affects student achievement. The first project – which focuses on mathematics performance – will randomise an extra math teacher across 150 schools in 9 large municipalities. The second project focusses on reading performance and will assign an extra teacher to a class within 150 participating schools in 30 different municipalities. While these experiments are sure to provide valuable new insights, they share the same limitations with the full-time teaching aide treatment arm in Project STAR.

Studies that are based on maximum class-size rules also have limitations. In some cases there are clear indications of manipulation of school enrolment (Angrist et al., 2017b; Urquiola and Verhoogen, 2009), in many other cases it is unknown whether opening an extra class affected other inputs.

Rockoff (2009) advises experimenting with alternative research designs. As examples he suggests that primary schools with enough students enrolling in the same grade could organise classes of different sizes (16 versus 28), or that secondary schools could vary class size across subjects. To avoid opposition from teachers, Rockoff suggests setting up a multi-year experiment and rotating teachers across large and small classrooms. Such new research efforts could not only collect information on student outcomes, but also on how teachers teach in classes of different sizes, how the classroom environment varies with class size, and how students, teachers and parents change their expectations and behaviours.

Finally, the notion of a class-size effect depends on traditional modes of teaching and learning where we can think of class size as a well-defined and policy-relevant proxy for inputs in schools' production function. While today's schools are remarkably similar to the schools of our grandparents, technology and pedagogical innovations may change the nature of classrooms and thereby the relationship between class size and inputs. Estimates of class-size effects are likely to have expiration dates when the production function is changing,

References

Akerhielm, K. (1995). Does class size matter? *Economics of Education Review*, 14:229–241.

Albornoz, F., Berlinski, S., and Cabrales, A. (2017). Motivation, resources and the organization of the school system. *Journal of the European Economic Association*, page Forthcoming.

Altinok, N. and Kingdon, G. (2012). New evidence on class size effects: A pupil fixed effects approach. *Oxford Bulletin of Economics and Statistics*, 74(2):203–234.

Angrist, J., Battistin, E., and Vuri, D. (2017a). In a small moment: Class size and moral hazard in Italian primary schools. *American Economic Journal: Applied Economics*, forthcoming.

Angrist, J. D. and Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114(2):533–575.

Angrist, J. D., Lavy, V., Leder-Luis, J., and Shany, A. (2017b). Maimonides rule redux. National Bureau of Economic Research, Working Paper 23486.

Bertoni, M., Brunello, G., and Rocco, L. (2013). When the cat is near, the mice won't play: The effect of external examiners in italian schools. *Journal of Public Economics*, 104:65–77.

Betts, J. R. and Shkolnik, J. L. (1999). The behavioral effects of variations in class size: The case of math teachers. *Educational Evaluation and Policy Analysis*, 21(2):193–213.

Bingley, P., Jensen, V. M., and Walker, I. (2010). The effect of class size on education and earnings: Evidence from Denmark. Unpublished Working Paper.

Bonesrønning, H. (2003). Class size effects on student achievement in Norway: Patterns and explanations. *Southern Economic Journal*, 69(4):952–965.

Bonesrønning, H. (2004). The determinants of parental effort in education production: Do parents respond to changes in class size? *Economics of Education Review*, 23:1–9.

Bonhomme, S., Jolivet, G., and Leuven, E. (2016). School characteristics and labor turnover: Assessing the role of preferences and opportunities. *Economic Journal*, 126(594):1342–1371.

Boozer, M. and Rouse, C. (2001). Intraschool variation in class size: Patterns and implications. *Journal of Urban Economics*, 50(1):163–189.

Bressoux, P., Kramarz, F., and Prost, C. (2009). Teachers training, class size and students' outcomes: Learning from administrative forecasting mistakes. *The Economic Journal*, 119(536):C540–C561.

Browning, M. and Heinesen, E. (2007). Class size, teacher hours and educational attainment. *Scandinavian Journal of Economics*, 109(2):415–438.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from project STAR. *Quarterly Journal of Economics*, 126(4):1593–1660.

Chingos, M. M. (2012). Class size and student outcomes: Research and policy implications. *Journal of Policy Analysis and Management*, 32(2):411–438.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., and York, R. (1966). *Equality of Educational Opportunity.* U.S. Government Printing Office, Washington, DC.

Correa, H. and Gruver, G. (1987). Teacher-student interaction: A game-theoretic extention of the economic theiory of education'. *Mathematical Social Sciences*, 13:19–47.

Dearden, L., Ferri, J., and Meghir, C. (2002). The effect of school quality on educational attainment and wages. *The Review of Economics and Statistics*, 84(1):1–20.

Denny, K. and Oppedisano, V. (2013). The surprising effect of larger class sizes: Evidence using two identification strategies. *Economics of Education Review*, 23:57–65.

Dieterle, S. G. (2015). Class-size reduction policies and the quality of entering teachers. *Labour Economics*, 36:35–47.

Ding, W. and Lehrer, S. F. (2010). Estimating treatment effects from contaminated multiperiod education experiments: The dynamic impacts of class size reductions. *Review of Economics and Statistics*, 92(1):31–42.

Ding, W. and Lehrer, S. F. (2011). Experimental estimates of the impacts of class size on test scores: robustness and heterogeneity. *Education Economics*, 19(3):229–252.

Dustmann, C., Rajah, N., and van Soest, A. (2003). Class size, education, and wages. *Economic Journal*, 113:F99–F120.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., and Willms, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest*, 2(1):1–30.

Falch, T., Sandsør, A., and Strøm, B. (2017). Do smaller classes always improve students' long-run outcomes? *Oxford Bulletin of Economics and Statistics*, 79(5):654–688.

Finn, J. D., Gerber, S. B., and Boyd-Zaharias, J. (2005). Small classes in the early grades, academic achievement, and graduating from high school. *Journal of Educational Psychology*, 97(2):214–223.

Fredriksson, P., Öckert, B., and Oosterbeek, H. (2013). Long-term effects of class size. *Quarterly Journal of Economics*, 128(1):249–285.

Fredriksson, P., Öckert, B., and Oosterbeek, H. (2014). Inside the black box of class size: Mechanisms, behavioral responses, and social background. IZA Discussion Paper No. 8019.

Fredriksson, P., Öckert, B., and Oosterbeek, H. (2016). Parental responses to public investments in children: Evidence from a maximum class size rule. *Journal of Human Resources*, 51(4):832–868.

Gary-Bobo, R. J. and Mahjoub, M. B. (2013). Estimation of class-size effects, using Maimonides' rule and other instruments: The case of French junior high schools. *Annals of Economics and Statistics*, 111-112:193–225.

Gladwell, M. (2013). *David and Goliath: Underdogs, Misfits, and the Art of Battling Giants*. Little, Brown & Company., New York.

Heinesen, E. (2010). Estimating class-size effects using within school variation in subject-specific classes. *The Economic Journal*, 120:737–760.

Hoxby, C. (1998). The effects of class size and composition on students' achievement: New evidence from natural population variation. National Bureau of Economic Research, Working Paper 6869.

Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics*, 115(4):1239–1285.

Imai, K. (2017). *Quantitative Social Science*. Princeton University Press.

Iversen, J. M. and Bonesrønning, H. (2013). Disadvantaged students in the early grades: Will smaller classes help them? *Education Economics*, 21:305–324.

Jakubowski, M. and Sakowski, P. (2006). Quasi-experimental estimates of class size effect in primary schools in Poland. *International Journal of Educational Research*, 45:202–215.

Jepsen, C. and Rivkin, S. (2009). Class size reduction and student achievement: The potential tradeoff between teacher quality and class size. *Journal of Human Resources*, 44(1):223–250.

Krassel, K. F. and Heinesen, E. (2014). Class-size effects in secondary school. *Education Economics*, 22(4):412–426.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, 114(2):497–532.

Krueger, A. B. and Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project STAR. *Economic Journal*, 111:1–28.

Lazear, E. (2001). Educational production. *Quarterly Journal of Economics*, 116(3):777–803.

Leuven, E. and Løken, S. (2017). Long term impacts of class size in compulsory school. IZA Discussion Paper No 10594.

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

Li, W. and Konstantopoulos, S. (2016). Class size effects on fourth-grade mathematics achievement: Evidence from TIMSS 2011. *Journal of Research on Educational Effectiveness*, 9(4):503–530.

Lindahl, M. (2005). Home versus school learning: A new approach to estimating the effect of class size on achievement. *Scandinavian Journal of Economics*, 107(2):375–394.

Nandrup, A. B. (2016). Do class size effects differ across grades? *Education Economics*, 24(1):83–95.

OECD (2016). *Education at a Glance 2016: OECD Indicators*. OECD Publishing, Paris.

Rockoff, J. (2009). Field experiments in class size from the early twentieth century. *Journal of Economic Perspectives*, 23:211–230.

Schanzenbach, D. W. (2014). Does class size matter? National Education Policy Center.

Shen, T. and Konstantopoulos, S. (2017). Class size effects on reading achievement in Europe: Evidence from PIRLS. *Studies in Educational Evaluation*, 53:98–114.

Stock, J. H. and Watson, M. W. (2015). *Introduction to Econometrics*. Pearson, 3rd edition.

Todd, P. and Wolpin, K. I. (2003). Towards a unified approach for modeling the production function for cognitive achievement. 113:F3–F33.

Urquiola, M. and Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review*, 99:179–215.

Wößmann, L. (2005). Educational Production in Europe. *Economic Policy*, 20(43):445–504.

Wößmann, L. and West, M. (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in timss. *European Economic Review*, 50(3):695–736.

EENEE Analytical Reports

33	Edwin Leuven Hessel Oosterbeek	Class size and student outcomes in Europe
32	Michel Vandenbroeck Karolien Lenaerts Miroslav Beblavý	Benefits of early childhood education and care and the conditions for obtaining them
31	Holger Bonin	The Potential Economic Benefits of Education of Migrants in the EU
30	Giorgio Brunello Maria De Paola	School Segregation of Immigrants and its Effects on Educational Outcomes in Europe
29	Mette Trier Damgaard Helena Skyt Nielsen	The use of nudges and other behavioural approaches in education
28	Marius Busemeyer Philipp Lergetporer Ludger Woessmann	Public Opinion and the Acceptance and Feasibility of Educational Reforms
27	Maria de Paola Giorgio Brunello	Education as a tool for the economic integration of migrants
26	Daniel Münich Steven Rivkin	Analysis of incentives to raise the quality of instruction
25	Elena Del Rey Ioana Schiopu	Student Debt in Selected Countries
24	Maria Knoth Humlum Nina Smith	The impact of school size and school consolidations on quality and equity in education
23	Torberg Falch Constantin Mang	Innovations in education for better skills and higher employability
22	Francis Kramarz Martina Viarengo	Using Education and Training to Prevent and Combat Youth Unemployment
21	Jo Blanden Sandra McNally	Reducing Inequality in Education and Skills: Implications for Economic Growth
20	Ludger Woessmann	The Economic Case for Education

19	Daniel Münich George Psacharopoulos	Mechanisms and methods for cost-benefit / cost-effectiveness analysis of specific education programmes
18	Reinhilde Veugelers Elena Del Rey	The contribution of universities to innovation, (regional) growth and employment
17	Giorgio Brunello Maria de Paola	The costs of early school leaving in Europe
16	Samuel Muehlemann Stefan C. Wolter	Return on investment of apprenticeship systems for enterprises: Evidence from cost-benefit analyses
15	Hessel Oosterbeek	The Financing of Adult Learning
14	Susanne Link	Developing key skills: What can we learn from various national approaches?
13	Marc Piopiunik Paul Ryan	Improving the transition between education/training and the labour market: What can we learn from various national approaches?
12	Daniel Münich Erik Plug George Psacharopoulos Martin Schlotter	Equity in and through Education and Training: Indicators and Priorities
11	Adrien Bouguen Marc Gurgand	Randomized Controlled Experiments in Education
10	Torberg Falch Hessel Oosterbeek	Financing lifelong learning: Funding mechanisms in education and training
9	Reinhilde Veugelers	A Policy Agenda for Improving Access to Higher Education in the EU
8	Giorgio Brunello Martin Schlotter	Non Cognitive Skills and Personality Traits: Labour Market Relevance and their Development in E&T Systems
7	Eric A. Hanushek Ludger Woessmann	The Cost of Low Educational Achievement in the European Union
6	George Psacharopoulos Martin Schlotter	Skills for Employability, Economic Growth and Innovation: Monitoring the Relevance of Education and Training Systems

5	Martin Schlotter Guido Schwerdt Ludger Woessmann	Methods for Causal Evaluation of Education Policies and Practices: An Econometric Toolbox
4	Martin Schlotter	Origins and Consequences of Changes in Labour Market Skill Needs
3	Martin Schlotter Guido Schwerdt Ludger Woessmann	The Future of European Education and Training Systems: Key Challenges and their Implications
2	George Psacharopoulos	The Costs of School Failure – A Feasibility Study
1	Ludger Woessmann Gabriela Schuetz	Efficiency and Equity in European Education and Training Systems

NC-02-17-766-EN-N

Publications Office

doi: 10.2766/28340
ISBN: 978-92-79-70085-9