The London School of Economics and Political Science

Causal Inference in Social Policy
Evidence from Education, Health, and Immigration

Gabriel Heller Sahlgren

A thesis submitted to the Department of Social Policy of the London School of Economics and Political Science for the degree of Doctor of Philosophy

London, July 2019

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

Chapter 2 of the thesis has been published in the *Journal of Health Economics*, Vol. 54, July 2017, pp. 66–78.

Chapter 3 of the thesis has been published in the *Economics of Education Review*, Vol. 62, February 2018, pp. 66–81.

I declare that my thesis consists of 59,762 words.

Statement of inclusion of previous work

I confirm that Chapter 2 is a heavily revised and extended version of the paper I submitted for the MSc degree in Social Policy (Research) at the London School of Economics and Political Science, awarded in November 2015.

Abstract

Separating causation from correlation in empirical studies is crucial for drawing the right conclusions for social-policy development. In the last decade, the emergence of increasingly sophisticated econometric techniques has opened up new ways to draw causal inferences in studies analysing observational data. This thesis contains four papers employing such techniques to answer important research questions in three different areas of social policy: education, health, and immigration.

The first paper analyses the impact of retirement on mental health in ten European countries. It exploits thresholds created by state-pension ages in an individual-fixed effects instrumental-variable set-up, borrowing intuitions from the regression-discontinuity design literature, to deal with endogeneity in retirement behaviour. The results display no short-term effects of retirement on mental health, but a large negative longer-term impact. This impact survives a battery of robustness tests, and applies to women and men as well as people of different educational and occupational backgrounds similarly. The findings suggest that reforms inducing people to postpone retirement are not only important for making pension systems solvent, but with time could also pay a mental-health dividend among the elderly and reduce public health-care costs.

The second paper studies whether independent-school competition involves a trade-off between pupil wellbeing and academic performance. To test this hypothesis, it analyses data covering pupils across the OECD, exploiting historical Catholic opposition to state schooling for exogenous variation in independent-school enrolment shares. The paper finds that independent-school competition decreases pupil wellbeing but raises achievement and lowers educational costs. The analysis and balancing tests indicate these findings are causal. In addition, it finds several mechanisms behind the trade-off, including more traditional teaching and stronger parental achievement pressure.

The third paper analyses the impact of refugee inflows on voter turnout in Sweden in a period when shifting immigration patterns made the previously homogeneous country increasingly heterogeneous. Analysing individual-level panel data and exploiting a national refugee placement programme to obtain plausibly exogenous variation in immigration, it finds that refugee inflows significantly raise the probability of voter turnout. Balancing tests on initial turnout as well as placebo tests regressing changes in turnout on future refugee inflows support the causal interpretation of our findings. The

results are consistent with group-threat theory, which predicts that increased out-group presence spurs political mobilisation among in-group members.

The fourth paper investigates the impact of adult education and training (AET) on employment outcomes in Sweden. Exploiting unusually rich data from the Programme for the International Assessment of Adult Competencies and using an inverse-probability weighted regression-adjustment estimator to deal with selection bias, it finds that AET raises the probability of doing paid work by 4 percentage points on average. This impact is entirely driven by non-formal, job-related AET, such as workshops and on-the-job training. The paper also finds that the effect – which increases with training intensity – is very similar across different types of non-formal, job-related AET. Specification and robustness tests indicate the estimates are causal. The results suggest that policies stimulating relevant AET take-up have promise as a way to secure higher employment rates in the future.

Acknowledgements

I am grateful to my supervisors, Julian Le Grand and Olmo Silva, for their guidance and support throughout the PhD programme. In addition, I am indebted to Magnus Henrekson and Henrik Jordahl at the Research Institute of Industrial Economics for advice and helpful discussions. Financial support from the Arvid Lindman Foundation, the Axel and Margaret Ax:son Johnson Foundation, the Economic and Social Research Council, and the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

On a more personal note, I would like to thank my parents for helping (and sometimes forcing) me to appreciate the value of education, my siblings and friends for being excellent sparring partners in discussions of relevance to my research, and my partner, Charlotte Rose Fox, for putting up with me.

Contents

List of tables

1. Introduction

Separating causation from correlation in empirical studies is crucial for drawing the right conclusions for social policy. To do so, the key is to obtain variation in the independent variable of interest that is not itself affected by the dependent variable and is not related to other factors – observable or unobservable – that in turn affect the dependent variable. Without research that establishes a causal connection between interventions and outcomes, it is impossible to know their effects, thus hampering the advancement of cost-effective policy. Indeed, crafting social policy without access to empirical research from which it is possible to draw causal inferences could be described as tantamount to taking a leap in the dark.

Yet this is often not reflected in policy development. One reason for this is simply the studies that have been available to policymakers; historically, social-policy research has generally not utilised rigorous impact-evaluation methods. This became especially evident after the EPI Centre, established at the Institute of Education to map social-policy research, published the results of systematic reviews of such research in the late 1990s and early 2000s. The goal was to create a database of studies from the 1940s onwards to inform policy development in the field, in a way similar to what the Cochrane Collaboration previously had done in medicine. However, the results were disappointing: one review revealed that only 2.5 per cent of thousands of studies could be classified as methodologically sound. Furthermore, the review found that studies using stronger methodologies were less likely to find a positive impact of the interventions under investigation. In other words, there was a correlation between optimistic findings and lower-quality methodology (Oakley et al. 2005). Thus, an important reason why strong empirical social-policy research has not been used broadly in policymaking is likely to be the lack of such research in a historical perspective.

However, another reason is that politicians have often chosen to rely on research that is insufficient from a causal standpoint because it happens to support their policies. Rather than shaping policy after robust evidence, there has been a tendency to exploit evidence – irrespective of quality – that supports the policy. An important example concerns education. With the emergence of large-scale international surveys, it has become easier than ever to benchmark countries against each other in terms of test-score performance. This, in turn, has enabled policymakers to engage in 'policy borrowing' by attempting to copy practices from high-performing countries, often supported by high-profile case studies in the area (e.g. Sahlberg 2014; Tucker 2011). An interesting example of such policy borrowing is displayed by Scotland's education reforms in the early 2000s, which were partly justified by comparisons with the Finnish education system. For example, an important reform in 2003 abolished statutory standardised tests and the associated league tables among pupils aged 5–14 (Peterkin 2003; Scott 2003), the lack of which is a feature of the current Finnish system that is often highlighted internationally (see Heller-Sahlgren 2015). Indeed, the decisions taken by the Scottish government 'reflect[ed] similar decisions in countries such as Finland where age and stage testing is limited to the final two years of secondary education' (Scottish Labour 2014, p. 23). More generally, policymakers worldwide have sought lessons from the OECD's (2016) report on policies and practices that correlate with higher education performance cross-nationally.

Yet such an approach is risky. This is because case studies of what countries do – as well as high-level cross-national correlations – can mislead us in understanding how the policies of interest contribute to pupil performance. There are many differences between countries that may explain differences in both policy and outcomes. For this reason, policies that are present in high-performing countries, or are positively related to performance, may have a negative causal impact, while, vice versa, policies that are present in low-performing countries, or are negatively related to performance, may have a positive causal effect. For example, with respect to Finland, research moving beyond high-level correlations suggests that factors often attributed to the country's success in international surveys, including the lack of standardised testing and league tables, are negative for pupil performance in such surveys (e.g. Bergbauer et al. 2018; Burgess et al. 2013; Heller-Sahlgren 2015). Interestingly, Scottish pupils' performance has also fallen in domestic and international sample-based tests in the past decades, following the reforms (IEA 2018; OECD 2016; SSLN 2017). While it is impossible to attribute the falling performance to the reduction of standardised testing and abolition of league tables without proper research, it is consistent with the more rigorous evidence that already exists on the topic – and inconsistent with the less rigorous evidence. And, in fact, following intense debate about Scotland's declining performance, the government responded by re-introducing national testing for pupils aged 5–14 (McIvor 2017).

With respect to the OECD's education analyses: while its reports are full of caveats regarding causal inference in the small print, the organisation's conclusions often ignore these caveats. An important illustration of this tendency is the organisation's claim that 'there is no evidence to suggest that [independent] schools help to raise the level of performance of the school system, as a whole' (OECD 2011, p. 4). More recently, it has also claimed that the evidence in this respect is 'unclear' since 'PISA shows no relationship between competition and results in cross-country comparisons' (OECD 2019, p. 92). While both statements hold true in the organisation's own analyses, which consist of unadjusted correlations between countries' independent-school enrolment shares and their average performances, this is not particularly informative given the potential for bias in such analyses. Indeed, more rigorous research suggests that independent-school competition – as captured by the independent-school enrolment shares at the country level – has causal positive effects on PISA scores at the system level (see Heller-Sahlgren 2018; West and Woessmann 2010).[1] Yet by drawing conclusions from simple cross-country correlations, and ignoring the academic research on the subject, the OECD has contributed to misinterpretations of the evidence in the policy community worldwide in terms of the system-level impact of school competition in international surveys.

Perhaps even more conspicuously, the organisation has advised the Swedish government that 'school choice and competition likely weakened school performance over time [in Sweden's compulsory education sector]' (OECD 2019, p. 76), as the country's results in international tests have fallen in the same period as choice and competition were introduced. This is despite the fact that the strongest available research analysing the effects of these particular reforms, on domestic performance metrics as well as international test scores, points in the opposite direction (Böhlmark and Lindahl 2015; Edmark et al. 2014; Wondratschek et al. 2013). In other words, by ignoring the academic evidence on the topic in favour of a simple correlation between the introduction of the reforms and changes in performance, the OECD has confused politicians seeking to pursue evidence-based policy.

Overall, there is thus little doubt that a lack of methodologically sound research, and a refusal to utilise such research in policymaking, hamper the development of social policy that is fit for purpose. As noted, the key aspect of methodologically sound quantitative

[1] Note that this conclusion is valid for research analysing system-level effects of independent-school competition on international test scores specifically; the broader literature on the effects of competition and independent schools is more mixed (see Heller-Sahlgren 2013).

research is that it in one way or the other obtains variation in the independent variable of interest that is not itself affected by the dependent variable and is not related to other factors – observable or unobservable – that affect the dependent variable. Traditionally, such variation has often been obtained by running randomised trials, which remain the supposed 'gold standard' of policy-relevant research. Certainly, randomised trials are important tools in obtaining internally-valid estimates of the effects of policy-relevant interventions, but they are not free from weaknesses (see Deaton and Cartwright 2018). For example, it is often difficult to obtain sufficiently large and diverse samples to obtain high external validity in the findings, which in turn decreases their significance for policy purposes. Also, they are often expensive and in some areas too politically sensitive to be a viable option, especially when studying systemwide structural interventions.

Fortunately, in the last decade, the emergence of 'big data' and increasingly sophisticated econometric techniques has opened up new ways to draw causal inferences also in observational studies (see Angrist and Piscke 2009). However, these opportunities are yet to be exploited widely in social-policy research, which this thesis seeks to partly rectify. To do so, the thesis presents four papers employing different econometric techniques to answer important questions in three areas of social policy: education, health, and immigration. In doing so, the thesis not only makes a significant contribution to the literature in each of the specific areas under investigation, but, by exploiting different datasets and methods, it highlights how recent methodological advances can be applied to draw causal conclusions in social-policy research more generally. In this sense, it informs both researchers in the respective fields and policymakers seeking to develop effective policy.

The second chapter – 'Retirement Blues' – studies the short- and long-run impact of retirement on mental health in Europe, using the Survey of Health, Ageing, and Retirement in Europe. Decreasing labour-force participation rates have put pressure on the sustainability of pension systems in many developed countries, while depressing savings rates and investment levels. Politicians have thus begun to reform state-pension systems to incentivise the elderly to postpone retirement, for example by increasing the regular retirement age at which state-pension benefits may be drawn. However, these policies may have unintended consequences, which must be taken into account to understand the reforms' total utility. One such issue concerns how retirement affects mental health, which has become an increasingly important policy goal in the past decades.

To analyse this issue, the chapter exploits an individual-fixed effects instrumental variable design, combined with intuitions borrowed from the regression-discontinuity literature, using state-pension thresholds as instruments to separate causation from correlation. It is the first paper to do so in a framework that separates the short- from longer-run effects. Its headline finding is that retirement has no impact on mental health in the short run, but a large negative longer-term effect. The study thus contributes significantly to our understanding of the effects of retirement, while also combining and highlighting methodological innovations that could be used in other areas of social-policy research.

The third chapter – 'Smart but Unhappy: Independent-school Competition and the Wellbeing-efficiency Trade-off in Education' – analyses whether independent-school competition involves a trade-off between pupil wellbeing and academic performance. While educational theory assumes that wellbeing and academic achievement go hand in hand, there is little empirical evidence supporting this assumption. It is crucial for policymakers to understand whether education interventions involve trade-offs between these goals, and, using international pupil-level data from 34 OECD countries, the chapter provides new evidence of relevance in this respect to independent-school competition specifically – the general effects of which have become a fiercely debated topic worldwide.

To obtain conditionally exogenous variation in independent-school competition at the country level, the chapter builds on and further develops an instrument based on Catholic resistance to state schooling in the late 19th and early 20th centuries to predict enrolment shares in independently-operated schools today. It also analyses whether the assumption of conditional exogeneity holds, including by using balancing tests on pupil-background characteristics. The study's headline finding is that independent-school competition decreases pupil wellbeing but raises achievement and lowers educational costs. In addition, it finds several mechanisms behind the results and hence the trade-off, including more traditional teaching and stronger parental achievement pressure. Overall, the paper thus breaks new ground in the study of markets in education – but also displays how historical events and large datasets can be exploited to obtain causal estimates in social policy more generally.

The fourth chapter – 'Group Threat and Voter Turnout: Evidence from a Refugee Placement Programme' – analyses the effects of refugee inflows on voter turnout in Sweden in a period when shifting immigration patterns made the previously homogeneous country increasingly heterogeneous. One important possible consequence of immigration could be altered political engagement among natives. However, the identification problems involved in analysing the causal impact of immigration on political outcomes are severe because of potential endogeneity in settlement and mobility patterns – among both immigrants and natives – as well as other sources of unobserved heterogeneity. This study addresses these problems by analysing individual-level panel data and exploiting a national refugee placement programme to obtain variation in refugee inflows that is free from bias due to endogenous settlement and mobility patterns, and, once adjusting for municipal-fixed effects, also plausibly exogenous to changes in individual-level turnout more generally.

The chapter finds that refugee inflows significantly raise the probability of voter turnout. The results are consistent with group-threat theory, which predicts that increased out-group presence should spur political mobilisation among in-group members. Balancing tests on initial turnout as well as placebo tests regressing changes in turnout on future refugee inflows support the causal interpretation of the findings. Given the relative scarcity of convincing research on the relevance of group threat for understanding individual-level voter turnout in general, the paper provides an important contribution to the literature of how real-world demographic changes affect political engagement. More generally, it highlights how the combination of individual-level panel data and a natural experiment may be exploited to get around difficult endogeneity problems in social-policy research.

The fifth chapter – 'Lifelong Learning and Employment Outcomes: Evidence from Sweden' – investigates the impact of adult education and training (AET) on employment outcomes in Sweden. It is plausible that AET could help promote the knowledge and skills necessary for ensuring high employment and productivity levels in the future. Yet most existing research uses linear regression methods that ignore possible selection bias and does not analyse the effects of different types of AET separately. Exploiting data from the Programme for the International Assessment of Adult Competencies, this paper adjusts for an unusually large number of important observable characteristics, including cognitive skills, formal education levels, and work history. Combined with an inverse-probability weighted regression-adjustment estimator – which assumes that assignment to AET is as good as random conditional on the covariates, but does not make assumptions about the functional form of the relationship between those covariates and the outcome analysed – this increases the probability that the estimates reflect a causal impact.

The chapter finds that AET raises the probability of doing paid work, an impact that is entirely driven by non-formal, job-related AET, such as workshops and on-the-job training. It also finds that the effect – which increases with training intensity – is very similar across different types of non-formal, job-related AET. Although it is not possible to conclusively rule out omitted-variable bias, several tests suggest this is not a serious problem for the findings. Overall, the paper thus provides important new evidence on the effects of AET on employment outcomes and illustrates how social-policy researchers can separate causation from correlation when no quasi-experimental variation is available, or when such quasi-experimental variation is not relevant for the population of interest.

Finally, the thesis concludes with a summary of the papers' findings and a discussion of the strengths and weaknesses of the methods utilised, and the papers more generally, while also providing directions for future research and considering the studies' policy implications. Apart from providing significant contributions to the literature in each of the separate areas under investigation, by studying different questions of relevance to social policy and employing different econometric techniques, with different strengths and weaknesses, the thesis provides important case studies of how recent scientific advances can be exploited to obtain causal estimates – given different contexts and data availability – in the social-policy field specifically. In doing so, it hopes to aid the development of more evidence-based social policy worldwide.

References

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton, NJ: Princeton University Press.

Bergbauer, Annika B., Eric A. Hanushek, and Ludger Woessmann. 2018. 'Testing.' NBER Working Paper No. 24836. National Bureau of Economic Research, Cambridge, MA.

Böhlmark, Anders and Mikael Lindahl. 2015. 'Independent Schools and Long-run Educational Outcomes: Evidence from Sweden's Large-scale Voucher Reform.' *Economica* 82 (327):508–551.

Burgess, Simon, Deborah Wilson, and Jack Worth. 2013. 'A Natural Experiment in School Accountability: The Impact of School Performance Information on Pupil Progress.' *Journal of Public Economics* 106:57–67.

Deaton, Angus and Nancy Cartwright. 2018. 'Understanding and Misunderstanding Randomized Controlled Trials.' *Social Science & Medicine* 210:2–21.

Edmark, Karin, Verena Wondratschek, and Markus Frölich. 2014. 'Sweden's School Choice Reform and Equality of Opportunity.' *Labour Economics* 30:129–142.

Heller-Sahlgren, Gabriel. 2013. *Incentivising Excellence: School Choice and Education Quality.* London: Centre for Education Economics.

Heller-Sahlgren, Gabriel. 2015. *Real Finnish Lessons: The True Story of an Education Superpower.* London: Centre for Policy Studies.

Heller-Sahlgren, Gabriel. 2018. 'Smart but Unhappy: Independent-school Competition and the Achievement-wellbeing Trade-off in Education.' *Economics of Education Review* 62:66–81.

IEA. 2018. Data retrieved from the IEA's TIMSS database on 12 December 2018, https://www.iea.nl/data.

McIvor, Jamie. 2017. 'No League Tables from New School Assessments', *BBC News*, 22 September, https://www.bbc.co.uk/news/uk-scotland-41363093.

Oakley, Ann, David Gough, Sandy Oliver, and James Thomas. 2005. 'The Politics of Evidence and Methodology: Lessons from the EPPI-Centre.' *Evidence & Policy* 1(1):5–31.

OECD. 2011. 'Private Schools: Who Benefits?' PISA in Focus, Report, Paris.

OECD. 2016. 'PISA 2015 Results Volume II: Policies and Practices for Successful Schools.' Report, Paris.

OECD. 2019. 'OECD Economic Surveys: Sweden.' Report, Paris.

Peterkin, Tom. 2003. '"Out of Date" Exam Lists Dropped in Scotland.' *The Telegraph*, 26 September, https://www.telegraph.co.uk/news/uknews/1442488/Out-of-date-exam-lists-dropped-in-Scotland.html.

Sahlberg, Pasi. 2014. *Finnish Lessons 2.0: What Can the World Learn from Educational Change in Finland?* New York: Teachers College Press.

Scott, Kirsty. 2003. 'Scotland to Scrap Tests on 5 to 14s.' *The Guardian,* 26 September, https://www.theguardian.com/uk/2003/sep/26/scotland.schools.

Scottish Labour. 2014. 'Mind the Gap: Tackling Educational Inequality in Scotland.' Report, Edinburgh.

SSLN. 2017. 'The Scottish Survey of Literacy and Numeracy – High Level Summary of Statistics.' Scottish Government, https://www2.gov.scot/Topics/Statistics/Browse/School-Education/ScottishSurveyOfLiteracyNumeracy.

Tucker, Marc, ed. 2011. *Surpassing Shanghai: An Agenda for American Education Built on the World's Leading Systems.* Cambridge, MA: Harvard Education Press.

West, Martin R. and Ludger Woessmann. 2010. '"Every Catholic Child in a Catholic School": Historical Resistance to State Schooling, Contemporary Private Competition and Student Achievement across Countries.' *Economic Journal* 120(546):F229–F255.

Wondratschek, Verena, Karin Edmark, and Markus Frölich. 2013. 'The Short- and Long-term Effects of School Choice on Student Outcomes: Evidence from a School Choice Reform in Sweden.' *Annals of Economics and Statistics* 111/112:508–551.

2. Retirement Blues*†

Published in the Journal of Health Economics, Vol. 54, July 2017, pp. 66–78.

Abstract

This paper analyses the short- and longer-term effects of retirement on mental health in ten European countries. It exploits thresholds created by state-pension ages in an individual-fixed effects instrumental-variable set-up, borrowing intuitions from the regression-discontinuity design literature, to deal with endogeneity in retirement behaviour. The results display no short-term effects of retirement on mental health, but a large negative longer-term impact. This impact survives a battery of robustness tests, and applies to women and men as well as people of different educational and occupational backgrounds similarly. Overall, the findings suggest that reforms inducing people to postpone retirement are not only important for making pension systems solvent, but with time could also pay a mental health dividend among the elderly and reduce public health care costs.

* This study uses data from SHARE wave 1 and 2, release 2.5.0, as of 24th May 2011, and SHARE wave 4 release 1.0.0, as of 30th November 2012. The SHARE data collection has been primarily funded by the European Commission through the 5th framework programme (project QLK6-CT-2001- 00360 in the thematic programme Quality of Life), through the 6th framework programme (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5- CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th framework programme (SHARE-PREP, 211909 and SHARE-LEAP, 227822). Additional funding from the US National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 and OGHA 04- 064, IAG BSR06-11, R21 AG025169) as well as from various national sources is gratefully acknowledged (see www.share-project.org for a full list of funding institutions).

† The author thanks Henrik Jordahl, Julian Le Grand, Ellen Meara, Olmo Silva, and two anonymous referees for useful comments and discussions.

2.1. Introduction

In the post-World War II period, the combination of increasing life expectancy, decreasing fertility rates, and the normalisation of old-age retirement has induced lower overall labour-force participation rates in developed countries. The rise of retirement was in turn in large part due to incentives built into public pension systems, which have induced people to exit the labour market voluntarily (e.g. Gruber and Wise 1999, 2004; Hurd et al. 2012). Decreasing labour-force participation rates have put pressure on the sustainability of pension systems, while depressing savings rates and investment levels. Politicians have begun to reform state-pension systems to incentivise the elderly to postpone retirement, for example by increasing the official retirement age at which state-pension benefits may be drawn.

However, these policies may have unintended consequences, which must be taken into account to understand the reforms' total utility. One such issue concerns how retirement affects mental health. In the past decades, public policy has become more concerned with improving people's wellbeing in general. Thus, if retirement has a positive impact on mental health, attempts to increase the effective retirement age may thwart this policy goal, while possibly also leading to higher sick-leave rates and rising health expenditures. On the other hand, if retirement has negative effects on mental health, policymakers who seek to incentivise people to postpone retirement could, if they are successful, produce a virtuous circle in which public pensions systems are made sustainable, health expenditures decreased, and mental health among the population improved.

Previous research analysing health effects of retirement yields mixed results, possibly reflecting both methodological choices and a general failure to distinguish between short- and longer-term effects. This study investigates the impact of retirement on mental health in Europe in both a short- and longer-term perspective. Utilising several waves of panel data from the Survey of Health, Ageing, and Retirement in Europe covering ten European countries, it exploits thresholds created by state-pension ages in an individual-fixed effects instrumental-variable (FE-IV) set-up, borrowing intuitions from the regression-discontinuity design (RDD) literature, to deal with endogeneity in retirement behaviour. The idea is that these thresholds create strong economic incentives to retire once crossed, but should not affect mental health in other ways once age effects are held constant in a flexible way. Previous research has utilised similar strategies – most often finding positive or zero short-term effects of retirement on mental health – but tends to ignore common pitfalls that threaten their validity and the possibility that the effects of retirement are not instant. To the best of our knowledge, this paper is the first to analyse both short- and longer-term, lagged effects of retirement on mental health in a set-up exploiting age-dependent discontinuities to predict retirement.

The results display no short-term impact of retirement on mental health, but strong negative effects that become apparent a couple of years following the event. The longer-term effect does not differ between men and women, or between people with different educational and occupational backgrounds. It also survives a battery of robustness tests, which include analyses of narrower age windows and using inverse probability weighting to deal with panel attrition. The differences compared with similar research are attributed to the fact that this study differentiates between short- and long-term effects as well as uses a methodology that provides a cleaner estimate of the impact of retirement per se.

Overall, the findings indicate that politicians do not face a trade-off between increasing state-pension ages and improving wellbeing. Inducing people to postpone retirement is not only necessary to make pension systems sustainable, but can also be a way to improve mental health among the elderly. While pension reforms may have immediate negative mental health effects prior to retirement as some research suggests – at least if these reforms postpone eligibility rules late in people's lives – this paper indicates that they may pay a mental health dividend after some time by delaying the negative longer-term impact of retirement per se.

The study proceeds as follows. Section 2.2 discusses the theoretical mechanisms potentially linking retirement to mental health; Section 2.3 discusses the previous literature; Section 2.4 discusses the data utilised; Section 2.5 outlines the paper's research strategy; Section 2.6 presents the results; and Section 2.7 concludes.

2.2. Theory

Why and how might retirement affect mental health? One way to think about the relationship between the two variables is through an economic lens in which individuals seek to maximise their utility. In Grossman's (2000) human capital model, health acts both as a direct consumption good, since it is important for people's wellbeing, and as an investment, since individuals must be in good health to be able to work and increase their lifetime earnings. Retirement may affect these properties differently: the incentive to be in good health for investment purposes is no longer present in retirement, but since individuals have more free time as retirees, the consumption value of health may increase. In the end, the theoretical net effect then depends on whether the overall marginal utility of health decreases or increases after retirement – which is far from straightforward to predict (Dave et al. 2006).

A similarly ambiguous story concerns other explanations that do not necessarily rely on rational choice. For example, retirement may affect individuals' social capital and networks, which research suggests have positive effects on health (e.g. d'Hombres et al. 2010; Folland 2008; Rocco et al. 2014; Ronconi et al. 2012). Yet it is theoretically unclear how retirement affects social interactions: people may lose work colleagues, but they also have more time to create a new, voluntarily established, social network. Perhaps reflecting this theoretical ambiguity, research finds mixed effects of retirement on the size of individuals' social networks (Börsch-Supan and Schuth 2014; Fletcher 2014). While retirement could potentially induce couples to spend more time together, it may also increase the probability of divorce (see Stancanelli 2014). A similarly equivocal story applies to stress. While retirement may decrease work-related stress, it is in itself a life event that can be very stressful. And retirement is often accompanied by a decrease in income and consumption (e.g. Finnie and Spencer 2013), which may affect mental health negatively directly and via increased stress. In fact, such ambiguous stories apply to most theoretical mechanisms potentially linking retirement to mental health.[1]

Furthermore, it is important to note that retirement may have different effects in the short- and longer-term perspective. This is partly because the effect of investments in health is likely to operate with a lag, which means that lower/higher investments do not necessarily bring negative/positive effects until after some time. Similarly, in the beginning, retirement may be perceived more like a holiday rather than as permanent labour market exit (e.g. Atchley 1976). If so, one may also expect people's mental health to improve – or at least not deteriorate – during this period. In a longer-term perspective, however, the holiday effect may fade out and be replaced by mechanisms

[1] Of course, there is a plethora of other theoretical mechanisms through which retirement could potentially affect mental health in different ways, such as changes in health insurance status, although these are also often related to the rational choice perspective.

generating lower mental health. Alternatively, it may be the case that retirement increases stress and dissatisfaction in the short run, which then subside in the long run as people acclimatise. All this is related more generally to the 'hedonic treadmill' hypothesis (Brickman and Campbell 1971), which stipulates that life events only affect wellbeing in the short term as individuals adapt with time. Thus, it is clearly important to take into account that short- and longer-term effects of retirement may differ, although it is difficult to predict how and in what ways.

2.3. Previous literature

The impact of retirement on health has become a topic of increasing interest among researchers in economics and other fields. Correlational studies analysing the association between retirement and mental health find mixed effects (e.g. Dave et al. 2008; Jokela et al. 2010; Lindeboom et al. 2002; Mein et al. 2003; Mosca and Barrett 2014; Oksanen et al. 2011; Vo et al. 2015; Westerlund et al. 2009). However, since the act of retirement is not random, these studies cannot tease out its causal effects on mental health.

Improving the methodology, some researchers have employed IV models using eligibility ages at which state-pension benefits can be drawn to predict retirement. The idea is that reaching these eligibility ages gives rise to significant economic incentives to retire. At the same time, the argument goes, there is no reason why reaching the threshold per se should affect mental health apart from via retirement, once smooth effects of age are held constant. This gives rise to the potential to use these thresholds as instruments for retirement in IV or fuzzy RDD frameworks. Another approach has been to utilise pension reforms in difference-in-difference set-ups, comparing individuals affected by the reforms with individuals who are not. Overall, studies using either of these research strategies tend to find no or positive effects of retirement on mental health (e.g. Behncke 2012; Blake and Garrouste 2012; Charles 2004; Coe and Zamarro 2011; Eibich 2015; Fé and Hollingsworth 2012; Fonseca et al. 2014; Johnston and Lee 2009; Latif 2013; Mazzonna and Peracchi 2014; Neuman 2008).[2]

[2] Using a regular IV set-up, Mazzonna and Peracchi (2014) also find weak evidence in robustness tests that longer time spent in retirement increases the likelihood of depression on average. However, the instrument used is the distance of respondents' actual age from the relevant state-pension thresholds. As discussed in Section 2.5.1, this variable should normally be included as a control to allow the age trend to differ on both sides of the threshold and thus decrease the risk that the binary retirement indicator, which in their study is only used to estimate the immediate impact of retirement, picks up non-linear

However, this research also suffers from some limitations. First, it tends to include a number of 'bad controls' that are endogenous to retirement, which means that it controls for some of the causal pathways through which retirement may affect mental health (Angrist and Pischke 2009). Such bad controls include consumption, marital status, and income – all of which may both affect mental health and be affected by retirement (e.g. Finnie and Spencer 2013; Haider and Stephens 2007; Stancanelli 2014). Second, most studies ignore potential differences between immediate and lagged effects of retirement, which, as noted in Section 2.2, may be quite different. Third, studies evaluating pension reforms in difference-in-difference set-ups ignore that such reforms often impact on behaviour, and mental health, before individuals retire (e.g. Bertoni et al. 2016; de Grip et al. 2012; Montizaan et al. 2010). This violates the assumption that the reforms used for identification affect mental health solely through retirement, thereby casting doubt on the studies' internal validity.

Another potential problem in most previous research is that it uses instruments constructed from both regular and early retirement ages. This neglects potential self-selection into jobs where individuals are more likely to be able to retire early, which could undermine the validity of the findings. Furthermore, since early retirement ages often differ depending on vocation in European countries, it is difficult to find out which threshold that applies to which segment of the population. While discontinuities arising at the state-pension age should help ameliorate measurement error in retirement status as observed in survey data – since the discontinuities are shaped by institutional rules and are therefore uncorrelated with potential measurement error – it is thus also possible that discontinuities at (alleged) early retirement ages induce new measurement error. This, in turn, is likely to produce attenuation bias in the estimates (Angrist and Pischke 2009). Of course, it is also not clear whether early and regular retirement events have the same effects; as countries are in the process of increasing regular state-pension ages specifically, it is important to disentangle the separate impact of retirement at these ages.

Another general problem with previous IV studies based on RDD intuitions is that they often ignore common pitfalls associated with such designs. For example, they do not present results in which the impact of age is allowed to differ on each side of the

---

effects of age. This is not a trivial concern, especially since the authors use a relatively broad age window and control for a linear age trend only.

discontinuities used as instruments (Lee and Lemieux 2010). Similarly, most existing studies do not analyse the sensitivity of the findings by narrowing the range of data analysed around the discontinuities (Angrist and Pischke 2015).[3] Instead, they choose a rather wide range of data, without sufficiently exploring non-linear effects of age or the results' sensitivity to the specific data range.[4]

Overall, therefore, while previous research most often finds no or positive effects of retirement on mental health, it suffers from some limitations. Perhaps most important is that previous studies do not generally separate short- from longer-term effects of retirement. This study aims to remedy the shortcomings highlighted and provide a more rigorous evaluation of the impact of retirement on mental health in Europe. Section 2.5 discusses this strategy in detail.

2.4. Data

This study utilises data from the first, second, and fourth waves of the Survey of Health, Ageing, and Retirement in Europe (SHARE), conducted at different points in 2004–05, 2006–07, and 2011–12 respectively. In these waves, SHARE provides information on a wide range of background and outcome variables from representative samples of individuals who are aged 50 and over in ten European countries (see Börsch-Supan et al. 2013): Austria, Belgium, Denmark, France, Germany, Italy, the Netherlands, Spain, Sweden, and Switzerland.[5] Analysing panel data spanning over several SHARE waves allows us to investigate both the short- and longer-term effects of retirement on mental health. Table 2.1 displays the timeline for interviews over the period analysed.

As Table 2.1 shows, the second interview is held on average two years and three months after the first interview, whereas the fourth interview is held on average four years and four months after the second interview, and six years and seven months after the first interview.[6] In our main analysis, we study the impact of a change in retirement

[3] Much previous multi-country research has also ignored the possibility that the impact of age on mental health differs across countries. If such differential age effects are correlated with the state-pension ages utilised in the analysis, estimates may be biased.
[4] The exception is Eibich's (2015) paper, which finds a short-term positive impact of retirement on mental health in Germany.
[5] The SHARE dataset has been used widely in related economic research (e.g. Coe and Zamarro 2011; Godard 2016; Mazzonna and Peracchi 2012; Rohwedder and Willis 2010).
[6] In unreported robustness tests, we restricted the sample to (1) only individuals with a time span between the first and second waves of minimum two years and maximum three years, and (2) only individuals with a time span between the second and fourth waves of minimum four years and maximum five years. Overall, the main results were very similar.

status over the first and second waves on the change in mental health over the second and the fourth waves. Since the fourth interview takes place several years after the change in retirement status took place between the first and the second waves, we believe our set-up captures a lagged, longer-term effect well. We also seek to compare and contrast this impact with the effect of a change in retirement status between the first and second waves on changes in mental health over the same period, which captures the immediate, short-term impact. The methodology utilised to tease out the different effects is discussed in detail in Section 2.5.

Table 2.1: Timeline for interviews

	Mean	Min	Max
Wave 1	October 2004	March 2004	December 2005
Wave 2	January 2007	September 2006	October 2007
Wave 4	May 2011	February 2011	March 2012
Time passed (Waves 1–2)	2 years, 3 months	11 months	3 years, 4 months
Time passed (Waves 2–4)	4 years, 4 months	3 years, 5 months	5 years, 6 months
Time passed (Waves 1–4)	6 years, 7 months	5 years, 6 months	7 years, 10 months

Note: the timeline refers to the main sample within the ten-year age window over and below the relevant state-pension age. It is essentially identical for the sample within the three-year age window.

2.4.1. Sample

The main sample includes individuals who were interviewed in the first, second, and fourth wave of SHARE and who were 50–75 years old at the second wave interview.[7] As discussed in Section 2.5, the study exploits discontinuities in retirement that occur at the state-pension age to obtain exogenous variation in retirement behaviour. The state-pension ages for the ten European countries utilised in the study – which take into account pension reforms in recent decades and thus vary across cohorts within countries – are displayed in Table 2.A.1 in the Appendix. The threshold varies between 55 and 67 years, depending on country, gender, and cohort. However, individuals reaching their state-pension age between the first and the second SHARE waves – which is the relevant threshold for the set-up outlined in Section 2.5 – were between 60 and

[7] Since the third wave was a special survey (SHARELIFE) that did not enquire respondents about their mental health, it cannot be used in this paper.

65.[8] This gives an age window of about ten years over and under the lowest and highest thresholds utilised, calculated at the second wave, which is in line with Moreau and Stancanelli (2013) and Stancanelli and Van Soest (2012).

The working sample thus consists of maximum 8,566 individuals across the ten countries, all observed three times, with the exact sample size depending on the restrictions discussed in Section 2.4.2. Table 2.2 provides summary statistics for the working sample as well as differences in means between individuals assigned to the treatment group (those who crossed the state-pension age between the first and second interviews), and individuals assigned to the control group (those who did not cross the state-pension age between the first and second interviews).[9] In robustness tests, we also utilise a narrower age window of three years around the lowest and highest state-pension ages, which means that people who were 57–68 years old at the second wave are included. This decreases the sample size to maximum 4,704 individuals, again all observed three times.

2.4.2. *Retirement*

The study employs three different definitions of retirement. First, it classifies individuals as retired if they claim to be retired or give a date at which they retired that precedes the interview date. Given the methodology outlined in Section 2.5, this means that we analyse the impact of retirement compared to the status of employed/self-employed, homemaker, unemployed, being engaged in other activities, as well as the permanently ill or disabled. This definition is useful since all non-retirees, not only those who are engaged in paid work, may respond to retirement incentives at the state-pension age. This is because they may still be able to access age-dependent benefits at that point, which could trigger permanent withdrawal from the labour market and

[8] The only exception is in Denmark, where individuals born before July 1939 instead faced a retirement age of 67. To ensure that individuals of the same age are analysed in all countries, the right-hand side of the age window is calculated from age 65 also for the Danish sample. However, all results are essentially identical if the Danish age window is instead extended by two years on the right-hand side, which only increases the sample by about 30 respondents. Results are also robust to extending the right-hand side of the age window to 67 for all countries.

[9] The FE-IV design described in Section 2.5 analyses compliers only: respondents assigned to the treatment group who retired between waves 1 and 2 and respondents assigned to the control group who did not retire in that period. In robustness tests, we simultaneously analyse the lagged impact of retiring between waves 1 and 2 and the direct effect of retiring between waves 2 and 4, both in the full sample and when excluding individuals who were retired and/or above the state-pension age at the first interview. As Table 2.A.5 shows, the results are very similar in these analyses.

make them regard themselves to be retired.[10] Furthermore, for policy purposes, the mental health impact of retirement is relevant regardless from which category respondents officially retire.

However, to ensure that the above definition of retirement does not drive the results, we also employ an alternative definition based on retirement from the labour force only. In this definition, which is similar to Coe and Zamarro's (2011), homemakers, respondents who report being permanently ill or disabled, and those who are engaged in other activities are instead included in the retirement category, as long as they do not also report having done any paid work in the past four weeks. Respondents in these categories who report having done paid work are still included as non-retirees. Finally, in the third definition, the sample is simply restricted to individuals who claim to be retired or to be working. The latter category includes employed/self-employed respondents and other non-retired individuals who report having done paid work in the past four weeks.

2.4.3. *Mental health*

Following previous research in the field, our primary measure of mental health is the Euro-D scale. The scale was developed specifically as a common depression gauge in the European Union and has shown to be valid for research purposes (see Prince et al. 1999). The scale ranges from 0 to 12, with higher values indicating stronger depressive tendencies, counting whether or not respondents reported having had problems with the following in the past month: appetite, concentration, depression, enjoyment, fatigue, guilt, interest, irritability, pessimism, sleep, suicidality, and tearfulness. Furthermore, we also analyse the likelihood that respondents have reached the conventional threshold for clinical depression, which is defined as a score of 4 or more on the Euro-D scale. In both cases, therefore, higher values indicate worse mental health.

2.4.4. *Control variables*

If the design discussed in Section 2.5 produces random variation in respondents' retirement behaviour, the only necessary covariates to ensure a causal interpretation of the estimates are flexible controls for respondents' age. However, to test robustness of

[10] This argument is supported by the fact that the F statistic for the instrument is the largest when utilising this definition, indicating that respondents who are not engaged in paid work prior to retirement do respond to the incentives in the regular state-pension system.

the estimates, it is also useful to include lagged mental health status to ensure that mean reversion does not bias the findings. Doing so may also increase precision in the estimates. It is also possible to interact certain background variables with the retirement indicator in order to investigate potential heterogeneous effects. These issues are discussed more formally in Section 2.5.1.

2.4.5. Attrition

It is important to consider potential selection problems due to non-random panel attrition, which is often ignored by studies exploiting panel surveys to evaluate various outcomes of retirement (e.g. Bonsang et al. 2012; Eibich 2015). Like in most panel surveys, attrition in SHARE is substantial: about 50 per cent of people interviewed in the first wave disappear by the fourth wave. Of course, this has no bearing on the results if attrition is unrelated to how retirement affects mental health, but this cannot be established a priori. However, it is possible to include individual-fixed effects, which control for time-invariant variables that affect both respondents' propensity to remain in the panel as well as their retirement behaviour and mental health. This should at least mitigate the attrition problem.

As a further robustness check, the study also exploits inverse probability weighting, which allows attrition to be non-random as long as its causes are captured by individuals' observable characteristics at the time before they drop out of the panel (Moffit et al. 1999). This means that we estimate the probability that individuals who were interviewed in the first wave remain in the fourth wave, from variables observed in the final wave in which they participated prior to the fourth wave. These variables include age, employment status, marital status, education level, and a battery of self-assessed, mental, and physical health variables as well as indicators for cognitive achievement.[11] In addition, country-fixed effects are included to ensure that differential attrition across countries does not bias the findings.

The inverse of the probability of remaining in the panel, as predicted by this model, is then used as weight in robustness regressions analysing the impact of retirement on

[11] The health and cognitive indicators include: self-assessed health, the Euro-D scale, the number of drugs taken, the number of diagnosed conditions, the number of limitations with activities of daily living, the number of mobility limitations, as well as memory and numeracy scores.

mental health.[12] If attrition poses a serious problem for the study, one would expect the results from the weighted models to differ significantly compared with the ones that exclude weights. If the results are very similar, on the other hand, it is unlikely that attrition poses a serious problem. Although it is impossible to demonstrate conclusively an absence of attrition bias, the combination of individual-level fixed effects and inverse probability weighting leaves little room for any remaining bias.

Table 2.2: Descriptive statistics

Variable	*Working sample*				*Treatment group*	*Control group*
Wave 1	Mean	SD	Min	Max	Mean	Mean
Age	60.63	6.65	47.23	74.46	62.53	60.42
Retired (1)	0.44	0.50	0	1	0.53	0.43
Retired (2)	0.59	0.49	0	1	0.70	0.57
Retired (3)	0.52	0.50	0	1	0.64	0.51
Euro-D	2.16	2.10	0	12	1.98	2.19
Depression	0.23	0.42	0	1	0.20	0.23
Wave 2	Mean	SD	Min	Max	Mean	Mean
Age	62.95	6.65	50	75.92	64.95	62.73
Retired (1)	0.51	0.50	0	1	0.81	0.48
Retired (2)	0.65	0.48	0	1	0.91	0.62
Retired (3)	0.60	0.49	0	1	0.90	0.57
Euro-D	2.09	2.06	0	11	1.91	2.11
Depression	0.21	0.41	0	1	0.19	0.21
Wave 4	Mean	SD	Min	Max	Mean	Mean
Age	67.24	6.66	54.08	80.77	69.20	67.02
Retired (1)	0.65	0.48	0	1	0.87	0.62
Retired (2)	0.78	0.42	0	1	0.96	0.75
Retired (3)	0.74	0.44	0	1	0.96	0.72
Euro-D	2.27	2.11	0	12	2.25	2.27
Depression	0.24	0.43	0	1	0.25	0.24
n	8,566				849	7,717

Note: Respondents assigned to the treatment group crossed the state-pension age between waves 1 and 2, while respondents assigned to the control group did not cross the state-pension age in that period. Retired (1): respondents who report to be retired or give a retirement date preceding the interview date. Retired (2): retired (1) plus homemakers, the permanently ill or disabled, and those engaged in other activities (as long as they do not do paid work). Retired (3): retired (1) but excluding all non-workers in the non-retired category (for the last definition: n = 7,113 in the total sample, n = 692 assigned to the treatment group and n = 6,421 assigned to the control group).

2.5. Research design

As noted in Section 2.3, any valid research design used to evaluate the causal effect of retirement on mental health must take into account that the former is likely endogenous to the latter. This section discusses the research design employed in the study to deal with endogeneity in retirement behaviour.

[12] The regression predicting the probability to remain in the panel is estimated using a probit model, but all weighted results are essentially identical if we use a linear probability model instead.

2.5.1. *Obtaining exogenous variation in retirement behaviour*

When analysing the impact of retirement on mental health, the easiest strategy is to estimate:

$$mh_i = \alpha + \beta_1 r_i + \beta_2 x_i + \varepsilon_i \tag{1}$$

where mh_i is the mental health outcome analysed; r_i is a dummy variable taking either the value 1 (retired) or 0 (not retired); and x_i is a vector of control variables assumed to affect both retirement and mental health.

The critical assumption is that $Cov(r_i, \varepsilon_i | x_i) = 0$. But if x_i does not include all factors that affect both r_i and mh_i, or if mh_i has an independent impact on r_i, the results will be plagued by endogeneity. In addition, measurement error in r_i may generate attenuation bias. Any of these issues would mean that $Cov(r_i, \varepsilon_i | x_i) \neq 0$ (Angrist and Pischke 2009). As noted in Section 2.3, the critical assumption is not likely to hold. To ensure a causal interpretation of the paper's findings, it is thus key to obtain exogenous variation in retirement behaviour.

To do so, the paper proposes an individual-fixed effects IV design, with intuitions borrowed from the RDD literature. The design is based on retirement ages in European pension systems, which create age thresholds at which economic incentives to retire increase substantially. In this set-up, the discontinuities act as instruments for individuals' employment status in a 2SLS model, with age as the continuous variable determining the discontinuities (Angrist and Piscke 2009; Imbens and Wooldridge 2009). As noted in Section 2.3, this idea has been exploited in previous research, but not in a way that allows researchers to distinguish between short- and longer-term effects. Because of the potential problems that may arise by using the early retirement age, as discussed in Section 2.3, the paper focuses solely on the regular state-pension age.

The idea behind the design is formalised as follows:

$$P(r_i = 1 | age_i) = \begin{cases} f_1(age_i) \text{ if } age_i \geq sp_i \\ f_0(age_i) \text{ if } age_i < sp_i \end{cases}, \text{where } f_1(sp_i) \neq f_0(sp_i)$$

where sp_i is the applicable eligibility age. For this paper's purposes, the assumption is that $f_1(sp_i) > f_0(sp_i)$, since economic incentives raise the likelihood that individuals retire when they reach the eligibility age. Thus, the probability of $r_i = 1$ as a function of age_i can be written:

$$P(r_i = 1|age_i) = f_0(age_i) + [f_1(age_i) - f_0(age_i)]\,\overline{sp}_i$$

where $\overline{sp}_i$ is a dummy variable with the value of 1 if $age_i \geq sp_i$ and 0 if $age_i < sp_i$. The strategy is dependent on the ability to separate smooth effects of age_i from the impact at $\overline{sp}_i$, which serves as instrument for r_i. The paper follows previous studies and assumes a quadratic age trend in the baseline estimates. Recent research shows that estimates including higher-order polynomials in similar designs may be misleading and that it is therefore preferable to reduce the range of data around the threshold (Gelman and Imbens 2014).

The principal difference between our set-up and a regular fuzzy RDD is the inclusion of individual-level fixed effects, which means that we focus on the variation within individuals across time rather than the variation between individuals. This also means that the identification assumptions are different. While a traditional fuzzy RDD would hinge on the assumption that people who are on different sides of, but close to, the state-pension age only differ in terms of the probability of being retired, once controlling flexibly for the direct impact of age, our individual fixed-effects IV estimator hinges on the assumption that merely crossing the threshold serving as instrument does not impact an individual's mental health around the threshold apart from via retirement. We note that similar designs, combining individual-fixed effects with intuitions from the RDD literature, have been utilised in previous research in this and other fields (e.g. Eibich 2015; Lemieux and Milligan 2008; Petterson-Lidbom 2012).

We thus model the lagged, longer-term impact of retirement on mental health using a 2SLS model. Unlike most previous research, we allow the impact of age to differ on both sides of the eligibility threshold used as instrument. To do so, we centre age and its polynomial – by subtracting the state-pension age from the respondent's age – which ensures that the coefficient of the retirement indicator still measures the jump in the dependent variable at the threshold serving as instrument (see Angrist and Pischke 2009, 2015). We also take into account cross-country differences in age effects, below and above the threshold.[13] The estimation then reads:

[13] Thus, a key difference between this paper's set-up and the one used by Coe and Zamarro (2011) is the fact that we exploit the longitudinal aspect of SHARE, allowing us to include individual-fixed effects and to separate short- from longer-term effects of retirement. Other differences include the fact that they analyse only men, do not take into account that the effect of age may differ across countries, ignore the fact that the impact of age may differ below and above the threshold serving as instrument, use both early

$$r_{it-1} = \alpha + \beta_1 \overline{sp}_{it-1} + \beta_2 \widetilde{age}_{it-1} + \beta_3 \widetilde{age}_{it-1}{}^2 + \beta_4 \overline{sp}_{it-1}(\widetilde{age}_{it-1})$$
$$+\beta_5 \overline{sp}_{it-1}(\widetilde{age}_{it-1}{}^2) + \gamma_c(\widetilde{age}_{it-1}) + \gamma_c(\widetilde{age}_{it-1}{}^2) + \gamma_c[\overline{sp}_{it-1}(\widetilde{age}_{it-1})]$$
$$+\gamma_c[\overline{sp}_{it-1}(\widetilde{age}_{it-1}{}^2)] + \delta_i + \mu_t + \varrho_t + \varepsilon_{it} \quad (2)$$

$$mh_{it} = \alpha + \beta_1 \widehat{r_{it-1}} + \beta_2 \widetilde{age}_{it-1} + \beta_3 \widetilde{age}_{it-1}{}^2 + \beta_4 \overline{sp}_{it-1}(\widetilde{age}_{it-1})$$
$$+\beta_5 \overline{sp}_{it-1}(\widetilde{age}_{it-1}{}^2) + \gamma_c(\widetilde{age}_{it-1}) + \gamma_c(\widetilde{age}_{it-1}{}^2) + \gamma_c[\overline{sp}_{it-1}(\widetilde{age}_{it-1})]$$
$$+\gamma_c[\overline{sp}_{it-1}(\widetilde{age}_{it-1}{}^2)] + \delta_i + \mu_t + \varrho_t + \varepsilon_{it} \quad (3)$$

where $\widehat{r_{it-1}}$ is the predicted values of r_{it-1} from the first stage with $\overline{sp}_{it-1}$ as the excluded instrument; $\widetilde{age}_{it-1}$ and $\widetilde{age}_{it-1}{}^2$ denote $(age_{it-1} - sp_{it-1})$ and $(age_{it-1} - sp_{it-1})^2$ respectively; δ_i denotes individual-fixed effects; and μ_t and ϱ_t represent separate year- and month-fixed effects respectively. Including interactions between the age variables and γ_c, which denote country dummies, means that the effect of age is allowed to differ across countries.[14]

The model thus effectively analyses the impact of a change in retirement status between the first and second waves (on average spanning two years and three months) on the change in mental health between the second and fourth waves (on average spanning four years and four months). Because the shift in retirement status may occur at any point between the two interviews, but mental health is measured at the exact point of the interviews, the model focuses only on the change in mental health that occurs after the shift in retirement status has taken place. Of course, the set-up thus also risks ignoring potential immediate effects of a shift in retirement status between the first and the second waves. Yet if we also control for lagged mental health, any difference between compliers in the treatment and control groups due to differential mental health trends between the first and the second waves is ignored. Also, in robustness tests, we simply analyse the change in mental health between the first and fourth waves, thus taking into account any short-term impact directly.

An important rationale behind the study is also to investigate whether the short-term effect of retirement differs from the longer-term impact. Thus, in models analysing the short-term effect of retirement, we estimate equations (2) and (3) but with all variables

and regular state-pension ages as instruments, and control for potentially endogenous variables, such as marital status and income.

[14] The results are very similar if we also include separate indicators for $\widetilde{age}_{it}$ and $\widetilde{age}_{it}{}^2$ and their interactions with γ_c when analysing longer-term effects. This is expected if the strategy induces random variation in retirement behaviour.

measured at t instead of $t-1$.[15] Effectively, we then analyse the impact of retiring between the first and the second waves on the change in mental health over the same period. In this way, by altering the observation window across two and four waves respectively, it is possible to compare and contrast the short- and longer-term effects of retirement on mental health using the same spike in retirement that occurs at the state-pension age as instrument, while at the same time holding constant differential linear and non-linear age trends under and above the thresholds.[16]

We also investigate potential heterogeneous effects depending on gender, education level, as well as physical and psychological occupational burden. We create one dummy that takes the value 1 for women and 0 for men, and one dummy that takes the value 1 for education levels equal to or below lower-secondary school and 0 otherwise. We then utilise Kroll's (2011) indexes measuring the physical burden (OPB) and psychosocial burden (OSB) of different occupations, calculated from their ISCO-88 codes, which we link to the SHARE dataset.[17] The indexes range from 1 to 10, with higher values indicating higher occupational strain. We create two dummies indicating physical and psychosocial occupational strain above the value of 5. By interacting these variables with the retirement indicator, and in turn instrumenting the product with the interaction between the variables and the state-pension threshold, it is possible to analyse potential heterogeneous effects within the above framework.

2.5.2. Assumptions

A useful instrument must first of all be relevant, which in this case means that it should correlate with retirement. It must also be valid, in this case meaning that it must be exogenous to mental health, and further satisfy the monotonicity requirement, which here means that there cannot be people who choose not to retire because they reach the

[15] In estimations analysing short-term effects, the sample is restricted to observations in the first and the second waves. This ensures that the short- and longer-term effects can be compared among the same individuals. However, results for the direct association and impact over the second and fourth waves are presented in Tables 2.A.2, 2.A.3, and 2.A.5 in the Appendix.

[16] Our modelling approach thus differs from Mazzonna and Perrachi's (2014) attempt to separate shorter- and longer-term effects of (early and regular) retirement on general health. Whereas these authors effectively use the interaction between $\overline{sp}_i$ and $(age_{it} - sp_{it})$ as instrument for time spent in retirement, we include the lagged version of this variable and its polynomial as well as their interactions with country dummies as controls to ensure that differential age trends above the thresholds are held constant.

[17] We use the ISCO-88 code for respondents' last job in the first instance. For some respondents, the code is only available for their second or first job. To maximise the number of observations, we thus use the code for the second job in the second instance and the code for the first job in the third instance.

state-pension age (Imbens and Angrist 1994; Hahn et al. 2001). Previous research shows that age-dependent financial incentives in public pension systems are important for retirement behaviour (e.g. Börsch-Supan et al. 2009; Gruber and Wise 1999, 2004; Hurd et al. 2012). It is also unlikely that reaching the state-pension age would induce some people not to retire, which would violate the monotonicity requirement. And since we hold constant individual-fixed effects and control very flexible for the continuous variable from which the binary instrument is constructed – in ways similar to RDD strategies – we believe the instrument satisfies the validity requirement in our set-up.[18]

Furthermore, having access to panel data also means that it is possible to investigate more thoroughly whether our set-up produces random variation in retirement. Indeed, if this is the case, the retirement coefficient should not differ much when including lagged mental health as independent variable, although precision may increase (Lee and Lemieux 2010). Including lagged mental health also makes it possible to test and control for potential mean reversion, which may affect the findings (e.g. Angrist and Pischke 2009).[19]

Finally, we note that the age window utilised may impact the findings. Choosing the window involves a trade-off between consistency and efficiency: a smaller window decreases the likelihood of bias, but fewer observations simultaneously increase the variance (Lee and Lemieux 2010). We thus restrict the age window in robustness tests to ensure that the results do not hinge on the one utilised in the main set-up.

2.6. Results

Table 2.3 displays estimates from OLS models that include individual-fixed effects, but do not correct for endogeneity in retirement behaviour, using the ten-year age window calculated at the second wave. The coefficients are insignificant for both measures analysed. This holds true when analysing the short-term association between the first and the second waves in the first panel, and when analysing the longer-term association in the second panel. Meanwhile, the results in Table 2.A.2 show a negative direct association between the second and fourth waves, indicating a positive relationship

[18] As always, the estimates capture a local average treatment effect (LATE) of retirement on mental health (Imbens and Angrist 1994). In this study, the LATE is relevant for individuals who retire because they reach the relevant state-pension eligibility age.

[19] Because individual-fixed effects are included, lagged mental health is mechanically correlated with ε_{it} (Nickell 1981), but this is not a problem for this study's purposes, as long as lagged mental health is orthogonal to the instrument – which the exercise is supposed to test.

between retirement and mental health over that period. The OLS estimates thus support previous findings of zero or positive effects of retirement on mental health.

Table 2.3: Estimates from individual-fixed effects OLS models

	FE-OLS	FE-OLS	FE-OLS	FE-OLS	FE-OLS	FE-OLS
Retirement definition	1	2	3	1	2	3
	Euro-D	*Euro-D*	*Euro-D*	*Clinical depression*	*Clinical depression*	*Clinical depression*
Short-term associations						
r_{it}	-0.04	-0.03	-0.06	0.00	0.00	0.00
	(0.07)	(0.07)	(0.08)	(0.01)	(0.02)	(0.02)
Long-term associations						
r_{it-1}	-0.03	0.02	-0.11	0.00	0.01	-0.01
	(0.07)	(0.07)	(0.08)	(0.02)	(0.02)	(0.02)
n	8,566	8,566	7,113	8,566	8,566	7,113

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the individual level are in parentheses. All regressions include individual-, year-, and month-fixed effects, a quadric age trend, and interactions with country dummies.

Turning to our main research strategy to deal with endogeneity, Table 2.4 displays the estimates from the FE-IV model in equations (2) and (3). The first stage results show that the instrument is strong, with the F statistics always displaying values considerably higher than 23.1, which is the relevant threshold when using cluster-robust standard errors (Olea and Pflueger 2013). The coefficients indicate that reaching the state-pension age threshold increases the likelihood of retirement by 10–17 percentage points, depending on the definition of retirement, and are always statistically significant at the 1 per cent level. It is thus clear that there is sufficient variation in retirement behaviour in the data, and that it can be predicted well by the state-pension age threshold.

The second-stage results in the first panel, in turn, display no evidence that a change in retirement status over the first and second waves, measured by the coefficient of $\widehat{r_{it}}$, has any short-term effects on changes in mental health over the same period. As displayed in Tables 2.A.3 and 2.A.5, this also holds true in the FE-IV analyses of the direct impact over the second and fourth waves, in sharp contrast to the OLS estimates in Table 2.A.2.

However, the results in the second panel uniformly indicate that a change in retirement status over the first and second waves, measured by the coefficient of $\widehat{r_{it-1}}$, has a large negative longer-term impact on changes in mental health between the second and fourth waves. The coefficients display that retirement increases the overall Euro-D score by 1.55–2.44 points, relative to the control group, depending on which retirement definition that is utilised. Based on the descriptive statistics for the fourth wave in Table 2.2, this corresponds to 0.81–1.30 standard deviations. Meanwhile, retirement increases the likelihood of remaining or becoming clinically depressed, defined as scoring 4 or higher on the Euro-D scale, in the longer term by 31–50 percentage points, relative to the control group. This corresponds to an impact of 0.72–1.16 standard deviations, which is slightly lower compared with the effect size obtained when analysing the Euro-D index.[20] Thus, the initial results indicate that retirement has considerable negative longer-term effects on mental health.[21]

Since we do not detect any short-term effects on mental health, the results imply that the longer-term coefficients pick up a negative effect of retirement on mental health rather than merely a reversion following positive short-term effects. This interpretation receives support from models including lagged mental health in the third panel, which display almost identical, but slightly more precise, longer-term estimates. Also, as Table 2.A.4 shows, the results are very similar when analysing the impact of a change in retirement status between the first and second waves on the change in mental health between the first and fourth waves, thus incorporating any short-term effects directly. Overall, therefore, the FE-IV estimates point in one direction: retirement has no impact on mental health in the short run, but a large negative effect in the longer term.[22]

Meanwhile, in models evaluating longer-term effects, the Hausman tests always reject the null hypothesis of no endogeneity, indicating that OLS estimates are biased downward. This is unsurprising given the results in Table 2.3, but may be unexpected

---

[20] The results are plausible since, as displayed in Table 2.2, the average Euro-D score increased from 1.91 to 2.25 (sd = 2.05), while the share of depressed individuals increased from 0.19 to 0.25 (sd = 0.46), between waves 2 and 4 among respondents who crossed the state-pension age between waves 1 and 2. This should be compared with an increase in the average Euro-D score from 2.11 to 2.27 (sd = 2.17), and the share of depressed individuals from 0.21 to 0.24 (sd = 0.48), between waves 2 and 4 in the group who did not cross the state-pension age between waves 1 and 2. The impact of retirement on mental health is effectively identified from these differences.

[21] The reduced-form effect of crossing the state-pension age threshold is to generate an increase of 0.287 Euro-D points (standard error = 0.088) and raise the likelihood of depression by 0.053 (standard error = 0.019), using the specifications equivalent to columns 1, 2, 4, and 5 in panel 2. In the sample analysed in columns 3 and 6 in panel 2, the reduced-form impact is 0.253 Euro-D points (standard error = 0.092) and 0.056 (standard error = 0.021) respectively.

[22] As Table 2.A.5 shows, the findings are very similar when analysing the lagged impact of retirement and controlling for the direct effect between waves 2 and 4, both in the main sample and when excluding respondents who were retired and/or above the state-pension age in wave 1. Overall, these results indicate that our main strategy is appropriate for capturing the longer-term effect of retirement on mental health.

since potential reverse causality is often thought to bias OLS estimates in the opposite direction, with poor mental health raising the probability of retirement. Yet the differences are plausible since omitted variables and measurement error may bias estimates downward more than potential reverse causality biases estimates upward.[23]

Table 2.4: Estimates from FE-IV models

	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV
Retirement definition	1	2	3	1	2	3
	Euro-D	*Euro-D*	*Euro-D*	*Clinical depression*	*Clinical depression*	*Clinical depression*
Short-term effects						
$\widehat{r_{it}}$	-0.16	-0.26	-0.44	-0.04	-0.06	-0.09
Second stage	(0.49)	(0.78)	(0.65)	(0.11)	(0.18)	(0.15)
$\overline{sp}_{it}$	0.17***	0.10***	0.13***	0.17***	0.10***	0.13***
First stage	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
F statistic	88.56	38.32	48.62	88.56	38.32	48.62
Hausman test	0.81	0.77	0.56	0.72	0.77	0.50
Longer-term effects						
Excluding lagged mental health						
$\widehat{r_{it-1}}$	1.70***	2.74***	1.90**	0.31**	0.50**	0.42**
Second stage	(0.55)	(0.95)	(0.74)	(0.12)	(0.21)	(0.17)
$\overline{sp}_{it-1}$	0.17***	0.10***	0.13***	0.17***	0.10***	0.13***
First stage	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
F statistic	88.46	38.25	48.73	88.46	38.25	48.73
Hausman test	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Including lagged mental health						
$\widehat{r_{it-1}}$	1.61***	2.60***	1.69**	0.29***	0.47**	0.37**
Second stage	(0.50)	(0.87)	(0.67)	(0.11)	(0.18)	(0.15)
$\overline{sp}_{it-1}$	0.17***	0.10***	0.13***	0.17***	0.10***	0.13***
First stage	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
F statistic	88.42	38.23	48.67	88.46	38.24	48.75
Hausman test	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
n	8,566	8,566	7,113	8,566	8,566	7,113

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the individual level are in parentheses. All models include the variables in equations (2) and (3).

2.6.1. Robustness tests

2.6.1.1. Three-year age window

How sensitive are the results to the specific age window around the threshold? Table 2.5 displays estimates from models with the sample restricted to individuals aged within approximately three years over and under the lowest and highest thresholds at the second interview. We then include a linear age trend. Again, there is little evidence

[23] Indeed, research analysing the impact of retirement on cognitive ability also finds that OLS results are biased in the same way (Bonsang et al. 2012).

of any short-term effects. However, the longer-term effects are considerable, despite the fact that 45 per cent of the main sample is dropped. Estimates in Table 2.A.6 also show that the long-term effect remains when including a quadratic age trend, despite the narrower age window. Overall, therefore, the results are robust to using a narrower age window, which further supports the idea that our research design captures causal effects.

Table 2.5: Estimates from FE-IV models (3-year age window)

	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV
Retirement definition	1	2	3	1	2	3
	Euro-D	*Euro-D*	*Euro-D*	*Clinical depression*	*Clinical depression*	*Clinical depression*
			Short-term effects			
$\widehat{r_{it}}$	-0.04	-0.06	-0.22	0.01	0.01	-0.01
Second stage	(0.40)	(0.60)	(0.49)	(0.09)	(0.14)	(0.11)
$\overline{sp}_{it}$	0.20***	0.13***	0.17***	0.20***	0.13***	0.17***
First stage	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
F statistic	123.34	60.33	78.85	123.34	60.33	78.85
Hausman test	0.98	0.99	0.82	0.98	0.91	0.89
			Longer-term effects			
			Excluding lagged mental health			
$\widehat{r_{it-1}}$	1.30***	1.98***	1.18**	0.25**	0.39**	0.27**
Second stage	(0.44)	(0.69)	(0.53)	(0.09)	(0.15)	(0.12)
$\overline{sp}_{it-1}$	0.20***	0.13***	0.17***	0.20***	0.13***	0.17***
First stage	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
F statistic	123.66	60.64	79.27	123.66	60.64	79.27
Hausman test	<0.01	<0.01	0.02	<0.01	<0.01	0.01
			Including lagged mental health			
$\widehat{r_{it-1}}$	1.28***	1.95***	1.08**	0.26***	0.40***	0.27**
Second stage	(0.40)	(0.63)	(0.49)	(0.09)	(0.14)	(0.11)
$\overline{sp}_{it-1}$	0.20***	0.13***	0.17***	0.20***	0.13***	0.17***
First stage	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
F statistic	123.63	60.62	48.67	123.67	60.64	79.28
Hausman test	<0.01	<0.01	0.01	<0.01	<0.01	<0.01
n	4,704	4,704	3,818	4,704	4,704	3,818

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the individual level in parentheses.

2.6.1.2. Inverse probability weighting

As noted in Section 2.4.5, it is important to investigate whether panel attrition threatens the validity of the findings. Table 2.A.7 displays results from the same specifications as those estimated in the third panels in Tables 2.4 and 2.5, with the exception that respondents are weighted by their inverse probability to remain in the panel, as predicted by the baseline characteristics discussed in Section 2.4.5. All estimates are very similar to the ones obtained without weights, irrespective of retirement definition utilised and outcome analysed. In fact, the coefficients become larger, even though a few observations are lost because of missing values on the variables used to predict the probability that respondents remain in the panel. The conclusion from this exercise is thus that it is unlikely that selective attrition poses a threat to the study's findings.

2.6.2. Heterogeneous effects

While all results so far indicate that retirement has a negative average impact on mental health in the longer-term perspective, this does not necessarily mean it affects everybody in the same way. Table 2.A.8 shows results from models that allow the long-term impact of retirement to differ depending on gender, educational background, and physical as well as psychosocial occupational strain, as discussed in Section 2.5.1. There is no evidence of statistically significant heterogeneous effects.[24] The coefficients are generally small and/or inconsistent across retirement definitions. In unreported regressions, we found no evidence of heterogeneous short-term effects between the first and second waves either. Overall, this indicates that the zero immediate and negative longer-term effects of retirement on mental health generally apply similarly to men and women as well as individuals with different socio-economic and occupational backgrounds.[25]

2.7. Conclusion

As policymakers worldwide have begun to reform state-pension systems to induce higher labour-force participation among the elderly, research investigating the causal impact of retirement on health and wellbeing has become increasingly important. While previous studies analysing mental health have generally found positive or no effects, they suffer from limitations. Perhaps most conspicuous is that nobody thus far has separated short- from longer-term effects of retirement in a rigorous framework that exploits discontinuities in retirement arising from state-pension ages. This is an

[24] This also applies when analysing the sample within the three-year age window. We found some evidence that the negative impact of retirement was only detectable among respondents who were living with a partner at the time of the first wave. However, the number of single individuals being affected by the instrument threshold is small in our sample, which makes it difficult to draw strong conclusions from this exercise.

[25] In unreported regressions, we also estimated separate models for the different groups, but again found little consistent evidence of heterogeneity; no differences were statistically significant and no subgroups appeared to benefit from retirement.

important shortcoming since there are good theoretical reasons to believe that the short- and longer-term effects of retirement differ.

This study has sought to remedy these issues by investigating the short- and longer-term effects of retirement on mental health in ten European countries. Analysing panel data from the Survey of Health, Ageing, and Retirement in Europe, it utilised an individual-fixed effects IV approach and age-based discontinuities within state-pension systems as instruments. Although the results show no impact of retirement in the short run, there is strong evidence of a considerable negative lagged effect that appears within a couple of years' time. This effect, which survives a range of robustness tests, is apparent both when analysing the Euro-D scale as well as the cut-off point for clinical depression. It applies to women and men similarly and also appears to operate independently of individuals' educational background and level of occupational strain. This indicates retirement affects people of different socio-economic backgrounds and professions similarly in terms of mental health.

Yet while the study has found a negative longer-term effect of retirement on mental health, it is silent on the mechanisms through which this effect operates. Policymakers and practitioners would certainly benefit from understanding these mechanisms when attempting to counter the negative long-run impact; identifying the specific mechanisms linking retirement to declining mental health in a longer-term perspective remains an important topic for future research to investigate.

Nevertheless, overall, this study's findings indicate that policymakers do not face a trade-off between making state-pension systems solvent and improving mental health among the elderly. Certainly, as displayed by other research, reforms affecting eligibility to state pensions may have immediate negative mental health effects operating independently of retirement, at least if these reforms affect people late in their lives. With time, however, our findings indicate that such reforms not only are necessary to make pension systems sustainable, but may also be an efficient way to improve mental health among the elderly by delaying the negative longer-term effect of retirement per se.

References

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2015. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton, NJ: Princeton University Press.

Atchley, Robert C. 1976. *The Sociology of Retirement*. New York: John Wiley & Sons Inc.

Behncke, Stefanie. 2012. 'Does Retirement Trigger Ill Health?' *Health Economics* 21:282–300.

Bertoni, Marco, Giorgio Brunello, and Gianluca Mazzarella. 2016. 'Does Postponing Minimum Retirement Age Improve Healthy Behaviours Before Retirement? Evidence from Middle-Aged Italian Workers.' IZA Discussion Paper No. 9834, Institute for the Study of Labor (IZA), Bonn.

Blake, Hélène and Clémentine Garrouste. 2012. 'Collateral Effects of a Pension Reform in France.' Working Paper No. 2012–25, Paris School of Economics, Paris.

Bonsang, Eric, Stéphane Adam, and Sergio Perelman. 2012. 'Does Retirement Affect Cognitive Functioning?' *Journal of Health Economics* 31(3):490–501.

Börsch-Supan, Axel, Martina Brandt, Christian Hunkler, Thorsten Kneip, Julie Korbmacher, Frederic Malter, Barbara Schaan, and Stephanie Stuck. 2013. 'Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE).' *International Journal of Epidemiology* 42(4):992–1001.

Börsch-Supan, Axel, Agar Brugiavini, and Enrica Croda. 2009. 'The Role of Institutions and Health in European Patterns of Work and Retirement.' *Journal of European Social Policy* 19(4):341–358.

Börsch-Supan, Axel and Morten Schuth 2014. 'Early Retirement, Mental Health and Social Networks.' Pp. 225–250 in *Discoveries in the Economics of Aging*. Chicago: University of Chicago Press.

Börsch-Supan, Axel and Christina B. Wilke 2006. 'The German Public Pension System: How It Will Become an NDC System Look-Alike.' Pp. 573–610 in *Pension Reform: Issues and Prospects for Non-Financial Defined Contribution (NDC) Schemes*, edited by Robert Holzmann and Edward Palmer. Washington, DC: World Bank.

Brickman, Philip and Donald Campbell 1971. 'Hedonic Relativism and Planning the Good Society.' Pp. 287–302 in *Adaptation-level Theory: A Symposium*. New York: Academic Press.

Charles, Kerwin K. 2004. 'Is Retirement Depressing? Labor Force Inactivity and Psychological Well-Being in Later Life.' *Research in Labor Economics* 23:269–299.

Coe, Norma B. and Gema Zamarro. 2011. 'Retirement Effects on Health in Europe.' *Journal of Health Economics* 30(1):77–86.

Dave, Dhaval, Inas Rashad, and Jasmina Spasojevic. 2006. 'The Effects of Retirement on Physical and Mental Health Outcomes.' NBER Working Paper No. 12123, National Bureau of Economic Research, Cambridge, MA.

Dave, Dhaval, Inas Rashad, and Jasmina Spasojevic. 2008. 'The Effects of Retirement on Physical and Mental Health Outcomes.' *Southern Economic Journal* 75(2):497–523.

de Grip, Andries, Maarten Lindeboom, and Raymond Montizaan. 2012. 'Shattered Dreams: The Effects of Changing the Pension System Late in the Game.' *Economic Journal* 122(559):1–25.

d'Hombres, Beatrice, Lorenzon Rocco, Marc Suhrcke, and Martin McKee. 2010. 'Does Social Capital Determine Health? Evidence from Eight Transition Countries.' *Health Economics* 19(1):56–74.

Eduardo, Fé. and Bruce Hollingsworth. 2012. 'Estimating the Effect of Retirement on Mental Health via Panel Discontinuity Designs.' Working Paper, Munich Personal RePEc Archive.

Eibich, Peter. 2015. 'Understanding the Effect of Retirement on Health: Mechanisms and Heterogeneity.' *Journal of Health Economics* 43:1–12.

Finnie, Ross and Bryon G. Spencer. 2013. 'How do the Level and Composition of Income Change after Retirement? Evidence from the LAD.' Working Paper No. 114, Canadian Labour Market and Skills Researcher Network, Vancouver.

Fletcher, Jason M. 2014. 'Late Life Transitions and Social Networks: The Case of Retirement.' *Economics Letters* 125:459–462.

Folland, Sherman. 2008. 'An Economic Model of Social Capital and Health.' *Health Economics, Policy and Law* 3:333–348.

Fonseca, Raquel, Arie Kapteyn, Jinkook Lee, Gema Zamarro, and Kevin Feeney. 2014. 'A Longitudinal Study of Well-Being of Older Europeans: Does Retirement Matter?' *Journal of Population Ageing* 7:21–41.

Gelman, Andrew and Guido Imbens. 2014. 'Why High-order Polynomials Should not be Used in Regression Discontinuity Designs.' NBER Working Paper No. 20405, National Bureau of Economic Research, Cambridge, MA.

Godard, Mathilde. 2016. 'Gaining Weight Through Retirement? Results from the SHARE Survey.' *Journal of Health Economics* 45:27–46.

Grossman, Michael 2000. 'The Human Capital Model.' Pp. 347–408 in *Handbook of Health Economics*, edited by Joseph P Newhouse and Anthony J Culyer. New York: Elsevier.

Gruber, Jonathan and David A. Wise, eds. 1999. *Social Security and Retirement around the World*. Chicago, IL: University of Chicago Press.

Gruber, Jonathan and David A. Wise, eds. 2004. *Social Security Programs and Retirement around the World: Micro-Estimation*. Chicago, IL: University of Chicago Press.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. 'Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design.' *Econometrica* 69(1):201–209.

Haider, Steven J. and Melvin Stephens. 2007. 'Is There a Retirement-Consumption Puzzle? Evidence Using Subjective Retirement Expectations.' *Review of Economics and Statistics* 89(2):247–264.

Hurd, Michael, Pierre-Carl Michaud, and Susann Rohwedder. 2012. 'The Displacement Effect of Public Pensions on the Accumulation of Financial Assets.' *Fiscal Studies* 33(1):107–128.

Imbens, Guido W. and Joshua D. Angrist. 1994. 'Identification and Estimation of Local Average Treatment Effects.' *Econometrica* 62(2):467–475.

Imbens, Guido W. and Jeffrey W. Wooldridge. 2009. 'Recent Developments in the Econometrics of Program Evaluation.' *Journal of Economic Literature* 47:5–86.

Johnston, David W. and Wang-Sheng Lee. 2009. 'Retiring to the Good Life? The Short-Term Effects of Retirement on Health.' *Economics Letters* 103(1):8–11.

Jokela, M, JE Ferrie, D Gimeno, T Chandola, MJ Shipley, J Head, J Vahtera, H Westerlund, MG Marmot, and M Kivimäki. 2010. 'From Midlife to Early Old Age: Health Trajectories Associated with Retirement.' *Epidemiology* 21(3):284–290.

Kroll, Lars E. 2011. 'Job Exposure Matrices (JEM) for ISCO and KldB (Version 1.0).' Dataset, GESIS Data Archive. http://doi.org/10.7802/1097.

Latif, Ehsan. 2013. 'The Impact of Retirement on Mental Health in Canada.' *Journal of Mental Health Policy and Economics* 16(1):36–46.

Lee, David S. and Thomas Lemieux. 2010. 'Regression Discontinuity Designs in Economics.' *Journal of Economic Literature* 48:281–355.

Lemieux, Thomas and Kevin Milligan. 2008. 'Incentive Effects of Social Assistance: A Regression Discontinuity Approach.' *Journal of Econometrics* 142(2):807–828.

Lindeboom, Maarten, France Portrait, and Gerard J. van den Berg. 2002. 'An Econometric Analysis of the Mental-Health Effects of Major Events in the Life of Older Individuals.' *Health Economics* 11:505–520.

Litwin, Howard. 2007. 'Does Early Retirement Lead to Longer Life?' *Ageing & Society* 27:739–754.

Mazzonna, Fabrizio and Franco Peracchi. 2012. 'Ageing, Cognitive Abilities and Retirement.' *European Economic Review* 56(4):691–710.

Mazzonna, Fabrizio and Franco Peracchi. 2014. 'Unhealthy Retirement?' EIEF Working Paper 09/14, Einaudi Institute for Economics and Finance, Rome.

Mein, G, P Martikainen, H Hemingway, S Stansfeld, and M Marmot. 2003. 'Is Retirement Good or Bad for Mental and Physical Health Functioning? Whitehall II Longitudinal Study of Civil Servants.' *Journal of Epidemiology & Community Health* 57:46–49.

Moffit, Robert, John Fitzgerald, and Peter Gottschalk. 1999. 'Sample Attrition in Panel Data: The Role of Selection on Observables.' *Annales d'Économie et de Statistique* 55–56:129–152.

Montizaan, Raymond, Frank Cörvers, and Andries de Grip. 2010. 'The Effects of Pension Rights and Retirement Age on Training Participation: Evidence From a Natural Experiment.' *Labour Economics* 17(1):240–247.

Moreau, Nicolas and Elena Stancanelli. 2013. 'Household Consumption at Retirement: A Regression Discontinuity Study on French Data.' IZA Discussion Paper No. 7709, Institute for the Study of Labor (IZA), Bonn.

Mosca, Irene and Alan Barrett. 2014. 'The Impact of Voluntary and Involuntary Retirement on Mental Health: Evidence from Older Irish Adults.' Discussion Paper No. 8723, Institute for the Study of Labor, Bonn.

Neuman, Kevin. 2008. 'Quit Your Job and Get Healthier? The Effect of Retirement on Health.' *Journal of Labor Research* 29:177–201.

Nickell, Stephen J. 1981. 'Biases in Dynamic Models with Fixed Effects.' *Econometrica* 49(6):1417–1426.

Oksanen, Tuula, Jussi Vahtera, Hugo Westerlund, Jaana Pentti, Noora Sjösten, Marianna Virtanen, Ichiro Kawachi, and Mika Kivimäki. 2011. 'Is Retirement Beneficial for Mental Health? Antidepressant Use Before and After Retirement.' *Epidemiology* 22(4):553–559.

Olea, José L. M. and Carolin E. Pflueger. 2013. 'A Robust Test for Weak Instruments.' *Journal of Business & Economic Statistics* 31(3):358–369.

Petterson-Lidbom, Per. 2012. 'Does the Size of the Legislature Affect the Size of Government? Evidence from Two Natural Experiments.' *Journal of Public Economics* 96(3–4):269–278.

Prince, M J., F Reischies, A T. Beekman, R Fuhrer, C Jonker, S L. Kivela, B A. Lawlor, A Lobo, H Magnusson, M Fichter, H van Oyen, M Roelands, I Skoog, C Turrina, and J R. Copeland. 1999. 'Development of the EURO-D Scale – A European Union Initiative to Compare Symptoms of Depression in 14 European Centres.' *British Journal of Psychiatry* 174(4):339–345.

Rocco, Lorenzo, Elena Fumagalli, and Marc Suhrcke. 2014. 'From Social Capital to Health – and Back.' *Health Economics* 23(5):586–605.

Rohwedder, Susann and Robert J. Willis. 2010. 'Mental Retirement.' *Journal of Economic Perspectives* 24(1):119–138.

Ronconi, Lucas, Timothy T. Brown, and Richard M. Scheffler. 2012. 'Social Capital and Self-Rated Health in Argentina.' *Health Economics* 21(2):201–208.

SSA. 2006. 'Social Security Programs Throughout the World: Europe, 2006.' SSA Publication No. 13–11801, Social Security Administration, Washington, DC.

Stancanelli, Elena. 2014. 'Divorcing Upon Retirement: A Regression Discontinuity Study.' Discussion Paper No. 8117, Institute for the Study of Labor, Bonn.

Stancanelli, Elena and Arthur Van Soest. 2012. 'Joint Leisure Before and After Retirement: A Double Regression Discontinuity Approach.' IZA Discussion Paper No. 6698, Institute for the Study of Labor (IZA), Bonn.

Westerlund, Hugo, Mika Kivimäki, Archana Singh-Manoux, Maria Melchoir, Jane E. Ferrie, Jaana Pentti, Markus Jokela, Constanze Leineweber, Marcel Goldberg, Marie Zins, and Jussi Vahtera. 2009. 'Self-Rated Health Before and After Retirement in France (GAZEL): A Cohort Study.' *The Lancet* 374:1889–1896.

Vo, Kha, Peta M. Forder, Meredith Tavener, Bryan Rodgers, Emily Banks, Adrian Bauman, and Julie E. Byles. 2015. 'Retirement, Age, Gender and Mental Health: Findings from the 45 and Up Study.' *Aging & Mental Health* 19(7):647–657.

Appendix

Table 2.A.1: State-pension ages across countries, gender, and cohorts

Country	*Men*	*Women*
Austria	65	60
Belgium	65	60-63
Denmark	65-67	65-67
France	60	60
Germany	65	60-62
Italy	60-65	55-60
Netherlands	65	65
Spain	65	65
Sweden	65	65
Switzerland	65	62-63

Note: The state-pension ages are based on those provided by Mazzonna and Perrachi (2012), with slight adjustments based on data from other sources (see Börsch-Supan and Wilke 2006; SSA 2006). The state-pension age varies by country, gender, and cohort (as indicated by the age ranges in the columns).

Table 2.A.2: The short-term association between retirement and mental health (waves 2–4)

	FE-OLS	FE-OLS	FE-OLS	FE-OLS	FE-OLS	FE-OLS
Retirement definition	1	2	3	1	2	3
	Euro-D	*Euro-D*	*Euro-D*	*Clinical depression*	*Clinical depression*	*Clinical depression*
r_{it}	-0.13**	-0.15**	-0.17**	-0.03**	-0.03**	-0.04**
	(0.06)	(0.06)	(0.07)	(0.01)	(0.01)	(0.01)
n	8,551	8,551	7,228	8,551	8,551	7,228

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the individual level are in parentheses. The model specifications correspond to those in the first panel in Table 2.3, but instead analysing changes between waves 2 and 4. The number of individuals differs slightly compared to the models in the paper. This is because of a few instances of missing data, and, in the case of the models using the third retirement definition, because more individuals reported themselves to be either working or retired between the second and fourth waves.

Table 2.A.3: The short-term impact of retirement on mental health (waves 2–4)

	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV
Retirement definition	1	2	3	1	2	3
	Euro-D	*Euro-D*	*Euro-D*	*Clinical depression*	*Clinical depression*	*Clinical depression*
$\widehat{r_{it}}$	-0.04	-0.06	-0.04	0.07	0.10	0.10
	(0.35)	(0.50)	(0.44)	(0.08)	(0.11)	(0.10)
F statistics	165.35	80.49	97.02	165.36	80.51	97.00
Hausman	0.97	0.99	0.97	0.23	0.27	0.24
n	8,551	8,551	7,228	8,551	8,551	7,228

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the individual level are in parentheses. The model specifications correspond to those in the first panel in Table 2.4, but instead analysing changes between waves 2 and 4, with all controls measured at t, while also including lagged mental health.

Table 2.A.4: Longer-term effects between waves 1 and 4

	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV
Retirement definition	1	2	3	1	2	3
	Euro-D	*Euro-D*	*Euro-D*	*Clinical depression*	*Clinical depression*	*Clinical depression*
$\widehat{r_{it-1}}$	1.52***	2.46***	1.47**	0.27**	0.44**	0.33*
	(0.56)	(0.94)	(0.75)	(0.12)	(0.20)	(0.17)
F statistic	88.46	38.25	48.73	88.46	38.25	48.73
Hausman	<0.01	<0.01	0.02	0.03	0.02	0.04
n	8,566	8,566	7,113	8,566	8,566	7,113

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the individual level in parentheses. The specification corresponds to the one used in the second panel in Table 2.4, but instead analysing changes in mental health between waves 1 and 4.

Table 2.A.5: Controlling for short-term effects between waves 2 and 4 (excluding and including further restrictions on the control group)

	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV
Retirement definition	1	2	3	1	2	3
	Euro-D	*Euro-D*	*Euro-D*	*Clinical depression*	*Clinical depression*	*Clinical depression*
Full working sample						
$\widehat{r_{it-1}}$	1.91***	2.67***	1.62*	0.45***	0.67***	0.51**
	(0.62)	(1.03)	(0.89)	(0.14)	(0.24)	(0.21)
$\widehat{r_{it}}$	0.33	0.06	-0.02	0.18	0.17	0.12
	(0.56)	(0.81)	(0.73)	(0.13)	(0.19)	(0.17)
n	8,550	8,550	7,011	8,550	8,550	7,011
Excluding individuals who were retired at wave 1						
$\widehat{r_{it-1}}$	1.66***	1.61**	1.20**	0.32***	0.34**	0.30**
	(0.52)	(0.63)	(0.59)	(0.12)	(0.14)	(0.13)
$\widehat{r_{it}}$	0.26	0.38	0.15	0.15	0.10	0.10
	(0.44)	(0.53)	(0.48)	(0.10)	(0.12)	(0.11)
n	4,757	3,535	3,334	4,757	3,535	3,334
Excluding individuals who were retired and/or above the state-pension age at wave 1						
$\widehat{r_{it-1}}$	2.50***	2.04**	1.64**	0.53***	0.40**	0.38**
	(0.94)	(0.82)	(0.77)	(0.20)	(0.18)	(0.17)
$\widehat{r_{it}}$	-0.03	-0.05	-0.18	0.10	0.03	0.04
	(0.54)	(0.70)	(0.61)	(0.12)	(0.15)	(0.14)
n	4,310	3,416	3,224	4,310	3,416	3,224

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the individual level are in parentheses. The specifications correspond to the one used in the third panel in Table 2.4, while also including $\widehat{r_{it}}$ (with $\overline{sp}_{it}$ as instrument). The minimum F statistic is 18.45. The minimum Cragg-Donald F statistic is 14.29. The Hausman test always displays values lower/higher than 0.1 for the lagged/non-lagged coefficient

Table 2.A.6: Combining a quadratic age trend with the 3-year age window

	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV
Retirement definition	1	2	3	1	2	3
	Euro-D	*Euro-D*	*Euro-D*	*Clinical depression*	*Clinical depression*	*Clinical depression*
Excluding lagged mental health						
$\widehat{r_{it-1}}$	2.61***	3.76***	2.75**	0.50***	0.72**	0.64***
	(0.84)	(1.33)	(1.07)	(0.18)	(0.28)	(0.24)
F statistics	45.92	27.43	28.71	45.92	23.43	28.71
Hausman	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Including lagged mental health						
$\widehat{r_{it-1}}$	2.27***	3.27***	2.25**	0.39**	0.57**	0.48**
	(0.75)	(1.18)	(0.94)	(0.16)	(0.24)	(0.21)
F statistics	45.82	23.33	28.50	46.01	23.37	28.76
Hausman	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
n	4,704	4,704	3,818	4,704	4,704	3,818

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the individual level are in parentheses.

Table 2.A.7: Estimates from models using inverse probability weighting

	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV
Retirement definition	1	2	3	1	2	3
	Euro-D	*Euro-D*	*Euro-D*	*Clinical depression*	*Clinical depression*	*Clinical depression*
Weighted estimates from the 3rd panel in Table 2.4						
$\widehat{r_{it-1}}$	2.21***	3.48***	2.67***	0.39***	0.61***	0.57***
	(0.61)	(1.05)	(0.86)	(0.11)	(0.22)	(0.19)
F statistic	71.95	34.00	39.62	72.01	34.02	39.70
Hausman	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
n	8,493	8,493	7,050	8,493	8,493	7,050
Weighted estimates from the 3rd panel in Table 2.5						
$\widehat{r_{it-1}}$	1.68***	2.49***	1.67***	0.33***	0.48***	0.38***
	(0.47)	(0.72)	(0.58)	(0.10)	(0.16)	(0.12)
F statistic	106.12	58.20	69.37	106.10	58.23	69.41
Hausman	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
n	4,661	4,661	3,782	4,661	4,661	3,782

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the individual level are in parentheses.

Table 2.A.8: Heterogeneous effects

	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV	FE-IV
Retirement definition	1	2	3	1	2	3
	Euro-D	*Euro-D*	*Euro-D*	*Clinical depression*	*Clinical depression*	*Clinical depression*
			Gender			
$\widehat{r_{it-1}} * female$	0.07	1.53	0.04	-0.06	0.16	-0.05
	(0.48)	(0.98)	(0.60)	(0.10)	(0.20)	(0.14)
$\widehat{r_{it-1}}$	1.57***	2.04***	1.67**	0.33***	0.41***	0.39***
	(0.54)	(0.71)	(0.67)	(0.12)	(0.15)	(0.15)
n	8,566	8,566	7,113	8,566	8,566	7,113
			Education			
$\widehat{r_{it-1}} * low\ education$	-0.53	-0.28	-0.66	-0.10	-0.06	-0.14
	(0.57)	(0.81)	(0.69)	(0.12)	(0.17)	(0.16)
$\widehat{r_{it-1}}$	2.02***	2.86***	2.22**	0.36**	0.51**	0.48**
	(0.66)	(1.02)	(0.92)	(0.15)	(0.22)	(0.21)
n	8,502	8,502	7,056	8,502	8,502	7,056
			Occupational Physical Burden			
$\widehat{r_{it-1}} * high\ OPB$	0.04	0.14	0.24	-0.10	-0.11	-0.04
	(0.49)	(0.69)	(0.60)	(0.11)	(0.15)	(0.13)
$\widehat{r_{it-1}}$	1.55***	2.33***	1.76**	0.34***	0.49***	0.38**
	(0.53)	(0.84)	(0.69)	(0.12)	(0.19)	(0.16)
			Occupational Psychosocial Burden			
$\widehat{r_{it-1}} * high\ OSB$	0.21	0.17	0.57	0.07	0.08	0.13
	(0.48)	(0.68)	(0.57)	(0.11)	(0.15)	(0.13)
$\widehat{r_{it-1}}$	1.43**	2.28**	1.47*	0.24*	0.39*	0.27
	(0.60)	(1.00)	(0.80)	(0.14)	(0.22)	(0.18)
n	7,607	7,607	6,665	7,607	7,607	6,665

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the individual level are in parentheses. The minimum F statistic for the instrumented interaction is 61.70 and the minimum F statistics for $\widehat{r_{it-1}}$ is 18.70. The minimum Cragg-Donald F statistic is 23.77. The Hausman test always displays p-values lower than or equal to 0.1. All models include lagged mental health. OPB = Occupational Physical Burden. OSB = Occupational Psychosocial Burden.

3. Smart but Unhappy: Independent-school Competition and the Wellbeing-efficiency Trade-off in Education*

Published in the Economics of Education Review, Vol. 62, February 2018, pp. 66–81.

Abstract

We study whether independent-school competition involves a trade-off between pupil wellbeing and academic performance. To test this hypothesis, we analyse data covering pupils across the OECD, exploiting historical Catholic opposition to state schooling for exogenous variation in independent-school enrolment shares. We find that independent-school competition decreases pupil wellbeing but raises achievement and lowers educational costs. Our analysis and balancing tests indicate these findings are causal. In addition, we find several mechanisms behind the trade-off, including more traditional teaching and stronger parental achievement pressure.

* The author thanks Sarah Cohodes, Susan Dynarski, Henrik Jordahl, Julian Le Grand, Olmo Silva, Dinand Webbink, two anonymous referees, and participants at the conference 'Efficient provision of public services' hosted by the Research Institute of Industrial Economics in Vaxholm, Sweden for useful comments and discussions.

3.1. Introduction

The extent to which independently-operated schools improve pupil outcomes has become a fiercely debated topic in the economics of education. An important motive behind reforms designed to increase independent-school access, such as vouchers, is that such schools will increase competition and thus generate improvements in pupil performance at the system level (e.g. Friedman 1962; Le Grand 2007; Neal 2002). In the past decades, research has begun to evaluate whether or not this holds true in different contexts.

However, the existing literature focuses mostly on academic outcomes. Certainly, such outcomes are important given their links with labour-market success, non-pecuniary long-term outcomes, and economic development (e.g. Atherton et al. 2013; Brunello et al. 2016; Card 1999; Hanushek et al. 2015; Hanushek and Woessmann 2012, 2016; Oreopoulos and Salvanes 2011). But there are also important non-cognitive outputs of schooling. These outputs include pupil wellbeing, which has become an increasingly emphasised policy goal in western countries, justified by the fact that wellbeing in childhood and adolescence is an important predictor of risky behaviour, adult wellbeing, and a range of other outcomes (e.g. Carneiro et al. 2007; Frijters et al. 2014; Jones 2013; Layard et al. 2014; Lévy-Garboua et al. 2006; Takakura et al. 2010). As school is a key part of youngsters' lives, it is perhaps unsurprising that measures of wellbeing at school also predict a range of more general wellbeing and behavioural indicators (e.g. Gibbons and Silva 2011; Huebner and Gilman 2006; Huebner and Diener 2008; Huebner et al. 2014; Locke and Newcomb 2004). Furthermore, it may be easier to positively affect pupil wellbeing and other non-cognitive indicators at school, compared with cognitive performance (e.g. Heckman and Kautz 2013; Payton et al. 2008).

Importantly, however, it is not clear that interventions improving academic efficiency, in terms of academic output per dollar spent, also have positive effects on wellbeing at school. Progressive pedagogical theory, characterised by child-centred ways of working, generally assumes the two go hand in hand (Christodoulou 2014; Mintz 2012), an idea that is often highlighted in policy debates. For example, Public Health England (2014:4) argues: '[P]romoting the health and wellbeing of pupils and students within schools and colleges has the potential to improve their educational outcomes *and* their health and wellbeing outcomes'. Yet there is little rigorous empirical evidence supporting this assumption. In fact, research suggests that policies improving academic performance also often appear to make learning and school life less joyful (e.g. Falch and Rønning 2012; Jürges and Schneider 2010; Warton 2001). If this is the case, policies that raise academic efficiency may also produce lower pupil wellbeing – thus generating a wellbeing-efficiency trade-off in education. We hypothesise that market incentives, which are likely to induce stronger focus on academic efficiency, involve such a trade-off. Given the widespread belief in pedagogical and policy circles that wellbeing and academic achievement are positively, and causally, related – as well as the considerable interest paid to the effects of market reforms in education in general – this is an important issue to investigate in its own right.

Utilising pupil-level data from the Programme for International Student Assessment (PISA) covering 15-year old pupils across 34 OECD countries, we test our hypothesis by analysing the system-level effects of independent-school competition on pupil wellbeing and academic efficiency.[1] We build on prior research – the most relevant of which is West and Woessmann's (2010) – and use an instrumental-variable (IV) strategy exploiting Catholic resistance to state schooling in the 19th and early 20th centuries to predict enrolment shares in independently-operated schools today. As school secularisation gained ground, Catholics tended to push for access to independent schools in countries without Catholicism as state religion. We thus use Catholic population shares in 1900, interacted with an indicator for whether or not Catholicism was the state religion, as instrument for contemporary independent-school enrolment shares. Controlling for detailed regional-fixed effects and a number of relevant pupil-, school-, and country-level controls, including the contemporary version of the instrument itself, it is plausible that this variation is exogenous. This is especially true since we account for a number of other important historical factors affecting the extent to which Catholic resistance did in fact generate higher independent-school competition, and, if it did, the extent to which this competition has survived to this day. If anything, other research and our analysis suggest the strategy may bias the results against our hypothesis.

We find that independent-school competition has considerable negative effects on pupil wellbeing. The effects are just as conspicuous when restricting the sample to pupils in state schools, indicating that the impact depends on system-level competition

[1] For a discussion about the advantages and disadvantages of using international data, see Hanushek and Woessmann (2011).

rather than on the direct impact of independent-school attendance and/or pupil sorting. We further confirm positive effects of competition on PISA scores and a negative impact on education expenditures found in previous research (see West and Woessmann 2010), thus supporting the idea of a wellbeing-efficiency trade-off. Balancing tests on pupil-background variables support the causal interpretation of our findings.

Analysing potential mechanisms behind the trade-off, we find that competition induces more traditional teaching, instructional time, and homework, which prior research suggests raise achievement and lower wellbeing. Also, competition makes pupil-teacher relations more hierarchical and increases parental achievement pressure, two other relevant mechanisms behind the wellbeing-efficiency trade-off.

Finally, based on our findings and other research comparing the longer-term returns to cognitive achievement and wellbeing in adolescence, we carry out a basic back-of-the-envelope cost-benefit analysis. This indicates that the positive effects of independent-school competition on academic efficiency outweigh the negative impact on pupil wellbeing from an economic standpoint. However, using adult life satisfaction as the unit of measurement rather than money, the costs of competition appear to outweigh its benefits. While further research is necessary to draw strong conclusions in this respect, the analysis at least suggests a more general trade-off between the traditional goals of education policy and the wellbeing agenda to which policymakers should pay attention.

The paper proceeds as follows. Section 3.2 discusses the theoretical link between school competition and wellbeing as well as related research; Section 3.3 outlines the data analysed; Section 3.4 describes the methodology; Section 3.5 presents the results and a tentative back-of-the-envelope cost-benefit analysis; and Section 3.6 concludes.

3.2. Theory and related literature

Theoretically, the system-level effects of independently-operated schools should depend on parental demand for different outcomes. If parents perceive the marginal utility of wellbeing at school to be high, we might expect independent-school competition to have positive effects in this respect. For example, increasing access to independent schools expands school choice, which may allow for a better match between pupil and school (e.g. Adnett and Davies 2002). Additionally, independent schools may be more responsive to pupil needs and have more capacity to innovate (e.g.

Chubb and Moe 1988). Finally, with more opportunities for choice, schools must compete to attract pupils (e.g. Hoxby 2003). The competitive pressures, in turn, would force schools that produce low wellbeing to either improve or go out of business. Overall, the result would be higher pupil wellbeing on average throughout the system.

However, this story hinges on the assumption that parents value pupil wellbeing, and, if there are trade–offs between wellbeing and other types of school quality, that they value the former more than the latter. The research in this respect is admittedly scarce, but it does not support this assumption. In England, Gibbons and Silva (2011) find that peer quality and school value added dominate pupil wellbeing as predictors of parental satisfaction with schools. And whereas peer quality and school value added are capitalised into house prices, average pupil happiness is not. This indicates that parents prefer academic and peer quality over pupil wellbeing, thus generating stronger incentives for schools to focus on the former rather than the latter.

Certainly, progressive pedagogical theory, characterised by child-centred ways of working, highlights the importance of wellbeing for improving achievement (Christodoulou 2014; Mintz 2012). Yet there is little evidence in favour of this hypothesis. On the contrary, cognitive research suggests that memorisation, repetition, and teaching of facts – activities that are not necessarily fun or inspiring – are key to learning (Christodoulou 2014; Ingvar and Eldh 2014). Furthermore, research has found that educational methods and interventions promoting higher performance, including traditional teaching methods and central exit examinations, also often appear to make learning and school life less joyful (e.g. Algan et al. 2013; Bietenback 2014; Jiang and McComas 2015; Jürges and Schneider 2010; Jürges et al. 2012; Kirschner et al. 2006; Regh 2012; Schwerdt and Wuppermann 2011; Sweller et al. 2007). Similar stories apply to time spent in school, instructional time, and time spent doing homework (Aucejo and Romano 2014; Csikszentmihalyi and Hunter 2003; Falch and Rønning 2012; Gustafsson 2013; Kuehn and Landeras 2012; Lavy 2015; Rivkin and Schiman 2015; Warton 2001). In other words, in contrast to the assumptions of progressive pedagogical theory, practices that produce higher academic efficiency also often seem to generate lower pupil wellbeing.[2]

[2] In an interesting contribution less related to wellbeing, West et al. (2016) find that Boston charter schools that produce high cognitive achievement appear to have negative effects on various self-reported non-cognitive measures. However, the latter impact may be due to reference bias, since charter-school pupils report having considerably stricter and more hierarchical school environments characterised by

A potential reason explaining these results is that interventions with positive effects on achievement make pupils work harder, which may in turn increase their stress levels and thus decrease their wellbeing. Another possibility is that the interventions decrease wellbeing via raised stress levels that induce pupils to focus more on their schoolwork – which, in turn, raise achievement. Yet another possible reason is that achievement and wellbeing affect each other positively, but that the net effect of the interventions on wellbeing is still negative due to other mechanisms that operate independently of achievement. Regardless of the mechanism, the cause of the differential effects is in any case the interventions per se – which appear to involve a causal net trade-off between achievement and wellbeing.

However, whether or not such a trade-off applies to market incentives in education is an open question. The literature analysing the effects of school choice, autonomy, and competition is mixed, but often finds small-to-moderate positive effects on academic outcomes and overall efficiency (e.g. Böhlmark and Lindahl 2015; Chakrabarti 2008; Eyles and Machin 2015; Lavy 2010; Muralidharan and Sundararaman 2015).[3] For this paper's purposes, the most relevant research is West and Woessmann's (2010) study, which uses similar data and instrument as we do. They find that larger independent-school enrolment shares improve academic efficiency by raising performance in PISA and lowering per-pupil expenditures.

Nevertheless, to the best of our knowledge, only one paper analyses effects on pupil wellbeing. Utilising Spanish high-school data, with an identification strategy based on independent-school availability, Green et al. (2014) find negative effects on pupil satisfaction of attending independently-operated schools. The authors speculate that this negative impact may be due to a stronger focus on academic achievement in such schools. However, they do not empirically investigate potential trade-offs directly or study the system-level effects of independent-school competition.

Overall, therefore, while the theoretical impact of competition from independently-operated schools on pupil wellbeing is somewhat ambiguous, it appears more reasonable to predict a negative effect. However, it also appears reasonable to predict that this negative impact will be accompanied by a positive effect on academic efficiency. We therefore expect a trade-off in terms of how school competition affects

high expectations. By studying competition at the country level, we minimise the potential for similar reference bias in pupil wellbeing.

[3] See Heller-Sahlgren (2013) for a comprehensive review and assessment of this literature.

pupil wellbeing and academic efficiency. The next section describes the data utilised to investigate these issues.

3.3. Data

To study how independent-school competition affects pupil wellbeing and academic efficiency, as well as mechanisms behind a potential trade-off, we exploit pupil-level data from the 2012 round of the OECD's PISA survey. PISA was created as a reliable metric of pupil knowledge, and has been carried out every three years since 2000. In the 2012 round, representative samples of pupils aged between 15 years and three months and 16 years and two months in 34 OECD countries – as well as in 31 other partner countries or economies – were tested in mathematical, reading, and scientific literacy.

In addition to sitting the tests, pupils complete questionnaires with rich details on their background characteristics and personal views, which we use to obtain indicators for pupil wellbeing. While the total sample across the 34 OECD countries covers about 295,000 pupils, the rotating design of the questionnaire means that the sample size when analysing wellbeing is about 190,000 pupils.[4] To obtain information on ownership structure and funding sources, we also make use of the school-level questionnaire, which enquired headteachers about school-background information. Table 3.A.1 outlines the descriptive statistics of the variables used in the analysis.

3.3.1. Data

In PISA 2012, pupils were for the first time asked about their happiness at school, or more specifically to what extent they agree with the following statement: 'I feel happy at school'. Pupils were asked to choose one of the following options: (1) 'strongly agree', (2) 'agree', (3), 'disagree', or (4) 'strongly disagree', which we recode so that higher values indicate higher wellbeing. Research indicates that similar measures of subjective wellbeing are valid and reliable, both across and within countries, for children and adults alike (e.g. Alesina et al. 2004; Frey and Stutzer 2002; Gilman and Huebner 2003; Huebner 2004; Krueger and Schkade 2008; Veenhoven 2012). While our preferred measure may to some extent also pick up general wellbeing, this is not necessarily a

[4] In PISA 2012, the questionnaire was divided into one common part, which covers background variables and was administered to all pupils, and one rotating part, which includes additional question sets that were randomly allocated to pupils within schools. The design, which follows the one for the cognitive test itself, means that about two thirds of pupils answered all questions in the rotating part (see OECD 2014).

problem. This is because happiness at school is likely to affect wellbeing more generally. Indeed, previous research suggests similar measures predict general wellbeing indicators, such as depression, anxiety, life satisfaction, and suicidal ideation (Gibbons and Silva 2011; Huebner and Diener 2008; Huebner and Gilman 2006; Huebner et al. 2014; Locke and Newcomb 2004). For our purposes, it makes most conceptual sense to study wellbeing at school specifically since the independent variable of interest is likely to affect wellbeing primarily via the school environment, and because we are particularly interested in the potential trade-off between pupil wellbeing and academic achievement. By focusing on wellbeing at school specifically, we therefore study the parameter of wellbeing that is most relevant to education policy per se. As highlighted by the OECD (2013 p. 33): 'As schools are a, if not *the*, primary social environment for 15-year olds, these subjective evaluations [of pupil happiness] provide a good indication of whether education systems are able to foster or hinder overall student well-being.' We thus believe our principal wellbeing measure is highly relevant for the purpose of this paper. Nevertheless, in robustness tests, we also consider alternative wellbeing metrics that are less likely to pick up pupils' attitudes to the school itself, including peer relations.

3.3.2. Academic efficiency

While our principal focus is on pupil wellbeing, we also analyse PISA scores in all subjects as well as cumulative per-pupil expenditures between ages 6 and 15, which we obtain from the OECD (2016a).[5] This allows us to investigate whether or not the positive effects on academic achievement and negative impact on educational expenditures, found in previous research using a similar methodology (West and Woessmann 2010), are detected also in our extended sample of countries in PISA 2012 and with the methodological alterations described in Section 3.4.1 and Appendix B. This is important for ensuring that our interpretation of a potential trade-off is correct.

3.3.3. Potential mechanisms

In addition, we consider potential mechanisms through which independent-school competition may operate. One plausible mechanism could be the way teachers interact

[5] For this analysis, we use the expenditure data reported in the PISA 2009 report since the data for Greece is missing in the PISA 2012 report. However, results are essentially identical if we instead use the latter data and exclude Greece.

with children and specifically their teaching methods. As noted in Section 3.2, research finds that pupil-centred methods generate lower achievement, while at the same time making learning more enjoyable. If competition sharpens incentives to raise academic efficiency, teachers may thus use more traditional methods as a means to compete. To study this issue, we use pupils' views regarding the extent to which their mathematics teachers are student oriented, according to the OECD's (2014) taxonomy: 'The teacher gives different work to classmates who have difficulties learning and/or to those who can advance faster'; 'The teacher assigns projects that require at least one week to complete'; 'The teacher has us work in small groups to come up with a joint solution to a problem or task'; and 'The teacher asks us to help plan classroom activities or topics'. Pupils are asked to choose one of the following options: (1) 'Every lesson', (2) 'Most lessons', (3) 'Some lessons', or (4) 'Never or hardly ever'. We recode the responses so that higher values indicate more use of pupil-centred methods.

Furthermore, we also consider a related potential mechanism: hierarchical school environments. Research on Knowledge is Power Program (KIPP) schools indicates that school models predicated on hierarchical environments boost pupil performance (e.g. Abdulkadiroğlu et al. 2016; Angrist et al. 2013). However, more hierarchical school environments may lower pupil wellbeing via worsened pupil-teacher relations. To study these issues, we use responses to the statement 'Most of my teachers really listen to what I have to say' as a proxy for the degree of hierarchy in pupils' relationships with teachers, and responses to the statement 'Students get along well with most teachers' as a general measure of pupil-teacher relations. In these cases, pupils were asked to choose one of the following options: (1) 'strongly agree'; (2) 'agree'; (3) 'disagree'; or (4) 'strongly disagree', which we recode so that higher values indicate less hierarchical and better pupil-teacher relations.

In addition, we investigate the effects of competition on parental achievement pressure. Such pressure could be positively related to both competition and performance, while also having negative effects on wellbeing. Thus, we consider headteachers' appraisals of the level of parental pressure to achieve high academic achievement: (1) 'There is constant pressure from many parents, who expect our school to set very high academic standards and to have our students achieve them; (2) 'Pressure on the school to achieve higher academic standards among students comes from a minority of parents'; or (3) 'Pressure from parents on the school to achieve higher academic standards among students is largely absent'. We recode the responses so that higher values indicate stronger parental achievement pressure.[6]

Finally, we analyse instructional time and time spent on homework. As noted in Section 3.2, these variables have been found to be positive for academic achievement, while also being associated with lower wellbeing. We thus analyse pupil responses to the question: 'In a normal, full week at school, how many class periods do you have in total?'. We further consider the number of class periods per week in each of the test subjects.[7] Unlike the previous statements, these are open questions and pupils are thus asked to write down the total number of class periods per week, instead of choosing from different sets of options. To analyse the total impact on time spent doing homework, we instead use the number of hours per week pupils report that they spend on 'Homework or other study set by your teachers'. Again, this question is open rather than closed.

3.3.4. Independent-school competition

In order to capture independent-school competition at the system level, we use the proportion of 15-year old pupils who attend independently-operated schools in each country, calculated from the PISA 2012 school questionnaire. In this questionnaire, headteachers were asked to report whether or not their school is a 'private school', defined as a school managed directly or indirectly by a non-government organisation, such as a church, trade union, business, or other private institution, or a 'public school', defined as a school managed directly or indirectly by a public education authority, government agency, or governing board appointed by government or elected by public franchise. The aggregate share of pupils attending independently-operated schools is a useful measure to capture the level of independent-school competition in an education system and has been used in similar research (see Hanushek and Woessmann 2011).

3.3.5. Control variables

[6] Since the sampling procedure of schools was designed to optimise sampling of pupils rather than schools, the OECD recommends that researchers 'analyse school-level variables as attributes of students rather than as elements in their own right' (OECD 2014, p. 398). This means that we analyse the effects of headteachers' responses at the pupil level rather than the school level.

[7] Since class periods vary in length, we also analysed the average period length in each of the test subjects in robustness tests. The results are briefly discussed in footnote 31 in this chapter.

We obtain relevant control variables from the school and pupil questionnaires. First, we include controls for a range of pupil-background characteristics: gender, age, immigrant background (first and second generation), an index of home possessions, the highest occupational status of parents, and the highest educational level of parents, expressed in years of schooling.[8] We also include indicators for whether or not schools are located in a village, small town, town, city, or large city.[9] In addition, we control for pupils' school starting age and grade attended. Since sampling is based on pupils' age at test, these variables may reflect important institutional characteristics of different education systems, which could potentially correlate with both our outcome variables and the instrument discussed in Section 3.4.1 through mechanisms other than competition.[10]

Finally, we also control for a number of country-level variables, including (log) GDP per capita in 2011, obtained from the OECD (2016b), and regional dummies for Oceania, East Asia, Europe, Middle East, Latin America, and North America in the baseline estimates. In most models, however, we further include dummies for Anglo-Saxon Europe, Northern Europe, Western Continental Europe, Eastern Europe, and Southern Europe. This allows us to control for fine-grained regional-fixed effects to ensure that cross-national cultural differences associated with both the instrument discussed in Section 3.4.1 and the outcomes are less likely to bias the findings.[11] In addition, we control for other relevant country-level variables discussed in Section 3.4 and Appendix B to strengthen our instrumental-variable strategy.

---

[8] Foreign-born pupils with two foreign-born parents are classified as first-generation immigrants, whereas native-born pupils with at least one foreign-born parent and foreign-born pupils with one native-born parent are classified as second-generation immigrants. Thus, pupils with two native-born parents are classified as natives. The index of home possessions, the highest occupational status of parents, and the highest years of schooling of parents compose the broader ESCS index (see OECD 2014).

[9] The average socio-economic and ethnic background of school peers may also affect the outcomes. However, since independent-school competition may increase school segregation (e.g. Hsieh and Urquiola 2006; Böhlmark et al. 2016), peer background is a potential mechanism through which competition may affect wellbeing as well as academic efficiency and is thus a 'bad control' (Angrist and Pischke 2009). However, as displayed in Table 3.A.5, the results are robust to controlling for the school-level mean of all pupil-level variables.

[10] As in most surveys, the PISA dataset contains some missing values for pupil- and school-level variables, although this problem is minor for any single control in our analysis. Nevertheless, we impute values for missing control variables using the weighted mean at the school or country level, always using the lowest level available. For dummy variables, we assign a value of 1 or 0 depending on which category the mean is closest to. In order to ensure that the results are not biased by this procedure, we also include dummies for missing values and interactions between them and the imputed values. Similar techniques are used widely in previous research analysing PISA data (see Hanushek and Woessmann 2011). Results are essentially identical if we instead drop observations with any control containing missing values.

[11] Note, however, that we refrain from controlling for input variables, such as education spending and class size, which are 'bad controls' and should be left out of the equation.

3.4. Empirical set-up

To study the impact of independent-school competition on the outcomes discussed in Section 3.3, our starting point is the following OLS model:

$$o_{psc} = \alpha + \beta_1 sp_c + \beta_2 x_{psc} + \beta_3 y_{sc} + \beta_4 z_c + \varepsilon_{psc} \tag{1}$$

where o_{psc} is the outcome of pupil p in school s in country c; sp_c denotes the share of pupils attending independently-operated schools in each country; x_{psc} is a vector of pupil-level predictors; y_{sc} is a vector of school-level predictors; and z_c is a vector of country-level predictors.

The model's assumption is that $Cov(sp_c, \varepsilon_{psc} | x_{psc}, y_{sc}, z_c) = 0$. However, if x_{psc}, y_{sc}, and z_c together do not include all variables that impact both o_{psc} and sp_c, or if o_{psc} affects sp_c directly, it would mean that $Cov(sp_c, \varepsilon_{psc} | x_{psc}, y_{sc}, z_c) \neq 0$ and the results will suffer from endogeneity bias (Angrist and Pischke 2009). That is, the level of independent-school competition may in itself be affected by the outcomes, generating reverse causality, and/or omitted variables may affect both the level of independent-school competition and the outcomes. The direction of bias arising from these issues is theoretically unclear, and partly depends on relative parental demand for different types of school quality per the discussion in Section 3.2.

3.4.1. Obtaining exogenous variation in independent-school competition

To address potential endogeneity, we seek to obtain exogenous variation in independent-school competition by exploiting historical resistance to state schooling among Catholics. This strategy has previously been used to predict independent-school enrolment shares within and between countries (Allen and Vignoles 2015; Cohen-Zada 2009; Cohen-Zada and Elder 2009; Falck and Woessmann 2013; West and Woessmann 2010). The idea is that in countries where Catholicism was not the de facto state religion in the late 19th and early 20th centuries, Catholics fiercely resisted the growing state monopolisation of education.

This is because in countries where Catholics could not ensure that teaching in state schools was consistent with their doctrine, such as Belgium, they worked to establish independent schools and pushed governments to obtain public funding for them. In some countries, Catholics joined forces with other groups that sought access to independently-operated schools. For example, in the Netherlands, Catholics teamed up with Calvinists against secular forces in the *Schoolstrijd*, which only ended in 1917 when equal funding for independent and state schools was enshrined in the Dutch constitution. As a general rule, however, Protestants were more accepting of state monopolisation of education and rarely engaged in the same struggles. Nevertheless, when Catholics were successful, the laws implemented often supported funding for other independent schools as well (see Glenn 1989, 2011). We discuss the intuition, and historical features that interfere with its logic, in more detail in Appendix B.

Thus, Catholic population shares in the early 20th century in countries where Catholicism was not the de facto state religion should be a useful instrument for enrolment shares in independently-operated schools today, once controlling for other relevant predictors discussed below. We obtain Catholic population shares in 1900 and 2010 from Brown and James (2015) and indicators for whether or not Catholicism was the state religion in 1900, 1970, and 2000 from Barrett et al. (2001).[12] Our instrument is then constructed from the interaction between Catholic population shares in 1900 and a dummy variable taking the value of 0 for countries in which Catholicism was the de facto state religion in 1900 as well as immediately before World War II and 1 otherwise.[13] The latter restriction is applied because the political dynamic in the education sphere in countries that permanently disestablished the Catholic Church early in the 20th century was often similar to those that had done so by 1900.[14] This historical instrument allows us to control directly for its contemporary version: Catholic population shares in 2010 interacted with an indicator taking the value of 0 if

[12] The only adjustments we make to the data obtained from Barrett et al. (2001) are: (1) Ireland is treated as not having Catholicism as state religion in 1900, since it was then part of the non-Catholic United Kingdom (see Barro and McCleary 2005), and (2) Austria is treated as having Catholicism as its de facto state religion in 1900. Although the region that became Austria in 1918 did not officially have any state religion since the Austro-Hungarian Compromise of 1867, the state provided an essential Catholic monopoly in the public education system, also after the formation of the state in 1918 until Nazi Germany's annexation of the country in 1938 (Glenn 2011; Kaiser and Wohnout 2004).

[13] For Slovakia, we use Catholic population shares in 1910 since this is the first year for which data are available for the country in Brown and James's (2015) data series.

[14] For example, while Chile only abolished Catholicism as state religion in 1925, the public education system had become secularised and centralised already in the mid-to-late 1800s, much to the denigration of the country's Catholics who consequently began pushing for access to independent schools (Barr-Melej 2001; Collier 1997; Gauri 1998). As discussed in Appendix B, a similar story applies to France prior to the abrogation of Napoleon's Concordat of 1801 in 1905.

Catholicism was the state religion in 1900, 1970, and 2000, and 1 otherwise.[15] This means that we control for any direct impact on the dependent variable that our instrument may pick up, and that the exogenous variation we exploit stems only from interactions in historical Catholic population shares and state religion that are unrelated to the contemporary interaction between these variables – that is, the change in the interaction – when holding constant the other variables in the model.

Also, in addition to the controls discussed in Section 3.3.5, we take further precautions by adjusting for the following variables: Calvinist population shares in 1900; population size in 1900; Communist and post-Soviet backgrounds; indicators for early defeat of the Catholic Church in countries where Catholicism was not the state religion; national bans on Jesuits in the late 19th and early 20th centuries; indicators for countries or regions that were de facto annexed into Nazi Germany during World War II; indicators for pro-Catholic allies or client states of Nazi Germany; and indicators for countries or regions that recently implemented voucher programmes in which for-profit operators participate on an equal basis, or carried out reforms that encouraged mass conversions of state schools to independently-operated status.

The general idea behind the inclusion of these variables is to control for sources of current independent-school enrolment shares that cannot be attributed to the instrument and thus maximise its relevance and increase confidence in its validity. To save space, we discuss the complete rationales for each variable in Appendix B and only provide a few short illustrations here. One example concerns Calvinists, who in some countries joined the Catholics' more general resistance to secular state schooling, giving the latter a probability of success that was disproportionate to the relative size of their community. We take this into account by controlling for the share of Calvinists in 1900. Another example is the role of the Society of Jesus in the establishment of independent schools, as the first teaching order of the Catholic Church. During the struggle between secular and religious forces in the 19th century, Jesuits and their associate orders were often banned from certain territories for longer periods of time, often specifically because of their educational influence. We thus control for these bans in our set-up. A third example relates to the impact of Nazi Germany during World War II. Being part of, or de facto annexed into (but not occupied by), the Greater German Reich meant severe

[15] The overall results are similar if we instead simply control for Catholic population shares in 2010 in all countries, by itself or together with a separate indicator for whether or not Catholicism was the de facto state religion in 1900, 1970, and 2000.

persecution of the Catholic Church and closure of all independent schools. We take this into account by controlling for an indicator of Nazi takeover and de facto annexation. Again, detailed accounts of all additional adjustments are provided in Appendix B.[16]

3.4.2. IV specification

Thus, to obtain plausibly exogenous variation in independent-school enrolment shares across OECD countries, we then estimate the following 2SLS model:

$$sp_c = \alpha + \beta_1 1900cs_c + \beta_2 x_{psc} + \beta_3 y_{sc} + \beta_4 z_c + \varepsilon_{psc} \tag{2}$$

$$o_{psc} = \alpha + \beta_1 \overline{sp}_c + \beta_2 x_{psc} + \beta_3 y_{sc} + \beta_4 z_c + \varepsilon_{psc} \tag{3}$$

where $\overline{sp}_c$ is the predicted values of sp_c from the first stage, while $1900cs_c$ represents the excluded instrument, outlined in Section 3.4.1. The vectors x_{psc}, y_{sc}, and z_c denote the pupil-, school-, and country-level controls discussed in Sections 3.3.5 and 3.4.1, including $2010cs_c$, which denotes the contemporary version of the instrument. The variation in the first stage is thus driven by the interaction between historical Catholic population shares and the state religion indicator when holding constant the contemporary interaction between these variables – thus obtaining identification from the change in the interaction over time – and only comparing countries within the regions controlled for by the regional-fixed effects. We cluster the standard errors at the country level, weight all regressions by pupils' inverse sampling probability, and give each country equal aggregate weight in the regressions.[17]

Of course, it is impossible to prove conclusively whether or not the above model captures the effects of competition on pupil wellbeing and academic efficiency, rather than the instrument's effects through a different channel. This may seem especially important since we study several outcomes, which as discussed in Section 3.2 might

[16] Without the inclusion of at least some of these variables, the F statistics drops radically, suggesting the instrument becomes too weak. For example, if we only include broader-regional fixed effects together with the pupil-level and school-location controls, the F statistic in the first stage drops from about 46 to 3. Adding (log) GDP per capita and the Communist indicator only increases the F statistics to 5, but adding indicators for post-Soviet background and national Jesuit bans raises the F statistics to 22. Adding the other variables then strengthens the instrument further. Excluding the additional controls also makes the data less balanced. Specifically, the instrument then significantly predicts the index of home possessions, which is the best pupil-level predictor of pupil happiness and a key predictor of test scores. Overall, the controls thus add considerable value by increasing the relevance and validity of the instrument.

[17] When analysing PISA scores, we follow the OECD's (2014) recommendation and account for the fact that scores are estimated from five separate 'plausible values', created via an item response theory model.

affect each other, but the assumptions are the same. Indeed, if our strategy does ensure bona fide exogenous variation in school competition, the fundamental cause of the effects on the separate outcomes is independent-school competition rather than the instrument itself – irrespectively of whether the outcome variables then affect each other as a result of this competition. The crucial aspect is thus to investigate, as far as possible, whether the instrument affects the outcome variables apart from via school competition or any mechanism through which school competition operates. We explain how we seek to do so below.

3.4.3. Catholicism and wellbeing

The specification above hinges on that historical Catholic population shares are conditionally unrelated to contemporary pupil wellbeing, in countries where Catholicism was not the state religion, apart from via contemporary independent-school competition. We believe this is a tenable assumption. If anything, the strategy may bias results against supporting our hypothesis that competition decreases pupil wellbeing. This is because research finds Catholic population shares to have positive spill-over effects on the wellbeing of Catholics and most non-Catholics, including Protestants (Clark and Lelkes 2009).[18] It is thus more likely that our research strategy would bias results in the opposite direction compared with what our hypothesis predicts. We can also indirectly test whether or not this is correct. If the assumption holds true, the contemporary version of the instrument, which is controlled for in equation (3), should be positively associated with pupil wellbeing. If the contemporary version's relationship with pupil wellbeing is positive, the instrument is likely to be only *negatively* related to pupil wellbeing via its impact on the level of independent-school competition today, once we adjust for the control variables in the model.

3.4.4. Catholicism and academic efficiency

[18] This does not necessarily mean that Catholics have higher wellbeing than other people. Controlling for a range of background characteristics, Clark and Lelkes (2009) find that Catholics have equally high wellbeing as Protestants, but that both Catholics and Protestants have higher wellbeing than adherents of other religions and non-religious people. Other research finds similar albeit slightly different results (see Graham and Crown 2014). On the other hand, Becker and Woessmann (2015) find that Protestants are more likely to commit suicide than Catholics, which they argue is due to Catholics' stronger levels of social cohesion. Regardless, for our purposes, the direct association between religious affiliation and wellbeing is less important since our instrument is based on historical Catholic population shares – which, if anything, appear to be positively related to wellbeing.

Our analysis of a wellbeing-efficiency trade-off hinges upon a causal interpretation also of the effects of independent-school competition on academic efficiency. Again, the strategy is more likely to work against finding evidence for our hypothesis. This is because Catholics historically emphasised cognitive achievement less compared with other groups, as indicated by the direct negative correlation between Catholic population shares and literacy rates in the early 20th century (see West and Woessmann 2010). This makes it likely that our estimates for academic outcomes will be negatively biased. Since the instrument's logic partly hinges on that Catholics historically lobbied governments to increase public funding for independent schools, our strategy is also more likely to bias effects upward in the analysis of per-pupil educational expenditures (and thus against our hypothesis). Again, we can indirectly test whether or not these intuitions are correct. If so, we expect the contemporary version of our instrument, which is controlled for in equation (3), to be negatively (positively) related to achievement (expenditures) today. If this holds true, it suggests that our historical instrument is only positively (negatively) related to contemporary PISA scores (expenditures) through independent-school competition. If anything, the results should then be biased against finding evidence supporting our hypothesis.

3.4.5. Balancing tests

Another way to explore whether or not the instrument is exogenous, once controlling for the other relevant country- and school-level variables, is to carry out balancing tests on the pupil-background characteristics that are included in the models analysing the effects of independent-school competition on pupil wellbeing and academic efficiency. We do so by swapping these indicators as dependent variables for each of the pupil-background variables included in the main regressions, while simultaneously excluding all other pupil-level variables on the right-hand side of the equation.[19] If the variation in independent-school competition predicted by the instrument is not significantly related to the pupil-background indicators, once adjusting for the other country- and school-level variables included in the model, it indicates that the instrument is indeed likely to be exogenous.

[19] Note that we do not use any imputed data in these analyses. In the models analysing immigrant background, we also exclude pupils from the other immigrant category to ensure that we compare each category with natives only.

3.5. Results

3.5.1. Pupil happiness

As a starting point, Columns 1 and 2 in Table 3.1 show the results from the OLS model when analysing our measure of pupil wellbeing: happiness at school. The estimates indicate that independent-school competition is negatively associated with pupil happiness, regardless of whether we control only for broader regional-fixed effects or also include controls for regions within Europe. Thus, there is a negative correlation between competition and pupil wellbeing across OECD countries.

Turning to the IV model, the first-stage results suggest our instrument is strong, with the F statistics displaying values of about 46. Meanwhile, the second stage displays that the coefficient for independent-school competition increases in size compared with the OLS estimates, indicating that the latter are biased downwards. Our preferred specification, which includes within-European regional-fixed effects, indicates that a 10 percentage-point increase in independent-school enrolment shares lowers pupil happiness by 0.13 points on the ordinal 1–4 scale, which corresponds to 0.17 standard deviations (SD). The estimate is not very precise, but we can rule out an effect size lower than 0.09 SD. The reduced-form estimates in the lower-left panel further support the results.

One potential reason why the OLS models underestimate the causal impact of independent-school competition on pupil wellbeing may be that competition emerges as a response to low test scores (see Hoxby 1994; West and Woessmann 2010). This, in turn, may be due to lower focus on academic achievement and higher focus on wellbeing, as suggested by the trade-off hypothesis discussed in Section 3.2 and analysed further in Section 3.5.4. If so, one would expect OLS estimates to bias the effects of competition towards zero both when analysing pupil wellbeing and test scores, albeit from different directions.

Thus, our analysis shows that independent-school competition has an important negative impact on pupil wellbeing. In contrast, we note that the association between the contemporary version of the instrument and pupil happiness is positive. This supports the idea that our strategy if anything may bias the estimates against finding evidence in favour of our hypothesis.[20]

Table 3.1: The impact of independent-school competition on pupil happiness

	(1)	(2)		(3)	(4)
	OLS	OLS	*Second stage*	IV	IV
Independent-school share	-0.52**	-0.69***	Independent-school share	-0.79***	-1.28***
	(0.24)	(0.17)		(0.22)	(0.32)
Catholic share 2010 *no state religion	0.22**	0.36***	Catholic share 2010 *no state religion	0.33***	0.49***
	(0.11)	(0.10)		(0.08)	(0.10)
Within-European regional-fixed effects	NO	YES		NO	YES
	(5)	(6)			
Reduced form	OLS	OLS	*First stage*		
Catholic share 1900 *no state religion	-0.24***	-0.29***	Catholic share 1900 *no state religion	0.31***	0.23***
	(0.06)	(0.08)		(0.05)	(0.03)
Catholic share 2010 *no state religion	0.15***	0.21*	Catholic share 2010 *no state religion	0.22***	0.21***
	(0.06)	(0.11)		(0.05)	(0.08)
			F statistic on the excluded instrument	46.33	45.87
n	190,348	190,348		190,348	190,348
Countries	34	34		34	34

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the country level in parentheses. All regressions include the controls described in Sections 3.3.5 and 3.4.1.

3.5.2. Robustness tests

In Table 3.2, we display results from several robustness tests.[21] Column 1 only includes pupils in government-operated schools. Note that these results reflect both competition effects and the impact of potential differential pupil sorting into the independent and state sectors. It is plausible that state schools could be especially sensitive to independent-school competition, since independent schools themselves may to some extent always face competition from the government sector. We find that the effect on pupils in state schools is in fact very similar compared to the overall impact, suggesting that the main results primarily reflect system-level effects of independent-school competition rather than the impact of attending an independent school per se.

[20] As displayed in Table 3.A.2, we find similar effects on alternative measures of pupil wellbeing, such as school satisfaction and peer relations.

[21] In unreported robustness tests, we also included additional country-level controls, including the Gini coefficient, the share of population in urban areas, and the relative size of the immigrant population. The results were very similar compared to the baseline models.

Next, in Column 2, we exclude all non-European countries in the equation, dropping ten countries and 36 per cent of the total pupil sample. The results are robust to this exercise, despite controlling for within-European regional-fixed effects. In Column 3, we instead exclude Belgium and the Netherlands, which are perhaps the most canonical examples of successful Catholic struggles for independent-school access, to ensure these countries do not drive the results. The findings are essentially identical when excluding these countries.[22] We thus conclude that the results are robust to excluding a considerable and important part of the sample.[23]

Table 3.2: Robustness tests

	(1)	(2)	(3)	(4)
	Only pupils in state schools	Only Europe	Excluding Belgium and the Netherlands	Excluding pupil-background variables
Independent-school share	-1.37***	-1.27***	-1.30***	-1.24***
	(0.38)	(0.43)	(0.32)	(0.33)
Catholic share 2010 *no state religion	0.46***	0.54***	0.47***	0.52***
	(0.09)	(0.13)	(0.08)	(0.10)
F statistic	48.62	42.27	42.5	45.23
n	150,231	121,050	182,146	190,348
Countries	34	24	32	34

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the country level in parentheses. All regressions include the controls outlined in Sections 3.3.5 and 3.4.1, including within-European regional-fixed effects. The exception is Column 4, which excludes pupil-background variables.

Finally, in Column 4, we exclude all pupil-background controls and find that the coefficient is essentially identical to the baseline, although it becomes slightly less precise. This is expected if some or all of the excluded background characteristics both affect the outcome variable, which unreported estimates show is indeed the case, and are uncorrelated with our instrument (Angrist and Pischke 2009). Overall, these results thus support the idea that the instrument is not significantly correlated with the pupil-background characteristics, once controlling for the other variables.[24]

[22] When dropping Ireland, another example of successful Catholic resistance, the coefficient increased in size somewhat and remained significant at the 1 per cent level, while the F statistics drops just under 20. Similarly, excluding Ireland together with Belgium or the Netherlands – or excluding all three countries at the same time – generated similar results. The same holds true when excluding the within-European regional-fixed effects and the F statistic never fell under 25 in these instances.

[23] In unreported regressions, we also dropped all countries one by one – and various combinations of countries – and the estimates were always robust to this exercise.

[24] We also tested the idea that our instrument isolates similar types of independent-school competition across countries by adding relevant variables that may affect the type of competition. The results are reported in Table 3.A.3. All estimates are basically identical compared to the relevant baseline models.

3.5.3. Balancing tests

As discussed in Section 3.4.5, to further explore the exogeneity of the instrument, we analyse the pupil-background characteristics as dependent variables instead of controls, while simultaneously excluding all other pupil-level variables on the right-hand side of the equation. We do not expect the variation in independent-school competition that is explained by our instrument to be significantly related to these variables, once other country-level factors are held constant, at least not in a direction that would bias the estimates in favour of finding evidence of our hypothesis.[25] The results from the IV models in Table 3.3 display that this is indeed the case. We are unable to predict any of the background variables using the variation in independent-school enrolment shares that is explained by our instrument. In other words, there is little evidence that the instrument is significantly correlated with potentially important predictors of wellbeing in a direction that would bias the estimates in favour of our hypothesis.[26]

Table 3.3: Balancing tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Gender	Age	Index of home possessions	Parental occupational status	Parental education	Immigrant (1st gen)	Immigrant (2nd gen)
Independent-school share	0.02	0.06	0.01	-1.41	3.07	-0.07	-0.16
	(0.04)	(0.10)	(0.49)	(9.69)	(2.36)	(0.12)	(0.11)
Catholic share 2010*no state religion	-0.02	0.00	0.38***	1.81	-0.95	-0.05	-0.05
	(0.02)	(0.03)	(0.13)	(2.15)	(0.62)	(0.05)	(0.04)
F statistic	47.19	47.18	46.99	46.71	46.54	46.27	44.73
n	295,416	295,330	291,731	280,796	285,877	244,043	269,794
Countries	34	34	34	34	34	34	34

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the country level in parentheses. All regressions include the controls described in Sections 3.3.5 and 3.4.1, including within-European regional-fixed effects, apart from pupil-level variables.

[25] The validity of our results hinges on the instrument being exogenous once the country-level variables are held constant, which we test by studying the pupil-background characteristics as dependent variables. Nevertheless, in unreported robustness tests, we also analysed the correlation between the instrument and relevant country-level variables, such as (log) GDP per capita, income inequality, the share of the population who live in urban areas, and the relative size of the immigrant population, including and excluding the contemporary version of the instrument and regional-fixed effects. The results did not indicate consistently significant correlations.

[26] In contrast, we note that the contemporary version of the instrument is positively correlated with the index of home possessions, which is the pupil-level variable that has the strongest positive association with the wellbeing measures of all pupil-level controls in the regressions in Tables 3.1–3.2.

3.5.4. *The trade-off with academic efficiency*

Thus far, we have shown that independent-school competition has a negative causal impact on pupil wellbeing. Since previous research using a similar strategy finds positive effects on pupil performance in PISA and a negative impact on educational expenditures (West and Woessmann 2010), this is sufficient to provide general support for our hypothesis of a wellbeing-efficiency trade-off. Still, since we analyse data from PISA 2012 rather than from PISA 2003, and use a modified IV set-up, we also explore the effects of competition on PISA test scores and per-pupil expenditure, using our preferred specification for the analysis of pupil-wellbeing.

The upper panel in Table 3.4 shows OLS estimates, which indicate that independent-school competition does not have a statistically significant relationship with pupil performance, apart from being marginally correlated with reading literacy, but that it is negatively associated with educational expenditures. However, turning to the preferred IV estimates in the lower panel, the coefficients increase in size and become strongly significant in all models: a 10 percentage-point increase in independent-school enrolment shares raises mathematical literacy by 21 PISA points (0.23 SD), reading literacy by 26 PISA points (0.28 SD), and scientific literacy by 18 PISA points (0.19 SD). Simultaneously, it lowers expenditures by $13,155 (0.48 SD). These effects are larger than those found by West and Woessmann (2010), which is mainly due to our inclusion of within-European regional-fixed effects. Indeed, when excluding these dummies, the effects are very similar to their results.[27] We also note that the contemporary version of the instrument is positively associated with expenditures, while its association with achievement is negative but not statistically significant.[28] Overall, the results are in line with West and Woessmann's (2010) findings and thus support our hypothesis that school competition involves a causal wellbeing-efficiency trade-off.[29]

[27] They obtain point estimates of 58.99-121.69 and -45,736 for PISA scores and expenditures respectively in their equivalent analyses (see Column 2 in their Tables 2 and 5 and Columns 2 and 5 in their Table 4), whereas we obtain 68.96–126.65 and -74,201 respectively when excluding within-European regional effects. None of the differences are statistically significant.

[28] Again, Table 3.A.4 shows that the results comfortably survive the other robustness tests conducted in regard to pupil wellbeing in Table 3.2. Unreported regressions also showed that the findings were robust to adding the additional variables in Table 3.A.3. In contrast, the coefficient of the contemporary instrument changes depending on the specification and sample (see Tables 3.A.4 and 3.A.5). This further supports the idea that the historical instrument is exogenous, while its contemporary version is not.

[29] We also carried out a basic placebo test on per-capita military expenditures in 2011, obtained from SIPRI (2016) and adjusted for different price levels using 2011 GDP per capita PPPs. Military

Table 3.4: The impact of independent-school competition on academic efficiency

	Mathematics	Reading	Science	Educational expenditures/pupil
	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
Independent-school share	46.86	56.70*	41.84	-37,214***
	(33.30)	(31.24)	(32.82)	(13,064)
Catholic share 2010 *no state religion	21.97	9.04	23.43	14,466
	(20.71)	(24.51)	(17.01)	(12,918)
	(5)	(6)	(7)	(8)
	IV	IV	IV	IV
Independent-school share	209.57***	262.10***	177.03***	-131,546***
	(51.62)	(69.56)	(50.14)	(28,377)
Catholic share 2010 *no state religion	-12.13	-34.01	-4.90	34,235***
	(20.59)	(27.97)	(18.94)	(10,801)
n	295,416	295,416	295,416	295,416
Countries	34	34	34	34

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the country level in parentheses. The regressions include the controls described in Sections 3.3.5 and 3.4.1, including within-European regional-fixed effects. F statistics in Columns 5–8: 46.54.

3.5.5. Potential mechanisms behind the trade-off

Finally, we turn to the potential mechanisms discussed in Section 3.3.3. The results in Table 3.5 indicate that a 10 percentage-point increase in independent-school enrolment shares induces less individualisation of teaching, corresponding to 0.16 points on the 1–4 ordinal scale (0.15 SD); less project work, corresponding to 0.09 points (0.10 SD); and less group work, corresponding to 0.08 points (0.08 SD). However, there is no impact on the extent to which teachers ask pupils to help plan classroom activities. Also, a 10 percentage-point increase in independent-school enrolment shares decreases perceptions that pupils get along with teachers by 0.08 steps on the 1–4 ordinal scale (0.12 SD) and perceptions that teachers listen to what pupils have to say by 0.07 steps (0.09 SD).[30] Meanwhile, it raises parental achievement pressure by 0.19 steps on the 1–3 ordinal scale (0.26 SD). This indicates that competition makes teaching more traditional and pupil-teacher relations more hierarchical, while sharpening parents' focus on achievement – which are clear mechanisms behind the wellbeing-efficiency trade-off.

expenditures are unlikely to be related to educational expenditures and appear to be an appropriate placebo outcome. We found no evidence indicating that the variation in independent-school competition that is explained by our instrument was related to military expenditures.

[30] In unreported regressions, we found very similar effects on the overall index of pupil-teacher relationships and headteachers' perceptions of such relationships.

Table 3.5: The impact on potential mechanisms behind the trade-off

	Teaching practices			
	(1)	(2)	(3)	(4)
	Individualisation of teaching	Project work	Group work	Pupils help to plan
Independent-school share	-1.57***	-0.90**	-0.81***	1.10
	(0.37)	(0.43)	(0.28)	(1.04)
Catholic share 2010 *no state religion	-0.05	-0.17	0.07	-0.60
	(0.10)	(0.13)	(0.10)	(0.36)
n	191,806	191,799	191,865	191,832
Countries	34	34	34	34
	Pupil-teacher relations, parental achievement pressure, and homework			
	(5)	(6)	(7)	(8)
	Pupils get along with teachers	Teachers listen to pupils	Parental achievement pressure	Hours of homework
Independent-school share	-0.84***	-0.68***	1.86***	8.18**
	(0.17)	(0.16)	(0.46)	(3.30)
Catholic share 2010 *no state religion	-0.02	-0.01	-0.47**	1.95
	(0.08)	(0.06)	(0.23)	(1.24)
n	187,146	191,320	282,606	187,146
Countries	34	34	34	34
	Instructional time			
	(9)	(10)	(11)	(12)
	Class periods (total)	Class periods (mathematics)	Class periods (test language)	Class periods (science)
Independent-school share	23.43***	4.53***	5.00**	-2.83***
	(6.75)	(1.47)	(1.94)	(0.98)
Catholic share 2010 *no state religion	2.32	-1.12*	-0.71	-1.31***
	(2.09)	(0.59)	(0.73)	(0.40)
n	162,430	184,354	183,030	179,223
Countries	34	34	34	34

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the country level in parentheses. The regressions include the controls described in Sections 3.3.5 and 3.4.1, including within-European regional-fixed effects. The F statistic ranges between 40.46 and 46.06 in all regressions.

We also find that competition increases instructional time and homework. An increase in independent-school enrolment shares by 10 percentage points increases the total number of class periods by 2.43 periods per week (0.30 SD), while inducing pupils to complete 0.82 hours (0.16 SD) more homework per week. However, there is heterogeneity in terms of the effects on the number of class periods in the subjects tested in PISA. While the number of periods in mathematics and test language increases by 0.45 periods (0.32 SD) and 0.50 periods (0.34 SD) per week respectively, the number of class periods decreases in science by 0.28 periods (0.13 SD). This indicates that competition increases instructional time in mathematics and test language, but decreases it in science.[31] This could suggest that competition increases schools' focus on core subjects to the detriment of other subjects. Still, instruction in the test language and mathematics (and other subjects) may improve performance also in science.[32] Overall, we thus conclude that competition has positive effects on instructional time and homework, two plausible mechanisms behind the wellbeing-efficiency trade-off.

3.5.6. A tentative cost-benefit analysis

Ultimately, the study's findings demand the question: should policymakers increase independent-school competition and thus raise academic efficiency or should they ignore such reforms and instead prioritise pupil wellbeing? The answer depends on the relative long-term societal and economic value of pupil wellbeing versus cognitive achievement in adolescence. In this section, we thus provide a basic back-of-the-envelope calculation to analyse whether the benefits of competition in terms of academic efficiency outweigh its costs in terms of pupil wellbeing.

Recent research indicates that cognitive achievement in childhood and adolescence is a much better predictor than wellbeing in childhood and adolescence of adult income. According to Layard et al.'s (2014) estimates, one standard deviation higher cognitive achievement in childhood and adolescence predicts 0.14 SD higher income at the age of 34, while such an increase in youth wellbeing is associated with 0.07 SD higher income at the same age. Our estimates indicate that a 10 percentage-point increase in independent-school competition raises average test scores by 0.23 SD and decreases pupil wellbeing by 0.17 SD. One would thus expect a benefit in terms of adult income by 0.03 SD via higher test scores and a cost of 0.01 SD via lower pupil wellbeing. Since we also find that independent-school competition decreases per-pupil cumulative education expenditures between ages 6–15, such competition thus appears to make sense from an economic perspective.

At the same time, Layard et al. (2014) also find that youth wellbeing is considerably more important than cognitive achievement for adult life satisfaction. A cost-benefit analysis using adult subjective wellbeing rather than money as unit of measurement

[31] In unreported regressions, we found no effects on average minutes per period in any of the test subjects, supporting the idea that our estimates capture the impact of competition on total learning time.

[32] However, note that the point estimate in Table 3.4 is smaller when analysing science scores. The negative impact we find on instructional time in science may thus lower the positive effects of competition.

would suggest that a 10 percentage-point increase in independent-school competition should generate 0.01 SD higher life satisfaction via higher cognitive achievement – but this is outweighed by the cost of 0.03 SD via lower pupil wellbeing.[33] In other words, if we hold subjective wellbeing as the primary goal of policy, the costs of independent-school competition may outweigh its benefits.

Certainly, given the tentative nature of the above cost-benefit analysis, it is important to pursue further research before drawing strong conclusions regarding the potential longer-term effects of independent-school competition on adult wellbeing and labour-market outcomes.[34] Yet the analysis at least indicates that the attractiveness of school competition as an education-reform strategy may depend on which goals policymakers seek to advance, which is beyond this paper to determine.

3.6. Conclusion

As governments worldwide have sought to inject competition from independent providers into their countries' education systems, an expanding literature has begun to evaluate the effects of such competition. Yet existing research focuses on academic outcomes and no one has thus far analysed how independent-school competition affects pupil wellbeing, which has become an increasingly important policy goal recently. Since effective learning involves many activities that are not necessarily fun or inspiring, and since market incentives are likely to sharpen schools' focus on academic achievement, it is plausible that competition involves a trade-off between wellbeing and education performance.

Analysing pupil-level PISA data across 34 OECD countries, this paper has sought to investigate the existence of such a trade-off and potential mechanisms behind it. It utilised an IV strategy based on Catholic resistance to state schooling in the 19th and early 20th centuries to predict enrolment shares in independently-operated schools today, while simultaneously controlling for the contemporary version of the instrument itself and other important variables that threaten its validity. We found that

[33] Note that the calculation is based on the direct correlation between youth wellbeing/cognitive achievement and adult life satisfaction, which means that any effects that operate via higher income in adulthood are incorporated in the calculation automatically. Similarly, the calculation regarding the impact of independent-school competition on adult income incorporates the latter's effect on life satisfaction automatically.

[34] For example, the cost-benefit analysis treats wellbeing in school as equal to the more general child wellbeing metrics employed by Layard et al. (2014). Further research is necessary to establish to what extent this matters for the results.

independent-school competition has a sizeable negative impact on pupil wellbeing, which survives a number of robustness tests. The paper further confirmed a positive effect on PISA scores and a negative impact on education spending found in previous research, thus providing clear evidence of a trade-off.

We also showed that balancing tests on pupil-background variables support the causal interpretation of our findings. In fact, if anything, there are more indications that our strategy may bias estimates against finding evidence in favour of our hypothesis. Nevertheless, future research should investigate whether or not alternative data and identification strategies generate similar results.

Analysing relevant mechanisms behind the wellbeing-efficiency trade-off, we found that independent-school competition makes teaching more traditional and pupil-teacher relationships more hierarchical, while also increasing parental achievement pressure. In addition, we found positive effects on instructional time and time spent on homework. These are all features that previous research suggests generate higher achievement and lower wellbeing. Future research should investigate other mechanisms linking competition to lower wellbeing and higher academic efficiency – and to what extent similar trade-offs apply to other education-reform strategies.

A tentative back-of-the-envelope calculation indicated that the economic benefits of independent-school competition via its positive impact on cognitive achievement appear to outweigh its cost via lower pupil wellbeing. At the same time, the calculation also indicates that the costs of competition may outweigh its benefits when using adult life satisfaction as the unit of measurement. While more research into this issue is necessary, justifying the higher direct and indirect costs of a non-competitive education system may hinge on upholding subjective wellbeing as a primary goal for public policy. While we refrain from drawing strong conclusions in this respect, our results highlight the potential for a more general trade-off between the traditional goals of education policy and the wellbeing agenda to which policymakers should pay attention – regardless of what goals they ultimately choose to pursue.

References

Abdulkadiroğlu, Atila, Joshua D. Angrist, Peter D. Hull, and Parag A. Pathak. 2016. 'Charters Without Lotteries: Testing Takeovers in New Orleans and Boston.' *American Economic Review* 106(7):1878–1920.

Adnett, Nick and Peter Davies. 2002. *Markets for Schooling: An Economic Analysis*. London: Routledge.

Alesina, Alberto, Rafael Di Tella, and Robert MacCulloch. 2004. 'Inequality and Happiness: Are Europeans and Americans Different?' *Journal of Public Economics* 2009–2042.

Algan, Yann, Pierre Cahuc, and Andrei Shleifer. 2013. 'Teaching Practices and Social Capital.' *American Economic Journal: Applied Economics* 5(3):189–210.

Allen, Rebecca and Anna Vignoles. 2015. 'Can School Competition Improve Standards? The Case of Faith Schools in England.' *Empirical Economics* doi: 10.1007/s00181-015-0949-4.

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters. 2013. 'Explaining Charter School Effectiveness.' *American Economic Journal: Applied Economics* 5(4):1–27.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Atherton, Paul, Simon Appleton, and Michael Bleaney. 2013. 'International School Test Scores and Economic Growth.' *Bulletin of Economic Research* 65(1):82–90.

Aucejo, Esteban M. and Teresa F. Romano. 2014. 'Assessing the Effect of School Days and Absences on Test Score Performance.' CEP Discussion Paper No 1302.

Böhlmark, Anders, Helena Holmlund, and Mikael Lindahl. 2016. 'Parental Choice, Neighbourhood Segregation or Cream Skimming? An Analysis of School Segregation after a Generalized Choice Reform.' *Journal of Population Economics* 29(4):1155–1190.

Böhlmark, Anders and Mikael Lindahl. 2015. 'Independent Schools and Long-Run Educational Outcomes: Evidence from Sweden's Large-Scale Voucher Reform.' *Economica* 82(327):508–551.

Barrett, David B., George T. Kurian, and Todd M. Johnson. 2001. *World Christian Encyclopedia*. 2nd ed. Oxford: Oxford University Press.

Barr-Melej, Patrick M. 2002. *Reforming Chile: Cultural Politics, Nationalism, and the Rise of the Middle Class*. Chapel Hill, NC: University of North Carolina Press.

Barro, Robert J. and Rachel M. McCleary. 2005. 'Which Countries Have State Religions?' *Quarterly Journal of Economics* 120(4):1331–1370.

Becker, Sascha O. and Ludger Woessmann. 2015. 'Social Cohesion, Religious Beliefs, and the Effect of Protestantism on Suicide.' CESifo Working Paper No. 5288, Munich.

Bietenback, Jan. 2014. 'Teacher Practices and Cognitive Skills.' *Labour Economics* 30:143–153.

Bol, Thijs and Herman G. Van de Werfhorst. 2013. 'The Measurement of Tracking, Vocational Orientation, and Standardization of Educational Systems: A Comparative Approach.' GINI Discussion Paper No. 81.

Brandt, Nicola. 2010. 'Climbing on Giants' Shoulders: Better Schools for all Chilean Children.' Economics Department Working Paper No. 784, Paris.

Brown, Davis and Patrick James. 2015. 'Religious Characteristics of States Dataset, Phase 1: Demographics (RCS).' Dataset, Department of Political Science, Maryville University of St Louis.

Brunello, Giorgio, Margherita Fort, Nicole Schneeweis, and Rudolf Winter-Ebmer. 2016. 'The Causal Effect of Education on Health: What is the Role of Health Behaviors?' *Health Economics* 25(3):314–336.

Card, David. 1999. 'The Causal Effect of Education on Earnings.' *Handbook of Labor Economics* 3(A):1801–1863.

Carneiro, Pedro, Claire Crawford, and Alissa Goodman. 2007. 'The Impact of Early Cognitive and Non-cognitive Skills on Later Outcomes.' London.

Chakrabarti, Rajashri. 2008. 'Can Increasing Private School Participation and Monetary Loss in a Voucher Program Affect Public School Performance? Evidence from Milwaukee.' *Journal of Public Economics* 92(5–6):1371–1393.

Christodoulou, Daisy. 2014. *Seven Myths About Education*. London: Routledge.

Chubb, John E. and Terry M. Moe. 1988. 'Politics, Markets and the Organization of Schools.' *American Political Science Review* 82(4):1065–1087.

Clark, Andrew E. and Orsolya Lelkes. 2009. 'Let Us Pray: Religious Interactions in Life Satisfaction.' Working Paper No. 2009–01.

Cohen-Zada, Danny. 2009. 'An Alternative Instrument for Private School Competition.' *Economics of Education Review* 28(1):29–37.

Cohen-Zada, Danny and Todd Elder. 2009. 'Historical Religious Concentrations and the Effects of Catholic Schooling.' *Journal of Urban Economics* 66(1):65–74.

Collier, Simon 1997. 'Religious Freedom, Clericalism, and Anticlericalism in Chile, 1820–1920.' Pp. 302–338 in *Freedom and Religion in the Nineteenth Century*, edited by Richard Helmstadter. Stanford, California: Stanford University Press.

Csikszentmihalyi, Mihaly and Jeremy Hunter. 2003. 'Happiness in Everyday Life: The Uses of Experience Sampling.' *Journal of Happiness Studies* 4(2):185–199.

Eyles, Andrew and Stephen Machin. 2015. 'The Introduction of Academy Schools to England's Education.' CEP Discussion Paper No 1368.

Falch, Torberg and Marte Rønning. 2012. 'Homework Assignment and Student Achievement in OECD Countries.' Discussion Paper No. 711, Oslo.

Falck, Oliver and Ludger Woessmann. 2013. 'School Competition and Students' Entrepreneurial Intentions: International Evidence Using Historical Catholic Roots of Private Schooling.' *Small Business Economics* 40(2):459–478.

Frey, Bruno S. and Alois Stutzer. 2002. 'What Can Economists Learn from Happiness Research?' *Journal of Economic Literature* 40(2):402–435.

Friedman, Milton. 1962. *Capitalism and Freedom*. Chicago: University of Chicago Press.

Frijters, Paul, David W. Johnston, and Michael A. Shields. 2014. 'Does Childhood Predict Adult Life Satisfaction? Evidence from British Cohort Surveys.' *Economic Journal* 124(580):F688–F719.

Gauri, Varun. 1998. *School Choice In Chile: Two Decades of Educational Reform*. Pittsburgh, PA: University of Pittsburgh Press.

Gibbons, Stephen and Olmo Silva. 2011. 'School Quality, Child Wellbeing and Parents' Satisfaction.' *Economics of Education Review* 30(2):312–331.

Gilman, Rich and Scott Huebner. 2003. 'A Review of Life Satisfaction Research with Children and Adolescents.' *School Psychology Quarterly* 18(2):192–205.

Glenn, Charles L. 1989. *Choice of Schools in Six Nations: France, Netherlands, Belgium, Britain, Canada, West Germany*. Washington, DC: US Department of Education.

Glenn, Charles L. 2011. *Contrasting Models of State and School: A Comparative Historical Study of Parental Choice and State Control*. New York: Continuum.

Graham, Carol and Sarah Crown. 2014. 'Religion and Wellbeing Around the World: Social Purpose, Social Time, or Social Insurance?' *International Journal of Wellbeing* 4(1):1–27.

Green, Colin P., Navarro-Paniagua, Domingo P. Ximénez-de-Embún, and María-Jesús Mancebón. 2014. 'School Choice and Student Wellbeing.' *Economics of Education Review* 38:139–150.

Gustafsson, Jan-Eric. 2013. 'Causal Inference in Educational Effectiveness Research: A Comparison of Three Methods to Investigate Effects of Homework on Student Achievement.' *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice* 24(3):275–295.

Hanushek, Eric A., Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann. 2015. 'Returns to Skills Around the World: Evidence from PIAAC.' *European Economic Review* 73:103–130.

Hanushek, Eric A. and Ludger Woessmann. 2011. 'The Economics of International Differences in Educational Achievement.' *Handbook of the Economics of Education* 3:89–200.

Hanushek, Eric A. and Ludger Woessmann. 2012. 'Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation.' *Journal of Economic Growth* 17:267–321.

Hanushek, Eric A. and Ludger Woessmann. 2016. 'Knowledge Capital, Growth, and the East Asian Miracle.' *Science* 351(6271):344–345.

Heckman, James J. and Tim Kautz. 2013. 'Fostering and Measuring Skills: Interventions That Improve Character and Cognition.' NBER Working Paper No. 19656, Cambridge, MA.

Heller-Sahlgren, Gabriel. 2013. *Incentivising Excellence: School Choice and Education Quality*. London: CMRE & IEA.

Hoxby, Caroline M. (1994). Do Private Schools Provide Competition for Public Schools? NBER Working Paper No. 4978.

Hoxby, Caroline M. 2003. 'School Choice and School Competition: Evidence From the United States.' *Swedish Economic Policy Review* 10(2):9–65.

Hsieh, Chang-Tai and Miguel Urquiola. 2006. 'The Effects of Generalized School Choice on Achievement.' *Journal of Public Economics* 90:1477–1503.

Huebner, Scott. 2004. 'Research on Assessment of Life Satisfaction of Children and Adolescents.' *Social Indicators Research* 66(1):3–33.

Huebner, E. Scott and Carol Diener 2008. 'Research on Life Satisfaction of Children and Youth: Implications for the Delivery of School-Related Services.' Pp. 393–413 in *The Science of Subjective Well-Being*, edited by Michael Eid and Randy J Larsen. New York: Guildford Press.

Huebner, E. Scott and Rich Gilman. 2006. 'Students Who Like and Dislike School.' *Applied Research in Quality of Life* 1(2):139–150.

Huebner, E. Scott, Kimberly J. Hills, Xu Jiang, Rachel F. Long, Ryan Kelly, and Michael D. Lyons 2014. 'Schooling and Children's Subjective Well-Being.' Pp. 797–819 in *Handbook of Child Well-Being: Theories, Methods and Policies in Global Perspective*, edited by Asher Ben-Arieh, Ferran Casas, Ivar Frønes, and Jill E Korbin. Dordrecht: Springer.

Ingvar, Martin and Gunilla Eldh. 2014. *Hjärnkoll på skolan*. Stockholm: Natur & Kultur.

Jürges, Hendrik and Kerstin Schneider. 2010. 'Central Exit Examinations Increase Performance. But Take the Fun Out of Mathematics.' *Journal of Population Economics* 23(2):497–517.

Jürges, Hendrik, Kerstin Schneider, Martin Senkbeil, and Claus H. Carstensen. 2012. 'Assessment Drives Learning: The Effect of Central Exit Exams on Curricular Knowledge and Mathematical Literacy.' *Economics of Education Review* 31(1):56–65.

Jiang, Feng and William F. McComas. 2015. 'The Effects of Inquiry Teaching on Student Science Achievement and Attitudes: Evidence from Propensity Score Analysis of PISA Data.' *International Journal of Science Education* 37(3):554–576.

Jones, Peter. 2013. 'Adult Mental Health Disorders and their Age at Onset.' *British Journal of Psychiatry* 202:5–10.

Kaiser, Wolfram and Helmut Wohnout, eds. 2004. *Political Catholicism in Europe 1918–45*. London: Routledge.

Kirschner, Paul A., John Sweller, and Richard E. Clark. 2006. 'Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching.' *Educational Psychologist* 41(2):75–86.

Krueger, Alan B. and David A. Schkade. 2008. 'The Reliability of Subjective Well-being Measures.' *Journal of Public Economics* 92(8–9):1833–1845.

Kuehn, Zoe and Pedro Landeras. 2012. 'Study Time and Scholarly Achievement.' MPRA Paper No. 49033, Munich.

Lévy-Garboua, Louis, Youenn Lohéac, and Bertrand Fayolle. 2006. 'Preference Formation, School Dissatisfaction and Risky Behavior of Adolescents.' *Journal of Economic Psychology* 27(1):165–183.

Lavy, Victor. 2010. 'Effects of Free Choice Among Public Schools.' *Review of Economic Studies* 77:1164–1191.

Lavy, Victor. 2015. 'Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries.' *Economic Journal* 125(588):F397–F424.

Layard, Richard, Andrew E. Clark, Francesca Cornaglia, Nattavudh Powdthavee, and James Vernoit. 2014. 'What Predicts a Successful Life? A Life-course Model of Well-being.' *Economic Journal* 124(580):F720–F738.

Le Grand, Julian. 2007. *The Other Invisible Hand: Delivering Public Services Through Choice and Competition*. Princeton, NJ: Princeton University Press.

Locke, Thomas F. and Michael E. Newcomb. 2004. 'Adolescent Predictors of Young Adult and Adult Alcohol Involvement and Dysphoria in a Prospective Community Sample of Women.' *Prevention Science* 5(3):151–168.

Mintz, Avi I. 2012. 'The Happy and Suffering Student? Rousseau's Emile and the Path not Taken in Progressive Educational Thought.' *Educational Theory* 62(3):249–265.

Muralidharan, Karthik and Venkatesh Sundararaman. 2015. 'The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India.' *Quarterly Journal of Economics* 130(3):1011–1066.

Neal, Derek. 2002. 'How Vouchers Could Change the Market for Education.' *Journal of Economic Perspectives* 16(4):25–44.

OECD. 2009. 'OECD Economic Surveys: Mexico.' Report, Paris.

OECD. 2013. 'PISA 2012 Results: Ready to Learn.' Report, Paris.

OECD. 2014. 'PISA 2012 Technical Report.' Report, Paris.

OECD. 2016a. Data retrieved from the OECD's PISA database on 20 December 2015: http://www.oecd.org/pisa/pisaproducts/.

OECD. 2016b. Data retrieved from the OECD database on 20 December 2015: https://stats.oecd.org/Index.aspx?DataSetCode=PDB_LV.

Oreopoulos, Philip and Kjell G. Salvanes. 2011. 'Priceless: The Nonpecuniary Benefits of Schooling.' *Journal of Economic Perspectives* 25(1):159–184.

Payton, John, Roger P. Weissberg, Joseph A. Durlak, Allison B. Dymnicki, Rebecca D. Taylor, Kriston B. Schellinger, and Molly Pachan. 2008. 'The Positive Impact of Social and Emotional Learning for Kindergarten to Eighth-Grade Students: Findings from Three Scientific Reviews. Technical Report.' Research Report, Chicago, IL.

Public Health England. 2014. 'The Link between Pupil Health and Wellbeing and Attainment: A Briefing for Head Teachers, Governors and Staff in Education Settings. Report, Public Health England, London.

Regh, Mariella. 2012. 'Primary School Teaching Practices and Social Capital.' Thesis, Sciences Po.

Rivkin, Steven G. and Jeffrey C. Schiman. 2015. 'Instruction Time, Classroom Quality, and Academic Achievement.' *Economic Journal* 125(588):F425–F448.

Schwerdt, Guido and Amelie C. Wuppermann. 2011. 'Is Traditional Teaching All That Bad? A Within-student Between-subject Approach.' *Economics of Education Review* 30(2):365–379.

SIPRI. 2016. 'SIPRI Military Expenditure Database.' Database available at: http://www.sipri.org/research/armaments/milex/milex_database, Stockholm.

Sweller, John, Paul A. Kirschner, and Richard E. Clark. 2007. 'Why Minimally Guided Teaching Techniques Do Not Work: A Reply to Commentaries.' *Educational Psychologist* 42(2):115–121.

Takakura, Minoru, Norie Wake, and Minoru Kobayashi. 2010. 'The Contextual Effect of School Satisfaction on Health-Risk Behaviors in Japanese High School Students.' *Journal of School Health* 80(11):544–541.

Warton, Pamela M. 2001. 'The Forgotten Voices in Homework: Views of Students.' *Educational Psychologist* 36(3):155–165.

Veenhoven, Ruut. 2012. 'Cross-national Differences in Happiness: Cultural Measurement Bias or Effect of Culture?' *International Journal of Wellbeing* 2(4):333–353.

West, Martin R. and Ludger Woessmann. 2010. ''Every Catholic Child in a Catholic School': Historical Resistance to State Schooling, Contemporary Private Competition and Student Achievement across Countries.' *Economic Journal* 120(546):F229–F255.

West, Martin R., Matthew, A. Kraft, Amy S. Finn, Rebecca E. Martin, Angela L. Duckworth, Chrtistopher F. O. Gabrieli, and John D. E. Gabrieli. 2016. 'Promise and Paradox: Measuring Students' Non-Cognitive Skills and the Impact of Schooling.' *Educational Evaluation and Policy Analysis* 38(1):148–170

Appendix A: Additional tables

Table 3.A.1: Descriptive statistics

Wellbeing and academic efficiency					
	Mean	SD		Mean	SD
Happiness at school	3.00	0.76	PISA science	501.14	93.78
PISA mathematics	494.03	93.93	Educational expenditure/pupil	69,130	28,217
PISA reading	496.45	93.91			
Country- and regional-level variables			*Pupil-background variables*		
	Mean	SD		Mean	SD
Independent-school share	0.19	0.21	Girl	0.50	0.50
Catholic share 1900 (no state religion)	0.29	0.36	Age	15.77	0.29
Catholic share 2010 (no state religion)	0.27	0.30	Index of home possessions	0.00	1.00
(log) GDP/capita 2011	10.43	0.35	Parental occupational status	50.66	21.61
Population 1900	13,800,000	18,200,000	Parental education	13.49	3.04
Calvinist share 1900	0.06	0.13	Immigrant (1st generation)	0.05	0.21
Early Catholic defeat (soft)	0.08	0.27	Immigrant (2nd generation)	0.15	0.36
Early Catholic defeat (hard)	0.06	0.24	*Pupil-level variables (institutional characteristics)*		
			Grade	9.62	0.73
Nazi annexation	0.18	0.38	School starting age	6.10	0.85
			School location		
Pro-Catholic Nazi ally	0.10	0.29	Village	0.09	0.29
Jesuit ban	0.16	0.36	Small town	0.21	0.41
Communist	0.18	0.38	Town	0.35	0.48
Post-Soviet	0.03	0.17	City	0.24	0.43
For-profit voucher/ mass conversion	0.08	0.28	Large city	0.11	0.31
Mechanisms					
	Mean	SD		Mean	SD
Individualisation of teaching	1.94	1.06	Achievement pressure (headteacher)	1.88	0.73
Project work	1.65	0.89	Class periods (total)	31.03	7.82
Group work	1.84	0.96	Class periods (language)	4.13	1.48
Help planning	1.66	0.88	Class periods (mathematics)	4.16	1.40
Get along with teachers	3.01	0.67	Class periods (science)	3.76	2.13
Teachers listen to pupils	2.89	0.74	Hours of homework	4.89	4.69

Note: The descriptive statistics display each variable's international mean and standard deviation (weighted by sampling probabilities with all countries given equal weight) without any imputed values.

Table 3.A.2: Alternative measures of pupil wellbeing

	General pupil wellbeing				
	Pupil happiness	Satisfaction with school	Things are ideal at school	Belong at school	Overall wellbeing index
Independent-school share	-1.28***	-1.31***	-1.93***	-1.69***	-2.22***
	(0.32)	(0.22)	(0.52)	(0.48)	(0.60)
Catholic share 2010*no state religion	0.49***	0.40***	-0.07	0.30	0.55**
	(0.10)	(0.11)	(0.25)	(0.19)	(0.24)
F statistic	45.87	45.88	45.92	45.83	45.97
N	190,348	190,616	190,585	190,639	191,913
Countries	34	34	34	34	34
	Peer relations/specific reasons for general pupil wellbeing				
	Outsider at school	Make friends easily at school	Feel awkward at school	Liked by other pupils	Lonely at school
Independent-school share	0.81***	-0.69***	0.36	-0.76***	0.53**
	(0.28)	(0.18)	(0.29)	(0.24)	(0.24)
Catholic share 2010*no state religion	-0.40***	0.23***	-0.23**	0.12	-0.22***
	(0.11)	(0.06)	(0.10)	(0.09)	(0.08)
F statistic	45.64	45.94	45.92	45.77	45.95
N	191,058	191,282	190,762	190,521	190,905
Countries	34	34	34	34	34

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the country level in parentheses. All regressions include the controls described in Sections 3.3.5 and 3.4.1, including within-European regional-fixed effects. Items analysed: (1) 'I am happy at school'; (2) 'I am satisfied with my school'; (3) 'Things are ideal in my school'; (4) 'I feel like I belong at school'; and (5) the overall wellbeing index. The overall wellbeing index is constructed from responses to all statements in Columns 1–4 as well as those in Columns 6-10, which tap into specific reasons behind the level of wellbeing, such as peer relations: 'I feel like an outsider (or left out of things) at school'; 'I make friends easily at school'; 'Other students seem to like me'; 'I feel awkward and out of place in my school'; and 'I feel lonely at school'.

Table 3.A.3: Further robustness tests for pupil happiness

	Control added for			
	Enrolment share of privately-funded independent schools	Share of state funding in independent schools	Average level of independent-school autonomy	Exit exams
Independent-school share	-1.29***	-1.02***	-1.20***	-1.25***
	(0.36)	(0.32)	(0.28)	(0.35)
Catholic share 2010 *no state religion	0.50***	0.40***	0.31***	0.47***
	(0.12)	(0.09)	(0.11)	(0.10)
Added control	-0.31	0.04	-0.18**	-0.03
	(1.57)	(0.22)	(0.09)	(0.08)
F statistic	33.09	39.10	23.27	25.63
n	190,348	184,292	187,217	190,348
Countries	34	32	33	34

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the country level in parentheses. All regressions include the controls described in Sections 3.3.5 and 3.4.1, including within-European regional-fixed effects. Data on centralised exit examinations are obtained from Bol and Van de Werfhorst (2013). This source lacks data on Chile and Mexico, which we obtain from Brandt (2010) and the OECD (2009) respectively. The other variables are obtained from OECD (2016a).

Table 3.A.4: Robustness tests for academic efficiency

	Mathematics	Reading	Science	Educational expenditures/pupil
		Only state schools		
Independent-school share	214.24***	255.56***	162.70**	-130,083***
	(67.01)	(85.30)	(63.43)	(28,477)
Catholic share 2010 *no state religion	-14.70	-39.18	-6.35	35,867***
	(25.13)	(32.71)	(22.56)	(9,998)
F statistic	49.73	49.73	49.73	49.73
n	233,309	233,309	233,309	233,309
Countries	34	34	34	34
		Only Europe		
Independent-school share	249.64***	285.28***	184.02***	-141,424***
	(49.06)	(57.99)	(48.29)	(27,638)
Catholic share 2010 *no state religion	-48.42**	-74.75***	-28.99	53,427***
	(22.24)	(24.12)	(20.45)	(10,285)
F statistic	42.52	42.52	42.52	42.52
n	188,173	188,173	188,173	188,173
Countries	24	24	24	24
		Belgium and the Netherlands excluded		
Independent-school share	252.03***	329.56***	230.67***	-157,986***
	(79.18)	(92.01)	(63.02)	(29,633)
Catholic share 2010 *no state religion	-3.44	-25.64	1.66	32,225***
	(26.81)	(34.08)	(22.72)	(11,545)
F statistic	42.87	42.87	42.87	42.87
n	282,359	282,359	282,359	282,359
Countries	32	32	32	32
		Excluding pupil-background characteristics		
Independent-school share	260.22***	321.27***	233.39***	-132,297***
	(58.22)	(76.71)	(53.85)	(28,644)
Catholic share 2010 *no state religion	-9.69	-33.75	-2.88	34,861***
	(22.42)	(31.19)	(20.72)	(10,927)
F statistic	45.88	45.88	45.88	45.88
n	295,416	295,416	295,416	295,416
Countries	34	34	34	34
		Reduced form		
Catholic share 1900 *no state religion	47.77***	59.74***	40.35***	-29,984***
	(10.24)	(14.45)	(10.61)	(3,929)
Catholic share 2010 *no state religion	33.08**	22.52	33.29**	5,862
	(16.39)	(18.49)	(15.51)	(4,545)
n	295,416	295,416	295,416	295,416
Countries	34	34	34	34

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the country level in parentheses. All regressions include pupil-, school-, and country-level controls described in Sections 3.3.5 and 3.4.1, including within-European regional-fixed effects.

Table 3.A.5: Including the school average of all pupil-level variables

	Pupil happiness	Mathematics	Reading	Science	Educational expenditures/pupil
Independent-school share	-1.26***	127.33***	187.35***	104.97***	-118,247***
	(0.32)	(37.23)	(54.44)	(38.92)	(24,941)
Catholic share 2010 *no state religion	0.43***	-14.75	-36.32*	-9.29	31,192***
	(0.10)	(14.71)	(21.80)	(13.72)	(9,143)
F statistic	46.04	45.96	45.96	45.96	45.96
n	190,348	295,416	295,416	295,416	295,416
Countries	34	34	34	34	34

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the country level in parentheses. All regressions include indicators for the school- and country-level variables described in Sections 3.3.5 and 3.4.1, including within-European regional-fixed effects. They also include the school-level shares of girls, first-generation immigrants, and second-generation immigrants, age, parental education, parental occupational status, the index of home possessions, grade attended, and school starting age.

Appendix B: Ensuring relevance and validity of the instrument

In order to maximise the relevance and validity of the instrument used in the third chapter, it is important to control for other historical factors, which have determined the extent to which Catholic resistance in the late 19th and early 20th centuries generated higher independent-school competition – and, if it did, the extent to which this competition has survived to this day. This problem is generally ignored in previous research, with the sole exception being the most obvious example: the rise of Communist regimes from the October Revolution in 1917 onwards, which undid most progress made by Catholics in the late 19th and early 20th centuries. We take this into account by controlling for countries' Communist background, obtained from Barro and McCleary (2005), and post-Soviet background.[1] The former applies Czech Republic, East Germany, Estonia, Hungary, Poland, Slovakia, and Slovenia, whereas the latter applies to Estonia only.[2]

But there are more nuanced historical issues, which are important to consider for the purpose of generating a relevant and valid instrument for independent-school competition based on the basic intuition described in Section 3.4.1. To begin with, we control for the share of Calvinists (reformed Protestants) in 1900, obtained from Brown and James (2015). This is to account for the fact that Calvinists in some countries, such as the Netherlands, joined the Catholics' more general resistance to secular state schooling (Glenn 2011, 1989). In those countries, with such reinforcements, Catholics could obtain successes in the educational sphere that were disproportionate to the relative size of their own community.

We also control for population size in 1900, obtained from Brown and James (2015), to account for the fact that Catholics in larger countries often faced more formidable coordination problems and higher transaction costs to mobilise successfully (e.g. Wilkinson et al. 2006). Further building on this intuition, we note that the success of

[1] The post-Soviet indicator is included to account for the fact that Soviet annexation throughout the latter part of the 20th century ensured an especially extreme form of centralisation and makeover of the education system, while also ensuring mass migration and a new parallel education system along linguistic lines (e.g. Krull and Trasberg 2006; Stevick 2006). Today, this may affect both demand and supply for independent schooling as well as pupil outcomes in ways that do not apply to post-Communist countries more generally.

[2] Whenever an indicator only affected parts of a country, we assign the share of the population affected. This strategy follows Barro and McClearly (2005) who assign a value of 0.204 for Germany in terms of its Communist background, representing the East German population share. The only exception is when the PISA data allow us to identify the relevant within-country regions, such as England in the UK, in which case we assign a value of 1 for all pupils in the region for which the indicator applies and 0 for the rest.

Catholics in defending their interests in countries with strong anticlerical currents depended on their actual success of early political mobilisation (see Kaiser and Wohnout 2004). In countries where Catholics failed to effectively mobilise politically against anti-clerical forces early on, and thus faced defeat, state monopolisation of schooling was more successful than in countries where Catholics mobilised more effectively at an early stage.

For example, whereas Belgian and Dutch Catholics were able to successfully lobby for access to and public funding for independent schools in the latter part of the 19th century – and German and Swiss Catholics ensured access to mostly public, but also independent, religious schools following decades of setbacks – due to the success of Catholic political mobilisation (e.g. Evans 1999; Glenn 1989, 2011; Kaiser and Wohnout 2004), French Catholics struggled to develop a coherent political strategy and were consequently less successful in this endeavour. Indeed, as the concept of *laïcité* came to dominate education policy in France in the latter part of the 19th century, French Catholics suffered consecutive defeats due to their inability to mount a successful political defence (Boyer 2004). Consecutive decrees in the 1880s decreased the role of the Church in state schooling considerably, and the 1904 law pushed through by Prime Minister Émile Combes sought to end it entirely in both the public and independent sectors. Many publicly-funded schools that had been maintained by congregations were thus reopened as fee-based schools that hired lay teachers but still maintained a 'Catholic character'. Overall, however, the 1904 law led to a considerable decrease in de facto independently-operated school enrolment shares. In 1902, 21.6 per cent of boys and 42 per cent of girls attended such schools; in 1912, these figures had declined to 12.8 per cent and 24.8 per cent respectively, not far from the situation at the end of the 20th century (Judge 2001). Furthermore, the abrogation of Napoleon's concordat with the Vatican in 1905, which finally marked the de jure separation of church and state in France, meant that a system of public funding for independent schools similar to those in Belgium and the Netherlands was never developed in the Third Republic (Teese 1986). Consequently, despite the fact that France had similar Catholic population shares as Belgium in the late 19th century, and considerably higher shares than the Netherlands, the inability of French Catholics to successfully mobilise politically at an early stage of state centralisation appears to be an important reason why independent-school competition never reached similar levels in France.

A similar, but more radical, story applies to Mexico, where Catholic political mobilisation only really took shape following the Mexican Revolution in 1910. However, this ended abruptly after the ouster of pro-Catholic Victoriano Huerta in 1914. In the 1917 constitution, religious institutions were banned from running independent schools, and the strict enforcement of this ban in the latter part of the 1920s – when Catholic schools were forcibly closed – was an important contributory factor to the Cristero War (see Curley 2008; Hamnet 2006; Schell 2003). In 1929, the Church finally caved on the issue of religious education in schools and agreed to carry it out in churches only (Fernández 2007). The ban on religious independent schools was not revoked until 1992, although its enforcement varied over the decades, and there is still essentially no public funding available (Blancarte 1993; OECD 2016). The early Catholic defeat and the inability to successfully mount a political defence later on thus had similar, albeit more severe, consequences for independent schooling in Mexico as in France. We thus control for an indicator of these significant early Catholic political defeats in France and Mexico in the late 19th and early 20th centuries, which appear important for today's levels of independent-school competition.

In other countries, early Catholic pressure for access to independent schools suffered less draconian defeats. Instead, laws were passed to ensure that independent confessional schools would simply not be eligible for public funding, precisely to decrease the alternative schooling opportunities for Catholics. For example, anti-Catholic sentiments in the US during the 1800s led to increased political pressure to legislate against public funding for parochial schools, while still maintaining essentially Protestant state schools. Thus, from the mid-1800s onwards, most US states began passing amendments to their constitutions, or joined the Union with such amendments, which banned government funding for independent religious schools. In the end, 41 states and the District of Columbia incorporated such measures in their constitutions at some point in time (Duncan 2003; Katz 2011), covering 86 per cent of the relevant population in the US. Similar developments occurred throughout Australia and New Zealand, albeit these changes were not constitutionally enshrined (Buckley et al. 2011; Wilkinson et al. 2006). Thus, we control for indicators for areas that experienced these less radical early Catholic defeats in the political realm separately.[3]

[3] That is, Australia, New Zealand, and the relevant part of the US. Note, however, that results are very similar if we merely include one indicator for all countries where Catholics suffered early political defeats.

Similarly, we also control for indicators of 19th century national bans on the Society of Jesus and its associate orders, as long as they remained in the early 20th century.[4] The Society of Jesus was the first teaching order of the Catholic Church and has been especially devoted to education from its inception, having founded hundreds of independent schools worldwide, while also inspiring other orders in this direction (Duminuco 2000). During the struggle between secular and religious forces in the 19th century, several countries banned Jesuits and their associate orders from their territories for longer periods of time, often specifically because of their educational influence (e.g. Chadwick 1998; Healy 2003). Due to the importance of the Jesuits and their associate orders in opening and maintaining independent schools, we control for these bans in our set-up to ensure maximum instrument relevance.

Furthermore, we take into account the unique impact of World War II on the independent-education systems in many countries. First, we control for indicators of Nazi takeover and de facto annexation of regions into the Greater German Reich. Nazi ideologues strongly emphasised the importance of education in socialising young people into their worldview, thus opposing independent or denominational schools and, indeed, any religious elements in education whatsoever. Inevitably, this led to a radical persecution of the Catholic Church in Germany as well as all territories that were either de jure or de facto annexed into the Reich. Indeed, in these areas, the Nazis closed down all denominational schools, public and independent (see Mariaux 1940; Pine 2010). For example, just months after the *Anschluss*, all independent schools in Austria were closed and taken over by the Nazi Party. *Reichskommissar* Josef Bürckel explained: 'We must take care of the preservation of our nation in this world. This only is possible if care is total care, therefore the school must belong to the state, upon which devolves the responsibility for the future' (Chicago Daily Tribune 1938, p. 1). Similarly, following the annexation of Alsace-Lorraine, the Vatican complained: 'There are no longer any Catholic private schools in Alsace. All Catholic educational institutions run by members of the Holy Order, priests or laymen, have been dissolved' (The Tablet 1941, p. 290). Similar fates afflicted other de facto annexed regions, including Eupen-Malmedy in Belgium, Luxembourg, Czech lands, most of Slovenia, and the whole of Poland, including

[4] More specifically, this applies to regions that belonged to the German Empire, France, Mexico, Norway, and Switzerland.

the quasi-colony in the eastern parts that became known as the General Government (see Lapomarda 2005; NCWC 1942; OUSSCCPAC 1946).[5]

However, the same story does not apply to areas that were solely under military or civil administrative control. For example, in Belgium and the Netherlands, 'the churches were given a degree of leeway that allowed them to maintain their influence and preserve confessional institutions' (Bank and Gevers 2016, p. 182). While the Nazis threatened to close Catholic schools in the Netherlands, due to anti-Nazi activities of the episcopacy, they refrained from doing so, leaving most schools operating normally during the occupation (Warmbrunn 1963). The fates of independently-operated schools in Belgium and the Netherlands nicely display the differences in regions and countries that were merely occupied by, compared with regions and countries that were de facto annexed into, the Greater German Reich.

At the same time, in pro-Catholic fascist and quasi-fascist countries allied to Nazi Germany, or acting as client states to Nazi Germany, the reverse situation occurred during and right before World War II: the Catholic Church again reached privileged status in public life, which meant that pressure for independent schools decreased for quite some time in countries that turned away from fascist and quasi-fascist ideology following World War II. The canonical example here is the clerico-fascist Slovak Republic, led by the Catholic priest Josef Tiso, who restored Church privileges in the public education system between 1938 and 1945 (Conway 1974; Ward 2013). Similar stories apply to Vichy France, Hungary, and Italy following the Lateran Treaty of 1929 (see Fazekas 2004; Sweets 1994; Wolff 1980).[6] Thus, whereas territories annexed into Nazi Germany experienced Catholic persecution, pro-Catholic regimes allied with the country rather defused such pressure for independent schools for quite some time. This often had implications for Catholic influence in education also after the war (e.g. Wolff 1992). We thus also include an indicator for pro-Catholic Nazi client states or allies,

[5] A small part of north-eastern Slovenia, covering about 6 per cent of today's population, was never de facto annexed by Nazi Germany, but was instead part of Hungary, a country we code as being pro Catholic. Nevertheless, the Hungarians were hardly pro Catholic in this annexed region: they closed all Slovenian schools, imprisoned Slovenian Catholic leaders, and made Protestants the new elite, since the latter were perceived to be more amenable to forced Magyarisation (Kranjc 2013). We thus do not code this small part of Slovenia as being pro Catholic, although results are unsurprisingly almost identical if we do.

[6] The Vichy regime did authorise communes to support independent schools financially, but this happened only rarely and, when it did, the subsidies were very small. Consequently, there was no increase in independent-school enrolment in France during the Vichy regime's tenure (Sweets 1994).

before and during World War II, which abandoned fascist and quasi-fascist ideology after the war.

Finally, we also include indicators for countries or regions that in the latter part of the 20th century implemented voucher programmes in which for-profit operators participate on an equal basis, or very recent reforms that enabled mass conversions of state schools to independently-operated status essentially overnight. There are two countries that allow for-profit operators on an equal basis – Chile and Sweden – and only one nation in one country that has allowed mass conversions of publicly-operated schools to independently-operated status: England. As a direct result of the 2010 Academies Act, which allowed essentially all English schools to become autonomous 'academies', the share of 15-year old pupils attending independently-operated schools in the United Kingdom increased from 6.31 per cent in 2009 to 45.16 per cent in 2012 (OECD 2016). Neither the enrolment growth in for-profit independently-operated schools nor such mass conversions has much to do with the independent-school competition that we aim to capture with the instrument based on historical Catholic resistance to state schooling in the 19th century.

References for Appendix B

Bank, Jan and Lieve Gevers. 2016. *Churches and Religion in the Second World War*. London: Bloomsbury.

Barro, Robert J. and Rachel M. McCleary. 2005. 'Which Countries Have State Religions?' *Quarterly Journal of Economics* 120(4):1331–1370.

Blancarte, Roberto J. 1993. 'Recent Changes in Church-State Relationships in Mexico: A Historical Approach.' *Journal of Church and State* 35(4):781–805.

Boyer, John W. 2004. 'Catholics, Christians and the Challenges of Democracy: The Heritage of the Nineteenth Century.' Pp. 7–45 in *Political Catholicism in Europe 1918–45*, edited by Wolfram Kaiser and Helmut Wohnout. London: Routledge.

Brown, Davis and Patrick James. 2015. 'Religious Characteristics of States Dataset, Phase 1: Demographics (RCS).' Dataset, Department of Political Science, Maryville University of St Louis.

Buckley, James J., Frederick C. Bauerschmidt, and Trent Pomplun, eds. 2011. *The Blackwell Companion to Catholicism*. Oxford: Wiley-Blackwell.

Chadwick, Owen. 1998. *A History of the Popes, 1830–1914*. Oxford: Oxford University Press.

Chicago Daily Tribune. 1938. 'Austrian Nazis Order Church Schools Closed.' *Chicago Daily Tribune*, September 2.

http://archives.chicagotribune.com/1938/09/02/page/1/article/austrian-nazis-order-church-schools-closed.

Clark, Andrew E. and Orsolya Lelkes. 2009. 'Let Us Pray: Religious Interactions in Life Satisfaction.' Working Paper No. 2009–01. Paris School of Economics.

Conway, John S. 1974. 'The Churches, the Slovak State and the Jews 1939–1945.' *Slavonic and East European Review* 52(126):85–112.

Curley, Robert. 2008. 'Political Catholicism in Revolutionary Mexico, 1900–1926.' Working Paper No. 349, Notre Dame, IN.

Duminuco, Vincent J., ed. 2000. *The Jesuit Ratio Studiorum: 400th Anniversary Perspectives*. New York: Fordham University Press.

Duncan, Kyle. 2003. 'Secularism's Laws: State Blaine Amendments and Religious Persecution.' *Fordham Law Review* 72(3):493–593.

Evans, Ellen L. 1999. *The Cross and the Ballot: Catholic Political Parties in Germany, Switzerland, Austria, Belgium and the Netherlands, 1785–1985*. Boston, MA: Brill.

Fazekas, Csaba 2004. 'Collaborating with Horthy: Political Catholicism and Christian Political Organizations in Hungary.' Pp. 195–216 in *Political Catholicism in Europe 1918–45*. London: Routledge.

Fernández, Eduardo C. 2007. *Mexican-American Catholics*. New York: Paulist Press.

Glenn, Charles L. 1989. *Choice of Schools in Six Nations: France, Netherlands, Belgium, Britain, Canada, West Germany*. Washington, DC: US Department of Education.

Glenn, Charles L. 2011. *Contrasting Models of State and School: A Comparative Historical Study of Parental Choice and State Control*. New York: Continuum.

Hamnet, Brian R. 2006. *A Concise History of Mexico*. Cambridge: Cambridge University Press.

Healy, Róisín. 2003. *The Jesuit Specter in Imperial Germany*. Boston, MA: Brill.

Judge, Harry. 2001. *Faith-based Schools and the State: Catholics in America, France, and England*. Oxford: Symposium Books.

Kaiser, Wolfram and Helmut Wohnout, eds. 2004. *Political Catholicism in Europe 1918–45*. London: Routledge.

Katz, Meir. 2011. 'The State of Blaine: A Closer Look at the Blaine Amendments and Their Modern Application.' *Engage* 12(1):111–120.

Kranjc, Gregor J. 2013. *To Walk with the Devil: Slovene Collaboration and Axis Occupation, 1941–1945*. Toronto: University of Toronto Press.

Krull, Edgar and Karmen Trasberg. 2006. 'Changes in Estonian General Education from the Collapse of the Soviet Union to EU Entry.' Working Paper, University of Tartu, Tartu.

Lapomarda, Vincent A. 2005. *The Jesuits and the Third Reich*. 2nd ed. Lewiston, NY: Edwin Mellen Press.

Mariaux, Walther, trans. 1940. *The Persecution of the Catholic Church in the Third Reich: Facts and Documents*. London: Burns Oates.

NCWC. 1942. ‘The Nazi War Against the Catholic Church.’ Conference Report, National Catholic Welfare Conference, Washington, DC.

OECD. 2016. Data retrieved from the OECD's PISA database on 20 December 2015: http://www.oecd.org/pisa/pisaproducts/.

OUSSCCPAC (Office of the United States Chief of Counsel for Prosecution of Axis Criminality). 1946. *Nazi Conspiracy and Aggression*. Washington, DC: US Government Printing Office.

Pine, Lisa. 2010. *Education in Nazi Germany*. Oxford: Berg.

Schell, Patience A. 2003. *Church and State Education in Revolutionary Mexico City*. Tucson, AZ: University of Arizona.

Stevick, E. Doyle. 2006. ‘Civic Education Policy and Practice in Post-Soviet Estonia, from Global Influences to Classroom Practice.’ PhD Dissertation.

Sweets, John F. 1994. *Choices in Vichy France: The French under Nazi Occupation*. Oxford: Oxford University Press.

Teese, Richard. 1986. ‘Private Schools in France: Evolution of a System.’ *Comparative Education Review* 30(2):247–259.

The Tablet. 1941. ‘The Church in Alsace.’ *The Tablet*, vol. 177, No. 5266. 12th April, http://archive.thetablet.co.uk/issue/26th-april-1941/10/14539#scanned.

Ward, James M. 2013. *Priest, Politician, Collaborator: Jozef Tiso and the Making of Fascist Slovakia*. Ithaca, NY: Cornell University Press.

Warmbrunn, Werner. 1963. *The Dutch under German Occupation, 1940–1945*. Stanford, CA: Stanford University Press.

West, Martin R. and Ludger Woessmann. 2010. ‘‘Every Catholic Child in a Catholic School’: Historical Resistance to State Schooling, Contemporary Private Competition and Student Achievement across Countries.’ *Economic Journal* 120(546):F229–F255.

Wilkinson, Ian R., Brian J. Caldwell, Richard Selleck, Jessica Harris, and Pam Dettman. 2006. ‘A History of State Aid to Non-government Schools in Australia.’ Report, Department of Education, Science and Training, Canberra.

Wolff, Richard J. 1980. ‘Catholicism, Fascism and Italian Education from the Riforma Gentile to the Carta Della.’ *History of Education Quarterly* 20(1):3–26.

Wolff, Richard J. 1992. ‘Italian Education During World War II: Remnants of Failed Fascist Education, Seeds of the New Schools.’ Pp. 73–83 in *Education and the Second World War: Studies in Schooling and Social Change*, edited by Roy Lowe. London: Falmer Press.

4. Group Threat and Voter Turnout: Evidence from a Refugee Placement Programme*

Abstract

We study the impact of refugee inflows on voter turnout in Sweden in a period when shifting immigration patterns made the previously homogeneous country increasingly heterogeneous. Analysing individual-level panel data and exploiting a national refugee placement programme to obtain plausibly exogenous variation in immigration, we find that refugee inflows significantly raise the probability of voter turnout. Balancing tests on initial turnout as well as placebo tests regressing changes in turnout on future refugee inflows support the causal interpretation of our findings. The results are consistent with group-threat theory, which predicts that increased out-group presence spurs political mobilisation among in-group members.

* The author thanks Michael Bruter, Karin Edmark, Björn Tyrefors Hinnerich, Henrik Jordahl, Julian Le Grand, Per Pettersson-Lidbom, and Olmo Silva for comments and discussions, as well as Karin Edmark and Per Pettersson-Lidbom for kindly sharing their data.

4.1. Introduction

In the past decades, immigration to European countries has increased considerably. Between the periods 1960–69 and 1990–99, net-migration flows increased from 1.1 million to about 10 million, generating large demographic shifts in many societies throughout the continent. The increased flows also reflect a new pattern of immigration since the 1980s. Following World War II, migration in Europe consisted mostly of intra-continental labour flows, especially from southern to northern countries, which were frequently reversed after employment contracts ended. However, since the end of the 1970s, such flows have been replaced by immigration from the developing world, which led to more permanent settlement in the host countries (Dustmann and Frattini 2012; Wanner 2002). In the latter part of the 20th century, previously ethnically homogeneous countries thus became increasingly heterogeneous – and the consequences of this change have been the subject of intense debate.

One important possible consequence of immigration could be altered political engagement among natives. In America, over half a century ago, Key (1949) noted a positive correlation between the presence of African Americans and voter turnout among whites; the perception of 'group threat' allegedly stimulated higher political mobilisation. However, other scholars have instead suggested that inter-ethnic contact under certain conditions may instead decrease existing perceptions of threat (e.g. Allport 1954; Williams 1947), implying that immigration if anything may reduce in-group bias as a source of political engagement and thus lower turnout among natives (e.g. Zingher and Thomas 2014). These opposing perspectives have generated an expanding empirical literature with mixed results. Yet identification problems involved in analysing the effect of ethnic diversity are severe because of potential endogeneity in settlement and mobility patterns as well as other sources of unobserved heterogeneity.

This paper provides new evidence on the impact of increasing ethnic diversity due to immigration on individual-level voter turnout in Sweden. In 1985, increasing refugee inflows led the government to enact a placement programme, through which most new refugees were contracted to the country's municipalities each year until the programme was dismantled in 1994. It is possible to exploit the municipal contracts to obtain variation in refugee inflows that is free from bias due to endogenous settlement patterns and measurement error, and, once we adjust for municipal-fixed effects, also plausibly exogenous to changes in individual-level turnout more generally.

Combining the municipal-level contracts with data from the Swedish National Election Studies Programme, carried out in conjunction with every Swedish election since 1956, the paper analyses how refugee inflows due to the placement programme affected changes in individual-level voter turnout between the national, municipal, and county elections in 1985 and 1988, 1988 and 1991, as well as 1991 and 1994. Two features of this survey are especially useful: (1) each individual is observed in two elections in a row, which allows us to adjust for individual-fixed effects and ensure that native mobility does not bias our findings, while (2) data on voter turnout are obtained directly from official records, which gets rid of potential response bias and minimises panel attrition. To ensure that that the sample analysed is in fact representative of the Swedish voting-age population, we weight each individual with the inverse probability of selection in the population of eligible voters. Still, since the survey was designed to be representative at the national rather than municipal level, the estimated effects should primarily be seen as valid for the randomly sampled population in each municipality.

Our study thus analyses whether refugee immigration in one period, spurred by the placement-programme contracts, altered individuals' propensity to vote over the same period. The results display that larger refugee inflows raise individuals' propensity to vote in national and local elections: a rise in the refugee inflow by 1 percentage point increases the likelihood of voter turnout by 5–7 percentage points. Balancing tests on initial turnout and placebo tests regressing changes in turnout on future refugee inflows support the causal interpretation of our findings, as do several robustness checks. In addition, supporting our expectations, we find that OLS estimates are biased towards zero, indicating that it is important to take into account endogenous settlement and mobility patterns and/or measurement error in refugee inflows.

Overall, our findings thus provide support for group-threat theory: natives appear to mobilise politically as a result of refugee immigration. While silent on the effects of involuntary contact, the study provides new evidence of how real-world demographic changes – which do not necessarily generate voluntary contact between in- and out-group members – affect political engagement. Given the relative scarcity of convincing research on the relevance of group threat for understanding individual-level voter turnout in general, the paper provides an important contribution to the literature.

The paper proceeds as follows. Section 4.2 discusses the theoretical mechanisms linking refugee immigration to voter turnout and reviews the empirical literature;

Section 4.3 describes the Swedish setting and the refugee placement programme; Section 4.4 discusses the data and methodology utilised; Section 4.5 outlines the estimation strategy; Section 4.6 presents the results; and Section 4.7 concludes.

4.2. Theory and literature review

Why would refugee immigration affect voter turnout? One theoretical mechanism rests on the social psychological concept of social identity, defined as 'that part of an individual's self-concept which derives from his knowledge of his membership in a social group' (Tajfel 1978:63). People's social identity may in turn be connected to the formation of their preferences, generating 'in-group bias'. Such group identification could be based on, for example, ethnicity, religion, nation, or class. The implication is that political mobilisation among in-group members may increase as a result of increased out-group presence, in accordance with group-threat theory (Key 1949).

However, it is not clear whether voting preferences as such would change in the equilibrium. In traditional explanations, increased diversity spurred by immigration may lower preferences for redistribution overall (e.g. Alesina and Glaeser 2004; Alesina and La Ferrara 2000; Costa-Font and Cowell 2014), which should mobilise voters in favour of right-wing parties. However, it is also possible that diversity increases natives' support for redistribution because of perceived threats of economic competition from immigrants (Brady and Finnigan 2014), thus making them more likely to vote for left-wing parties. Additionally, since immigrants are generally seen as the least deserving of welfare (van der Waal et al. 2010; van Oorschot 2006), populist right-wing parties that appeal to welfare chauvinism may benefit disproportionally.

Of course, these theoretical explanations are not necessarily incompatible since they may be relevant for different parts of the population. For example, high-income earners may be especially susceptible to decrease their support for redistribution (e.g. Dahlberg et al. 2012, 2017), and therefore mobilise for right-wing parties, while low-income earners, who are more directly affected by welfare policies, may be more likely to mobilise in favour of left-wing parties (Brady and Finnigan 2014). Alternatively, both may mobilise for anti-immigration, welfare chauvinist parties in support of welfare for in-group members only (see Bay et al. 2013; Harrison and Bruter 2011). However, most importantly for our purposes, also differential mobilisation patterns would be expected to increase turnout in the equilibrium if group-threat theory is correct.

Another potential mechanism linking group threat to higher turnout is via electoral competition from out-group members. This could occur if immigrants have different political preferences and vote for different parties compared with natives. Sweden has a comparatively high naturalisation rate, with 82 per cent of the total foreign-born population, and fully 94 per cent of immigrants from countries outside the European Economic Area, holding citizenship in 2007 (OECD 2011). Furthermore, the country also allows some immigrants without citizenship to vote in local elections. In this story, natives' propensity to vote would increase as a result of larger inflows of out-group members because of perceptions that the latter threaten to alter the political situation.

While group-threat theory predicts increased turnout as a result of immigration, there are other reasons to believe the impact may be the opposite. The 'contact hypothesis' holds that increased interaction between in- and out-group members under certain circumstances reduces prejudices and conflicts (Allport 1954; Williams 1947). If this holds true, existing perceptions of group threat and, in the end, political mobilisation among in-group members may in fact decrease as the share of out-group members increases (see Zingher and Thomas 2014). In this story, immigration thus lowers turnout: increased diversity stimulates greater inter-ethnic contact, which reduces perceptions of threat and in-group bias as sources of political engagement among natives. In the context of generally decreasing turnout across industrialised countries (Blais and Rubenson 2013; Gray and Caul 2000), this means that the decline is expected to be more significant in communities with larger inflows of immigrants compared to communities with smaller inflows. Natives in the former experience increased contact with out-group members, as a result of greater opportunities for such contact, which could lower in-group bias via lowered perceptions of threats, leading to lower mobilisation and turnout. In contrast, natives in the latter may be aware that their country is getting increasingly diverse, but have little opportunity for direct inter-ethnic day-to-day contact through which such diversity could help ameliorate negative perceptions. Consequently, their in-group bias and mobilisation remain strong, resulting in higher turnout relative to more diverse communities.

At a general level, empirical studies analysing the effects of diversity on inter-ethnic attitudes display mixed findings (e.g. Avery and Fine 2012; Bobo and Hutchings 1996; Dustmann and Preston 2001; Fox 2004; Hopkins 2010; Markaki and Longhi 2012; McLaren 2003; McLaren and Johnson 2007; Newman 2013; Oliver and Mendelberg

2000; Oliver and Wong 2003; Schlueter and Scheepers 2010). A similarly heterogeneous picture emerges from research exploring the effects of diversity and immigration on (1) public spending and attitudes towards redistribution and welfare (Costa-Font and Cowell 2014; Stichnoch and Van der Straeten 2011; Schaeffer 2013); (2) vote outcomes (Arzheimer 2009; Della Posta 2013; Gerdes and Wadensjö 2010; Giles and Buckner 1993; Harrison and Bruter 2011; Roch and Rushton 2008; Rydgren and Ruth 2011, 2013; Voss 1996; Voss and Miller 2001); and (3) political engagement and turnout (Bhatti et al. 2017; Fieldhouse and Cutts 2008; Hill and Leighley 1999; Leighley and Vedlitz 1999; Matthews and Prothro 1963; Schlichting et al. 1998; Zingher and Thomas 2014). Yet this research does not generally exploit plausibly exogenous variation in diversity and thus likely fails to isolate causal relationships.[1] For example, settlement patterns of ethnic minorities and immigrants are not random, but often depend on community characteristics (e.g. Bracco et al. 2018; Damm 2009) – which may also affect the outcomes under investigation.

Focusing on methodologically advanced research, studies analysing American college students who are randomly allocated to roommates of different ethnicities indicate that contact often breeds more positive inter-group attitudes (see Boisjoly et al. 2006; Carrell et al. 2015; Shook and Fazio 2008; Van Laar et al. 2005). However, the extent to which these findings are relevant for real-world demographic changes is questionable. This is because in-group members are usually not forced to interact with out-group members, and the choice to do so is endogenous to inter-group attitudes. In other words, for the contact hypothesis to be relevant for real-world demographic changes, diversity itself should stimulate greater inter-group contact and, in this way, decrease perceived group threat.[2]

Yet available evidence does not necessarily suggest this is the case. In one interesting study, Enos (2014) analyses the impact of randomly increasing the daily presence of

[1] This appears to include Dahlberg et al.'s (2012) study, which utilises the Swedish placement programme analysed here and finds negative effects of diversity on redistributive preferences. Yet Nekby and Pettersson-Lidbom (2012, 2017) show that the results are unreliable because of a partially endogenous instrument and sample selection bias due to considerable attrition (see Dahlberg et al. [2013, 2017] for rejoinders). As discussed in Section 4.4.2, we circumvent these problems by (1) using the endogenous instrument as our primary independent variable capturing refugee inflows, while (2) exploiting an alternative instrument developed by Nekby and Pettersson-Lidbom (2012, 2017) that is more likely to be exogenous, (3) analysing a dependent variable with very little attrition, and (4) weighting respondents by the inverse of their probability of being selected for the survey in the relevant election.

[2] While the effects of residential segregation may make studies analysing larger geographical areas less likely to pick up significant inter-group interaction (see Enos 2016; Zingher and Thomas 2014), voluntary contact is always endogenous regardless of geographical area analysed.

Hispanics at Boston train stations in homogeneously white neighbourhoods, finding that treatment induced stronger exclusionary attitudes among residents. This type of experiment is more relevant for understanding the effects of real-world demographic changes on exclusionary attitudes than studies analysing various forms of involuntary inter-ethnic interactions, but the external validity for settings outside the very local context of Boston train stations is clearly questionable.

In general, group-threat theory also often receives support in studies analysing aggregate vote outcomes, which tend to find that diversity boosts aggregate vote shares for right-wing and anti-immigration parties (Barone et al. 2016; Dustmann et al. 2016: Halla et al. 2017; Harmon 2017; Jofre-Monsey et al. 2011; Otto and Steinhardt 2014), although some research reaches different conclusions (Mendez and Cutillas 2014; Steinmayer 2016). At the same time, studies analysing aggregate turnout find mixed effects (Barone et al. 2016; Bratti et al. 2017; Dustmann et al. 2016: Mendez and Cutillas 2014). Regardless, research studying aggregate outcomes cannot normally separate voter responses from the effects of mobility due to immigration.[3]

While most studies focus on the effects of increasing diversity on vote outcomes and turnout, some research investigates the effects of decreasing diversity instead. In an interesting contribution, Enos (2016) analyses a natural experiment in Chicago when the reconstruction of public housing displaced African Americans living close to neighbourhoods predominantly inhabited by whites. As a result of the African-American outflow from nearby communities, turnout decreased by over 10 percentage points among whites, who also became less likely to vote Republican. These effects decrease the farther away voters lived from the housing projects. At the same time, turnout and party choices among African Americans nearby were unaffected.

Overall, the empirical literature on the effects of diversity on inter-group attitudes and political behaviour is thus mixed. Stronger studies analysing political behaviour tend to support the notion of group threat rather than the contact hypothesis more generally, but effects on turnout specifically are not consistent across these studies either. Furthermore, prior methodologically advanced research on the attitudinal and political effects of diversity generally focuses on very local settings, which makes it

[3] It is also often unclear whether the identification strategies work as intended. For example, some research exploits historical levels of immigration and housing stock as instruments (e.g. Barone et al. 2016; Harmon 2017), which may generate bias because of potential serial correlation in unobserved regional characteristics that both attract immigrants, or change the housing stock, and affect outcomes.

difficult to draw general conclusions, and/or analyses aggregate outcomes, which makes it impossible to separate voter responses from the effects of native mobility, which may also change as a result of diversity. Finally, the preponderance of research has focused on immigration or ethnic diversity in a broader sense rather than on refugee immigration specifically. In this paper, we exploit a national refugee placement programme as well as individual-level panel data in an attempt to mitigate these problems and gaps in the existing literature.

4.3. Swedish immigration and the refugee placement programme

To study the causal effects of refugee immigration on voter turnout, we exploit data from Sweden, which has seen significant changes in its immigration patterns since World War II. After a brief period of refugee influx from Eastern Europe in the late 1940s, intra-European labour migrants dominated inflows between the 1950s and the 1970s. These migrants came primarily from Nordic countries (especially Finland). However, as a result of more restrictive rules and a less favourable economic climate, economic immigration gradually decreased in the 1970s, while refugee immigration from the developing world increased, first gradually and then rapidly from the mid-1980s onwards (Lundh and Ohlsson 1999; Nilsson 2004). Indeed, Sweden accepted the highest number of refugees per capita in Europe each year between 1983 and 2003 (Ruist 2015). Unsurprisingly, this development has generated an increasingly heterogeneous population. In 1960, 4 per cent of the population were born abroad; by 2014, this figure had increased to 17 per cent (Statistics Sweden 2015).

As a consequence of the shifting migration patterns, the immigrant population has also become increasingly ethnically different from native Swedes. As late as 1970, only 5 per cent of the foreign-born population came from outside Europe, a figure that had increased to 12 per cent in 1980, 28 per cent in 1990, 39 per cent in 2000, and 48 per cent in 2012 (Aldén and Hammarstedt 2014). In other words, in the last thirty years of the 20th century, Sweden was transformed from an ethnically homogeneous country to an ethnically diverse country.

The refugee situation changed especially from 1986 onwards, which is the starting point for the period under investigation in this paper. Whereas on average fewer than 5,000 refugees arrived annually in the period 1982–85, this figure increased radically to 19,000 refugees annually in the period 1986–91, and further to 35,000 refugees annually in the period 1992–94 (Dahlberg et al. 2012). To cope with the unprecedented situation, the Swedish government implemented a refugee placement programme, which lasted between 1985 and July 1994, with the intention to distribute refugees more evenly across the country's then 284 municipalities and especially break their concentration to large cities. The idea was first to contract about 60 municipalities, but more came to participate as the number of refugees increased radically (Edin et al. 2003). According to the data utilised in this paper – which were extracted from the official contracts by Nekby and Pettersson-Lidbom (2012, 2017) – 196 municipalities agreed to accept at least some refugees in 1986, a figure that increased to 243 only a year later and to 279 in 1990. In other words, the placement programme affected essentially the entire country, although to different degrees.

It also appears clear that the goal to distribute incoming refugees more evenly was achieved: there is a clear trend break in 1985 in terms of the share of refugees settling in Stockholm, Gothenburg, and Malmö. Between 1982 and 1984, the share of new refugees who moved into these three municipalities increased from about 50 to 60 per cent. In 1985, however, the figure decreased to about 35 per cent, and it continued to decrease to just over 10 per cent in 1990, while it increased slightly again from 1991 to just below 20 per cent in 1994. Meanwhile, the share of incoming refugees who ended up in municipalities with fewer than 50,000 inhabitants increased from below 20 per cent in 1984 to over 50 per cent in 1989 (Dahlberg et al. 2012). Thus, the placement programme both altered settlement patterns and, in a context of rapidly increasing refugee inflows, induced considerable within-municipality variation over time.

Importantly, refugees who were allocated via the programme were assigned to municipalities based on contracts rather than being able to choose where to settle initially. While they were allowed to move after the initial assignment, by exploiting the number of contracted refugees as an instrument for refugee inflows in the municipalities, it is possible to circumvent the problem of endogenous settlement and mobility patterns during this period. More generally, while the programme did not place (or contract) refugees randomly across municipalities (see Dahlberg et al. 2013, 2017; Nekby and Pettersson-Lidbom 2012, 2017), for this paper's purposes it is only necessary that predicted refugee inflows based on the contracts are exogenous to changes in individual-level turnout once other relevant variables are held constant. We present evidence in support of this assumption.

4.4. Data and methodology

During the period analysed, elections occurred every three years, always on the third Sunday in September. All elections – national, municipal, and county – are held on the same day. Non-citizens are eligible to vote in municipal and county elections if they have been officially registered as living in Sweden for at least three years prior to the election.[4] Immigrants (refugees) are required to be officially registered as living in Sweden for five (four) years prior to being eligible to apply for Swedish citizenship, which gives them the right to vote also in national elections. The paper's estimation strategy, described formally in Section 4.5, hinges on being able to link data on individual-level turnout to municipal-level variables, including the measure we exploit to obtain plausibly exogenous variation in refugee inflows. This section describes the data and methodology utilised in the analysis.[5] The descriptive statistics are displayed in Table 4.1 at the end of this section.[6]

4.4.1. Voter turnout

Data on voter turnout are obtained from the Swedish National Election Studies Programme (SNES 2015), a survey that has been carried out in conjunction with every election since 1956.[7] Since 1973, the surveys are carried out as rotating panels in which individuals are interviewed at two elections in a row, with a new panel sample being drawn on each occasion. The panel sample is randomly selected and is representative of the Swedish population of eligible voters in national elections at the time of selection. The survey contains information on political and voting preferences as well as background characteristics of the respondents.

[4] Citizens of EU/Nordic countries must only be officially registered in Sweden at the time of the election to be able to vote in these elections.

[5] We thank Karin Edmark and Per Pettersson-Lidbom who kindly supplied data on most municipal-level variables used in the paper, which are obtained from Statistics Sweden, the Swedish Migration Agency, and the Swedish Public Employment Service. We obtained the remaining municipal-level variables, and updated some variables supplied by the authors, using Statistics Sweden's (2016) database.

[6] The negative minimum value for Δ%Refugees is due to (1) two observations where the municipal population increased faster during the period than the refugee inflow over that period and (2) 28 observations covering respondents who moved to a municipality with a smaller accumulated refugee inflow in the year of the second election than the initial accumulated refugee inflow in the home municipality. If these observations are excluded, the minimum is zero.

[7] The election-survey data were originally collected within a research project at the Department of Political Science, University of Gothenburg. For the main years analysed in this study, the principal investigators were Sören Holmberg and Mikael Gilljam. Neither bears any responsibility for the analyses and findings in this paper.

The paper analyses data from the elections in 1985, 1988, 1991, and 1994 in the main analysis, and exploits the rotating panel structure to create three different survey panels for pairwise samples: 1985/88, 1988/91, and 1991/94. Each panel thus includes respondents who were surveyed in both years only. This is necessary since we focus on the individual-level voter responses to immigration rather than effects on aggregate turnout, which risk mixing the impact on turnout with demographic changes across communities. If people who feel more threatened by refugee immigration 'vote with their feet' and move to municipalities with less immigration, or avoid moving to municipalities with higher immigration, the comparison groups become contaminated. This is not a trivial concern in this context since prior Swedish research suggests that non-European immigration induces both native flight and avoidance (Aldén et al. 2015), which is also confirmed in our data as shown in Table 4.A.1.[8] Focusing on respondents in the panel sample allows us to take into account mobility bias directly by assigning all movers – corresponding to 393 individuals or 8 per cent of the final sample – to the municipality in which they lived at the time of the first election. This ensures that the comparison groups are defined based on individuals' municipality of residence at the first election, which gets rid of any potential mobility bias due to differential treatment intensity (see Angrist and Pischke 2009).[9]

However, such an intention-to-treat analysis assumes that movers were exposed to the full refugee inflow in the municipality of origin, even though they relocated between the elections. Thus, as outlined in Section 4.5, we calculate the variable capturing refugee inflows using data from the municipality of residence at the time of each election, while calculating the instrument predicting that change using data from the municipality of origin only. In robustness tests, we also provide estimates from an

[8] For example, our results indicate that a 1 percentage point larger accumulated share of contracted refugees in the municipality of origin at $t-1$ is associated with a 10–11 percentage point higher likelihood of moving. Movers are also exposed to a 0.14–0.19 percentage point smaller increase in the accumulated contracted refugee share relative to non-movers, which amounts to an 18–24 per cent decrease compared with the mean. As the findings in Table 4.A.1 display, the results are very similar when studying the share of refugees for which municipalities received a grant. Also, unreported findings show that movers display more positive changes in turnout on average, are younger, and have lower income than non-movers. This shows the importance of accounting for mobility bias – which is only possible when analysing individual-level panel data.

[9] Of course, our choice to focus on individual-level turnout also means that the estimated effects cannot necessarily be extrapolated to the average municipal population. This is because the survey we exploit was designed to be representative of the national population rather than the municipal populations. Thus, the estimated effects should primarily be seen as valid for the randomly sampled population in each municipality.

intention-to-treat analysis as well as a reduced-form analysis where the instrument is used as the main predictor.

While all panel surveys normally contain significant non-response rates and attrition, we circumvent this problem by exploiting a useful feature of this particular survey: certain information is collected directly from administrative records, and this includes voter turnout in national, municipal, and county elections. As a consequence, data are also available for individuals who were sampled, but who did not end up participating in the survey. However, the probability of selection in the second survey of the rotating panel sample for respondents who did not respond in the first survey was 50 per cent, which means that half of them disappeared by the second election. In addition, a few observations, totalling less than 1 per cent of the sample, have missing voting records. Overall, therefore, the attrition rate is just 14 per cent out of a total panel sample of 5,571 individuals, which is comparatively low and means that endogenous sampling is unlikely to be a problem. This is especially true since the principal reason for attrition is that sampled individuals who did not end up participating in the first survey were randomly unselected as potential participants in the second survey.[10]

Yet to make sure that attrition does not threaten the validity of our findings, we also employ inverse probability weighting (Horvitz and Thompson 1952; Solon et al. 2015). This means that all individuals are weighted according to the inverse probability that they were randomly drawn from the entire population of eligible Swedish voters at the time of the second election in the panel.[11] By definition, if attrition or changes in the underlying population between two consecutive elections have made the sample unrepresentative, inverse probability weighting restores its representativeness. To ensure that attrition does not induce sample-selection bias, we always include such weights in the regressions.

Taking into account that there are three observations with missing values on our primary independent variable described in Section 4.4.2, the total sample in the analysis

[10] The available panel sample with voting records totalled 1,599 individuals in 1988 (compared with 1,901 individuals in the original 1985 panel sample), 1,700 individuals in 1991 (compared with 1,956 individuals in the original 1988 panel sample), and 1,481 in 1994 (compared with 1,714 individuals in the original 1991 panel sample), giving a total available panel sample of 4,780 individuals out of the original panel sample of 5,571.

[11] The inverse probability of selection is defined as: 1/(eligible voters in the municipality of origin/eligible voters in the country). We obtain data on the number of eligible voters, nationally and in the different municipalities, from Statistics Sweden (2016).

of the national and municipal elections includes 4,777 individuals from 284 municipalities, each observed in two elections in a row. For the analysis of county elections, the sample is reduced to 4,366 individuals because three municipalities – Gothenburg, Gotland, and Malmö – carried out the responsibilities of the counties in the period analysed, and their inhabitants therefore had no county elections in which to vote.

4.4.2. Refugee inflows and the placement programme as instrument

To capture refugee inflows, we utilise the number of refugees for whom municipalities received a grant from the government for the cost of initial integration. While this variable has previously been used as an exogenous measure for the placement programme itself (e.g. Dahlberg and Edmark 2008; Dahlberg et al. 2012), it in fact covers all newly arrived refugees rather than just those who arrived in the municipalities via the placement programme.[12] This means that the measure is likely to be partly endogenous since it suffers from some self-selection of refugees into the municipalities. In addition, since grants were sometimes paid out with a time lag, there is some measurement error in the variable in terms of precisely when the refugees arrived in the municipality (see Nekby and Pettersson-Lidbom 2012, 2017).[13] For these reasons, we prefer to treat the variable as an endogenous measure of refugee inflows rather than as an exogenous measure of such inflows.

In order to solve the problems of endogenous settlement patterns and measurement error, and hopefully other sources of omitted variable bias, we exploit the placement programme as an instrument, captured by the number of refugees that municipalities were contracted to receive over each period (Nekby and Pettersson-Lidbom 2012, 2017).[14] This variable by definition gets rid of any endogenous settlement patterns and

[12] This is supported by a comparison of the total number of refugees received in the municipalities and the total number of refugees for which municipalities received grant payments in the period 1991–94: in 1991, 1992, 1993, and 1994 respectively, a total of 18,961, 18,472, 25,300, and 61,500 refugees were received (Swedish Migration Agency 2018) and there were 18,842, 18,546, 25,218, and 62,853 refugees for which municipalities received grants, according to our data. The slight discrepancies between the two measures are likely explained by the fact that there was sometimes a time lag between refugee arrival and the grant payment, as noted below.

[13] As indicated by the previous footnote, measurement error due to the time lag between refugee arrival and the grant payment appears to be relevant for a relatively small proportion of refugees in total, at least in the period 1991–94, but we cannot rule out that there are differences between the municipalities in this respect.

[14] Further showing the differences between the contracts and grant payments, yearly changes in the absolute number of contracted refugees are not always strongly correlated with yearly changes in the

measurement error in refugee inflows, as it is solely based on pre-determined contracts covered by the placement programme. The paper thus exploits the number of refugees contracted to arrive in the municipalities during the panel periods as instrument for the number of refugees for which municipalities received grant payments during the same periods, both normalised by the municipality population.

Since we utilise two ratios with (almost) the same divisor as main independent variable and instrument, it is important to ensure that the correlation between them does not merely arise because of the common divisor (see Bazzi and Clemens 2013; Hunt and Clemens 2017).[15] We test whether or not this is the case by creating a placebo instrument, where we replace the number of refugees contracted to arrive in the municipalities during the panel periods as numerator with Poisson-distributed white noise with the same mean. If the findings are solely due to variation in the divisor, we expect them to be similar when using the placebo instrument. In addition, we test whether the results are robust to splitting the two ratios into separate variables, instrumenting the absolute inflow of refugees with changes in the absolute number of contracted refugees, while simultaneously adjusting for changes in the municipal populations (see Kronmal 1993). If the findings depend entirely on variation in the divisor rather than the numerator, one would expect them to change considerably when doing so.

4.4.3. Control variables

As noted in Section 4.3, the placement programme did not contract refugees randomly across different municipalities, but it is still potentially possible to extract variation that is exogenous to individual-level turnout, at least if we adjust for a couple of municipal-level controls. Apart from breaking the concentration of new refugees in the larger cities, one goal was to place refugees in municipalities with good housing and local labour-market conditions. In interviews with Swedish Migration Board officials, available housing has been upheld as a key factor (see Bengtsson 2002; Dahlberg et al.

absolute number of refugees. However, they are strongly correlated over the panel periods, which is what is relevant for this paper (see Dahlberg et al. 2013; Nekby and Pettersson-Lidbom 2012, 2017). Apart from the micro-level regressions reported here, we also found support for this relationship in unreported municipal-level panel regressions over the periods analysed.

[15] The ratios do not have exactly the same divisor because of how we calculate the endogenous variable for movers, as explained in Section 4.4.1.

2012).[16] Both may in turn impact turnout and are thus included as controls in our main models. The former is captured by the local unemployment rate, while the latter is (at least partly) captured by the rate of vacancies in public housing or rental flats, both measured as period averages.[17] Similarly, to account for the aim of the placement programme to distribute refugees more evenly around the country, we include the period average of the municipal-population size as control.

In addition, it is conceivable that preferences for redistribution in the different municipalities, possibly linked to the functioning of the welfare system, affected the number of refugees they were willing to contract. Indeed, refugees had to be supported by welfare benefits in the short run and were also generally more welfare prone in a longer-term perspective (Dahlberg and Edmark 2008). Differential preferences for, and ability to engage in, welfare spending may also be linked to turnout – and we therefore include the period average of per-capita welfare spending as a control.[18]

Furthermore, it is plausible that the local political situation affected the contracted number of refugees. Indeed, Folke (2014) finds that the local political makeup generally affected the influx of refugees into municipalities in the period 1985-2006. Thus, we include the shares of municipal government seats held by the Social Democratic Party, the Left Party, the Moderate Party, the New Democracy Party, the Liberal Party, the Centre Party, and the Christian Democrats during each election period in the analysis.[19] Finally, we also include the initial accumulated share of contracted refugees and the initial total share of foreign citizens to account for the possibility that municipalities may have been more or less likely to contract a higher number of refugees in the subsequent period if they already had contracted a large number of refugees to the

---

[16] The empirical evidence on whether available housing in fact was important for refugee placement is inconclusive (Dahlberg et al. 2013; Nekby and Pettersson-Lidbom 2012, 2017). However, we present evidence indicating that initial levels of available housing predict changes in the contracted refugee shares, suggesting that available housing at least played a role in determining the contracts.

[17] We use the rate of vacancies in public housing or rental flats in September each year, but results are essentially identical if we use the rate of vacancies in March each year.

[18] Data on welfare spending are missing for 38 respondents in five municipalities in 1991 and 1994 in our sample, which we deal with by inter- and extrapolating welfare spending based on the municipal-specific trend in such spending. However, results are unsurprisingly essentially identical if we instead drop these respondents from the sample. In unreported regressions, we also adjusted for the period average of the per-capita tax base, which theoretically may affect, and be affected by, the share of contracted refugees as well as voter turnout for similar reasons. However, the results were almost identical when doing so and the per-capita tax base was not related to refugee inflows in the first stage.

[19] The excluded category is thus the shares of municipal government seats held by the Green Party and small political parties that were never represented at the national level in the period analysed. Results are almost identical if we also include the Green Party seat share, with or without exclusion of some of the other parties' shares.

municipality and/or seen high overall immigration in the recent past – which may also affect changes in voter turnout.

To study whether or not the variables above are in fact relevant for understanding the placement programme, we regressed changes in the share of contracted refugees over the panel periods on the initial levels of the variables.[20] The results are reported in Table 4.A.2. As expected, when excluding municipal-fixed effects and the initial accumulated share of contracted refugees, higher rates of vacancies in public housing and per-capita welfare spending predict larger increases in contracted refugee shares, whereas larger populations and higher unemployment rates predict smaller increases in the contracted refugee shares. However, the initial share of foreign citizens appears unrelated to changes in contracted refugee shares.

There is also evidence that municipalities with larger shares of municipal government seats held by New Democracy, a populist anti-immigration party, experienced smaller increases in the shares of contracted refugees, while larger shares of municipal government seats held by the Christian Democrats predict larger positive changes in contracted refugee shares. Unsurprisingly, the joint F test for all variables combined thus strongly rejects the null hypothesis of no relationship. When adding the initial accumulated share of contracted refugees, which is strongly positively related to future changes in the contracted refugee shares, the coefficients of most of the other variables decrease in size and sometimes become statistically insignificant, but the joint F test continues to reject the null hypothesis.

However, when we include municipality dummies in the regression, the picture changes substantially. The only variables that remain statistically significant are the share of municipal government seats held by the New Democracy Party and (marginally) population size. Consequently, the joint F test now fails to reject the null hypothesis, albeit only marginally so ($p = 0.12$). Thus, while the initial levels of our control variables appear to matter for changes in the contracted refugee shares in the expected ways, these effects mostly disappear when including municipal-fixed effects.

Overall, this exercise suggests that the municipal-level variables discussed above are indeed relevant for understanding the distribution of changes in contracted refugee shares across the country. However, it also suggests that our instrument is much more

[20] Using the initial levels rather than period averages in this analysis is important to decrease the probability of reverse causality, which, as noted below, may be an issue.

likely to be exogenous once we condition on time-invariant unobserved differences across municipalities. Of course, despite doing so, it is possible that the municipal-level controls still have some impact on the evolution of the placement programme. In our main empirical strategy, formally outlined in Section 4.5, we thus adjust for these controls as well as municipal-fixed effects.

Certainly, the municipal-level controls, especially those that are averaged over the panel period, may to some extent be affected by immigration – and could thus potentially act as 'bad controls' in the regressions (Angrist and Pischke 2009). Yet since we are interested in turnout responses to refugee immigration itself – rather than indirect responses via changes in the economic and demographic environment – we believe the controls are relevant for retrieving the parameter of interest. Our main results thus most likely capture psychological effects on turnout of increasing diversity, once the impact on any potentially endogenous variables are held constant. Such psychological effects are also the most theoretically relevant for the implications of group-threat theory for voter turnout (Enos 2016).

Nevertheless, in robustness checks, we also report estimates from models in which we (1) only adjust for the pre-determined local political makeup and (2) exclude all municipal-level controls entirely. This serves as a test of the extent to which potentially endogenous controls affect the estimates, and more generally the likelihood that our instrument is exogenous. If potentially endogenous municipal-level controls are not crucial for our methodology, we expect the findings to be robust to these exercises.

Finally, we include a couple of individual-level variables measured at the first election in each panel as noise controls. These may affect future changes in turnout but should not be affected by our instrument, at least once we condition on the municipal-level controls.[21] More specifically, we adjust for respondents' year of birth, gender, marital status, labour-market status, blue-collar status, whether they live in villa areas or in high-rise apartments, and an indicator for high income.[22] The latter represents roughly the top 20 per cent in the income distribution at each election. We also include

[21] As noted in Section 4.6.3, the results are very similar when excluding most noise controls although they do become slightly less precise. This is expected since they should be related to changes in individual-level turnout, but at the same time not be related to our municipal-level instrument.

[22] There are some missing values on the marriage, labour-market, blue-collar, and housing indicators. We replace these with zero and include separate indicators for missing values in the regressions.

interactions between these variables and the year dummies.[23] In addition, as noted in Section 4.5, we include turnout at $t-1$ in some models to test for potential mean reversion and ensure that initial turnout at the individual level does not affect future treatment intensity.

Table 4.1: Descriptive statistics

	Mean	SD	Min	Max
ΔTurnout (national elections)	-0.006	0.317	-1.000	1.000
ΔTurnout (municipal elections)	-0.006	0.324	-1.000	1.000
ΔTurnout (county elections)	-0.006	0.328	-1.000	1.000
Δ%Refugees	0.823	0.489	-1.386	7.465
Δ%Contracted refugees	0.794	0.417	0.000	11.801
Average welfare spending per capita	317.288	183.995	34.813	938.155
Average population	121,883	185,723	2,958	693,719
%Average unemployment rate	3.263	2.202	0.207	10.080
%Average vacant public housing	1.805	2.632	0.000	21.128
%Foreign citizens	5.457	3.407	0.401	25.568
%Social Democratic Party seats	41.634	8.514	13.333	66.667
%Left Party seats	5.216	3.281	0.000	21.951
%New Democracy Party seats	0.930	1.968	0.000	15.385
%Moderate Party seats	20.533	8.196	3.226	61.224
%Liberal Party seats	11.414	3.652	2.041	29.268
%Centre Party seats	11.969	7.899	0.000	46.939
%Christian Democratic Party seats	3.022	3.303	0.000	20.000

Note: n = 4,777 for all variables apart from ΔTurnout (county elections), where n = 4,366. The variable denoting welfare spending is given in SEK. The descriptive statistics for the independent variables are calculated based on the sample for the national elections. Δ%Refugees is calculated using data from the municipality of residence at the time of each election, while Δ%Contracted refugees is calculated using data from the home municipality only.

4.5. Estimation strategy

The key problem in research studying the relationship between ethnic diversity and voter turnout consists of endogenous settlement patterns and other sources of omitted variable bias. To address these issues, as discussed in Sections 4.3 and 4.4, we exploit the placement programme contracts as instrument for refugee inflows, while taking advantage of the panel structure in the data to analyse the dynamics of individual-level voter turnout. The following are thus our formal baseline equations, estimated in a regular 2SLS set-up:

[23] This means that we allow the effects of all noise controls to differ depending on the year of the election. Since the income information provided is not entirely consistent across survey years, these indicators vary slightly across years. Also, the income information in the 1988 survey was calculated from 1986 rather than from the year of the election, as in the other surveys. Including interactions between the indicators and the year dummies picks up the effects of these slight differences across years.

$$\Delta r_{mt} = \beta_1 \Delta c_{mt} + \beta_2 \bar{x}_{mt} + \beta_3 p_{mt} + \beta_4 b_{im} + \beta_5 i_{mt-1} + \mu_t + \varepsilon_{imt} \quad (1)$$

$$\Delta v_{imt} = \beta_1 \widehat{\Delta r_{mt}} + \beta_2 \bar{x}_{mt} + \beta_3 p_{mt} + \beta_4 b_{im} + \beta_5 i_{mt-1} + \mu_t + \varepsilon_{imt} \quad (2)$$

where Δv_{imt} is the change in individual turnout between $t-1$ and t; Δc_{mt} denotes the change in the accumulated number of refugees contracted to arrive in the respondent's municipality of origin, normalised by the municipal population, which serves as the excluded instrument; Δr_{im} represents the change in the accumulated share of refugees for which municipalities received a grant, calculated using data from the respondent's municipality of residence at the time of each election; $\widehat{\Delta r_{mt}}$ denotes the predicted values from the first stage; $\bar{x}_{mt}$ denotes a vector including the period averages of population levels, public housing vacancies, the unemployment rate, and per-capita welfare spending in the municipality of origin; p_{mt} is a vector of variables denoting the shares of municipal government seats held by the Social Democratic Party, the Left Party, the New Democracy Party, the Moderate Party, the Liberal Party, the Centre Party, and the Christian Democrats, in the municipality of origin following the first election; b_{im} is a vector including the individual-level noise controls; i_{mt-1} is a vector including the initial share of foreign citizens and the initial accumulated share of contracted refugees in the municipality of origin in each period; and μ_t represents time-fixed effects, capturing nation-wide trends in the dependent variable. Standard errors are clustered at the municipality level to allow for correlation at the level at which the independent variable of interest is measured.

By differencing the dependent and main independent variables, the model effectively becomes a difference-in-difference instrumental-variable estimator with individual- and municipal-fixed effects.[24] The identification depends on the assumption that the predicted influx of refugees is exogenous to changes in individual turnout, when conditioning on the control variables outlined above. We test whether or not our assumptions are likely to hold by adding turnout at $t-1$. Lagged turnout is likely to be

[24] A possible alternative would be to aggregate the data at the municipal level and weight the observations by the number of respondents in the panel sample, or the total voting-age population, in each municipality for each election (see Angrist and Pischke 2009; Nekby and Pettersson-Lidbom 2017). However, this is not possible in models where we calculate Δr_{mt} for movers using the municipality of origin in $t-1$ and municipality of residence in t, since the primary independent variable then differs also between respondents originating in the same municipality. Furthermore, to increase precision, we include individual-level noise controls, which makes it more complicated to replicate the regressions in aggregate form (Angrist and Piscke 2009). Thus, since there is no inherent advantage in studying aggregated data, we stick with individual-level data and cluster the standard errors at the municipal level.

a strong predictor of changes in turnout and should thus increase the precision of our coefficient of interest. Yet, if the results are causal, we should not expect lagged turnout to affect the coefficient as such significantly (Angrist and Piscke 2009).[25] Including lagged turnout also means that we can test whether mean reversion is likely to bias our findings and is thus an important robustness check of our results.

Based on the above intuition, we would not expect that changes in diversity over the periods analysed should affect turnout at the first election in the rotating panel. In other words, the sample should be balanced on the dependent variable at the outset. We thus estimate equations (1) and (2) but swap Δv_{imt} for v_{imt-1} as dependent variable. On the other hand, we should not expect the same sample to be balanced at the second election following treatment, which we test by swapping Δv_{imt} for v_{imt}. If the samples are balanced at the first election but not at the second, this further supports the idea that our main estimates capture the causal effect of refugee inflows on voter turnout.

In addition, we estimate placebo tests in which we regress changes in turnout on future refugee inflows. The idea behind this exercise is simple: if our research design picks up the causal impact of refugee immigration on turnout, changes in refugee shares should not affect past changes in turnout. In the placebo test, we thus study whether changes in individual-level turnout between $t-1$ and t are affected by refugee inflows between t and $t+1$. We then use changes in the accumulated number of contracted refugees in the municipality of origin, normalised by the municipal population, between t and $t+1$ as an instrument for changes in the accumulated refugee share over the same period, calculated using data from the respondents' municipality of residence at the time of each election. Since we have access to data from the election study carried out in 1982, we are able to carry out also the placebo analysis using three panels – 1982/85, 1985/88, and 1988/91 – with samples of similar sizes as in the main analyses.

Overall, we believe our research design is likely to capture a causal impact of refugee immigration on individual-level turnout. Particularly, exploiting the contracts as instruments entirely eliminates the problem of self-selection of refugees into certain municipalities, whereas the panel structure of our data allows us to ensure that native mobility does not bias the findings. Importantly, we are able to test whether our assumptions hold using the balancing and placebo tests described. However, we note

[25] Note that the inclusion of voter turnout at $t-1$ means that the model no longer includes individual- and municipal-fixed effects. If the results are very similar without such effects included, while adjusting for individuals' initial turnout, it further supports the causal interpretation of the findings.

that if there is any remaining bias, it is likely to drive coefficients in a direction that would make it easier to reject the group-threat theory and confirm the contact hypothesis. This is because municipalities with populations who are less threatened and more tolerant towards refugees should be more likely to accept larger contracted refugee inflows, consequently resulting in lower political mobilisation.

4.6. Results

In this section, we report our main findings as well as results from the balancing, placebo, and robustness tests. We begin by noting that the results from OLS models in Table 4.2, which assume that exposure to refugee inflows is exogenous, display a positive correlation between such inflows and changes in voter turnout: a 1 percentage point increase in the refugee share predicts a 3 percentage point higher probability of voter turnout. This holds true whether or not we include lagged turnout in the models, which makes the estimates slightly more precise, and regardless of the election type analysed. However, since our main independent variable to some extent suffers from endogenous settlement patterns and measurement error, and since the variable also picks up the impact of changes in refugee exposure among movers, these results are likely to be biased towards zero.

Table 4.2: The relationship between refugee inflows and changes in turnout (OLS)

Panel 1: Municipal-level controls and individual-level noise controls			
Election type	National	Municipal	County
	(1)	(2)	(3)
Δ%Refugees	0.033**	0.032*	0.029*
	(0.016)	(0.016)	(0.017)
Panel 2: Add initial turnout			
	(7)	(8)	(9)
Δ%Refugees	0.033**	0.032**	0.030**
	(0.014)	(0.014)	(0.015)
n	4,777	4,777	4,366
Municipalities	284	284	281

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the municipality level in parentheses. All regressions include period-fixed effects. Respondents are weighted by the inverse of the probability of being selected for the survey in the relevant election.

Turning to the 2SLS models in Table 4.3, this is indeed what the results suggest. The estimates in the first panel display a positive impact of refugee inflows on voter turnout that is about twice as large as the OLS estimates: a 1 percentage point increase in the refugee inflow raises the probability of voter turnout by 5–6 percentage points. In the second panel, which adds lagged turnout, the effects are again very similar but more precise. This is expected if our strategy captures causal effects, as lagged turnout is a strong predictor of changes in turnout.

Meanwhile, the Hausman tests generally display statistically significant values, indicating that the OLS estimates in Table 4.2 are biased downwards, while the F tests display values considerably higher than 23.1, which is the valid threshold when utilising cluster-robust standard errors (Olea and Pflueger 2013). This indicates that refugees self-select into municipalities where natives are less likely to respond by mobilising politically, that natives leave/avoid municipalities with larger refugee shares (as our findings in Table 4.A.1 show), and/or that measurement error drives the OLS results towards zero. It also shows that our instrument based on contracted refugees is strong enough to be utilised. Indeed, the estimates indicate that a 1 percentage point increase in the share of contracted refugees in the home municipality raises the exposure to refugee inflows by 0.64 percentage points. Our results thus suggest that refugee immigration has causal positive effects on voter turnout.

Table 4.3: The causal effect of refugee inflows on changes in turnout (IV)

Panel 1: Municipal-level controls and individual-level noise controls			
Election type	National	Municipal	County
	(1)	(2)	(3)
Second stage			
Δ%Refugees	0.057***	0.047**	0.061**
	(0.022)	(0.021)	(0.026)
First stage			
Δ%Contracted refugees	0.638***	0.638***	0.637***
	(0.043)	(0.043)	(0.043)
Hausman test	0.095	0.312	0.061
F statistic	225.02	225.00	218.72
Panel 2: Add initial turnout			
	(4)	(5)	(6)
Second stage			
Δ%Refugees	0.063***	0.056***	0.065***
	(0.020)	(0.019)	(0.023)
First stage			
Δ%Contracted refugees	0.638***	0.638***	0.637***
	(0.043)	(0.043)	(0.043)
Hausman test	0.018	0.061	0.014
F statistic	224.93	224.77	218.72
n	4,777	4,777	4,366
Municipalities	284	284	281

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the municipality level in parentheses. All regressions include period-fixed effects. Respondents are weighted by the inverse of the probability of being selected for the survey in the relevant election.

4.6.1. *Balancing tests*

If our research design isolates exogenous variation in refugee inflows during the periods analysed, we do not expect future treatment to be related to initial turnout. That is, the sample should be balanced on the dependent variable at the first election in the panel, once we adjust for the other variables in the model. However, the same sample should not be balanced following treatment at the second election in the panel. Table 4.4 displays the findings from this exercise. There is little evidence that future treatment is related to initial turnout, suggesting that the sample is indeed balanced. However, treatment is related to turnout at the second election in the same sample. In fact, the results are essentially identical to our main findings, even though these models analyse the level of, rather than change in, turnout without initial turnout as a control. Overall, the results thus corroborate the idea that our main results do indeed reflect the causal effects of refugee immigration on voter turnout at the individual level.

Table 4.4: Balancing tests

Panel 1: Turnout at the first election (level)			
Election type	National	Municipal	County
	(1)	(2)	(3)
Second stage			
Δ%Refugees	0.012	0.017	0.008
	(0.020)	(0.020)	(0.021)
First stage			
Δ%Contracted refugees	0.640***	0.640***	0.638***
	(0.042)	(0.042)	(0.043)
Hausman test	0.500	0.388	0.713
F statistic	226.91	226.89	220.34
Panel 2: Turnout at the second election (level)			
	(4)	(5)	(6)
Second stage			
Δ%Refugees	0.069***	0.064***	0.069***
	(0.023)	(0.022)	(0.024)
First stage			
Δ%Contracted refugees	0.638***	0.638***	0.637***
	(0.043)	(0.043)	(0.043)
Hausman test	0.030	0.051	0.028
F statistic	225.02	225.00	218.72
n	4,777	4,777	4,366
Municipalities	284	284	281

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the municipality level in parentheses. All regressions include municipal-level controls, individual-level noise controls, and period-fixed effects. Respondents are weighted by the inverse of the probability of being selected for the survey in the relevant election.

4.6.2. Placebo tests

To provide further evidence on the causality of our findings, we carry out the formal placebo test in treatment, as described in Section 4.5. We do so by lagging the panel of voters by one period and pretend that refugee inflows occurred in the previous panel period. We add data from the election study in 1982 in order to carry out the placebo tests over three panels to make it as similar as possible to the main analysis. This means that changes in voter turnout in the periods 1982–85, 1985–88, and 1988–91 are regressed on refugee inflows due to the placement programme in the periods 1985–88, 1988–91, and 1991–94 respectively. If our assumptions hold true, we should not find any significant estimates in this analysis since treatment occurs in the periods after the change in individual turnout is observed. Reassuringly, the results in Table 4.5 display no relationship between changes in individual turnout and future immigration spurred by the placement programme. The point estimates are negative but close to zero and far from statistically significant. Thus, the results from the placebo tests support the causal interpretation of our main findings.

Table 4.5: Placebo tests regressing changes in turnout on future refugee inflows

Panel 1: Municipal-level controls and individual-level noise controls			
Election type	National	Municipal	County
	(1)	(2)	(3)
Second stage			
Δ%Refugees$_{t+1}$	-0.007	-0.009	-0.012
	(0.011)	(0.011)	(0.011)
First stage			
Δ%Contracted refugees$_{t+1}$	0.672***	0.672***	0.672***
	(0.081)	(0.081)	(0.082)
Hausman test	0.693	0.769	0.444
F statistic	68.30	68.22	67.61
Panel 2: Add initial turnout			
	(4)	(5)	(6)
Second stage			
Δ%Refugees$_{t+1}$	-0.005	-0.006	-0.008
	(0.015)	(0.015)	(0.015)
First stage			
Δ%Contracted refugees$_{t+1}$	0.673***	0.672***	0.672***
	(0.081)	(0.081)	(0.081)
Hausman test	0.457	0.452	0.542
F statistic	69.68	69.42	68.87
n	4,930	4,881	4,450
Municipalities	284	284	281

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the municipality level in parentheses. All regressions include period-fixed effects and weight respondents by the inverse of the probability of being selected for the survey in the relevant election.

4.6.3. Robustness tests

In this section, we provide evidence on the robustness of our findings. First, we test to what extent the results change when we utilise the placebo instrument described in Section 4.4.2. This is to ensure that the results are not merely a consequence of changes in the size of the municipal population, which serves as the divisor in both the main independent variable and the instrument. We also test whether the results are robust to splitting each of the two ratios into separate variables, instrumenting the absolute inflow of refugees with changes in the absolute number of contracted refugees, while simultaneously adjusting for changes in the municipal population. The results in Table 4.A.3 show that the second-stage results turn insignificantly negative when using the placebo instrument, with the F test in the first stage showing no correlation whatsoever between the placebo instrument and the main independent variable. Furthermore, when splitting the ratios into separate variables, the results are very similar to our main findings. Overall, these robustness tests thus confirm that our main results are due to the variation in refugee inflows spurred by the placement programme, rather than a spurious correlation arising from the common divisor.

In further robustness tests, as noted in Section 4.4.3, we analysed the extent to which the results were affected by only including the local political makeup as control at the municipal level, and by excluding all municipal-level controls entirely. We also tested omitting all individual-level noise controls apart from respondents' year of birth. Furthermore, we excluded municipalities with more than 50,000 inhabitants from the analysis, thus halving the sample size to 2,382–2,389 respondents in 244 municipalities. As one of the aims behind the placement programme was to redistribute refugee inflows to smaller municipalities, we expect results to be similarly pronounced in this analysis. Finally, as noted in Section 4.4.1, we carried out the intention-to-treat analysis, in which both changes in refugee shares and the instrument are calculated from the municipality of origin, as well as a reduced-form analysis. As Tables 4.A.4–4.A.9 show, the results are entirely robust to these exercises.[26]

[26] However, the results are not robust to using the reduced samples in Dahlberg et al.'s (2013) study, which analyses a survey response as dependent variable and consequently suffers from attrition. When analysing their samples, giving us up to 2,439 respondents in 276 municipalities, with the exact number depending on the sample restrictions, our estimates are closer to zero and not statistically significant. The bias towards zero is also stronger in proportion to the degree of attrition. These findings support Nekby and Pettersson-Lidbom's (2012, 2017) argument: significant non-random attrition is likely to pose a

Overall, all results thus point in one direction: (1) refugee inflows spurred by the placement programme had positive effects on voter turnout, and (2) strategies that do not properly deal with endogenous settlement and mobility patterns as well as measurement error are likely to bias estimates towards zero. More generally, the findings provide relatively strong evidence in favour of group-threat theory over the contact hypothesis, in the context of changing immigration patterns in Sweden in the last decades of the 20th century.

4.7. Conclusion

This paper has investigated the impact of refugee immigration on voter turnout in Sweden in the period 1985–94. Exploiting a placement programme through which refugees were contracted to the country's municipalities, and individual-level panel data almost entirely free from attrition, we have sought to obtain conditionally exogenous variation in refugee inflows, in order to be able to draw conclusions regarding causal inference. This is also our principal contribution to a literature that thus far mostly has provided associational evidence and/or analysed aggregate outcomes.

Our results showed that refugee inflows spurred by the placement programme increased the probability of voter turnout. Balancing tests on initial turnout, and placebo tests analysing whether refugee inflows predict prior changes in individual-level turnout, corroborate the causal interpretation of the results, as do several robustness checks. The main findings also differ from OLS estimates, which are considerably smaller. Endogenous settlement and mobility patterns and/or measurement error thus appear to bias results towards zero, highlighting the importance of obtaining variation in refugee inflows free from these problems in order to draw credible causal inferences.

Certainly, since the survey we exploit was designed to be representative at the country level rather than the municipal level, the external validity of the findings may be questionable; the results should primarily be seen as valid for the randomly sampled population in each municipality. It is also important to note that the effects are obtained in the specific context of increasing refugee immigration in a previously homogeneous

problem for studies linking the placement programme to survey data more generally. A key strength of this paper is that we are able to circumvent this problem almost entirely.

country; whether other types of immigration have similar effects is unclear. Furthermore, it is not clear whether the increased probability in individual-level turnout translates into higher aggregate turnout, which partly depends on how immigration affects overall demographic changes across communities. The extent to which these issues affect our conclusions is an important venue for future research.

Since group-threat theory and the contact hypothesis were formulated over half a century ago, a large body of research has sought to evaluate their relative relevance for understanding various social and political outcomes. Conceptually and empirically, this literature suggests the importance of distinguishing between the effects of voluntary and involuntary contact. Perceptions of threat may potentially be addressed through increased inter-group interactions, which have been found to improve inter-group attitudes in some contexts, but such interactions tend to be endogenous to existing attitudes. While silent on the effects of involuntary contact, our study provides evidence supporting group-threat theory in the context of real-world demographic changes that do not necessarily generate more voluntary contact between in- and out-group members. In light of intense debate regarding potential consequences of current European immigration patterns, this is an important finding to which policymakers should pay attention.

References

Aldén, Lina and Mats Hammarstedt. 2014. 'Utrikes födda på den svenska arbetsmarknaden – en översikt och en internationell jämförelse.' Report 2014:5, Centre for Labour Market and Discrimination Studies, Linnaeus University, Växjö.

Aldén, Lina, Mats Hammarstedt, and Emma Neuman. 2015. 'Ethnic Segregation, Tipping Behavior, and Native Residential Mobility.' *International Migration Review* 49(1):36–69.

Alesina, Alberto and Edward Glaeser. 2004. *Fighting Poverty in the U.S. and in Europe: A World of Difference*. New York: Oxford University Press.

Alesina, Alberto and Eliana La Ferrara. 2000. 'Participation in Heterogeneous Communities.' *Quarterly Journal of Economics* 115(3):847–904.

Allport, Gordon W. 1954. *The Nature of Prejudice*. Cambridge, MA: Perseus Books.

Angrist, Joshua A. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Arzheimer, Kai. 2009. 'Contextual Factors and the Extreme Right Vote in Western Europe, 1980–2002.' *American Journal of Political Science* 53(2):259–275.

Avery, James M. and Jeffrey A. Fine. 2012. 'Context Matters: The Effect of Racial Composition of State Electorates on White Racial Attitudes.' *American Review of Politics* 33:211–231.

Barone, Guglielmo, Alessio D'Ignazio, Guido de Blasio, and Paolo Naticchioni. 2016. 'Mr. Rossi, Mr. Hu and Politics: The Role of Immigration in Shaping Natives' Voting Behavior.' *Journal of Public Economics* 136:1–13.

Bay, Ann-Helén, Henning Finseraas, and Axel West Pedersen. 2013. 'Welfare Dualism in Two Scandinavian Welfare States: Public Opinion and Party Politics.' *West European Politics* 36(1):199–220.

Bazzi, Samuel and Michael A. Clemens. 2013. 'Blunt Instruments: Avoiding Common Pitfalls in Identifying the Causes of Economic Growth.' *American Economic Journal: Macroeconomics* 5(2):152–186.

Bengtsson, Marie. 2002. 'Stat och kommun i makt(o)balans: En studie av flyktingmottagandet.' PhD Dissertation, Department of Political Science, Lund University.

Bhatti, Yosef, Bolette Danckert, and Kasper M. Hansen. 2017. 'The Context of Voting: Does Neighborhood Ethnic Diversity Affect Turnout?' *Social Forces* 95(3):1127–1154.

Blais, André and Daniel Rubenson. 2013. 'The Source of Turnout Decline: New Values or New Contexts?' *Comparative Political Studies* 46(1):95–117.

Bobo, Lawrence and Vincent Hutchings. 1996. 'Perceptions of Racial Group Competition: Extending Blumers Theory of Group Position to a Multiracial Social Context.' *American Sociological Review* 61(6):951–972.

Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles. 2006. 'Empathy or Antipathy? The Impact of Diversity.' *American Economic Review* 96(5):1890–1905.

Bracco, Emanuele, Maria De Paola, Colin P. Green, and Vincenzo Scoppa. 2018. 'The Effect of Far Right Parties on the Location Choice of Immigrants: Evidence from Lega Nord Mayors.' *Journal of Public Economics* 166:12–26.

Brady, David and Ryan Finnigan. 2014. 'Does Immigration Undermine Public Support for Social Policy?' *American Sociological Review* 79(1):17–42.

Bratti, Massimiliano, Claudio Deiana, Enkelejda Havari, Gianluca Mazzarella, and Elena C. Meroni. 2017. 'What Are You Voting For? Proximity to Refugee Reception Centres and Voting in the 2016 Italian Constitutional Referendum.' IZA Discussion Paper No. 11060, Institute of Labor Economics, Bonn.

Carrell, Scott E., Mark Hoekstra, and James E. West. 2015. 'The Impact of Intergroup Contact on Racial Attitudes and Revealed Preferences.' NBER Working Paper No. 20940, National Bureau of Economic Research, Cambridge, MA.

Costa-Font, Joan and Frank Cowell. 2014. 'Social Identity and Redistributive Preferences: A Survey.' *Journal of Economic Surveys* Doi: 10.1111/joes.12061.

Dahlberg, Matz and Karin Edmark. 2008. 'Is There a 'Race-to-the-Bottom' in the Setting of Welfare Benefit Levels? Evidence from a Policy Intervention.' *Journal of Public Economics* 92(5–6):1193–1209.

Dahlberg, Matz, Karin Edmark, and Heléne Berg. 2017. 'Revisiting the Relationship between Ethnic Diversity and Preferences for Redistribution: Reply.' *Scandinavian Journal of Economics* 119(2):288–294.

Dahlberg, Matz, Karin Edmark, and Heléne Lundquist. 2013. 'Ethnic Diversity and Preferences for Redistribution: Reply.' IFN Working Paper No. 955, Research Institute of Industrial Economics, Stockholm.

Dahlberg, Matz, Karin Edmark, and Heléne Lundqvist. 2012. 'Ethnic Diversity and Preferences for Redistribution.' *Journal of Political Economy* 120(1):41–76.

Damm, Anna P. 2009. 'Determinants of Recent Immigrants' Location Choices: Quasi-experimental Evidence.' *Journal of Population Economics* 22(1):145–174.

Della Posta, Daniel J. 2013. 'Competitive Threat, Intergroup Contact, or Both? Immigration and the Dynamics of Front National Voting in France.' *Social Forces* 92(1):249–273.

Dustmann, Christian and Tommaso Frattini. 2012. 'Immigration: The European Experience.' Discussion Paper No. 2012–01, NORFACE Research Programme on Migration, London.

Dustmann, Christian and Ian Preston. 2001. 'Attitudes to Ethnic Minorities, Ethnic Context and Location Decisions.' *Economic Journal* 111:353–373.

Dustmann, Christian, Kristine Vasiljeva, and Anna P. Damm. 2016. 'Refugee Migration and Electoral Outcomes.' Discussion Paper CPD 19/16, Centre for Research and Analysis of Migration, University College London.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. 2003. 'Ethnic Enclaves and the Economic Success of Immigrants: Evidence from a Natural Experiment.' *Quarterly Journal of Economics* 118(1):329–357.

Enos, Ryan D. 2014. 'Causal Effect of Intergroup Contact on Exclusionary Attitudes.' *Proceedings of the National Academy of Sciences of the United States of America* 111(10):3699–3704.

Enos, Ryan D. 2016. 'What the Demolition of Public Housing Teaches Us about the Impact of Racial Threat on Political Behavior.' *American Journal of Political Science* 60(1):123–142.

Fieldhouse, Edward and David Cutts. 2008. 'Diversity, Density and Turnout: The Effect of Neighbourhood Ethno-religious Composition on Voter Turnout in Britain.' *Political Geography* 27(5):530–548.

Folke, Olle. 2014. 'Shades of Brown and Green: Party Effects in Proportional Election Systems.' *Journal of the European Economic Association* 13(5):1361–1395.

Fox, Cybelle. 2004. 'The Changing Color of Welfare? How Whites' Attitudes toward Latinos Influence Support for Welfare.' *American Journal of Sociology* 110(3):580–625.

Gerdes, Christer and Eskil Wadensjö. 2010. 'The Impact of Immigration on Election Outcomes in Danish Municipalities.' Working Paper 2010:3, The Stockholm University Linnaeus Center for Integration Studies (SULCIS), Stockholm.

Giles, Michael W. and Melanie A. Buckner. 1993. 'David Duke and Black Threat: An Old Hypothesis Revisited.' *Journal of Politics* 55(3):702–713.

Gray, Mark and Miki Caul. 2000. 'Declining Voter Turnout in Advanced Industrial Democracies, 1950 to 1997: The Effects of Declining Group Mobilization.' *Comparative Political Studies* 33(9):1091–1122.

Halla, Martin, Alexander F. Wagner, and Josef Zweimüller. 2017. 'Immigration and Voting for the Far Right.' *Journal of the European Economic Association* DOI: https://doi.org/10.1093/jeea/jvx003.

Harmon, Nikolaj A. 2017. 'Immigration, Ethnic Diversity, and Political Outcomes: Evidence from Denmark.' *Scandinavian Journal of Economics* DOI: 10.1111/sjoe.12239.

Harrison, Sarah and Michael Bruter. 2011. *Mapping Extreme Right Ideology: An Empirical Geography of the European Extreme Right*. New York: Palgrave Macmillan.

Hill, Kim Q. and Jan E. Leighley. 1999. 'Racial Diversity, Voter Turnout, and Mobilizing Institutions in the United States.' *American Politics Research* 27(3):275–295.

Hopkins, Daniel J. 2010. 'Politicized Places: Explaining Where and When Immigrants Provoke Local Opposition.' *American Political Science Review* 104:40–60.

Horvitz, D. G. and D. J. Thompson. 1952. 'A Generalization of Sampling Without Replacement From a Finite Universe.' *Journal of the American Statistical Association* 47(260):663–685.

Hunt, Jennifer and Michael A. Clemens. 2017. 'The Labor Market Effects of Refugee Waves: Reconciling Conflicting Results.' NBER Working Paper No. 23433, National Bureau of Economic Research, Cambridge, MA.

Jofre-Monseny, Jordi, Pilar Sorribas-Navarro, and Javier Vázquez-Grenno. 2011. 'Welfare Spending and Ethnic Heterogeneity: Evidence from a Massive Immigration Wave.' Working Paper 2011/34, IEB – Institute d'Economia de Barcelona.

Key, V. O. 1949. *Southern Politics in State and Nation*. New York: Alfred A. Knopf.

Kronmal, Richard A. 1993. 'Spurious Correlation and the Fallacy of the Ratio Standard Revisited.' *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 156(3):379–392.

Leighley, Jan E. and Arnold Vedlitz. 1999. 'Race, Ethnicity, and Political Participation: Competing Models and Contrasting Explanations.' *Journal of Politics* 61(4):1092–1114.

Lundh, Christer and Rolf Ohlsson. 1999. *Från arbetskraftsimport till flyktinginvandring*. Stockholm: Studieförbundet Näringsliv och Samhälle.

Markaki, Yvonni and Simonetta Longhi. 2012. 'What Determines Attitudes to Immigration in European Countries? An Analysis at the Regional Level.' Discussion Paper No. 2012–32, Norface Migration, University of Essex .

Matthews, Donald R. and James W. Prothro. 1963. 'Social and Economic Factors and Negro Voter Registration in the South.' *American Political Science Review* 57(1):24–44.

McLaren, Lauren M. 2003. 'Anti-Immigrant Prejudice in Europe: Contact, Threat Perception, and Preferences for the Exclusion of Migrants.' *Social Forces* 81(3):909–936.

McLaren, Lauren and Mark Johnson. 2007. 'Resources, Group Conflict and Symbols: Explaining Anti-Immigration Hostility in Britain.' *Political Studies* 55(4):709–732.

Mendez, Ildefonso and Isabel M. Cutillas. 2014. 'Has Immigration Affected Spanish Presidential Election Results?' *Journal of Population Economics* 27:135–171.

Nekby, Lena and Per Pettersson-Lidbom. 2012. 'Revisiting the Relationship between Ethnic Diversity and Preferences for Redistribution.' Working Paper, Department of Economics, Stockholm University.

Nekby, Lena and Per Pettersson-Lidbom. 2017. 'Revisiting the Relationship between Ethnic Diversity and Preferences for Redistribution: Comment.' *Scandinavian Journal of Economics* 119(2):268–287.

Newman, Benjamin J. 2013. 'Acculturating Contexts and Anglo Opposition to Immigration in the United States.' *American Journal of Political Science* 57(2):374–390.

Nilsson, Åke. 2004. 'Efterkrigstidens invandring och utvandring.' Demographic Reports 2004:5, Statistics Sweden, Stockholm.

OECD. 2011. *Naturalisation: A Passport for the Better Integration of Immigrants?* Paris: OECD.

Olea, José L. M. and Carolin Pflueger. 2013. 'A Robust Test for Weak Instruments.' *Journal of Business & Economic Statistics* 31(3):358–369.

Oliver, J. Eric. and Tali Mendelberg. 2000. 'Reconsidering the Environmental Determinants of White Racial Attitudes.' *American Journal of Political Science* 44(3):574–589.

Oliver, J. Eric. and Janelle Wong. 2003. 'Intergroup Prejudice in Multiethnic Settings.' *American Journal of Political Science* 47(4):567–582.

Otto, Alkis H. and Max F. Steinhardt. 2014. 'Immigration and Election Outcomes — Evidence from City Districts in Hamburg.' *Regional Science and Urban Economics* 45:67–79.

Roch, Christine H. and Michael Rushton. 2008. 'Racial Context and Voting over Taxes Evidence from a Referendum in Alabama.' *Public Finance Review* 36(5):614–634.

Ruist, Joakim. 2015. 'Refugee Immigration and Public Finances in Sweden.' Working Paper No. 613, Department of Economics, University of Gothenburg.

Rydgren, Jens and Patrick Ruth. 2011. 'Voting for the Radical Right in Swedish Municipalities: Social Marginality and Ethnic Competition?' *Scandinavian Political Studies* 34(3):202–225.

Rydgren, Jens and Patrick Ruth. 2013. 'Contextual Explanations of Radical Right-wing Support in Sweden: Socioeconomic Marginalization, Group Threat, and the Halo Effect.' *Ethnic and Racial Studies* 36(4):711–728.

SCB. 2015. 'Summary of Population Statistics 1960–2014.' Statistics Sweden, Stockholm. http://www.scb.se/en_/Finding-statistics/Statistics-by-subject-area/Population/Population-composition/Population-statistics/Aktuell-Pong/25795/Yearly-statistics--The-whole-country/26040/.

Schaeffer, Merlin. 2013. 'Ethnic Diversity, Public Goods Provision and Social Cohesion: Lessons from an Inconclusive Literature.' Discussion Paper SP VI 2013–103, WZB Berlin Social Research Center, Berlin.

Schlichting, Kurt, Peter Tuckel, and Richard Maisel. 1998. 'Racial Segregation and Voter Turnout in Urban America.' *American Politics Research* 26(2):218–236.

Schlueter, Elmar and Peer Scheepers. 2010. 'The Relationship Between Outgroup size and Anti-outgroup Attitudes: A Theoretical Synthesis and Empirical Test of Group Threat- and Intergroup Contact Theory.' *Social Science Research* 39(2):285–295.

Shook, Natalie J. and Russel H. Fazio. 2008. 'Interracial Roommate Relationships: An Experimental Field Test of the Contact Hypothesis.' *Psychological Science* 19(7):717–723.

SNES. 2015. 'The Swedish National Election Studies Program.' Department of Political Science, University of Gothenburg.

Solon, Gary, Steven J, Haider, and Jeffrey M. Wooldridge. 2015. 'What Are We Weighting For?' *Journal of Human Resources* 50(2):301–316.

Steinmayer, Andreas. 2016. 'Exposure to Refugees and Voting for the Far-Right: (Unexpected) Results from Austria.' IZA Discussion Paper No. 9790, Institute for Labor Economics, Bonn.

Stichnoth, Holger and Karine Van der Straeten. 2013. 'Ethnic Diversity, Public Spending, and Individual Support for the Welfare State: A Review of the Empirical Literature.' *Journal of Economic Surveys* 27(2):364–389.

Swedish Migration Agency. 2018. 'Kommunmottagna tidigare år.' Stockholm. https://www.migrationsverket.se/Om-Migrationsverket/Statistik/Oversikter-och-statistik-fran-tidigare-ar/Kommunmottagna--tidigare-ar.html.

Tajfel, Henri. 1978. *Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Relations*. London: Academic Press.

van der Waal, Jeroen, Peter Achterberg, Dick Houtman, Willem de Koster, and Katerina Manevska. 2010. ''Some Are More Equal than Others': Economic Egalitarianism and Welfare Chauvinism in the Netherlands.' *Journal of European Social Policy* 20(4):350–363.

Van Laar, Colette, Shana Levin, Stacey Sinclair, and Jim Sidanius. 2005. 'The Effect of University Roommate Contact on Ethnic Attitudes and Behavior.' *Journal of Experimental Social Psychology* 41:329–345.

van Oorschot, Wim. 2006. 'Making the Difference in Social Europe: Deservingness Perceptions among Citizens of European Welfare States.' *Journal of European Social Policy* 16:23–42.

Wanner, Philippe. 2002. 'Migration Trends in Europe.' European Population Papers Series No. 7, Council of Europe, Strasbourg.

Williams Jr, Robin M. 1947. *The Reduction of Intergroup Tensions: A Survey of Research on Problems of Ethnic, Racial, and Religious Group Relations*. New York: Social Science Research Council.

Voss, Stephen. 1996. 'Beyond Racial Threat: Failure of an Old Hypothesis in the New South.' *Journal of Politics* 58(4):1156–1170.

Voss, Stephen and Penny Miller. 2001. 'Following a False Trail: The Hunt for White Backlash in Kentucky's 1996 Desegregation Vote.' *State Politics & Policy Quarterly* 1(1):62–80.

Zingher, Joshua N. and M. Steen Thomas. 2014. 'The Spatial and Demographic Determinants of Racial Threat.' *Social Science Quarterly* 95(4):1137–1154.

Appendix

Table 4.A.1: Refugee inflows and native mobility

	The probability of moving		
	No controls	Municipal controls	All controls
	(1)	(2)	(3)
%Refugees (destination)$_t$	-0.050***	-0.046**	-0.040*
	(0.018)	(0.021)	(0.021)
%Refugees (origin)$_{t-1}$	0.073**	0.075**	0.072**
	(0.037)	(0.036)	(0.035)
	(4)	(5)	(6)
%Contracted refugees (destination)$_t$	-0.057**	-0.050**	-0.039*
	(0.023)	(0.023)	(0.022)
%Contracted refugees (origin)$_{t-1}$	0.099**	0.108**	0.104**
	(0.046)	(0.044)	(0.043)
The refugee/contracted refugee exposure among movers relative to non-movers			
		Δ%Refugees (exposure)	
	(7)	(8)	(9)
Mover	-0.180***	-0.157**	-0.146**
	(0.060)	(0.073)	(0.072)
		Δ%Contracted refugee (exposure)	
	(10)	(11)	(12)
Mover	-0.190***	-0.167**	-0.136**
	(0.054)	(0.066)	(0.064)
n	4,777	4,777	4,777
Municipalities	284	284	284

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the municipality level in parentheses. All regressions include period-fixed effects and weight respondents by the inverse of the probability of being selected for the survey in the relevant election. The municipal controls in the upper/lower panel are the same as in Column 1/2 in Table 4.A.2. All controls include the municipal controls as well as the individual controls outlined in Section 4.4.3.

Table 4.A.2: Predictors of changes in contracted refugee shares (% of municipal population)

	(1)	(2)	(3)
Welfare spending per capita/1000$_{t-1}$	1.096***	0.333	0.749
	(0.316)	(0.215)	(0.958)
%Vacant public housing$_{t-1}$	0.043**	0.032*	0.015
	(0.021)	(0.018)	(0.023)
%Unemployment$_{t-1}$	-0.052**	-0.035*	-0.097
	(0.024)	(0.020)	(0.116)
Population/1000$_{t-1}$	-0.001*	-0.000	-0.043*
	(0.006)	(0.000)	(0.023)
%Foreign citizens$_{t-1}$	-0.007	-0.005	0.077
	(0.009)	(0.008)	(0.064)
%Social Democratic Party seats	0.007	0.006	-0.002
	(0.006)	(0.004)	(0.012)
%Left Party seats	0.014	0.012*	0.030
	(0.009)	(0.007)	(0.027)
%New Democracy Party seats	-0.045***	-0.031**	-0.036**
	(0.016)	(0.012)	(0.015)
%Moderate Party seats	-0.006	-0.002	0.003
	(0.007)	(0.005)	(0.019)
%Liberal Party seats	0.008	0.002	0.001
	(0.009)	(0.007)	(0.017)
%Center Party seats	0.009	0.006	0.008
	(0.007)	(0.005)	(0.013)
%Christian Democratic Party seats	0.029***	0.021***	0.012
	(0.009)	(0.006)	(0.026)
%Contracted refugees$_{t-1}$		0.534***	-0.025
		(0.098)	(0.133)
Adjusted R^2	0.364	0.435	0.482
Joint F test of all variables (p-value)	<0.001	<0.001	0.120
Municipal-fixed effects	No	No	Yes
n	852	852	852
Municipalities	284	284	284

Note: Standard errors clustered at the municipality level in parentheses. All models include period-fixed effects.

Table 4.A.3: Using the placebo instrument and analysing levels instead of ratios

Placebo instrument			
Panel 1: Municipal-level controls and individual-level noise controls			
Election type	National	Municipal	County
	(1)	(2)	(3)
Second stage			
Δ%Refugees	-0.121	-0.138	-0.118
	(0.141)	(0.158)	(0.148)
First stage			
Δ%White noise	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Hausman test	0.053	0.042	0.069
F statistic	2.05	2.05	2.07
Panel 2: Add initial turnout			
	(4)	(5)	(6)
Second stage			
Δ%Refugees	-0.085	-0.099	-0.080
	(0.125)	(0.138)	(0.130)
First stage			
Δ%White noise	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Hausman test	0.085	0.065	0.128
F statistic	2.05	2.05	2.07
n	4,777	4,777	4,366
Municipalities	284	284	281
Analysing levels instead of ratios			
Panel 1: Municipal-level controls and individual-level noise controls			
Election type	National	Municipal	County
	(7)	(8)	(9)
Second stage			
ΔRefugees/1000	0.096**	0.080*	0.115**
	(0.044)	(0.045)	(0.051)
First stage			
ΔContracted refugees/1000	0.711***	0.711***	0.731***
	(0.137)	(0.137)	(0.112)
Hausman test	0.059	0.133	0.037
F statistic	26.88	26.88	42.37
Panel 2: Add initial turnout			
	(10)	(11)	(12)
Second stage			
ΔRefugees/1000	0.104***	0.091**	0.123***
	(0.039)	(0.038)	(0.045)
First stage			
ΔContracted refugees/1000	0.711***	0.711***	0.731***
	(0.137)	(0.137)	(0.112)
Hausman test	0.014	0.035	0.008
F statistic	26.87	26.88	42.31
n	4,777	4,777	4,366
Municipalities	284	284	281

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the municipality level in parentheses. All regressions include period-fixed effects and weight respondents by the inverse of the probability of being selected for the survey in the relevant election.

Table 4.A.4: Excluding all municipal-level controls apart from local political makeup

Panel 1: Local political makeup and individual-level noise controls			
Election type	National	Municipal	County
	(1)	(2)	(3)
<u>Second stage</u>			
Δ%Refugees	0.046**	0.041**	0.045**
	(0.019)	(0.019)	(0.021)
Hausman test	0.145	0.296	0.085
F statistic	179.27	179.26	175.58
Panel 2: Add initial turnout			
	(4)	(5)	(6)
<u>Second stage</u>			
Δ%Refugees	0.045***	0.042***	0.044**
	(0.016)	(0.016)	(0.017)
Hausman test	0.018	0.048	0.012
F statistic	179.79	179.60	176.08
n	4,777	4,777	4,366
Municipalities	284	284	281

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the municipality level in parentheses. All regressions include period-fixed effects and weight respondents by the inverse of the probability of being selected for the survey in the relevant election.

Table 4.A.5: Excluding all municipal-level controls

Panel 1: Individual-level noise controls			
Election type	National	Municipal	County
	(1)	(2)	(3)
<u>Second stage</u>			
Δ%Refugees	0.036**	0.032**	0.035**
	(0.016)	(0.015)	(0.017)
Hausman test	0.239	0.411	0.146
F statistic	185.60	185.63	179.38
Panel 2: Add initial turnout			
	(4)	(5)	(6)
<u>Second stage</u>			
Δ%Refugees	0.035***	0.033**	0.034**
	(0.013)	(0.014)	(0.014)
Hausman test	0.047	0.095	0.029
F statistic	186.10	185.90	179.79
n	4,777	4,777	4,366
Municipalities	284	284	281

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the municipality level in parentheses. All regressions include period-fixed effects and weight respondents by the inverse of the probability of being selected for the survey in the relevant election.

Table 4.A.6: Excluding all individual-level noise controls apart from year of birth

Panel 1: Municipal-level controls and respondents' year of birth			
Election type	National	Municipal	County
	(1)	(2)	(3)
Second stage			
Δ%Refugees	0.050**	0.041*	0.055**
	(0.022)	(0.022)	(0.027)
Hausman test	0.149	0.379	0.079
F statistic	233.26	233.23	226.30
Panel 2: Add initial turnout			
	(4)	(5)	(6)
Second stage			
Δ%Refugees	0.061***	0.054**	0.064***
	(0.022)	(0.021)	(0.024)
Hausman test	0.020	0.055	0.011
F statistic	232.93	232.72	226.13
n	4,777	4,777	4,366
Municipalities	284	284	281

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the municipality level in parentheses. All regressions include period-fixed effects and weight respondents by the inverse of the probability of being selected for the survey in the relevant election.

Table 4.A.7: Excluding municipalities with more than 50,000 inhabitants

Panel 1: Municipal-level controls and individual-level noise controls			
Election type	National	Municipal	County
	(1)	(2)	(3)
Second stage			
Δ%Refugees	0.055**	0.049**	0.058**
	(0.023)	(0.022)	(0.027)
Hausman test	0.095	0.264	0.082
F statistic	217.47	217.43	215.02
Panel 2: Add initial turnout			
	(4)	(5)	(6)
Second stage			
Δ%Refugees	0.066***	0.061***	0.065***
	(0.022)	(0.020)	(0.024)
Hausman test	0.016	0.046	0.020
F statistic	217.14	217.02	214.83
n	2,389	2,389	2,382
Municipalities	244	244	244

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the municipality level in parentheses. All regressions include period-fixed effects and weight respondents by the inverse of the probability of being selected for the survey in the relevant election.

Table 4.A.8: Intention-to-treat estimates

Panel 1: Municipal-level controls and individual-level noise controls			
Election type	National	Municipal	County
	(1)	(2)	(3)
<u>Second stage</u>			
Δ%Refugees	0.054***	0.045**	0.058**
	(0.020)	(0.020)	(0.024)
Hausman test	0.332	0.768	0.264
F statistic	143.82	143.80	143.69
Panel 2: Add initial turnout			
	(4)	(5)	(6)
<u>Second stage</u>			
Δ%Refugees	0.060***	0.053***	0.062***
	(0.019)	(0.018)	(0.021)
Hausman test	0.093	0.230	0.086
F statistic	143.96	143.88	143.89
n	4,777	4,777	4,366
Municipalities	284	284	281

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the municipality level in parentheses. All regressions include period-fixed effects and weight respondents by the inverse of the probability of being selected for the survey in the relevant election.

Table 4.A.9: Reduced-form estimates

Panel 1: Municipal-level controls and individual-level noise controls			
Election type	National	Municipal	County
	(1)	(2)	(3)
Δ%Contracted refugees	0.036**	0.030**	0.039**
	(0.015)	(0.014)	(0.017)
Panel 2: Add initial turnout			
	(4)	(5)	(6)
Δ%Contracted refugees	0.040***	0.036***	0.041***
	(0.014)	(0.013)	(0.015)
n	4,777	4,777	4,366
Municipalities	284	284	281

Note: Significance levels: *p<0.10; **p<0.05; ***p<0.01. Standard errors clustered at the municipality level in parentheses. All regressions include period-fixed effects and weight respondents by the inverse of the probability of being selected for the survey in the relevant election.

5. Lifelong Learning and Employment Outcomes: Evidence from Sweden*

Abstract

We study the impact of adult education and training (AET) on employment outcomes in Sweden. Exploiting unusually rich data from the Programme for the International Assessment of Adult Competencies and using an inverse-probability weighted regression-adjustment estimator to deal with selection bias, we find that AET raises the probability of doing paid work by 4 percentage points on average. This impact is entirely driven by non-formal, job-related AET, such as workshops and on-the-job training. We also find that the effect – which increases with training intensity – is very similar across different types of non-formal, job-related AET. Specification and robustness tests indicate the estimates are causal. Our results suggest that policies stimulating relevant AET take-up have promise as a way to secure higher employment rates in the future.

* The author thanks Henrik Jordahl, Julian Le Grand, Olmo Silva, Anders Stenberg, and two anonymous referees for comments and discussions.

5.1. Introduction

In the past decades, knowledge requirements on developed countries' labour markets have changed radically, as technological innovation has displaced many low-skilled jobs and increased the required skills and competencies in jobs that continue to exist (e.g. ILO 2011). Sweden, an advanced export-oriented economy with a munificent welfare state, serves as an important case study in this respect. Between the 1970s and the early 2000s, the share of low-educated workers in low-skilled jobs in Sweden decreased from 48 per cent to 11 per cent, while the share of high-educated workers in high-skilled jobs increased from 28 per cent to 58 per cent (Tåhlin 2007). In 2016, only 4.8 per cent of Swedes were employed in jobs with no or low education requirements, the lowest figure in the European Union (Schermer 2017). In addition, a comparatively large share of Swedish employees report that their jobs have changed because of structural and technological transformation (OECD 2013a). Forecasts suggest these trends will continue and accelerate in the coming decades (Cefedop 2015). Due to the structural changes on the labour market, it is not surprising that Sweden has the lowest share of over-skilled workers, and one of the largest shares of under-skilled workers, among all OECD countries for which there are comparable data (OECD 2013a, 2016).

Overall, these developments indicate the importance of knowledge and skills for ensuring high employment and productivity levels in the future. While the school system is likely to play a key role in this endeavour, adult education and training (AET) is also likely to be important. Indeed, people's knowledge and skills are not only developed in the school system but also at work and through learning later in life, which in turn may help people to maintain and continuously update their skills (Statistics Sweden 2013). And on the knowledge-intensive labour markets in the developed world, it is likely to become increasingly important for individuals to update their competencies during their working lives.

There are thus reasons to believe that lifelong learning could play an important role for ensuring well-functioning labour markets in the future. Indeed, while existing research display mixed findings, some studies do suggest that AET can have positive effects on individuals' labour-market outcomes. Yet no research investigates both formal and non-formal AET simultaneously or separates the effects of job-related and non job-related AET. This is important in the Swedish context: in 2010–12, fully 65 per cent of the adult population underwent some form of adult education and training, with

52 per cent participating in job-related AET and 13 per cent participating in non job-related AET. In addition, the most popular form of AET was non-formal, consisting of, for example, on-the-job training, courses, and workshops, which do not necessarily lead to official qualifications: 60 per cent of the adult population underwent non-formal AET and 14 per cent underwent formal AET.[1] Whether or not the former type is related to labour-market outcomes in the Swedish context has not yet been investigated. And, as noted, no research has thus far sought to disentangle the employment effects of different types of AET at a more general level.

In this paper, we analyse how lifelong learning affects individuals' employment prospects in Sweden. Using data from the 2012 round of the Programme for the International Assessment of Adult Competencies (PIAAC) allows us to adjust for an unusually large number of important observable characteristics, including cognitive skills, formal education levels, and work history. In combination with our exploiting an inverse-probability weighted regression-adjustment estimator, this increases the probability that our estimates reflect a causal impact. Our tests show that the weighting procedure balances the rich covariates across the treatment and control groups, suggesting our main model is well specified. We also show essentially identical results when we exclude all covariates apart from age, gender, and work history, suggesting differences in the other covariates – including social background, cognitive skills, and formal educational levels – that affect both the probability of engaging in AET and the probability of doing paid work are ironed out using this sparse specification. Although we cannot conclusively rule out omitted-variable bias, there is little in our analysis suggesting it is a serious problem for our findings. Still, some caution regarding the results is naturally warranted and future research should further investigate the effects of AET using quasi-experimental variation to address potential selection on unobservable characteristics.

[1] These calculations, which do not include students aged 16–24 who are in their first formal cycle of studies, are carried out using micro-level data from the OECD's (2017) PIAAC database, which we utilise in this paper. The data are derived from questions enquiring whether respondents studied for (1) any formal qualification at primary, secondary, university, or post-secondary level, and (2) whether they participated in non-formal education through 'Courses conducted through open or distance education, 'Organised sessions for on-the-job training or training by supervisors or co-workers', 'Seminars and workshops', and/or 'Other courses or private lessons'. Job-related AET is defined as training that individuals report having undergone for the purposes of improving their employment chances in general. Statistics Sweden (2014) reports very similar figures using other data.

Our findings show that individuals who participated in AET in the year prior to the survey are about 4 percentage points more likely to do paid work in the week before the survey took place than comparable individuals who did not undergo AET. Intriguingly, the findings are entirely driven by job-related AET, which raises the probability of working by 6 percentage points – and this impact is in turn entirely driven by non-formal, job-related AET, which raises the probability of working by about 8 percentage points. If anything, the positive impact is even larger when analysing the probability of working full time instead of the probability of doing any paid work at all. We further find that the effect is very similar across different types of non-formal, job-related AET, which also suggests that potential sources of omitted-variable bias that are specific to certain types of such AET do not drive the results. Interestingly, the effect of non-formal, job-related AET varies depending on training intensity, since the impact increases with the number of such AET activities in which individuals have participated, further indicating that omitted-variable bias does not drive the findings. Yet non-formal, non job-related AET and formal AET of either type have no effects on employment outcomes. Several robustness tests support these conclusions.

Of course, it is important to note the context of our findings: we analyse data collected in Sweden during the country's economic recovery following the 2008 financial crisis. Since the effects of AET on employment outcomes may well be country specific and interact with the business cycle, future research should investigate the generalisability of the results in these respects. Nevertheless, the findings indicate that policies stimulating higher non-formal, job-related AET take-up have promise as a way to increase employment in similar settings – and reforms to increase access to lifelong learning should be considered. However, more research into what works in this respect is necessary.

The paper proceeds as follows. Section 5.2 discusses the theoretical mechanisms linking AET and employment outcomes and reviews the empirical literature; Section 5.3 outlines the data analysed; Section 5.4 discusses the methodology employed; Section 5.5 presents the results; and Section 5.6 concludes.

5.2. Theory and prior literature

According to economic theory, individuals and companies invest in education and training in order to improve human capital and in this way raise their earnings and productivity respectively (Becker 1964). Empirical research has established that education has positive effects on labour-market outcomes, and there is much to suggest that an important part of this impact operates via higher human capital (e.g. Bhuller et al. 2017; Brunello et al. 2016). Similarly, research finds that average knowledge levels across countries' populations are strongly related to economic growth (Hanushek and Woessmann 2015). Since education does not only create value for the individuals who participate in it, there are compelling reasons for the government to finance and stimulate investments in knowledge and skill development (see McMahon 2010).

There are also reasons to believe that different forms of AET may be especially important for maintaining and developing individuals' human capital over the course of their working lives, once they have completed their formal education at school and university. AET makes the labour market and competence provision more effective, since it 'facilitates career shifts if there are changes to demand in the labour market or in individuals' health' (Stenberg 2016, pp. 20–21). Adult education does not have to be formal to improve individuals' human capital – also non-formal education is important. As Statistics Sweden (2013, p. 41) writes:

> The formal education system is not the only setting in which literacy, numeracy, and digital problem-solving skills are developed. Learning also occurs in several other settings, such as within the family, in the workplace, or during recreational activities and self-studies. Adults who do not utilise their skills sufficiently at work or during leisure time risk losing their competencies and abilities. The longer since a person completed his or her studies, the weaker is the direct relationship between the level of formal education and the person's skills. For older people, other factors than formal education levels matter greatly for skill development, such as the type of vocation, opportunities for learning in the workplace, and the social environment.

In other words, it is plausible that AET may have positive effects on human capital of relevance for individuals' labour-market outcomes. Indeed, some empirical research does suggest that formal adult education has positive effects on salaries and employment (see Stenberg 2016). However, the results are still mixed. For example, British research shows that formal education at lower- and upper-secondary levels in adulthood only generates higher incomes among men with low initial education. Still, specifically vocational education does seem to increase the probability of employment more generally (Jenkins et al. 2003). Recent Swedish studies using a similar methodology as this paper tend to find positive effects on labour-market outcomes of formal adult education at the primary and secondary levels as well as employment-training programmes, at least in a slightly longer-term perspective (Bergemann and van den Berg 2014; Stenberg 2016; Stenberg and Westerlund 2015). The evidence on non-formal learning, such as personnel training, is also somewhat mixed, but again sometimes shows positive effects (e.g. Albert et al. 2010; Bassanini et al. 2005; Blundell et al. 1999; Dearden et al. 2006; Haelermans and Borghans 2012; Georgiadis and Pitelis 2016; Konings and Vanormelingen 2015; Schwerdt et al. 2012; Vignoles et al. 2004).

However, to the best of our knowledge, no research has analysed both formal and non-formal AET simultaneously – or investigated the effects of non-formal AET in Sweden. Similarly, no studies separate the effects of job-related from non job-related AET in the same analysis. These gaps are important since the impact of different types of AET may differ. For example, AET that focuses on specifically job-related skills, which individuals undergo for job-related reasons, may have a more positive impact on employment outcomes compared with other types of AET. Also, formal education may better prepare individuals for the labour market than non-formal education, if the former provides more structured and rigorous training than the latter. On the other hand, non-formal education may be more practically oriented and thus generate skills that are more relevant on the labour market. To explore such potential heterogeneity, it is important to analyse the effects of different types of AET. More research is also necessary at a more general level, since most existing studies are quite old, which means that they are unlikely to reflect the effects of AET on today's labour markets, and often use regression methods that ignore potential selection bias.

5.3. Data

To study the impact of AET on employment outcomes in Sweden, we exploit micro-level data from PIAAC. PIAAC surveys the adult population's literacy, numeracy, and problem solving in technology-rich environments. It also collects rich information on respondents' backgrounds and how they utilise their skills. In the first round, which was carried out in 2012, there were 166,000 participants aged 16–65 from 24 countries. In the 2016 round, 14 additional countries participated.[2] Table 5.A.1 outlines the descriptive statistics of the data analysed.

The Swedish sample in PIAAC 2012 was composed of 10,000 individuals, who were randomly drawn from the adult population aged 16–65. Data were collected between August 2011 and May 2012. The non-response rate was 55 per cent, which means that the sample in the end was composed of 4,468 individuals.[3] Using the sample weights provided, it is nevertheless possible to ensure that this sample is representative of the targeted population. Indeed, despite the non-response rate, the OECD (2013b) found the Swedish results to be reliable. However, we exclude all students aged 16–24 who are in their first formal cycle of studies, since they do not form part of the AET population. This decreases the available sample to 3,888 individuals.[4]

5.3.1. Adult education and training

To analyse the effects of lifelong learning, we exploit a variable in the PIAAC database, which indicates whether respondents participated in some form of adult education and training in the 12 months before the survey was carried out. This variable is derived from questions enquiring whether respondents studied for (1) any formal qualification at primary, secondary, university, or post-secondary level, and (2) whether they participated in non-formal education through 'Courses conducted through open or distance education, 'Organised sessions for on-the-job training or training by supervisors or co-workers', 'Seminars and workshops', and/or 'Other courses or private lessons'. Individuals who did participate in any form of AET are given a value of 1, while those who did not are given a value of 0. This variable does not include education undergone by students aged 16–24 who are in their first formal cycle of studies, which means that it only picks up different types of AET. This is useful since the average effect of different types of AET has not previously been evaluated in Sweden.

Yet it is also important to investigate heterogeneous effects depending on the type of AET pursued. For example, it is plausible that job-related AET affects employment

[2] In contrast to many international assessments at the school level, PIAAC is not carried out continuously in the same countries. To date, each country has only participated once, with the exception of the United States (which participated in both the 2012 and 2016 rounds).
[3] Technically, the final sample was composed of 4,469 respondents, but due to one observation with no values in the database, the number available for analysis is 4,468.
[4] However, our preferred inverse-probability weighted regression-adjustment estimator, discussed in Section 5.4, automatically excludes a few individuals who have/have not undergone AET, but whose values on the covariates do not overlap with any of the individuals who have not/have undergone AET.

outcomes differently compared with non job-related AET. In PIAAC, respondents were asked whether the AET in which they participated was 'job related', which is defined as AET that individuals report having undergone for the purposes of improving their employment chances in general. The OECD has constructed two separate indicators for job-related and non job-related AET respectively, which we utilise in our analysis.[5] Similarly, the effects of AET may depend on whether it is delivered through the formal education system or provided informally outside that system. We thus separate job-related and non job-related AET into their formal and non-formal components, again based on the above statements, creating four different AET categories in total.[6] We also separate each non-formal AET category that is related to employment into its four components.[7] Finally, to study whether any detected effects vary by training intensity, we create a variable indicating the number of AET activities in which individuals participated, for each AET type that we find to be related to employment.

5.3.2. Employment outcomes

The respondents were asked whether they did any paid work in the week before the survey was carried out. Those who reported doing any paid work for at least an hour are given the value of 1, while those who did not are given a value of 0. We utilise this indicator as our main dependent variable. However, in one robustness test, we also use an indicator for whether respondents worked full time at the time of the survey. Those who report that they are full-time workers are given a value of 1 and those who do not – including part-time workers – are given a value of 0. This allows us to investigate

---

[5] It is not possible for individuals to simultaneously report that they underwent both job-related and non job-related, formal AET, or job-related and non job-related, non-formal AET, but it is possible for them to report that they underwent different combinations of formal and non-formal AET. The OECD's assignment of individuals to the job-related and non job-related categories is based on, firstly, the type of formal AET they underwent, and, secondly, the type of non-formal AET they underwent. For example, individuals who underwent non-job related, formal AET as well as job-related, non-formal AET are assigned to the non-job related AET category. In an unreported robustness test, we instead created a separate indicator for individuals who underwent both job-related, formal AET and non job-related, non-formal AET, or non job-related, formal AET and job-related, non-formal AET (about 2 per cent of the sample), but found little support for interaction effects in this respect.

[6] We follow the OECD's method of assigning individuals to the different categories by, firstly, the type of formal AET they underwent, and, secondly, the type of non-formal AET they underwent (see the previous footnote). In an unreported robustness test, we created a separate indicator for respondents who underwent both formal and non-formal AET (about 9 per cent of the sample), but found little support for interaction effects in this respect.

[7] In this analysis, to be able to distinguish differential effects, we exclude individuals who participated in more than one non-formal AET component. In an unreported robustness test, we further excluded individuals who also underwent some form of formal AET, but the results were essentially identical.

whether the effects of AET on the probability of working full-time differ from the effects on the probability of doing any work at all.

5.3.3. Covariates

In the analysis, we adjust for a number of relevant covariates. These include indicators for the respondents' background: age, gender, first- and second-generation immigration status, whether Swedish is their native language, years spent in Sweden, the number of books at home (6 intervals), maternal educational level (3 levels), paternal educational level (3 levels), and the number of people in the household (capped at 6 people). Also, we control for the participants' formal educational level (in years of schooling) and their literacy and numeracy scores in PIAAC. This is important since Swedes with higher levels of education and cognitive skills tend to be more likely to pursue AET than people with lower levels of education and cognitive skills (see Bussi and Pareliussen 2015).

Furthermore, in the preferred model, we include an indicator for whether or not respondents carried out paid work in the 12-month period prior to the survey. Since all respondents who worked in the week before the survey by definition also did so at some point in the year preceding it, our comparison effectively takes into account whether or not currently non-working respondents did paid work at some point in the previous year. We also include an indicator for whether or not respondents have done any paid work in their lives at all. In robustness tests, we further include industry dummies as well as indicators for the geographical region in which respondents live.[8]

Given that we are able to adjust for unusually rich data, we believe we are able to control for most factors that explain both AET and employment. Still, the functional form of the relationship between the covariates and employment is far from clear, making it important to create treatment and control groups that have comparable values on the observable characteristics that we seek to hold constant. The next section discusses the method we utilise to be able to do so.

5.4. Method

For the purposes of studying the total impact of all types of AET on employment outcomes, consider the following OLS model:

[8] Following previous research, we replace missing values for the covariates with the sample means and include separate indicators for missing values in the regressions (see Hanushek and Woessmann 2011).

$$e_i = \alpha + \beta_1 aet_i + \beta_2 x_i + \varepsilon_i \tag{1}$$

where e_i is the indicator for paid work in the week before the survey; aet_i denotes the AET dummy; x_i is a vector of observable covariates; and ε_i is the error term. The model's assumption is that $Cov(aet_i, \varepsilon_i|x_i) = 0$ so that the average treatment effect is given as $E[e_i|x_i, aet_i = 1] - E[e_i|x_i, aet_i = 0] = E[e_{1i} - e_{0i}|x_i]$. However, if x_i does not include all variables that impact both e_i and aet_i or if e_i affects aet_i directly, it would mean that $Cov(aet_i,\ \varepsilon_i|x_i) \neq 0$ and the results will suffer from endogeneity bias (Angrist and Pischke 2009). That is, AET may be affected by paid work, generating reverse causality, and/or omitted variables may affect both the probability to participate in AET as well as the likelihood of working.

In order to estimate causal effects, it is necessary to either obtain quasi-experimental variation in the take-up of AET or have access to enough observable characteristics to make the conditional independence assumption plausible. In our data, it does not seem feasible to obtain quasi-experimental variation, especially since we seek to investigate potential heterogeneous effects across different types of AET. However, we do believe that our dataset is rich enough to make the conditional independence assumption plausible. Still, research indicates that parametric models with rich controls may in fact increase bias, if the functional form of the relationship between the controls and the outcome is not adequately captured. This is especially true when the mean and variances in the controls differ significantly between the treatment and control groups (see Heckman et al. 1998; Rubin 2001, 2008; Rubin and Thomas 2000). In other words, it is likely important to balance the distribution of the control variables included in the model across the treatment and control groups, so as not to make assumptions of the functional form of the relationship between the controls and the outcome.

We propose to do so using an inverse-probability weighted regression-adjustment estimator, or IPWRA (see Wooldridge 2010). This estimator has recently been exploited for similar purposes, such as in analyses of the effects of vocational education and personality traits on labour-market outcomes (see Brunello and Rocco 2017; Mendolia and Walker 2015).[9] The fundamental problem in all econometric evaluations is that one cannot observe any given individual both treated and not treated at the same time. In our case, since we in some models seek to analyse different types of AET separately,

[9] Our outline of the estimator draws on Brunello and Rocco's (2017, pp. 338–342) discussion.

there are several different potential employment outcomes for each individual, yet we only know the outcome that followed the type of AET, if any, the individual actually pursued. The inverse-probability weighted regression-adjustment estimator allows us to solve this problem by predicting for each individual all potential outcomes, using information from individuals with similar observable characteristics who did not undergo any AET (or an alternative type of AET than the individual in question). This means that we can study all average treatment effects of the different forms of AET separately in the same model.

In other words, the conditional independence assumption for all possible AET types a gives us $E[e_i(a)|x] = E[e_i(a)|a_i(a) = 1, x]$, where $a_i(a) = 1$ denotes individuals undergoing AET type a. Once we adjust for x, the average treatment effect of a compared with the benchmark a' – which denotes either individuals who have not undergone any AET or those who have undergone other types of AET than a – is obtained by effectively comparing the conditional employment rate among individuals who underwent a with the conditional employment rate among individuals with benchmark a':

$$E[e_i(a)|a_i(a) = 1, x] - E[e_i(a')|a_i(a') = 1, x] = E[e_i(a) - e_i(a')|x] \qquad (2)$$

The inverse-probability weighted regression-adjustment estimator is thus composed of two steps. The first step estimates the probability of undergoing AET a from observable characteristics, using a logit model or, when analysing the effect of different types of AET, a multinomial logit model. The second step then estimates the effect of AET a on the employment outcomes using a linear model, with the inverse probability of undergoing AET a as weight, while also including all covariates.[10] In practice, the estimator thus compares individuals' actual employment outcomes after undergoing AET a with the counterfactual outcomes the same individuals should have obtained under benchmark a'.[11] To ensure that the weighting procedure balances the covariates across the main treatment and control groups, which is important for drawing valid

[10] Since the model is dependent on respondents having similar values on the covariates, it is only possible to study AET effects among people in the treatment and control groups for which the covariates overlap. That is, the cell where $x = X$, for all possible X, must include respondents undergoing both a and a'.

[11] While the assumptions are essentially the same as in propensity score matching, the latter does not allow analyses of multiple treatments simultaneously. In non-reported robustness tests, we instead used regular nearest matching to study the average effect of AET. The results were essentially identical to those reported in Column 6 in Table 5.1.

inferences (see Rubin 2008), we use Imai and Ratkovic's (2014) over-identification test for covariate balance and also present the raw and weighted differences as well as variance ratios for the two groups.[12]

Unlike simple regression-adjustment estimators, which only model the outcome directly based on covariates, and inverse-probability weighting estimators, which only model the treatment based on these covariates, the inverse-probability weighting regression-adjustment estimator involves both steps. Thus, it has the 'double-robust' property: the estimates will still be consistent if one of the equations is – but not if both equations are – incorrectly specified (Wooldridge 2010). To ensure that our estimates are relevant for the population from which the sample is drawn, we include the sample weights provided in the PIAAC database in the first step of the model.[13]

Using the methodology outlined above, our identification assumption is thus that assignment to different types of AET is as good as random conditional on the covariates outlined in Section 5.3.3, including individuals' work history. While we cannot conclusively rule out the possibility of selection bias due to unobservable characteristics, we test whether our assumption is likely to hold by carrying out a number of robustness tests. These include adding indicators for industry and geographical region to the equation, and excluding all covariates apart from age, gender, and the dummy indicating whether or not respondents carried out any paid work in the previous year. If the results are very similar despite adding potentially important predictors of employment outcomes, and when excluding predictors that are normally crucial for such outcomes – including socio-economic background, cognitive skills, and years of schooling – we believe it is reasonable to assume that the conditional independence assumption holds. We also analyse whether the effects of any AET type that we find to be related to employment outcomes vary depending on training intensity. If the relationship is 'dose dependent', it indicates that general selection into

[12] The over-identification test can only be carried out in analyses with one treatment group, and we thus only present these statistics for the main model analysing the average effect of all types of AET, as well as for models separately analysing the types of training that have effects on employment outcomes.

[13] While literacy and numeracy scores in PIAAC are estimated from ten 'plausible values' derived from multiple imputations, and replicate weights are used to adjust for sampling error (see OECD 2013b), we use the average of all plausible values for each subject in our models and regular robust standard errors. This is in order to estimate both the inverse-probability weighted regression-adjustment estimator and the covariate balance test correctly. However, the OLS results are identical if we estimate the regressions for each plausible value separately and use replicate weights to adjust the standard errors, suggesting these adjustments matters little – which is further supported by prior research analysing similar survey structures (see Jerrim et al. 2017).

the AET type in question is unlikely to bias estimates. Still, some caution regarding the results is naturally warranted and future research should further investigate the effects of AET using quasi-experimental variation to address potential selection on unobservable characteristics.

5.5. Results

As a starting point, Columns 1–3 in Table 5.1 show the results from a regular OLS model when including different combinations of control variables. The estimates indicate that AET is positively associated with paid work, regardless of which controls we include. In the full model in Column 3, undergoing some form of AET is associated with a 5 percentage point increased likelihood of doing paid work. Thus, our initial results indicate that there is a positive correlation between AET, as measured in PIAAC, and employment outcomes in Sweden.

Table 5.1: AET and the probability of doing paid work

	OLS	OLS	OLS	IPWRA	IPWRA	IPWRA
	(1)	(2)	(3)	(4)	(5)	(6)
AET	0.18***	0.14***	0.05***	0.17***	0.13***	0.04***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)
Background variables	Yes	Yes	Yes	Yes	Yes	Yes
Education and PIAAC scores	No	Yes	Yes	No	Yes	Yes
Paid work in the previous year	No	No	Yes	No	No	Yes
Never done paid work	No	No	Yes	No	No	Yes
R^2	0.08	0.12	0.48	N/A	N/A	N/A
n			3,884			

Note: Significance levels: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in parentheses. We always restrict the sample to observations that do not violate the overlap assumption in Column 6.

Turning to the estimates from the inverse-probability weighted regression adjustment model in Columns 4–6, we note that the results are in fact very similar compared to the OLS results: in the full model in Column 6, we find that undergoing some form of AET increases the probability of doing paid work by 4 percentage points. Since this model is most likely to pick up the causal impact of AET, we take this to be our main estimate. Nevertheless, the headline finding is in fact similar regardless of whether we assume that the relationships between the covariates and the probability to do paid work are linear, using the OLS model, or do not make any assumptions of the functional form of these relationships, using the IPWRA model.

However, as noted in Section 5.4, it is important to analyse whether the weighting estimator balances the covariates across the treatment and control groups. Table 5.2

presents results from the balance tests following the full model in Column 6 in Table 5.1. There is no doubt that the raw differences between treatment and control groups are significant in many cases, highlighting the importance of ensuring that our weighting procedure balances the groups appropriately. Indeed, the estimates indicate that our method worked as intended: the weighted differences in means and variances are small, and the over-identification test for covariate balance displays a value of 0.94, suggesting we comfortably fail to reject the null hypothesis of balanced covariates. Overall, the results thus support our drawing causal inferences from our preferred specification.

Table 5.2: Balance tests on covariates

	Standardised differences		Variance ratio	
	Raw	Weighted	Raw	Weighted
PIAAC numeracy score	0.50	-0.02	0.74	1.03
PIAAC literacy score	0.58	-0.02	0.75	1.04
First-generation immigrant	-0.13	-0.02	0.79	0.96
Missing dummy for immigrant background	0.05	-0.01	1.18	0.98
Second-generation immigrant	-0.02	0.02	0.90	1.13
Years in Sweden	-0.22	0.03	0.77	1.02
Mother's education	0.38	-0.04	1.34	0.99
Missing dummy for mother's education	-0.10	0.00	0.52	1.01
Father's education	0.34	-0.01	1.34	1.04
Missing dummy for father's education	-0.16	-0.01	0.48	0.95
Books at home	0.46	-0.02	0.90	0.91
Missing dummy for books at home	-0.10	0.00	0.17	0.94
Years of schooling	0.59	-0.01	0.93	1.16
Age	-0.44	0.03	0.92	0.99
Gender	0.09	-0.03	1.01	1.00
Paid work in the previous year	0.49	0.01	0.35	0.97
No paid work ever	-0.07	-0.01	0.62	0.92
Number of people in household	0.11	-0.02	0.96	0.96
Native language is Swedish	0.11	0.03	0.83	0.96
Test for covariate balance (p-value)		0.94		

Note: the table displays standardised differences between the treatment and control group and the variance ratios for these groups based on the model in Column 6 in Table 5.1. The raw (weighted) number of observations in the treatment/control group is 2,659/1,225 (1,994.7/1,889.3).

5.5.1. Does the impact differ depending on AET type?

Do the effects of AET depend on whether it is job-related or non job-related and whether it is formal or non-formal? Table 5.3 shows results from the inverse-probability weighted regression-adjusted model corresponding to Column 6 in Table 5.1. The first panel presents results from models analysing the effects of job-related and non job-related AET separately, while the second panel presents results from models further separating the effects of job-related and non job-related AET into their formal and non-formal components.

The results in the first panel show that the effect of AET on employment is entirely driven by job-related AET, which raises the probability of doing paid work by 6 percentage points. The coefficient for non job-related AET is small and statistically insignificant. Interestingly, when we separate these effects along the formal and non-formal dimension, we find that job-related, non-formal AET raises the probability of employment by 8 percentage points, while the other types have no impact: the coefficients for the two types of formal AET are negative, but not by a statistically significant margin. Overall, our findings thus suggest that job-related, non-formal AET, such as on-the-job training, dominates other types in terms of helping individuals improve their work prospects.

Table 5.3: Effects of different types of AET on the probability of doing paid work

Panel 1	(1)
Job-related AET	0.06***
	(0.01)
Non job-related AET	0.01
	(0.02)
n	3,872
Panel 2	(2)
Job-related, formal AET	-0.01
	(0.03)
Non job-related, formal AET	-0.09
	(0.07)
Job-related, non-formal AET	0.08***
	(0.01)
Non job-related, non-formal AET	0.02
	(0.02)
n	3,858
Panel 3	(3)
Job-related, formal AET	-0.01
	(0.03)
Non job-related, formal AET	-0.11
	(0.08)
Non job-related, non-formal AET	0.03
	(0.02)
On-the-job training (job-related, non-formal AET)	0.06***
	(0.02)
Seminars and workshops (job-related, non-formal AET)	0.08***
	(0.02)
Courses through open or distance education (job-related, non-formal AET)	0.09***
	(0.03)
Other courses or private lessons (job-related, non-formal AET)	0.06**
	(0.03)
n	2,770

Note: Significance levels: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in parentheses. The models include the same covariates as the one presented in Column 6 in Table 5.1.

But are all forms of job-related, non-formal AET equal? To investigate this, we separate the impact of job-related, non-formal AET into its four separate components, as outlined in Section 5.3.1. The results are presented in the third panel in Table 5.3. We find that the effect of job-related, non-formal AET is similar across the four components, suggesting that the impact is not driven by any particular component of such training. The point estimate is the largest for open or distance education and the smallest for on-the-job training and other courses or private lessons, but these differences are not statistically significant.[14] The homogeneous findings in this respect also suggest that possible sources of omitted-variable bias that are specific to certain types of job-related, non-formal AET are unlikely to drive the results.[15]

5.5.2. Robustness tests

To test the robustness of our findings, we (1) present results from models analysing the effects of the different types of AET on the probability to do full-time work, (2) restrict the sample to individuals aged 35 and over, (3) include industry and geographical dummies as covariates, and (4) exclude all covariates apart from age, gender, and the dummy indicating whether or not respondents carried out any paid work in the 12-month period prior to the survey. Finally, since we find that job-related, non-formal AET drives the overall positive impact on the probability to do paid work, we (5) exclude respondents who have undergone other types of AET, as this allows us to carry out the over-identification test for covariate balance for this particular type of training. The results are reported in Table 5.4.

Regardless of model, we obtain results that are consistent with our main estimates. Interestingly, the effects of AET on the probability of working full time in fact appear larger than on the probability of doing any paid work at all. It is especially noteworthy that the model that excludes all controls apart from age, gender, and the dummy

---

[14] This also holds true when excluding individuals who did not participate in any form of job-related, non-formal AET and using "On-the-job training" as benchmark category.

[15] For example, the findings largely rule out one potential source of bias: individuals undergoing on-the-job training as preparation for a new position for which they have already been selected. Such individuals undergo AET as a result of getting the new position rather than vice versa. While we believe adjusting for paid work in the previous year is sufficient to deal with this potential issue, the fact that we find very similar effects across all types of job-related, non-formal AET further indicates it is not an important problem. This conclusion is also supported by unreported analyses in which we found essentially identical effects of job-related, non-formal AET when excluding all individuals who underwent such AET mainly (1) because they were obliged to do so or (2) to do their jobs better and/or improve their career prospects, either of which would presumably apply if AET was undertaken as a result of getting a new position.

indicating whether or not respondents carried out any paid work in the previous 12-month period displays almost identical results as the main estimates, suggesting it effectively irons out differences in other covariates – including social background, cognitive skills, and formal educational levels – that generally are assumed to affect both the probability of engaging in AET and of doing paid work. This strengthens our assumption that the conditional independence assumption is likely to hold in our analysis. The impact of job-related, non-formal AET is also essentially identical when we exclude respondents who have undergone other types of AET. The test for covariate balance shows a p-value of 0.96, suggesting that the weighting procedure also works as intended for the type of training that drives our main findings. Overall, the findings thus further strengthen the idea that specifically job-related, non-formal AET raises the probability of doing paid work, while other types have no consistent effects in this respect.

Table 5.4: Robustness tests

	(1)	(2)	(3)	(4)
	Full-time work	People aged 35+	Add industry and region	Only age, gender, and paid work in previous year as covariates
Panel 1				
AET	0.07***	0.06***	0.04***	0.05***
	(0.02)	(0.01)	(0.01)	(0.01)
n	3,883	2,820	3,883	3,890
Panel 2				
Job-related AET	0.10***	0.07***	0.05***	0.06***
	(0.02)	(0.01)	(0.01)	(0.01)
Non job-related AET	-0.03	0.03	0.00	0.01
	(0.03)	(0.02)	(0.02)	(0.02)
n	3,871	2,809	3,847	3,878
Panel 3				
Job-related, formal AET	-0.05	0.04	-0.05	0.02
	(0.04)	(0.03)	(0.03)	(0.05)
Non job-related, formal AET	-0.03	0.01	-0.07	-0.03
	(0.09)	(0.11)	(0.04)	(0.05)
Job-related, non-formal AET	0.12***	0.07***	0.07***	0.08***
	(0.02)	(0.02)	(0.01)	(0.01)
Non job-related, non-formal AET	0.01	0.03	0.02	0.03
	(0.03)	(0.03)	(0.02)	(0.02)
n	3,857	2,749	3,722	3,878
	(5)			
Panel 4	Only job-related, non-formal AET compared with no AET			
Job-related, non-formal AET	0.08***			
	(0.01)			
Test for covariate balance (p-value)	0.96			
n	2,962			

Note: Significance levels: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in parentheses.

5.5.3. Does the impact vary depending on training intensity?

One may expect the effect of job-related, non-formal AET to vary depending on training intensity; participating in more AET activities may theoretically yield larger benefits on the labour market. To explore whether this is the case, we analyse the impact of participating in one job-related, non-formal AET activity, two such activities, three-to-five such activities, and more than five such activities separately.[16] Table 5.A.2 shows the results.[17] They indicate that there is a 'dose-dependent' relationship between the number of job-related, non-formal AET activities in which individuals have participated and the probability that they do paid work: the coefficient increases with each step change on the variable. For example, while participating in one activity increases the probability of doing paid work by 6.5 percentage points, participating in five or more activities increases the probability of doing paid work by 11.1 percentage points, a difference that is statistically significant. The results thus support the idea that the effects of job-related, non-formal AET depend on training intensity. Moreover, since we in this analysis focus our comparison on individuals who have all participated in at least one job-related, non-formal AET activity, they further suggest that general selection into this type of AET is unlikely to bias the estimates in Tables 5.3 and 5.4.[18]

5.6. Conclusion

As rapid technological development increases knowledge requirements on developed countries' labour markets, it is likely to become more important for people to continuously maintain and update their skills to ensure high employment rates in the future. There are thus reasons to believe that adult education and training could play a key role in ensuring well-functioning labour markets in the future.

In this paper, we have analysed how AET affects employment outcomes in Sweden, exploiting data from the international survey PIAAC, which allows us to obtain rich information on individuals' observable characteristics, including cognitive skills, formal

[16] These categories roughly correspond to observations in the 25th percentile and below, between the 25th and the 50th percentile, between the 50th and the 75th percentile, and in the 75th percentile and above, among people who participated in at least one job-related, non-formal activity.
[17] In order to accurately display the 'dose-dependent' relationship, we report coefficients with three decimal places, instead of two, in Table 5.A.2 specifically.
[18] In unreported analyses, we also studied the impact of participating in several types of job-related, non-formal AET, among individuals who participated in at least one such AET type, and found evidence of positive effects of participating in three or all four types (about 7 per cent of the sample), and a smaller and statistically insignificant impact of participating in two types (about 14 per cent of the sample), compared with the baseline category of participating in one type only (about 22 per cent of the sample).

education levels, and work history. In combination with our exploiting an inverse-probability weighted regression-adjustment estimator, this increases the probability that the estimates reflect causal effects.

We found that individuals who participated in AET in the year prior to the survey are about 4 percentage points more likely to do paid work in the week before the survey than comparable individuals who did not undergo AET. Intriguingly, the effect is entirely driven by job-related AET, which raises the probability of working by 6 percentage points – and this impact is in turn entirely driven by non-formal, job-related AET, which raises the probability of working by about 8 percentage points. We further found that the positive impact is very similar across different types of non-formal, job-related AET. If anything, this positive impact is even larger when analysing the probability of working full time instead of the probability of doing any paid work at all. Interestingly, the effect of non-formal, job-related AET appears to be 'dose dependent', as the impact increases with the number of such AET activities in which individuals have participated. Yet non-formal, non job-related AET and formal AET of either type are not related to the employment outcomes under investigation. While we cannot conclusively rule out that omitted-variable bias affects the results, our analysis does not suggest it is a crucial problem for our conclusions.

Certainly, it is important to note the context of our findings: we analyse data collected in Sweden during the country's economic recovery following the 2008 financial crisis. Since the effects of AET on employment outcomes may well be country specific and interact with the business cycle, future research should investigate the generalisability of the results in these respects. Nevertheless, the positive impact found on employment outcomes in a developed country in the aftermath of a serious economic downturn at least suggests that adult education and training could play a role in ameliorating negative employment effects in similar settings.

Overall, our findings thus support policies seeking to increase take-up of non-formal, job-related AET as a way to secure high employment, at least in similar countries and economic contexts. While reforms to increase access to lifelong learning should be considered, more research into what works in this respect is necessary before deciding on such reforms on a large scale. Finding out what works to efficiently raise relevant AET take-up is thus likely to be a fruitful venue of future research.

References

Albert, Cecilia, Carlos García-Serrano, and Virginia Hernanz. 2009. 'On-the-Job Training in Europe: Determinants and Wage Returns.' *International Labour Review* 149(3):315–341.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Bassanini, Andrea, Alison Booth, Giorgio Brunello, Maria De Paola, and Edwin Leuven. 2005. 'Workplace Training in Europe.' Discussion Paper No. 1640, Bonn.

Becker, Gary S. 1964. *Human Capital: A Theoretical Analysis with Special Reference to Education*. New York: Columbia University Press.

Bergemann, Annette and Gerard J. van den Berg. 2014. 'From Giving Birth to Paid Labor: The Effects of Adult Education for Prime-Aged Mothers.' Working Paper 2014:5, Uppsala.

Bhuller, Manudeep, Magne Mogstad, and Kjell G. Salvanes. 2017. 'Life-Cycle Earnings, Education Premiums, and Internal Rates of Return.' *Journal of Labor Economics* 35(4):993–1030.

Blundell, Richard, Lorraine Dearden, Costas Meghir, and Barbara Sianesi. 1999. 'Human Capital Investment: The Returns from Education and Training to the Individual, the Firm and the Economy.' *Fiscal Studies* 20(1):1–23.

Brunello, Giorgio and Lorenzo Rocco. 2017. 'The Effects of Vocational Education on Adult Skills, Employment and Wages: What Can We Learn from PIAAC?' *SERIEs* 8:315–343.

Brunello, Giorgio, Guglielmo Weber, and Christoph, T. Weiss. 2016. 'Books are Forever: Early Life Conditions, Education and Lifetime Earnings in Europe.' *Economic Journal* 127(600):271–296.

Bussi, Margherita and Jon K. Pareliussen. 2015. 'Skills and Labour Market Performance in Sweden.' Working Paper No. 1233, OECD Economics Department, Paris.

Cedefop. 2015. 'Skill Supply and Demand Up to 2025.' Forecast, Thessaloniki. http://www.cedefop.europa.eu/printpdf/publications-and-resources/country-reports/sweden-skills-forecasts-2025.

Dearden, Lorraine, Howard Reed, and John Van Reenen. 2006. 'The Impact of Training on Productivity and Wages: Evidence from British Panel Data.' *Oxford Bulletin of Economics and Statistics* 68(4):502–528.

Georgiadis, Andreas and Christos N. Pitelis. 2016. 'The Impact of Employees' and Managers' Training on the Performance of Small- and Medium-Sized Enterprises: Evidence from a Randomized Natural Experiment in the UK Service Sector.' *British Journal of Industrial Relations* 54(2):397–421.

Haelermans, Carla and Lex Borghans. 2012. 'Wage Effects of On-the-Job Training: A Meta-Analysis.' *British Journal of Industrial Relations* 50(3):502–528.

Hanushek, Eric A. and Ludger Woessmann. 2011. 'The Economics of International Differences in Educational Achievement.' *Handbook of the Economics of Education* 3:89–200.

Hanushek, Eric A. and Ludger Woessmann. 2015. *The Knowledge Capital of Nations: Education and the Economics of Growth*. Cambridge, MA: MIT Press.

Heckman, James J., Hidehiko Ichimura, and Petra Todd. 1998. 'Matching as an Econometric Evaluation Estimator.' *Review of Economic Studies* 65:261–294.

ILO. 2011. 'A Skilled Workforce for Strong, Sustainable and Balanced Growth: A G20 Training Strategy.' Report, International Labour Office, Geneva.

Imai, Kosuke and Marc Ratkovic. 2014. 'Covariate Balancing Propensity Score.' *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 76(1):243–246.

Jenkins, Andrew, Anna Vignoles, Alison Wolf, and Fernando Galindo-Rueda. 2003. 'The Determinants and Labour Market Effects of Lifelong Learning.' *Applied Economics* 35(16):1711–1721.

Jerrim, John, Luis A. Lopez-Agudo, Oscar D. Marcenaro-Gutierrez, and Nikki Shure. 2017. 'What Happens When Econometrics and Psychometrics Collide? An Example Using PISA Data.' Working Paper No. 17–04, Department of Quantitative Social Science, UCL Institute of Education, London.

Konings, Jozef and Stijn Vanormelingen. 2015. 'The Impact of Training on Productivity and Wages: Firm-Level Evidence.' *Review of Economics and Statistics* 97(2):485–497.

McMahon, Walter W. 2010. 'The External Benefits of Education.' in *The Economics of Education*. Oxford: Elsevier.

Mendolia, Silvia and Ian Walker. 2015. 'Youth Unemployment and Personality Traits.' *IZA Journal of Labor Economics* 4(19):1–26.

OECD. 2013a. 'OECD Skills Outlook 2013: First Results from the Survey of Adult Skills.' OECD Publishing, Paris.

OECD. 2013b. 'Technical Report of the Survey of Adult Skills (PIAAC).' Report, OECD Publishing, Paris.

OECD. 2016. 'Skills Matter: Further Results from the Survey of Adult Skills.' Report, OECD Skills Studies, OECD Publishing, Paris.

OECD. 2017. Micro-level data obtained from the OECD's public database: http://www.oecd.org/skills/piaac/publicdataandanalysis/.

Rubin, Donald B. 2001. 'Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation.' *Health Services and Outcomes Research Methodology* 2(3–4):169–188.

Rubin, Donald B. 2008. 'For Objective Causal Inference, Design Trumps Analysis.' *Annals of Applied Statistics* 2(3):808–840.

Rubin, Donald B. and Neal Thomas. 2000. 'Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates.' *Journal of the American Statistical Association* 95(450):573–585.

Schermer, Isabelle G. 2017. 'Enkla jobb – internationellt.' Stockholm. (https://www.ekonomifakta.se/Fakta/Arbetsmarknad/Sysselsattning/Lagkvalificerade-jobb-internationellt/).

Schwerdt, Guido, Dolores Messer, Ludger Woessmann, and Stefan C. Wolter. 2012. 'The Impact of an Adult Education Voucher Program: Evidence from a Randomized Field Experiment.' *Journal of Public Economics* 96:569–583.

Statistics Sweden. 2013. 'Den internationella undersökningen av vuxnas färdigheter.' Rapport 2013:2, Statistics Sweden, Stockholm. https://www.scb.se/contentassets/9d5f8334eb2a4787b2f2b05cfbc00b6b/uf0546_2013a01_br_a40br1302.pdf.

Statistics Sweden. 2014. 'Vuxnas deltagande i utbildning 2011/2012.' Temarapport 2014:3, Statistics Sweden, Stockholm.

Stenberg, Anders. 2016. *Att välja utbildning – Betydelse för individ och samhälle*. Stockholm: SNS Förlag.

Stenberg, Anders and Olle Westerlund. 2015. 'The Long-term Earnings Consequences of General vs. Specific Training of the Unemployed.' *IZA Journal of European Labor Studies* 4(22):1–26.

Tåhlin, Michael 2007. 'Överutbildningen i Sverige – utveckling och konsekvenser.' Pp. 70–89 in *Utbildningsvägen – vart leder den? Om ungdomar, yrkesutbildning och försörjning*, edited by Johan Olofsson. Stockholm: SNS Förlag.

Vignoles, Anna, Fernando Galindo-Rueda, and Leon Feinstein. 2004. 'The Labour Market Impact of Adult Education and Training: A Cohort Analysis.' *Scottish Journal of Political Economy* 51(2):266–280.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Appendix

Table 5.A.1: Descriptive statistics

Variable	Mean	SD	Min	Max
Participated in AET	0.65	0.48	0	1
Participated in AET (paid work in the previous year)	0.70	0.46	0	1
Participated in AET (no paid work in the previous year)	0.37	0.48	0	1
Participated in job-related AET	0.52	0.50	0	1
Participated in non job-related AET	0.13	0.34	0	1
Participated in formal AET	0.14	0.35	0	1
Participated in formal, job-related AET	0.10	0.30	0	1
Participated in formal, non job-related AET	0.04	0.20	0	1
Participated in non-formal AET	0.60	0.49	0	1
Participated in non-formal, job-related AET	0.49	0.50	0	1
Participated in non-formal, non job-related AET	0.11	0.31	0	1
On-the-job training (job related)	0.29	0.45	0	1
Seminars and workshops (job related)	0.28	0.45	0	1
Courses through open or distance education (job related)	0.12	0.32	0	1
Other courses or private lessons (job related)	0.15	0.36	0	1
Number of job-related, non-formal AET activities	2.45	5.31	0	83
1 job-related, non-formal AET activity	0.13	0.33	0	1
2 job-related, non-formal AET activities	0.09	0.28	0	1
3–5 job-related, non-formal AET activities	0.15	0.36	0	1
>5 job-related, non-formal AET activities	0.13	0.33	0	1
Paid work	0.73	0.44	0	1
Full-time work	0.61	0.49	0	1
Paid work in the previous year	0.85	0.35	0	1
No paid work ever	0.02	0.14	0	1
Age	43.53	13.01	16	65
Woman	0.49	0.50	0	1
Number of people in household	2.70	1.32	1	6
First-generation immigrant	0.19	0.39	0	1
Second-generation immigrant	0.03	0.17	0	1
Native language is Swedish	0.89	0.38	0	1
Years in Sweden	39.31	16.36	0	65
Number of books at home	3.79	1.40	1	6
Mother’s educational level	1.66	0.82	1	3
Father’s educational level	1.69	0.83	1	3
Years of schooling	12.21	2.55	6	20
PIAAC literacy score	278.09	49.04	23.57	415.64
PIAAC numeracy score	278.54	53.35	52.18	444.13

Note: all data are weighted by respondents’ sampling probability in PIAAC. Only observations without imputed values are used to calculate the descriptive statistics. The 21 industry and 8 geographical dummies, as well as dummies indicating missing values for covariates, are suppressed.

Table 5.A.2: Effects of job-related, non-formal AET depending on training intensity

	(1)	(2)
1 job-related, non-formal AET activity	0.065***	Benchmark
	(0.018)	
2 job-related, non-formal AET activities	0.070***	0.005
	(0.018)	(0.024)
3–5 job-related, non-formal AET activities	0.074***	0.020
	(0.018)	(0.021)
>5 job-related, non-formal AET activities	0.111***	0.045**
	(0.015)	(0.021)
n	3,799	2,023

Note: Significance levels: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in parentheses. The models include the same covariates as the one presented in Column 6 in Table 5.1. The first model includes individuals who did not participate in any job-related, non-formal AET activities and are therefore the baseline category. However, in the second model, such individuals are excluded and the baseline category is instead composed of individuals who participated in one job-related, non-formal AET activity.

6. Conclusion

Drawing causal inferences from empirical research is vital for allowing politicians to create and implement social policies that effectively and efficiently deal with the issues they are designed to address. Yet this is rarely reflected in the policy process, partly because much social-policy research historically has not employed methods with which it is possible to separate causation from correlation, and partly because policymakers often ignore the quality of research for political reasons. Rather than shaping policy after robust evidence, there has been a tendency to lean on research that supports policy and ideology.

A key aspect of methodologically sound quantitative research is that it, in one way or the other, retrieves variation in the variable that is not itself affected by the dependent variable and is not related to other factors – observable or unobservable – that affect the dependent variable. Historically, such variation has often been obtained by running randomised trials, earning them a reputation for being the 'gold standard' of policy-relevant research. Yet also randomised trials have weaknesses, such as the difficulties involved in obtaining sufficiently large and diverse samples to obtain high external validity in the findings, which in turn often decrease their usefulness for policy purposes. Also, they can be expensive and are sometimes too sensitive to be politically viable, especially when analysing the effects of large-scale interventions. For this reason, randomised trials are neither always possible nor desirable for the purposes of obtaining exogenous variation in the policy of interest for research purposes – which in turn makes it important to exploit econometric methods with which one can draw causal inferences also from observational data.

This thesis has presented four papers employing such methods to answer important research questions in three different areas of social policy: education, health, and immigration. To conclude the thesis, this chapter summarises the main findings of the papers, discusses the strengths and weaknesses of the methods utilised and the papers more generally, provides directions for future research, and outlines policy implications.

6.1. Retirement and mental health

The second chapter of the thesis analysed the short- and long-run impact of retirement on mental health in Europe, exploiting several waves of data from the Survey of Health, Ageing, and Retirement in Europe. Since the 1990s, several governments have reformed their state-pension systems to incentivise higher labour-force participation rates among the elderly, including by increasing the official state-pension age. Postponing retirement may theoretically carry both mental-health benefits and costs, which have important implications for the viability of the reforms that are currently underway. The paper used an individual-fixed effects instrumental-variable design with state-pension thresholds as instruments to obtain causal estimates of the impact of the retirement decision in both a short- and longer-run perspective, being the first study to separate the short- from longer-term effects in such a set-up. While the paper found no impact in the short run, it did find a large longer-term, lagged negative impact of retirement on mental health, which does not differ between men and women or people with different educational and occupational backgrounds.

An important strength of the study's methodology is its reliance on institutional differences in retirement incentives that are dependent on age thresholds. These thresholds are highly likely to be exogenous to mental health and therefore allow us to draw causal inferences from the results. The findings' credibility hinges on the assumption that merely crossing the state-pension age thresholds serving as instruments does not impact individuals' mental health around the thresholds, apart from via retirement, once adjusting flexibly for the impact of age. The study provided several robustness and sensitive tests to show that this assumption is likely to hold. In addition, the variation exploited is highly relevant for public policy, as the estimates concern individuals whose retirement are determined by the state-pension age thresholds – which in turn are those most likely to be affected by changes to those thresholds. In other words, the method employed is credible both from a research and policy perspective.

At the same time, the method's strength in this respect is to some extent also a weakness: the study's findings cannot necessarily be extrapolated to people who retire for other reasons than because they reach the regular state-pension age, including those who retire early or because of ill health. In other words, while the study's findings are highly relevant for understanding the mental-health effects of retirement due to one important policy-relevant parameter – the incentives created by the regular state-pension age – it is not necessarily relevant for understanding the effects of retiring for other policy-relevant reasons. Analysing the short- and longer-term mental-health effects of retirement using other sources of exogenous variation would therefore be a fruitful venue for future research. In addition, the study is silent on the mechanisms through which the negative longer-term effect of retirement at the regular state-pension threshold operates. As policymakers would benefit from understanding such mechanisms, for example when seeking to address the negative impact found in the paper, it is important for future research to identify mechanisms linking retirement to declining mental health. Also, while the study analyses the effects of retirement in the longer-term perspective, this perspective is defined as about 4–6 years following the retirement decision; future research should determine whether these effects persist further. Finally, the paper has identified an average effect of retirement across ten European countries; it has not sought to investigate possible heterogeneity across the countries and it is not clear whether the findings can be extrapolated to countries not included in the study.

Still, overall, the paper has important policy implications. Its findings suggest that European policymakers do not face a trade-off between increasing the regular state-pension age and improving mental health among the elderly, at least in a slightly longer-term perspective. On the contrary, the results indicate that politicians who raise the regular state-pension age could produce a virtuous circle in which pension systems are made sustainable, health expenditures decreased, and, with some time, mental health among the elderly improved.

6.2. School competition and the wellbeing-efficiency trade-off in education

The third chapter, in turn, investigated whether independent-school competition involves a trade-off between pupil wellbeing and academic performance. While much educational theory assumes that wellbeing and achievement go hand in hand, existing evidence indicates that interventions that have a positive impact on achievement often decrease pupil wellbeing. Yet this is the first study to investigate whether the trade-off applies to independent-school competition specifically, the general effects of which have become a fiercely debated topic worldwide. Using international pupil-level data, covering hundreds of thousands of pupils in 34 OECD countries, the paper exploited an instrument based on early Catholic resistance to state schooling to obtain exogenous variation in current enrolment shares in independently-operated schools. It found that independent-school competition decreases pupil wellbeing but raises achievement and lowers educational costs. It also found mechanisms explaining the trade-off, including more traditional teaching and stronger parental achievement pressure.

A key strength of the third chapter's methodology is its ability to estimate the system-level effects of independent-school competition on pupil wellbeing, achievement, and costs in the very long-term perspective. Studies analysing the effects of within-country reforms that increase independent-school competition often suffer from limited variation over time, making it difficult to draw conclusions about the general-equilibrium effects in the long run. In addition, within-country studies are more likely than cross-national ones to suffer from spill-over effects across the regional levels used to estimate the impact of competition. Exploiting cross-country data on pupil wellbeing, drawn from nationally representative samples of pupils, also enables the study to obtain estimates that are relevant for external-validity purposes in this respect, which otherwise is difficult because of the lack of pupil-wellbeing surveys that are representative at regional levels within countries. Another strength of the study more generally, which contrasts it with the other chapters in the thesis, is that it is able to explore and identify several mechanisms linking independent-school competition to lower pupil wellbeing and higher academic performance, which is somewhat rare in the empirical literature.

However, as in the second chapter, some of the strengths of the method employed may also imply possible weaknesses in other respects. For example, exploiting cross-national variation in independent-school competition means that the study must deal with unobserved heterogeneity at the country level. The study went at length to show that the methodology addresses this issue satisfactorily; if anything, it suggested that such heterogeneity may bias the results against the paper's conclusions. In other words, the effects of independent-school competition on pupil wellbeing and academic performance may in fact be larger than what the study finds. Yet another potential weakness is that the findings cannot necessarily be extrapolated to independent-school competition that is not linked to Catholic resistance to state schooling in the late 19th and early 20th centuries. For example, competition that arose because of voucher reforms in which for-profit operators have been allowed to participate, as in Chile and Sweden, or because of reforms that enabled mass conversions of state schools to independently-operated status essentially overnight, as in England, is unlikely to have much to do with the variation exploited. Future research should therefore employ alternative strategies and analyse different settings to analyse how other forms of independent-school competition affect pupil wellbeing and academic performance.

Nevertheless, overall, the third chapter carries important policy lessons. In contrast to the second chapter, it suggests that policymakers do face a trade-off between different goals – academic performance and children's subjective wellbeing at school – when considering reforms designed to increase independent-school competition. A tentative back-of-the-envelope calculation indicated that the economic benefits of independent-school competition via its positive impact on achievement are likely to outweigh its cost via lower pupil wellbeing, but also that the costs of competition may outweigh its benefits when using adult life satisfaction as the unit of measurement. The paper thus suggests the potential for a more general trade-off between the traditional goals of education policy and the wellbeing agenda, to which policymakers should pay attention.

6.3. Refugee immigration and voter turnout

Meanwhile, the fourth chapter analysed the effects of refugee inflows on voter turnout in Sweden in a period when shifting immigration patterns made the previously homogeneous country increasingly heterogeneous. An important possible consequence of immigration could be altered political engagement among natives, but it is not clear in which direction engagement should theoretically be affected – and disentangling cause from effect is very difficult. Analysing individual-level panel data and exploiting a national placement programme – which assigned refugees to the municipalities via contracts – to obtain plausibly exogenous variation in immigration, the paper found that refugee inflows significantly raise the probability of voter turnout. Balancing tests on initial turnout as well as placebo tests regressing changes in turnout on future refugee inflows support the causal interpretation of the findings. The results are consistent with group-threat theory, which predicts that increased out-group presence spurs political mobilisation among in-group members.

An important strength of the fourth chapter's methodology is that it addresses the difficult endogeneity issues involved in studying the effects of immigration on political outcomes, thus suggesting high internal validity of the estimates. The identification depends on the assumption that the influx of refugees into the municipalities arising due to the placement-programme contracts is exogenous to changes in individual-level turnout, when conditioning on variables of relevance for understanding the programme's dynamic. The study provided a battery of balancing, placebo, and robustness tests to show that this assumption is likely to hold. Overall, the methodology employed makes the study a significant contribution to a literature that thus far mostly has provided associational evidence and/or analysed aggregate outcomes. Additionally, the variation used to obtain exogenous variation in refugee inflows is policy relevant, since it springs from a government programme. Similarly to the method employed in the second chapter, the variation exploited in the fourth chapter is thus pertinent both from a research and policy perspective.

Certainly, the methodology also has limitations, perhaps the most important one of which is the potential lack of external validity. This is because the data analysed were designed to be representative at the national rather than municipal level, meaning that it is not necessarily possible to extrapolate the findings beyond the sampled population. Also, the findings are obtained in a situation when a previously homogeneous country rapidly became more heterogeneous because of refugee immigration; whether other types of immigration, or similar types of immigration in other contexts, would have the same effects is unclear. Similarly, it is not clear whether an increased probability in individual-level turnout translates into higher aggregate turnout at the municipal level, since this partly depends on how immigration affects overall demographic changes across communities. The extent to which these issues affect the study's conclusions is an important venue for future research. Finally, although the study highlights a prominent theoretical mechanism linking refugee immigration to turnout, it does not analyse the precise reasons explaining this mechanism, such as attitudinal changes towards immigrants. Exploring such reasons empirically is thus another important topic for future research.

In similarity with the third chapter, the fourth chapter's policy implications are not clear-cut. On the one hand, the study suggests that refugee immigration is likely to boost the probability of voter turnout, which is an important goal for policymakers worldwide. On the other hand, this boost is likely to be the result of 'group threat', which has been theorised to stimulate political mobilisation among natives. While perceptions of threat may potentially be addressed through increased contact between natives and immigrants, which has been found to improve inter-group relations in some contexts, such contact tends to be endogenous to inter-group attitudes – and real-world demographic changes do not necessarily generate more voluntary contact between in- and out-group members. Thus, while it is possible to view the impact of refugee inflows on voter turnout in a positive light, a more pertinent lesson is likely to be the importance of devising strategies to decrease natives' sense of threat in situations of high refugee immigration, perhaps through various forms of nudging to increase the probability of positive contact with refugees. Yet further empirical research to investigate what works in this respect is necessary.

6.4. Adult education and employment outcomes

Finally, the fifth chapter analysed the impact of adult education and training (AET) on employment outcomes in Sweden. It is plausible that AET could help promote the knowledge and skills necessary for ensuring high employment and productivity levels in the future. Exploiting unusually rich data from the Programme for the International Assessment of Adult Competencies and using an inverse-probability weighted regression-adjustment (IPWRA) estimator to deal with selection bias, the study found that AET raises the probability of doing paid work by 4 percentage points on average. This impact is entirely driven by non-formal, job-related AET, such as workshops and on-the-job training. The study also found that the effect – which increases with training intensity – is very similar across different types of non-formal, job-related AET.

Unlike the other studies in the thesis, the fifth chapter does not rely on quasi-experimental variation to obtain causal estimates. Instead, it assumes that assignment to the different forms of non-formal, job-related AET is as good as random conditional on the rich covariates included, including individuals' work history, but does not make assumptions about the functional form of the relationship between those covariates and the outcome analysed. The specification and robustness tests do suggest that the conditional independence assumption holds – which is further supported by the fact that the impact is very similar across different forms of non-formal, job-related AET and increases with training intensity – and if it does, the IPWRA model comes with the benefit of high external validity, since it estimates the average treatment effect in the population of interest. In contrast, as highlighted above, models relying on quasi-experimental variation estimate the local average treatment effect among individuals who respond to the specific variation that is exploited. This difference is especially important for this study since it concurrently analysed several forms of AET, for which it is difficult to find and simultaneously exploit separate sources of quasi-experimental variation. In this sense, the IPWRA model also holds an advantage over matching models, which do not allow analyses of multiple treatments at the same time.

Yet the method also has weaknesses that, as in previous chapters, to some extent are related to the sources of its strengths. For example, despite the specification and robustness tests, the study cannot entirely rule out that unobservable heterogeneity biases the estimates. Some caution regarding the results is thus warranted and future research should attempt to investigate the effects of separate forms of AET using different types of quasi-experimental variation to address potential selection on unobservable characteristics. At a more general level, the study's findings cannot necessarily be extrapolated beyond Sweden, a country with one of the most knowledge-intensive labour markets in the world, and the context of the country's economic recovery following the 2008 financial crisis. In addition, the paper is silent on potential mechanisms linking AET to better work prospects. An important venue for future research would thus be to analyse the impact of multiple types of AET in other contexts as well as explore mechanisms that could help explain any detected effects.

Still, the fifth chapter also has important policy implications. Specifically, its results indicate that policies stimulating take-up of relevant job-related, non-formal AET have promise as a way to secure higher employment rates in the future, at least on labour markets as knowledge intensive as the one in Sweden. Reforms to increase access to relevant lifelong learning and training should thus be considered and trialled.

6.5. Concluding thoughts

Overall, the papers presented in the thesis have together highlighted important trade-offs between different identification strategies frequently employed in observational studies designed to obtain causal estimates, as they rely on different assumptions and are able to answer different questions about the effects of the factors analysed. They have distinctive strengths and weaknesses, which must be carefully considered when drawing research and/or policy conclusions. In this sense, apart from making a significant contribution to the literature in each of the separate areas under investigation, the thesis has provided important case studies of how different econometric techniques, given varying contexts and data availability, can be utilised to answer different questions of relevance to social policy specifically – as well as displayed how important it is that politicians draw on this type of research more often than is currently the case. In doing so, it has sought to contribute to the development of more evidence-based social policy worldwide.