IFAU
Institute for Evaluation of Labour
Market and Education Policy

Essays in education and family economics

Krzysztof Karbownik

DISSERTATION SERIES 2014:6

Presented at the Department of Economics, Uppsala University

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

IFAU also provides funding for research projects within its areas of interest. The deadline for applications is October 1 each year. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The institute has a scientific council, consisting of a chairman, the Director-General and five other members. Among other things, the scientific council proposes a decision for the allocation of research grants. A reference group including representatives for employer organizations and trade unions, as well as the ministries and authorities concerned is also connected to the institute.

Postal address: P O Box 513, 751 20 Uppsala
Visiting address: Kyrkogårdsgatan 6, Uppsala
Phone: +46 18 471 70 70
Fax: +46 18 471 70 71
ifau@ifau.uu.se
www.ifau.se

This doctoral dissertation was defended for the degree of Doctor in Philosophy at the Department of Economics, Uppsala University, December 12, 2013. Essay 1 has been published by IFAU as Working paper 2014:13, Essay 2 as Working paper 2014:14 and Essay 3 as Working paper 2014:15.

ISSN 1651-4149

Dissertation presented at Uppsala University to be publicly examined in B115, Kyrkogårdsgatan 10 B, Uppsala, Thursday, 12 December 2013 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Erik Plug (University of Amsterdam).

Abstract

Karbownik, K. 2013. Essays in education and family economics. *Economic studies* 140. 182 pp. Uppsala: Department of Economics. ISBN 978-91-85519-47-7.

Paper 1: This paper examines the determinants of teacher turnover using matched employee-employer panel data from Swedish lower and upper secondary schools in a market-oriented institutional environment with a growing private sector and individually negotiated wages. I find statistically significant and robust negative correlations between mobility and monetary compensations. Unlike previous research, I do not find robust evidence that share of minorities correlates positively with turnover. The positive association exists; however, in the case of private and upper secondary institutions.

Paper 2: This paper examines the job mobility of teachers with different skills using matched employer-employee data from Swedish secondary schools. In addition to standard quality measures, I have access to population-wide data on cognitive and non-cognitive assessments of males born in 1951 or later. The results show that high-quality teachers are less mobile than others, and that there is no significant correlation between turnover and share of minority students. Interestingly, teachers with better skills are less likely to leave the profession, which suggests that the documented drop in the quality of inflowing teachers may partly be offset by a higher tendency for high quality teachers to stay in the profession.

Paper 3: This paper examines teachers' mobility in response to exogenous changes in the credentials of their students using data from Stockholm high schools. I explore a major admission reform that lead to the reshuffling of students between schools within the municipality of Stockholm. The results show that a 10-percentile-point increase in student quality decreases the probability of a separation by up to 9 percentage points. These effects are very similar across all types of teachers and are found mainly for mobility between schools rather than out of the profession. They are also present only in the lower half of the student quality distribution. Teachers react mostly to direct measures of student quality (grades from compulsory school) rather than to other characteristics that are correlated with student quality (immigrant status, parental income, paternal cognitive skills). Finally, I do not find any significant effects of changes in student quality on individual teacher's earnings or school hiring policies.

Paper 4: We examine the effects of child's gender on family expenditure patterns using data from the Polish Household Budget Survey 2003-2010. A first-born daughter compared to first-born son increases overall and child related spending on clothing. At the same time, it decreases spending on games and toys as well as, at the intensive margin, early inputs into human capital production function measured as pre-kindergarten and kindergarten expenditure. We show that these findings are not driven by effects of child gender on maternal labor supply and are unlikely to be driven by the negative effects of first-born girl on fertility decisions or marital stability. Our findings suggest that child gender leads to differential expenditure patterns, which in turn might lead to early assignment of social roles.

Paper 5: We estimate the effect of family size on female labour supply using data from the Polish Household Budget Survey, and instrumenting for family size with twinning at first and second birth. We identify a positive bias of OLS in the estimates of maternal labour supply on family size among highly educated and older mothers, and those who had their first child born after the age of 26. This unusual and counterintuitive result confirms the importance of accounting for heterogeneous treatment effects in the analysis of the relationship between labour market and family outcomes. Furthermore, among families with at least one child we identify the total average causal effect of an additional child on mother's employment to be -6.7 pp. We find no significant effects of having additional children on female employment among families with two or more kids.

Keywords: teacher mobility, labor supply, gender preferences

Krzysztof Karbownik, Department of Economics, Box 513, Uppsala University, SE-75120 Uppsala, Sweden.

For my Mother.

In the memory of my Father.

Acknowledgements

Wow! It was a long and bumpy road and, looking back, it reminds me of the bumpiness of an actual road next to my high school in Łódź, Poland, where this train initially departed. Now, the train is finally arriving at Uppsala Central and the last segment of the ride approaches its end. When I look back at my time as an undergraduate and a graduate student I can see all the possible combinations of colors. There were greys of tedious data work, bright spots when papers got accepted to conferences and blacks when they got rejected from journals. There were tears of both joy and misery, but most importantly there was an enormous amount of knowledge and experience that I was bestowed with. Now that I am nearly at the end of this long voyage I am certain about one thing – it was worth it! I am proud of my work and my achievements, and at the same time realize and acknowledge that this thesis and all other work I have done would not have been possible without help, patience and support of numerous people.

First and foremost, I would like to thank my advisor, Björn Öckert. His knowledge, patience, guidance and often down to Earth overview of the situation have been invaluable. Björn's empirical expertise and creativity cannot be overstated. When the deadlines were approaching, the results were not as expected, and I had already started panicking he was always a perfect example of the British war-time motto: Keep calm and carry on. Although we have not written any paper together, without his hard work on educational data collection for IFAU none of the three papers using Swedish registries would have been written. In that sense, he was the most generous advisor that a graduate student can have and did not hesitate to share the data with his advisee. Björn, you have been more than amazing, so expect a long line of potential advisees queuing in front of your office in the future.

I am also indebted to my two co-advisors, Per-Anders Edin and Mikael Lindahl. Both of them were extremely helpful, and their guidance before Björn took over as my main advisor provided me with a solid foundation for my thesis. Our countless conversations not only gave me the professional knowledge but also an outlook on how the life of a scholar looks like. The beers we shared with Mikael will always remind me that you can actually have fun with your advisors.

Many thanks go to my co-author on the last two papers in the thesis, Michał Myck, who shares my passion for introducing evidence-based policy in our home country. Michał, thank you for cleaning and sharing the data on consumption and family structure. I enjoyed working on these two papers and I think we proved that there is a demand for high quality economic research in Poland. I really cannot wait to kick off the other projects that we have already contracted.

I would like to thank Nina Andersson and Ann-Sofie Wettergren Djerf for the administrational assistance. Special thanks here must go to graduate studies coordinator – Katarina Grönvall. Without her flexibility, hard work and willingness to answer my e-mails on Sunday evenings it would have taken me much longer to complete this thesis. As an empirical researcher I am also highly dependent on the support of the IT staff. Thus, big thanks go to Uppsala's amazing IT duo: Jörgen Moen and Åke Qvarfort.

I have started my PhD at the University of Arizona so I would like to express my gratitude to the economics community at UoA, and to my undergraduate advisors whose recommendation letters got me the job in the desert. Had it not been for my illness, I would have probably continued graduate school in Tucson. Thus, I would like to express my deep gratitude to Martin Dufwenberg from UoA and Bertil Holmlund from Uppsala who understood that my medical condition should not stand in my way to a degree and helped me in transferring to Uppsala University. Thank you for believing in me and supporting me during these tough days. I would not be able to write this thesis if it weren't for your help.

Starting spring of 2012, I have been visiting Institute for Policy Research at Northwestern University. I am indebted to Jan Wallander and Tom Hedelius foundation whose initial grant permitted this visit to the United Stated. However, I am perhaps even more grateful to my advisors, the department of economics and Uppsala Center of Labor Studies who decided to continue the funding of my doctoral studies even though I spent most of my time at Northwestern. Big thanks go to David Figlio, who invited me to the IPR and encouraged to work with him on projects that use the most amazing micro dataset that has ever been constructed in the US. Thanks to Jon Guryan and Jeffrey Roth – our two amazing collaborators on this research agenda. I am also grateful to Kirabo Jackson, Diane Whitmore Schanzenbach, Joel Mokyr, Joe Ferrie and Lori Beaman for taking time to discuss my work. During my stay at Northwestern I also started collaboration on health-education project using historical data and I would like to thank my awesome co-author Anthony Wray for countless discussions about biases, matching and historical censuses.

During my years in Uppsala I have shared an office with Gabriella Chirico and Tove Eliasson. I have really enjoyed our time as colleagues and friends, and I cannot thank you enough for helping me in dealing with administrational issues and language barrier. Other people at the department who made this hard work almost as pleasant as vacations include, among others, Adrian, Spencer, Teodora, two Matthiases, two Jons, Rita, Johan, Glenn and Haishan. However, at the department we not only took care of our brains but also of our bodies (at least some did) so big thanks to Oscar Erixon, Rachatar Nilavongse and Mohammad Sepahvand for countless ping pong games and for letting me win so many times. Finally, I would like to thank Arizo Karimi, Martin Nilsson and Anna Persson for interesting and in depth political discussions. You guys nearly converted me to the left side of political spectrum!

I am also extremely grateful to my Polish friends. You guys were awesome, supportive and never forgot about me. We went together through ups and downs and you mercifully tolerated my obnoxiousness when things were not going so well. So Mateusz B., Piotr B., Ania Ch., Grzesiek K., Magda M., Asia O., Kalina O., Ewa P., Paweł P., Magda P., Tamara P. and Michał T. I owe you so much that I am not sure if I will ever be able to pay you back.

Last but definitely not least I have no words for explaining my gratitude towards my family. Although, they often did not understand my choices, they were always my biggest supporters throughout life. Without my parents, my brother, his wife and their wonderful daughter I would not have so much energy and passion in finishing the greatest project in my life thus far – this dissertation. You are my heroes and my success will always be partially yours.

Evanston, October 2013

Krzysztof Karbownik

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

- I Karbownik, K. (2013) The determinants of teacher mobility in Sweden
- II Karbownik, K. (2013) Job mobility among high-skilled and low-skilled teachers
- III Karbownik, K. (2013) Do changes in student quality affect teacher mobility? Evidence from an admission reform
- IV Karbownik, K., Myck, M. (2013) Who gets to look nice and who gets to play? Effects of child gender on household expenditure
- V Karbownik, K., Myck, M. (2013) For some mothers more than other: how children matter for labour market outcomes when both fertility and female employment are low

Reprints were made with permission from the respective publishers.

Contents

Introduction

This thesis consists of five essays and can be divided into two major parts. The first makes use of Swedish data registers to study the teacher labor market. The second brings the analysis to Poland and studies the effects of family size on parental labor supply as well as the effects of child gender on family outcomes. All of the essays are empirical, but the first two present descriptive evidence while the latter three attempt to gauge causal effects.

Essay I studies the institutional aspects of labor market for teachers and how they correlate with teacher turnover. Essay II documents turnover among Swedish teachers of different quality, where skills are measured not only using standard input-based quality measures (education or experience), but also cognitive and non-cognitive military assessments. Essay III investigates how student quality affects teacher mobility using exogenous variation in students' credentials. Essay IV examines the effects of child's gender on family consumption, stability, fertility and labor supply using data on Polish households. Essay V assesses the effects of family size on parental labor supply in a setting with low levels of female employment and a low fertility rate.

Even though all of the papers share some broad methodological attributes, they are clearly associated with two different sub-fields of applied microeconomics and use data from two utterly different countries. Therefore, this thesis is best described as first discussing job mobility of teachers in Sweden and subsequently the impact of children on family choices in Poland.

There have been numerous studies linking teacher mobility to both pecuniary and non-pecuniary characteristics of teaching profession. Greenberg and McCall (1974) were among the first to study mobility of teachers in San Diego school district. More studies followed, and they concentrated on the role of monetary compensations (Antos and Rosen, 1975), minorities (Falch and Strøm, 2005), outside opportunities for teachers (Dolton and van der Klaauw, 1995), general working conditions (Mont and Rees, 1996), distance from hometown (Barbieri et al., 2011), school resources (Feng et al., 2010), student quality (Hanushek et al., 2004) or teacher destinations (Lankford et al., 2002). These studies come from a variety of countries including Great Britain, Italy, Norway and United States but this phenomenon has never been studied in Sweden. However, analyzing teacher turnover in Sweden should be of general interest because due to its institutional setting which combines many of the features listed above and the richness of data that include, for example, explicit measures of teachers' skills.

The first part of this thesis attempts to fill in this gap in the literature. It has been established that teachers are one of the most important factors in the education production function (Rivkin et al., 2005). It is not clear *a priori*, however, whether policy makers would want to minimize the teacher turnover. On the one hand, they may want to lower turnover if it leads to lower student achievement (Ronfeldt et al., 2013) or introduces substantial transaction costs to municipality budgets. On the other hand, they may want to increase teacher turnover in order to improve teacher-school-student match quality (Abelson and Baysinger, 1994).

Beyond the discussion about teacher mobility, there is also a literature focusing on changes in the educational system. These include institutional changes as well as "natural" evolution of the profession. In the former categories over the past two decades researchers were interested primarily in school choice (Cullen et al., 2006), class size (Fredriksson et al., 2013), school voucher programs (Hsieh and Urquiola, 2006), student busing (Jackson, 2009), competition between public and private sector (Hensvik, 2012), changes in admission policies (Söderström and Uusitalo, 2010) or changes in funding sources (Fredriksson and Öckert, 2008). The latter category was mostly concerned with decreasing quality of teacher education and negative selection of weaker candidates into teaching (Fredriksson and Öckert, 2007). However, only few of these policies have been analyzed in the context of labor market for teachers; the reason for these gaps in the literature relates to data availability.

Sweden has multiple institutional features that are of importance from the view point of all the papers. First, there is a growing private sector in schooling (Björklund et al., 2006). In 1992, Sweden introduced a school voucher reform and this decision resulted in an increase in the fraction of privately owned schools. Second, Sweden has shifted the financial responsibility for schooling from the central government to municipalities (Fredriksson and Öckert, 2008). Third, teacher wages are determined at local level through individual wage bargaining between teacher and principal given the collective bargaining outcomes set at the national level (Skolverket, 2009). To my knowledge there are few countries that has introduced such a level of decentralization, choice and flexibility into their educational system, and thus, it is inherently interesting to study how estimates from such a market-oriented system compare to estimates from more rigid settings.

Essay I, "The determinants of teacher mobility in Sweden", examines teacher turnover using matched employee-employer panel data from Swedish lower and upper secondary schools in a market-oriented institutional environment. For example, I demonstrate that teacher turnover correlates negatively with teacher monetary compensations but that it does not correlate significantly with the fraction of minorities at school on average. The latter finding is novel in the literate, as thus far most studies estimated strong and positive correlations between teacher turnover and the level of minority enrollment. More importantly, I document substantial heterogeneity in this association and I show that it exists only for upper secondary and private schools. If share of minorities at school is associated with a disruptive behavior or not-fitting-in, and these behavioral problems grow in teenagehood, then my results suggest that the turnover rates in US high schools might actually be even higher than estimated thus far mostly for lower levels of educational system. Interestingly, the average association that cannot be found in Sweden is strong and robust in the case of another Nordic country – Norway (Falch and Strøm, 2005).

Furthermore, the completeness of Swedish registry data allows me to study associations in a setting with relatively flexible wages and private ownership of schools. I demonstrate that privately owned schools experience higher teacher turnover rates and that this correlation is weaker for upper secondary schools. I also find a negative relationship between earnings and teacher turnover, which decreases when I add control variables, and thus, a somewhat speculative interpretation of this negative result could be that it may be possible to influence teacher's mobility decisions through changes in their monetary compensations.

Essay II, "Job mobility among high-skilled and low-skilled teachers", examines turnover of teachers with different skills using matched employer-employee data from Swedish lower and upper secondary schools. This analysis proves to be crucial from the viewpoint of policy makers as studies have documented a deterioration of the skills in the pool of potential teachers to hire (Fredriksson and Öckert, 2007). Therefore, reducing turnover among high-quality teachers must probably be crucial for any principal wishing to sustain the competence level in their school. I show that university educated and experienced teachers are less likely to both leave their current school and the profession. Furthermore, using the unique enlistment records I show that teachers with high non-cognitive skills are less likely to change employers. At the same time, I do not find robust correlations for cognitive skills when I control for standard teacher quality measures like education or experience.

These findings may have two important policy conclusions that should be further examined. First, the drop in teacher quality documented by others may be partly offset by lower tendency for high-quality teachers to leave the profession. Second, although the quality measures used previously do a fairly good job in capturing teacher cognitive quality dimension, they do rather poor job in capturing the set of socio-emotional skills. The latter finding is also important because socio-emotional skills may be crucial in managing classroom full of teenagers.

Finally, I provide evidence on how school and teacher characteristics interact with teacher quality. For example, I do not find any support for the common view that schools serving minority students experience higher turnover rates of high-quality pedagogues. Moreover, I find a robust negative correlation between monetary compensations and teacher turnover, and when I add earnings to the equation, the coefficients on teacher quality measures decrease. Although speculative, due to correlational nature of this study, such results hint upon the possibility of influencing the mobility decisions of high-quality teachers through changes in their monetary compensations.

Essay III, “Do changes in student quality affect teacher mobility? Evidence from an admission reform?”, addresses the question of how exogenous changes in student quality affect teacher labor supply decisions. I explore a major school choice reform that was implemented in the municipality of Stockholm in school year 2000/2001. Prior to the reform students applied only for a program and the grades from lower secondary school decided on the admission. Students could state their school preferences, but the ones living closest to a particular facility enjoyed admission priority. The 2000 reform abolished all residence-based admission criteria and introduced a system that is solely based on lower secondary school performance. The intention of the reform was to undo the effects of residential segregation. Söderström and Uusitalo (2010) show that it induced large reshuffling of students.

Their findings can be treated as a “first stage” relation for the question posed in my paper. First, the reform generated an exogenous, from the perspective of a teacher, change in the composition of incoming students at the school level, but it did not alter the average composition of students in the municipality of Stockholm. In other words, some teachers ended up with lower quality pupils and some teachers ended up with higher quality pupils than they *a priori* expected. Therefore, if I observe separations in the subsequent periods, I can attribute them to changes in student quality. Such a design calls for difference-in-differences approach where the first difference compares teacher mobility before and after the reform and the second difference compares teacher mobility across changes in the distribution of student quality, i.e. dosage of the treatment.

The results suggest a strong and robust negative causal effect of student quality on teacher turnover. This effect is very similar across all types of teachers, so adversely shocked schools are equally likely to lose their high and low quality teachers. In line with a simple theoretical framework, I find that the reshuffling induces teachers to switch jobs within teaching profession rather than to quit in favor of other occupations. This finding yields an important policy conclusion, as the negative selection into teaching does not seem to be reinforced by high-quality teachers leaving the profession in response to a change in the student’s composition. Furthermore, I show that teachers react mostly to direct measures of student quality (grades) rather than to characteristics that are correlated with student quality (immigrant status, parental income and paternal intellectual capacity). Finally, I do not find any effects of the reshuffling on individual teacher's earnings or school hiring policies.

Summarizing, the three papers yield the following empirical findings: first, there is a negative relationship between teacher turnover and monetary compensations, a positive relationship between private school ownership and teacher turnover, and no relationship on average between fraction of minority students and teacher turnover; second, there is a negative relationship between the probability of job separation and teacher quality and teachers rather change schools within the profession than leave for other occupations; and third, there is strong and robust negative effect of student quality on teacher turnover that is driven by student aptitude rather than other socioeconomic characteristics correlated with student aptitude.

The second part of the thesis studies family decisions in Poland. Together with Michał Myck we study phenomena that have been investigated previously in the literature, but because of limited data availability have never been causally addressed in Poland or in any other Central and Eastern Europe country. Until 1990 Polish economy was centrally planned and characterized, among other features, by support to large families as the fundaments of the economic success of the nation. For example, there was a large chain of public childcare facilities, families with children were more likely to obtain a flat from the state and women could easily come back to their workplace after the maternity leave. Since the end of the 1980s, however, the fertility rate declined rapidly, which is generally attributed to high economic instability and uncertainty in the late pre-transformation period. This trend continued in the 1990s and for the most part of the 2000s Poland experienced population decline. At the same time, labor market experienced rapid declines in female labor force participation. In such an institutional environment and using data from years 2003 to 2010 we attempt the answer causally two empirical questions: first, how child's gender affect within family arrangements; and second, how family size affects parental labor supply decisions.

Essay IV (joint with Michał Myck), "Who gets to look nice and who gets to play? Effects of child gender on household expenditure", examines the effects of child's gender on family stability, fertility, labour supply and expenditure decisions. Previous research addressed the first three topics, however, to our knowledge there is not a single study from developed countries that touches upon the last outcome (Dahl and Moretti, 2008; Ichino et al., 2011). The identification strategy is based on the assumption that gender of a first-born child is randomly assigned. This assumption is much more plausible in Poland, in comparison to other countries, due to cultural and economic factors. Similarly to the previous literature, we find that a first-born girl decreases the probability of living with a father and mother being even married. It also decreases the probability of having two or more children. Unlike previous research, we do not find any effects of first child gender on probability of divorce or maternal labor supply.

Given that when we control for family observable characteristics the differences in spending are unlikely to be driven by partnership stability, fertility or family labor supply decisions, we interpret our estimates on consumption as causal documentation of differential treatment of sons and daughters. For example, we find that parents spend on average 12.8% less on games and toys when their first-born child is a girl versus when it is a boy. At the same time, they also increase spending on clothing by 9.7%. Furthermore, we show that households with first-born girls spend significantly more on adult female clothing and decrease spending on adult male clothing. We interpret these findings as reflecting Polish parents' gender stereotypical preferences. We also find that households in Poland spend on average 6% less on kindergarten and pre-kindergarten care if their first-born child is a girl versus a boy. We believe that these effects could translate into important disadvantages in school age and adult life.

Essay V (joint with Michał Myck), "For some mothers more than others: how children matter for labour market outcomes when both fertility and female employment are low", addresses an old empirical question of how family size affects parental labor supply (Rosenzweig and Wolpin, 1980; Angrist and Evans, 1998). Poland is an interesting case-study here due to its simultaneously low fertility and low female employment rates. The identification strategy is based on instrumental variables, where we instrument for family size with twinning at first and second birth. This instrument under standard assumptions provides exogenous variation in family size as the mother is expecting a single offspring as a result of a pregnancy, but ends up having two children.

The results indicate that among families with at least one child, the total average causal effect of an additional child on mother's employment is -6.7 percentage points. However, we do not find any statistically significant effects of having additional children on female employment among families with two or more children. We also do not find any effects of family size on paternal labor supply decisions. In most cases, OLS estimates exaggerate the negative effects of children on maternal labor supply, which is a standard finding in the literature, yet in the heterogeneity analysis we demonstrate that for some groups the effect of omitted variables may actually be reversed. Thus, the OLS for highly educated and older mothers, and those who had their first child born after the age of 26 underestimates the negative causal effects of children. To our knowledge such a positive bias in OLS has not been documented in previous studies, and it points towards the importance of accounting for heterogeneous treatment effects. The analysis suggests that for some groups of women good labor market prospects may be key determinants of their fertility decisions, while for some other groups it might be the unobserved lower labor market attachment.

Summarizing, the two papers yield the following empirical findings: first, Polish parents tend to differentiate their child-related consumption depending on the gender of a first-born child and this calls for a further longitudinal investigation in order to understand if such a gender stereotypical allocation of resources has long-run consequences; second, these same parents also differentiate the consumption of educational inputs which might hurt the girls in a long-run; and third, family size reduces maternal labor supply in Poland, but this effect is very heterogeneous and calls for targeted policies rather than uniform interventions.

References

Abelson, M.A., and D.B. Baysinger (1984), "Optimal and dysfunctional turnover: Toward an organizational level model", *Academy of Management Review*, 9:2, 331-341.

Angrist, D.J., and W.N. Evans (1998), "Children and their parent's labor supply: evidence from exogenous variation in family size", *American Economic Review* 88:3, 450-477.

Antos R. J., and S. Rosen (1975), "Discrimination in the market for public school teachers", *Journal of Econometrics*, 3:2, 123 – 150.

Barbieri, G., C. Rossetti, and P. Sestito (2011), "The determinants of teacher mobility: Evidence using Italian teachers' transfer applications", *Economics of Education Review*, 30:6, 1430-1444.

Björklund, A., M.A. Clark, P-A. Edin, P. Fredriksson, and A. Krueger (2006), "The market comes to education in Sweden: An Evaluation of Sweden's surprising school reforms", Russel Sage Foundation, New York.

Cullen, B.J., B.A. Jacob, and S.D. Levitt (2006), "The effects of school choice on participants: Evidence from randomized lotteries", *Econometrica*, 74:5, 1191-1230.

Dahl, G., and E. Moretti (2008), "The demand for sons" *Review of Economic Studies*, 75:4, 1085-1120.

Dolton, P., and W. van der Klaauw (1995), "Leaving teaching in the UK: a duration analysis", *Economic Journal*, 105:429, 431–444.

Falch, T., and B. Strøm (2005), "Teacher turnover and non–pecuniary factors", *Economics of Education Review*, 24:6, 611–631.

Feng, L., D. Figlio, and T. Sass (2010), "School accountability and teacher mobility", NBER Working Paper 16 070.

Fredriksson, P., and B. Öckert (2007), "The supply of skills to the teacher profession", Mimeo Uppsala University.

Fredriksson, P., and B. Öckert (2008), "Resources and student achievement – evidence from a Swedish policy reform", *Scandinavian Journal of Economics*, 110:2, 277-296.

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

Greenberg D., and J. McCall (1974), "Teacher Mobility and allocation", *Journal of Human Resources*, 9:4, 480 – 502.

Hanushek, E.A., F.J. Kain, and G.S. Rivkin (2004), "Why public schools loose teachers?", *Journal of Human Resources*, 39:2, 326–354.

Hensvik, L., (2012), "Competition, wages and teacher sorting: four lessons from a voucher reform", *Economic Journal*, 122:561, 799-824.

Hsieh, C-T., and M. Urquiola (2006), “The effects of generalized school choice on achievement and stratification: Evidence from Chile’s voucher program”, *Journal of Public Economics*, 90:8-9, 1477-1503.
Ichino, A., E-A. Lindström, and E. Viviano (2011), “Hidden consequences of a first-born boy for mothers” IZA Discussion Papers, No. 5649.
Jackson, C.K., (2009), “Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation”, *Journal of Labor Economics*, 27:2, 213-256.
Lankford, H., S. Loeb, and J. Wyckoff (2002), “Teacher sorting and the plight of urban schools: the importance of alternative labor market opportunities and non – pecuniary variation”, *Educational Evaluation and Policy Analysis*, 24:1, 37-62.
Mont, D., and D. Rees (1996), “The influence of classroom characteristics on high school teacher turnover”, *Economic Inquiry*, 34:1, 152 – 167.
Rivkin, S.G., E.A. Hanushek, and J.F. Kain (2005), “Teachers, schools, and academic achievement”, *Econometrica*, 73:2, 417–458.
Ronfeldt, N., S. Loeb, and J. Wyckoff (2013), “How teacher turnover harms student achievement”, *American Educational Research Journal*, 50:1, 4-36.
Rosenzweig, R.M., and K.I. Wolpin (1980), “Testing the quantity-quality fertility model: the use of twins as a natural experiment”, *Econometrica,* 48:1, 227-240.
Skolverket (2009), “Resursfördelning utifrån förutsättningar och behov”, Skolverket Rapport 330.
Söderström, M., and R. Uusitalo (2010), “School Choice and Segregation: Evidence from an Admission Reform”, *Scandinavian Journal of Economics*, 112:1, 55-76.

Essay 1

The determinants of teacher mobility in Sweden

1 Introduction

The effectiveness of schools is fundamentally important for future labour productivity and economic growth. It has been established that teachers are one of the most important factors in the education production function (Rivkin et al., 2005); however, their effectiveness depends on the quality of the match between a school and a teacher (Jackson, 2013), and teachers may leave schools when the match quality is low. It is not clear a priori whether policy makers would want to minimize the teacher turnover. On the one hand, they may want to reduce turnover if its high rates lead to lower student achievement. On the other hand, they may want to increase teacher turnover in order to improve teacher-school-student match quality. The few empirical studies do not help to resolve this issue (Guin, 2004). Correlational (Boyd et al., 2005) and more recent causal (Ronfeldt et al., 2013) studies reveal a negative relationship between teacher turnover and student achievement. At the same time, however, the organizational management literature suggests a positive relationship between personnel rotation and infusion of new ideas into organizations (Abelson and Baysinger, 1994). Finally, there is the evidence that more effective teachers are at least as likely and sometimes even more likely to stay in schools than their less effective peers (Hanushek and Rivkin, 2010; Boyd et al., 2011).

Given the importance of schooling, teacher turnover has attracted much attention in the last decade. This is likely related to many recent educational policies, such as alterations in teacher compensations, introduction of free schools, or broadening school choice, that affect both students and the labor market for teachers.[1] Furthermore, out-of-teaching mobility has been seen as a potential explanation of declining teacher quality (Fredriksson and Öckert, 2007; Grönqvist and Vlachos, 2008). Most of the articles studying the determinants of teacher mobility have been correlational, but some are causal. Studies show that teachers are generally discouraged by high fractions of poor, minority, and low-achieving students (Hanushek et al., 2004; Falch and Strøm, 2005; Scafidi et al., 2007; Jackson, 2009; Barbieri et al., 2011; Bonhomme et al., 2011). Furthermore, teachers are responsive to even small variation in wages (Figlio, 1997; Figlio, 2002; Feng 2011; Falch, 2011); however, this relationship failed to be robustly confirmed in a large cross-sectional data (Hanushek et al., 2004). The competition between publicly and privately run schools also affects teacher turnover (Jackson, 2012; Hensvik, 2012). Finally, it is important to understand the differences between the wages offered to teachers in education and in other sectors of the economy (Dolton and van der Klaauw, 1995, 1999; Brewer, 1996; Dolton and Marcenaro-Gutierrez, 2011).

[1] Free schools and school choice is studied by Cullen et al. (2005; 2006), Hsieh and Urquiola (2006), Jackson (2012) and Hensvik (2012) among others. Teacher compensations are studied by Figlio (1997), Figlio (2002), Lavy (2009), Falch (2011) and Fryer (2011) among others.

In this paper, I make use of the Swedish institutional setup and high quality administrative data to shed more light on the aforementioned associations. I present evidence on the relationship between teacher turnover and teacher compensations, employment in privately owned school, exposure to minorities, and employment in upper secondary school in a large repeated cross-section of lower and upper secondary schools for years 1996/1997 to 2006/2007. The correlations for compensation and private ownership could help in understanding teachers' decisions in an environment with much higher variation in compensations than in previous studies and with rapidly growing private sector. The relationship to educational system level should be of interest as vast majority of the aforesaid studies focus on relatively younger kids attending primary or middle school and not on teenagers whose school behavior might be more troublesome for teachers.

I find that teacher turnover correlates negatively with teacher monetary compensations but it does not correlate significantly with the fraction of minorities at school on average. More importantly, I document substantial heterogeneity in this association and I show that it exists only for upper secondary and private school. Furthermore, I demonstrate that privately owned schools experience higher teacher turnover rates and that this correlation is weaker for upper secondary schools. Finally, the relationship between earnings and teacher turnover becomes weaker when I add control variables, and thus, a somewhat speculative interpretation of this negative result could be that it may be possible to influence teacher's mobility decisions through changes in their monetary compensations.

The paper is organized as follows: section two briefly presents the institutional background, section three presents econometric modeling and data sources, section four presents descriptive evidence, section five contains the main results, section six includes heterogeneity analyses, and section seven concludes.

2 Swedish schooling system and institutions

The Swedish schooling system starts with preschool and continues with nine years of comprehensive school. Lower secondary school covers the grades 7 to 9. The academic grades in 9th grade determine student's chances to advance to upper secondary school. Swedish municipalities are obliged by law to provide upper secondary schooling to all students who successfully complete compulsory education. Upper secondary school consists of different programs (subject oriented tracks), lasts three years, and provides eligibility for post-secondary education.

Private schooling is growing in Sweden and is encouraged by the government. In 1992, Sweden introduced a school voucher reform that allowed both non-profit and for-profit independent schools. The municipality is obliged to pay the independent schools for each student they can attract, with an amount corresponding roughly to the average per student cost in the public schools.[2] Since the reform the fraction of private schools has risen, in particular at the upper secondary level. In the 2005/2006 school year there were 220 private upper secondary schools, which constituted 33.1% of all upper secondary schools in Sweden, a rise from 8.1% in 1996/1997. At the same time, the number of private lower secondary schools constituted only 15.8% of all schools at this level starting from 3.2% in 1996/1997.[3]

The teaching profession in Sweden is regulated with different required qualifications depending on the subject taught and the type of school. Teaching at the secondary school level requires completing special coursework beyond what is required from a compulsory school teacher. Individuals from other professions who want to become teachers need to supplement their professional degrees with a minimum of 1.5 years of preparation in pedagogy, didactics, and teaching practice.

Municipalities are the primary employers of teachers in Sweden, and thus, handle the responsibility of recruiting them.[4] In practice, however, the decisions regarding recruitment, selection, and employment of a teacher are made at the school level by a principal. Finally, teacher wages are determined at local level through individual bargaining between teacher and principal given the collective bargaining outcome set at the national level.[5]

One can distinguish several important underlying decisions related to job mobility in this summary of the institutional setting. Every year an individual teacher considers whether to leave their current school appointment or not. Then, a school principal can either let the teacher leave or re-employ them under the new conditions. If the teacher leaves, they can either seek employment at a different school or find a job in a different occupation. In the former case they negotiate a new contract with a new school principal. In both the case of re-employment and at a new hire the teacher and school determine salary in an individual bargaining. The decision to re-employ teachers seems to be important given that 21% of teachers are in temporary positions. Typically, teachers in temporary positions are employed under fixed-term contracts and are exposed to higher probabilities of job separa-

[2] An independent school receives around 85-95% of the average per student cost in public schools though amounts vary year to year. Some municipalities also have a socioeconomic gradient for the school voucher. The private schooling was effectively introduced at lower secondary level in 1992, and at upper secondary level in 1994 (Böhlmark and Lindahl, 2007, 2008).

[3] This information is based on registry data.

[4] For more information on the reform that shifted responsibility for schooling from the central government to municipalities see: Fredriksson and Öckert (2008). There is still a small fraction of schools run by county or state, however, those employ around 1% of all the teachers between 1996/1997 and 2005/2006. Those schools are excluded from the analysis since they have different sources of funding and their role is diminishing.

[5] Individualized pay was introduced in 1996 and is discussed in detail by Hensvik (2012) and in survey by Lindholm (2006).

tion. In the analysis I consider three types of separations: total turnover, within-teaching turnover, and out-of-teaching turnover.

3 Data sources and econometric modeling

This paper utilizes multiple Swedish population-wide registries. The main data source is the teacher registry that covers all teachers employed in Swedish schools in years 1996/1997 to 2006/2007. It contains information on teachers' education, specialization, experience, certification, place of work, type of contract (permanent vs. temporary), and workload. To these data I have matched background information on age, gender, immigration histories, education, employment, and income for all teachers in the registries. I use pupil registries for lower and upper secondary schools to obtain information on students in a given school. These allow linking children to their parents to schools, as well as obtaining the average percentiled GPA of the students. Administrative records on earnings and wages provide information on teachers' monetary compensations.[6] The details of the sample construction are discussed in the appendix.

This paper focuses on the relationship between pecuniary and non-pecuniary characteristics of jobs and teachers' decisions to stay at or leave their current employment. The main analysis is done using a series of binary choice models that attempt to capture the manifestation of teachers' job preferences with respect to how they value particular characteristics of the working environment. The paper is only descriptive, so I am not able to identify teacher's preferences in an econometric sense. Nonetheless, it should be intuitive that leaving employment j in favor of an alternative opportunity k is related to how teacher values employment j in comparison to k. Thus, I specify the following linear model. The dependent variable equals unity if a teacher leaves their current employment year to year, and such a decision is regressed on teacher's working environment and their own characteristics. These binary models show whether teachers who remain in their appointments have, on average, different characteristics than those who leave their jobs.

From a policy perspective, one should also investigate whether the factors associated with mobility differ by type of school. The uniqueness of Swedish system and the completeness of the data allow me to study differences by level of schooling (lower secondary vs. upper secondary) and type of ownership (private vs. public). Using the main specification, I also run separate regressions depending on the teacher's destination. In particular, I specify two distinct variables for transition: switching schools within the teaching

[6] Monthly wages are available for all public school teachers, and a sample of private school teachers.

profession and leaving teaching in favor of a different occupation.[7] This analysis could be of interest to policy makers, as loosing pedagogues in favor of other sectors of the economy may lead to worsening productivity of the educational system as a whole.

In order to simplify the interpretation of the results, the estimation strategy is based on the least squares using linear probability model.[8] The following econometric model is estimated:

$$y_{ijt} = \alpha_0 + \alpha_1 W_{ijt} + \alpha_2 X_{jt} + \alpha_3 P_{ijt} + \delta t \cdot c + \varepsilon_{ijt} \tag{1}$$

where y_{ijt} is equal to unity if teacher i leaves the current employer j at the period following t, W_{ijt} is the earnings or wages of teacher i at school j and time t, X_{ijt} is a vector of observable school characteristics at school j at time t (share of minorities, student quality, mean parental income, student's gender composition, school resources, and school size polynomial), P_{ijt} is a vector of personal characteristics of teacher i at school j and time t (age polynomial, gender, origin, marital status, education, specialization, type of employment, type of school, and workload) and ε_{ijt} is an error term that represents unobserved characteristics, which is clustered at school level. The clustering follows the idea that in a perfect experiment teachers would be randomly assigned to different schools and their mobility decisions would be observed conditional on school characteristics. Since the turnover variation occurs at the school level and I have an unbalanced panel of all lower and upper secondary schools in Sweden the errors should be clustered at the school level. The vector of δs captures county-by-year fixed effects.

Ideally, an analysis of teacher sorting between schools should be based on a complete characterization of the individual decision of occupational choice, the initial matching process with school, and the transition of teachers between schools and out of teaching. In this paper, I focus on the relationship between variety of characteristics of jobs and teachers' decisions whether to stay at their current appointment or not. Given the nature of the study, I am interested in the coefficients on earnings, school types and fraction of minorities; however, we can consider them jointly conditional on all the other control variables or separately. The analysis provides evidence about the univariate correlations of the characteristics and mobility as well as

[7] Switching to primary education or adult education is treated as a school-to-school mobility. Switching to kindergarten, pre-K, or university education is treated as leaving the profession. The results are robust to various definitions of school-to-school mobility and quits, and are available upon request.
[8] This method yields very similar estimates to the non–linear models. The regressions using logit and multinomial logit models with marginal effects evaluated at means are available upon request. The majority of correlations between explanatory variables are below 0.1 and the correlogram is available upon request.

multivariate correlations conditional on all other variables. The latter modeling is the preferred specification. The results survive in the univariate regressions.[9]

4 Descriptive evidence

The total turnover is split into turnover within teaching (school-to-school transitions) and quits (leaving teaching for a different occupation). Figures 1 to 3 provide descriptive evidence of turnover patterns by several characteristics of interest. The school-to-school turnover rate increases over 11 years from 4.3% to 4.9% (Figure 1). The quit rate increases from 6.3% in 1996 to 7.9% in 2005. This adds up to a total turnover increasing from 10.6% in 1996 to 12.8% in 2005. Over the same period of time the number of teachers increases by 33%.

Figure 1. Turnover and number of teachers over time.

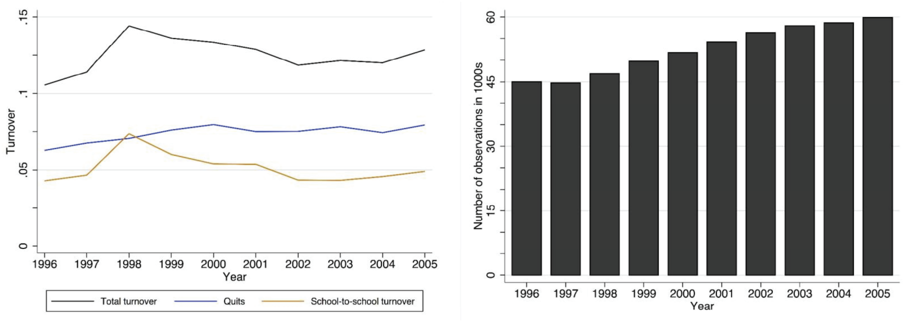


Figure 2. Turnover by minority enrollment and earnings.

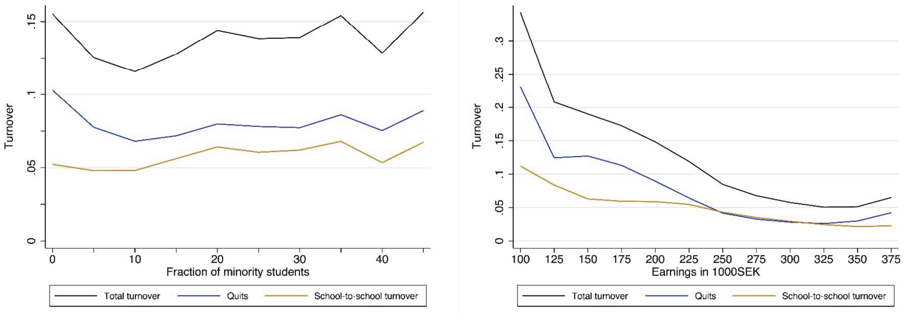


Figure 2 depicts the relationship between teacher turnover and share of minorities in the left panel (Falch and Strøm, 2005) as well as monetary com-

[9] The univariate regressions can be found in table A1 in the appendix.

pensations in the right panel (Falch, 2011). The school-to-school transitions are largely stable across schools with different number of minority students; however, the share of teachers leaving teaching in favor of other occupations is the largest in schools with zero minority enrolment. This initial decrease in the quit rate is in opposition to findings in Falch and Strøm (2005). Moreover, there is a negative relationship between earnings and turnover – the more teachers earn the lower the turnover due to both the school-to-school transitions and to quits. In fact, the two lines converge at about 230 000 SEK yearly. It is worth noting, that teachers who earn less than 125 000 SEK a year are likely to be temporarily employed, and thus I control for the type of employment in the regressions.

Figure 3. Turnover by school type.

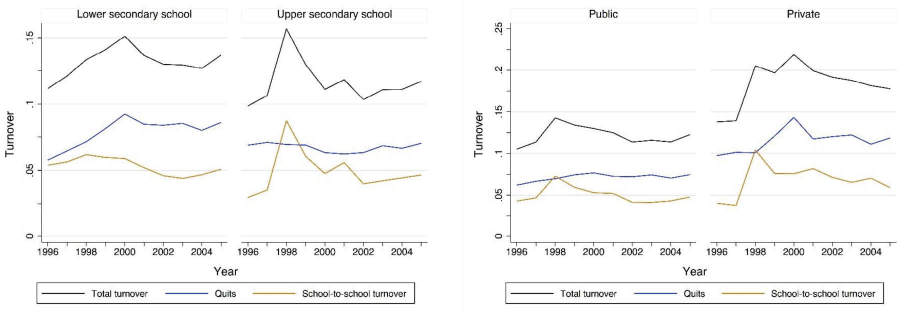


Figure 3 depicts the turnover rates separately for lower, upper, public and private institutions. There is a large jump in upper secondary school within-teaching turnover in 1998.[10] Lower and upper secondary schools teachers behave differently across years. In the upper secondary schools both school-to-school mobility and quits seem to be relatively flat whereas in lower secondary schools the quits increase over time from 5.8% in 1996 to 8.6% in 2005, while the school-to-school mobility is flat. Turnover levels are higher for privately owned institutions than for public schools.

Table 1 presents descriptive statistics of variables used in the econometric analysis. Panel A presents three turnover measures, panel B presents pecuniary and personal characteristics, and panel C presents average school-level characteristics. Total turnover rate, is at 12.5%, which is lower than the overall turnover rate across all the occupations in Sweden (Edin et al., 2009). While studies based on the US registry data have higher school-to-school mobility than quit rates, Sweden has higher percentage of turnover due to

[10] The jump is due to the adult education expansion reform proposed in the mid-1990s. If job-to-job mobility is of interest then this variation should be kept as teachers indeed change their jobs voluntarily. They simply prefer moving to adult education over staying in their current employment when such an opportunity occurs. If I exclude these transitions to adult education then the jump vanishes. This graph is available upon request.

leaving the profession than switching between schools. In Swedish schools 56% of teachers are women, 6.8% come from non-Nordic countries, 20.8% are employed on temporary contracts, and their average yearly earnings equals to 221 866 SEK. There is 15.6% science, 13.8% vocational and 6.6% remedial education teachers and 67% of teachers are university graduates.[11] The fraction of teachers employed in private schools during the study period rose from around 2% in 1996 to 10.5% in 2005. Panel C shows the student-teacher ratio in full-time equivalence, which is a proxy for school resources, is 11.5% and the average number of pupils is 574.[12] There is on average 8.3% non-Nordic immigrants in Swedish schools. This number is larger than the one reported for Norway (Falch and Strøm, 2005).

Table 1. Descriptive statistics.

Variable	Mean	Standard deviation
Panel A: Teacher turnover		
Total mobility	0.125	(0.331)
Within profession mobility	0.051	(0.220)
Out of profession mobility	0.074	(0.262)
Panel B: Personal and pecuniary characteristics		
Log yearly earnings (1000SEK)	5.290	(0.586)
Log monthly wages*	9.952	(0.161)
Upper secondary	0.437	(0.496)
Private	0.056	(0.230)
Age	44.115	(9.668)
Female	0.562	(0.496)
Foreign	0.068	(0.251)
Married	0.572	(0.495)
University diploma	0.674	(0.469)
Science	0.156	(0.363)
Vocational	0.138	(0.345)
Remedial	0.066	(0.249)
Temporary	0.208	(0.406)
Workload	86.488	(23.273)
Panel C: School characteristics		
Share of foreign students	0.083	(0.086)
Student's percentiled GPA	48.175	(6.708)
Students' parents income in 1000SEK	380.201	(96.397)
Share of girls	0.482	(0.100)
Student-teacher ratio in full time equivalence	11.511	(3.241)
Number of students/100	5.739	(4.574)
N	525076	

Note: mean values, standard errors in parentheses.
*N = 475 505.

[11] Remedial education teacher (Speciallarare) works with students in need of special assistance concerning learning and development. Special teacher training is a postgraduate education in the regular teacher training and includes 90 credits. Special education teachers focus on either language or mathematics. A university graduate is defined as an individual graduating three, four, or five year long university (hogskoleutbildning) education or individual with a research degree. Note that other forms of post-secondary education (eftergymnasial) education are not treated as university graduates.

[12] Number of students in lower-secondary school is measured as the sum of pupils attending grades 7 to 9 and it is provided in compulsory school registry by Statistics Sweden. Number of students in upper secondary school is measured based on the registry of students enrolled in grades 1 to 3 in upper secondary schools.

5 Main results

The estimates presented in this section correspond to the model outlined in section 3. I estimate a binary linear regression model with county-by-year fixed effects and the dependent variable equals to unity if the teacher leaves a particular school between year t and year t+1, and zero otherwise.[13] Column (1) shows estimates on types of schools controlling only for personal characteristics. Column (2) adds average school-level characteristics to the estimates from column (1) and additionally displays coefficient on fraction of immigrants. Column (3) adds annual earnings to the specification from column (2). This allows me to understand if the differences in mobility by type of school and school characteristics are driven by differences in earnings. Column (4) estimates column (3) on the sample of public school teachers, which is then used in column (5), where I substitute the log yearly earnings with the log monthly salary. This exercise is performed to investigate how covariates in model from column (3) change when the sample is restricted to public school teachers for whom the monthly wage data are available. Column (3) which includes all personal, pecuniary, and school-level characteristics is the preferred specification. In addition to the main coefficients of interest in this paper the tables also report some other coefficients that might be of interest to the readers (gender, temporary employment, foreign and science teacher indicators, and average school-level student GPA).

The results in columns (1), (2) and (3) suggest that private schools experience higher teacher turnover. Working in private school is associated with 1.5 to 2.6 percentage points (pp) higher turnover depending on the specification. Teaching at upper secondary school has a negative association with turnover when I do not control for school characteristics, but a positive association in the sample of public schools with all of the controls.

Column (3), where earnings are added, suggests a negative relationship between monetary compensations and the probability that a teacher is going to leave their employment in the following year. The significant and negative estimate of 6.4 pp indicates that principals may have a scope for changing the turnover through manipulation of monetary compensations; however, the limitations of descriptive methods mean that there well might be other explanations to the observed pattern. The results on earnings combined with the Swedish institutional flexibility in pay negotiations are in line with the causal findings from Falch (2011) that even small changes in teacher wages can result in lower turnover rates. My estimates are smaller than those in Hanushek et al. (2004) – another correlational study looking at teachers' compensations. However, their measure of monetary compensation is differ-

[13] Specifications with only year, or only county, or only year and county, or using municipality instead of county fixed effects have also been estimated and yield similar results. Including school fixed effects removes some of the variation that is of interest in the heterogeneity analyses in this paper.

ent than mine. In Hanushek et al. (2004) virtually all the salary associations vanish when school district fixed effects are applied, whereas, here the coefficients on both log earnings and log monthly salaries are stable qualitatively and quantitatively across various fixed effects specifications.

Table 2. Baseline estimation results. The dependent variable is equal to unity if the teacher changes job.

VARIABLES	(1) Mobility	(2) Mobility	(3) Mobility	(4) Mobility	(5) Mobility
Log yearly earnings (1000SEK)			-0.064***	-0.069***	
			(0.002)	(0.002)	
Log monthly salary					-0.153***
					(0.008)
Upper-secondary school	-0.009***	-0.003	0.002	0.003	0.007**
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Private school	0.019***	0.026***	0.015**		
	(0.005)	(0.006)	(0.006)		
Share of immigrant students		0.009	0.015	-0.004	-0.005
		(0.016)	(0.016)	(0.016)	(0.016)
GPA		-0.001***	-0.001***	-0.001**	-0.001**
		(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.005***	-0.004***	-0.009***	-0.008***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Foreign born	0.013***	0.012***	0.005	0.004	0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Science	0.009***	0.010***	0.011***	0.009***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Temporarily employed	0.224***	0.224***	0.195***	0.197***	0.211***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
R-squared	0.129	0.130	0.138	0.136	0.130
Observations	525,076	525,076	525,076	475,505	475,505

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). All regressions include county-by-year fixed effects. In addition to the displayed variables in column (1) I control for teacher's age, marital status, university education, vocational and special education indicator variables and workload. In column (2) on top of column (1) I control for student-teacher ratio in full time equivalence, share of female students, mean parental income, second order polynomial in school size and indicator for schools with less than 100 students. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7th that teachers work with.

The coefficients on log yearly earnings in columns (3) and (4) range from -6.9 to -6.4 pp, and the coefficient on log monthly wages in column (5) among public school teachers is -15.3 pp. This difference in the size of earnings vs. wage coefficients in columns (4) and (5) can be attributed to different definitions of both monetary compensation measures. At the same time, since they give the same results qualitatively and monthly wages are not available for all teachers further analyses are conducted on the full sample using log yearly earnings.[14]

The additional covariates displayed in table 2 are: gender, immigrant status, science teacher, and temporary employment indicators. The gender indicator suggests that female teachers experience lower turnover rates. Specializing in science and being employed on temporary contract are associated with higher turnover rates. There is a positive relationship between being foreign born and mobility when I do not control for monetary compensations. Finally, student quality is negatively associated with teacher turnover.

[14] Estimates for public school teachers and monthly wages are available upon request. The main findings remain unchanged.

The results in table 2 show no relationship between the share of minorities and teacher turnover, and this is in contrast to other research from the US (Hanushek et al., 2004), Norway (Falch and Strøm, 2005), Italy (Barbieri et al., 2011) or Netherlands (Bonhomme et al., 2011). All of the coefficients in columns (2) to (5) are statistically insignificant and substantively small. I further explore this relationship in table 3 by grouping minorities into students coming from European and non-European countries (panel A) and interacting the share of minority students with an immigrant teacher dummy variable (panel B). In panel A there is no indication for any heterogeneity in the association depending on the geographical and cultural origin of the immigrants. In panel B there is suggestive evidence that immigrant teachers cluster with immigrant students, which is in line with prior research (Hanushek et al., 2004; Jackson, 2009). Table 3 also suggests a positive correlation between an indicator for a foreign born teacher and their mobility, which may reflect either lower quality of matches between immigrant teachers and schools or generally increased occupational mobility among immigrants (Green, 1999). At the same time, the coefficient on the level of minority students at school is consistently small and insignificant.

Table 3. Minorities at school. The dependent variable is equal to unity if the teacher changes job.

VARIABLES	(1) Mobility Full sample
Panel A: Split analysis	
Share of European students (2.6%)	0.045
	(0.036)
Share of other immigrant students (5.7%)	-0.000
	(0.022)
R-squared	0.138
Panel B: Interaction analysis	
Immigrant teacher	0.012***
	(0.004)
Share of immigrant students	0.024
	(0.016)
Share of immigrant students*Immigrant teacher	-0.070**
	(0.028)
R-squared	0.138
Observations	525,076

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). Estimates based on specification from column (3) in table 2. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7th that teachers work with.

Univariate regressions (table A1 in the appendix) shed more light on the relative contributions of included covariates. Type of contract is the factor that explains the most of the variation in teacher turnover and monetary compensation (earnings or wages) is the second.[15] Considering the variables grouped into personal, pecuniary, and school-level characteristics the amount of explained variation in total turnover is the following: personal

[15] When all control variables are analyzed then a factor that explains the most of the variation is the type of employment, followed by pecuniary characteristics and workload. The univariate regressions for all covariates used in the analysis are available upon request.

(R^2=0.13), pecuniary (R^2=0.08) and school-level (R^2=0.01). When comparing just the monetary vs. school-level characteristics conditional on personal observables, the former one (R^2=0.14) explains slightly more variation in the total turnover than the latter one (R^2=0.13). Thus, it is not trivial to quantitatively gauge the relative importance of either of these groups for teacher turnover. However, it seems that teachers in Sweden are less sensitive to school characteristics than teachers in other countries.

Finally, I can only observe mobility if teachers leave their school between one year and the next, but I do not know if this mobility is voluntary or not. In particular, there can be reshuffling of teachers between schools in municipalities due to the fact that employment protection is based on an employment in municipality and not at the school (this does not apply to privately owned institutions). It could also be the case that if one school has an opening for a teacher and there are other schools in the same municipality laying off teachers, there might be bargaining and reshuffling of teachers within the municipality. To address this issue I restrict the analysis to the sample of municipalities that never experienced reductions in teacher stock by more than 5% over the studied period.

Table A2 in the appendix replicates table 2 using this restricted sample. The sample size is reduced to a quarter of the full sample size, however the coefficients on earnings and wages remain negative and significant, and are roughly of the same magnitude, and the estimates on the minority enrollment remain insignificant and cannot be statistically distinguished from the ones presented in table 2. The associations between school ownership and teacher turnover are now insignificant and smaller than in table 2 though they remain positive. Overall, these estimates indicate that the differences in mobility should not be driven by selective lay-offs when schools are down-sizing.

A final question is whether it is reasonable to pool 11 years of data in one equation (Falch and Strøm, 2005). It might be questionable, as teachers who come into the sample in the later years have a smaller window in which they can make mobility decisions than the more experienced teachers. As a further robustness check I estimate columns (1) to (3) from table 2 using only teachers that were present in the sample in the first year of the study. The results are reported in table A3 in the appendix. The sample size is reduced by approximately 48%, however the results do not change substantively. The coefficient on earnings decreases while the ones on school ownership increase. Furthermore, the coefficients on upper secondary school become insignificant in column (1) and turn positive and significant in columns (2) and (3). Similarly to all previous results, I do not find a statistically significant relationship between minority enrollment and teacher turnover.

6 Heterogeneity analysis

The results presented so far suggest that schools in Sweden experience higher teacher turnover rates in privately owned institutions, have a negative relationship between teacher compensations and turnover, and do not have an association between minority enrolment and teacher mobility decisions. In table 4, I investigate how these characteristics differ by level of school and by school ownership. In table 5, I further document how the estimates differ depending on whether a teacher transfers to another school or transitions out-of-teaching.

Table 4. Heterogeneity analysis by school types. The dependent variable is equal to unity if the teacher moves.

VARIABLES	(1) Lower secondary school	(2) Upper secondary school	(3) Private school	(4) Public school
Log yearly earnings (1000SEK)	-0.073***	-0.056***	-0.051***	-0.067***
	(0.003)	(0.003)	(0.005)	(0.002)
Upper-secondary school			-0.026**	0.002
			(0.012)	(0.003)
Private school	0.022***	0.014		
	(0.008)	(0.009)		
Share of immigrant students	-0.013	0.110***	0.131***	-0.008
	(0.017)	(0.039)	(0.049)	(0.017)
GPA	-0.001**	-0.001***	-0.001**	-0.001**
	(0.000)	(0.000)	(0.001)	(0.000)
Female	-0.014***	-0.003	-0.010*	-0.009***
	(0.001)	(0.002)	(0.005)	(0.001)
Foreign born	-0.003	0.013***	0.013	0.002
	(0.004)	(0.004)	(0.010)	(0.003)
Science	0.007***	0.013***	0.025***	0.010***
	(0.002)	(0.002)	(0.006)	(0.001)
Temporarily employed	0.219***	0.164***	0.139***	0.199***
	(0.004)	(0.004)	(0.009)	(0.003)
R-squared	0.149	0.132	0.100	0.142
Observations	295,454	229,622	29,520	495,556

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). Estimates based on specification from column (3) in table 2. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7th that teachers work with.

Columns (1) and (2) in table 4 present estimates for lower and upper secondary schools, respectively. The association between monetary compensation and turnover is significantly larger (p-value: 0.000) in lower secondary schools which suggests that the cost of retaining a teacher through changes in earnings could be lower in these schools. There is also a positive and significant correlation between school ownership and turnover at the lower level of schooling. Although this relationship is not significant in upper secondary schools, I cannot rule out the equality of the coefficients in both schools (p-value: 0.483). Finally, column (2) points towards a strong correlation between the share of minorities and teacher turnover in upper secondary schools in Sweden. Prior research has focused mostly on relatively younger kids, enrolled in elementary or lower secondary education, and found significant results for minority enrollment (Hanushek et al., 2004; Falch and Strøm, 2005; Scafidi et al., 2007; Bonhomme et al., 2011). This is not evident in the lower-secondary schools in Sweden (column (1)), however, the

10.9 pp estimate for upper secondary schools is similar in size to Hanushek et al. (2004) elementary schools' results. Given that upper secondary school covers ages when pupils go through adolescence, which is often strongly connected to increased disruptive behavior, then the positive correlations found for younger children in Netherlands, Norway, and the US may be even larger in the upper secondary schools in these countries. Interestingly, Barbieri et al. (2011) find similar coefficients on fraction of minorities for primary, lower secondary, and upper secondary school teachers using Italian data from the mid-2000s. However, instead of turnover rates they use applications for transfers.

There is a positive relationship between the share of minorities and teacher turnover in privately owned institutions, though not in public schools (columns (3) and (4); p-value: 0.007). At the same time, however, there is a negative association between teaching in upper secondary private school and individual mobility, which can explain why the coefficient on minority enrollment is larger in private schools given the findings from columns (1) and (2) that point towards relationship between minorities and turnover only in upper-secondary schools. Finally, the estimate on earnings is significantly larger (p-value: 0.005) for public schools. This suggests that the cost of retaining a teacher by increasing earnings could be lower in public schools than in private schools.

Table 5. Heterogeneity analysis by different destinations.

VARIABLES	(1) Within teaching mobility	(2) Out-of-teaching mobility
Log yearly earnings (1000SEK)	-0.009***	-0.055***
	(0.001)	(0.002)
Upper-secondary school	0.005**	-0.003*
	(0.002)	(0.002)
Private school	0.003	0.012***
	(0.004)	(0.004)
Share of immigrant students	0.008	0.006
	(0.011)	(0.009)
GPA	-0.001***	-0.000*
	(0.000)	(0.000)
Female	-0.002***	-0.006***
	(0.001)	(0.001)
Foreign born	0.009***	-0.004
	(0.003)	(0.003)
Science	0.007***	0.003***
	(0.001)	(0.001)
Temporarily employed	0.057***	0.138***
	(0.001)	(0.002)
R-squared	0.032	0.115
Observations	525,076	525,076

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). Estimates based on specification from column (3) in table 2. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7th that teachers work with.

The models used so far pool all destinations of teachers leaving the school together, however, there is research indicating that the correlations with teacher characteristics differ depending on the destination (Lankford et al., 2002). To investigate whether the relationship between teacher quality and teacher turnover depends on destination, in table 5, I estimate the baseline specification from column (3) in table 2 for mobility within teaching profession and for mobility out-of-teaching.

The estimates on earnings are negative for both school-to-school and out-of-teaching mobility, however, the latter one is significantly more negative. This suggests that if indeed principals can retain teachers by increasing their compensations, then it is relatively cheaper to encourage teachers to stay in the profession rather than to stay with their current school. Furthermore, the estimated relationship between upper secondary school indicator and school-to-school transitions is positive while it is negative in the case of out-of-teaching transitions. This indicates that upper secondary school teachers are more mobile within teaching but they are less likely to leave the profession for an alternative occupation. I also find positive and significant association between school ownership and leaving teaching. This suggests that private schools are more likely than public schools to lose teachers in favor of alternative jobs. Finally, in neither the case of within teaching nor out-of-teaching transitions I find statistically significant relationship between minority enrollment and turnover. This is in stark contrast to Hanushek et al. (2004) and Falch and Strøm (2005), whose results point towards quitting the profession rather than changing schools within the same geographical unit or occupation.

7 Conclusions

The contemporary literature on teacher mobility lacked a detailed study using high quality data in an environment for which the economists usually argue for i.e., with individual-level variation in wages and relatively large and growing private sector (Björklund et al., 2006). Furthermore, most of the aforementioned studies focus on rather younger kids attending primary and middle school and we know relatively little about the teacher turnover in high schools. This paper attempts to fill in these gaps in the literature on teacher turnover using unusually rich data on teachers from Swedish lower and upper secondary schools covering years 1996/1997 to 2006/2007.

The results indicate that, in Sweden unlike in US, Italy, the Netherlands, and Norway, schools with higher shares of minorities on average do not seem to experience higher turnover rates. At the same time, I document substantial heterogeneity in this association. In particular, I show that this relationship exists for upper secondary and private schools and is roughly of the same magnitude as the one documented for lower levels of schooling in Hanushek et al. (2004). If share of minorities at school is associated with the disruptive behavior or not-fitting-in and these behavioral problems grow in a teenagehood, then my results suggest that the turnover estimates in US high schools might actually be even higher. I also find support for the hypothesis that privately owned institutions experience higher teacher turnover and that this correlation is smaller for upper secondary schools. On the other hand, I do not find any support for the fact that turnover differs by level of schooling. The average differences in turnover in lower vs. upper secondary schools and small and insignificant. Finally, a somewhat speculative interpretation of the negative results found for earnings and wages is that it may be possible to influence teacher's mobility decision through changes in their monetary compensations.

References

Abelson, M.A., and D.B. Baysinger (1984), "Optimal and dysfunctional turnover: Toward an organizational level model", *Academy of Management Review*, 9:2, 331-341.

Barbieri, G., C. Rossetti, and P. Sestito (2011), "The determinants of teacher mobility: Evidence using Italian teachers' transfer applications", *Economics of Education Review*, 30:6, 1430-1444.

Björklund, A., M.A. Clark, P-A. Edin, P. Fredriksson, and A. Krueger (2006), "The market comes to education in Sweden: An Evaluation of Sweden's surprising school reforms", Russel Sage Foundation, New York.

Bonhomme, S., G. Jolivet, and E. Leuven (2011), "Estimation of preferences for job attributes using job–to–job mobility: teacher turnover and school characteristics", mimeo CREST.

Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff (2005), "Explaining the short careers of high – achieving teachers in school with low performing students", *American Economic Review*, 95:2, 166–171.

Boyd, D., H. Lankford, S. Loeb, M. Ronfeldt and J. Wyckoff (2011), "The role of teacher quality in retention and hiring: Using applications-to-transfer to uncover preferences of teachers and schools", *Journal of Policy and Management*, 30:1, 88-110.

Böhlmark, A., and M. Lindahl (2007), "The impact of school choice on pupil achievement, segregation and costs: Swedish evidence", IZA Working Paper 2786.

Böhlmark, A., and M. Lindahl (2008), "Does school privatization improve educational achievement? Evidence from Sweden's voucher reform", IZA Working Paper 3691.

Brewer, J.D., (1996), "Career paths and quit decisions: evidence from teaching", *Journal of Labor Economics*, 14:2, 313–339.

Cullen, B.J, B.A. Jacob, and S.D. Levitt (2005), "The impact of school choice on student outcomes: An analysis of the Chicago Public Schools", *Journal of Public Economics*, 89:5-6, 729-760.

Cullen, B.J., B.A. Jacob, and S.D. Levitt (2006), "The effects of school choice on participants: Evidence from randomized lotteries", *Econometrica*, 74:5, 1191-1230.

Dolton, P., and W. van der Klaauw (1995), "Leaving teaching in the UK: a duration analysis", *Economic Journal*, 105:429, 431–444.

Dolton, P., and W. van der Klaauw (1999), "The turnover of teachers: a competing risk explanation", *Review of Economics and Statistics*, 81:3, 543–552.

Dolton, P., and O.D. Marcenaro-Gutierrez (2011), "If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance", *Economic Policy*, 26:65, 5-55.

Edin, P-A., B. Holmlund, and O. Nordström Skans (2009), "Wage Dispersion Between and Within Plants: Sweden 1985-2000", in E. Lazear and K. Shaw (eds.) *The Structure of Wages: an International Comparison*, University of Chicago Press.

Falch, T., and B. Strøm (2005), "Teacher turnover and non–pecuniary factors", *Economics of Education Review*, 24:6, 611–631.

Falch, T., (2011), "Teacher mobility responses to wage changes: Evidence from a quasi – natural experiment", *American Economic Review*, 101:3, 460-465.

Feng, L., (2009), "Opportunity wages, classroom characteristics, and teacher mobility", *Southern Economic Journal*, 75:4, 1165-1190.

Figlio, D., (1997), "Teacher salaries and teacher quality", *Economics Letters*, 55:2, 267-271.

Figlio, D., (2002), "Can public schools buy better-qualified teachers?", *Industrial and Labor Relations Review*, 55:4, 686-699.

Fredriksson, P., and B. Öckert (2007), "The supply of skills to the teacher profession", Mimeo Uppsala University.

Fredriksson, P., and B. Öckert (2008), "Resources and student achievement – evidence from a Swedish policy reform", *Scandinavian Journal of Economics*, 110:2, 277-296.

Fryer, G.R., (2011), "Teacher incentives and student achievement: Evidence from New York City public schools", NBER Working Paper 16850.

Green, A.D., (1999), "Immigrant occupational attainment: Assimilation and mobility over time", *Journal of Labor Economics*, 17:1, 49-79.

Grönqvist, E., and J. Vlachos (2008), "One size fits all? The effects of teacher cognitive and non-cognitive abilities on student achievement", IFAU Working Paper 25.

Guin, K., (2004), "Chronic teacher turnover in urban elementary schools", *Educational Evaluation and Policy Analysis*, 12:42, 1-25.

Hanushek, E.A., F.J. Kain, and G.S. Rivkin (2004), "Why public schools loose teachers?", *Journal of Human Resources*, 39:2, 326–354.

Hanushek, E.A., and G.S. Rivkin (2010), "Constrained job matching: Does teacher job search harm disadvantaged urban schools?", NBER WP 15816.

Hensvik, L., (2012), "Competition, wages and teacher sorting: four lessons from a voucher reform", *Economic Journal*, 122:561, 799-824.

Hsieh, C-T., and M. Urquiola (2006), "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program", *Journal of Public Economics*, 90:8-9, 1477-1503.

Jackson, C.K., (2009), "Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation", *Journal of Labor Economics*, 27:2, 213-256.

Jackson, C.K., (2013), "Match quality, worker productivity and worker mobility: Direct evidence from teachers", *Review of Economics and Statistics*, forthcoming.

Jackson, C.K., (2012), "School competition and teacher labor markets: Evidence from charter school entry in North Carolina", *Journal of Public Economics*, 96:5-6, 431-438.

Lankford, H., S. Loeb, and J. Wyckoff (2002), "Teacher sorting and the plight of urban schools: the importance of alternative labor market opportunities and non

– pecuniary variation”, *Educational Evaluation and Policy Analysis*, 24:1, 37-62.
Lavy, V., (2009), “Performance pay and teachers’ effort, productivity and grading ethics”, *American Economic Review*, 99:5, 1979-2011.
Lindholm, A., (2006), “Lärarförbundet och lärarnas riksförbund medlemsundersökning om lön och arbetstid”.
Rivkin, S.G., E.A. Hanushek, and J.F. Kain (2005), “Teachers, schools, and academic achievement”, *Econometrica*, 73:2, 417–458.
Ronfeldt, N., S. Loeb, and J. Wyckoff (2013), “How teacher turnover harms student achievement”, *American Educational Research Journal*, 50:1, 4-36.
Scafidi, B., D.L. Sjoquist, and T.R. Stinebrickner (2007), “Race, poverty, and teacher mobility”, *Economics of Education Review*, 26:2, 145–159.

Appendix

Tables

Table A1. Estimation results from univariate OLS models.

VARIABLES	(1) Mobility	R^2	(2) Within teaching mobility	R^2	(3) Out-of-teaching mobility	R^2
Log-earnings	-0.154*** (0.002)	0.079	-0.038*** (0.001)	0.018	-0.115*** (0.002)	0.068
Log-wages	-0.559*** (0.008)	0.060	-0.152*** (0.005)	0.017	-0.407*** (0.006)	0.050
Upper secondary	-0.015*** (0.002)	0.006	-0.003* (0.002)	0.008	-0.012*** (0.001)	0.003
Private	0.058*** (0.006)	0.008	0.015*** (0.003)	0.008	0.043*** (0.004)	0.004
Share of immigrant students	0.030** (0.014)	0.006	0.014 (0.009)	0.008	0.015* (0.009)	0.002
GPA	-0.001*** (0.000)	0.007	-0.001*** (0.000)	0.008	-0.001*** (0.000)	0.002
Female	0.002 (0.001)	0.006	0.002** (0.001)	0.008	-0.000 (0.001)	0.002
Foreign born	0.072*** (0.004)	0.009	0.024*** (0.003)	0.009	0.048*** (0.002)	0.004
Science	0.002 (0.002)	0.006	0.008*** (0.001)	0.008	-0.005*** (0.001)	0.002
Temporarily employed	0.266*** (0.003)	0.112	0.072*** (0.001)	0.026	0.193*** (0.002)	0.091

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). All regressions include only county-by-year fixed effects. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7th that teachers work with. All models except for wages regressions are based on 525 076 observations. Regressions for wages are based on 475 505 observations.

Table A2. Estimation results on a sample of municipalities with limited reductions in teacher stock. The dependent variable is equal to unity if the teacher changes job.

VARIABLES	(1) Mobility	(2) Mobility	(3) Mobility	(4) Mobility	(5) Mobility
Log yearly earnings (1000SEK)			-0.070***	-0.070***	
			(0.004)	(0.006)	
Log monthly salary					-0.174***
					(0.017)
Upper-secondary school	-0.021***	-0.009	-0.007	-0.000	0.005
	(0.004)	(0.008)	(0.008)	(0.009)	(0.009)
Private school	0.013	0.017	0.007		
	(0.009)	(0.011)	(0.010)		
Share of immigrant students		0.034	0.040	0.017	0.008
		(0.028)	(0.028)	(0.027)	(0.028)
GPA		-0.002***	-0.001***	-0.001**	-0.001*
		(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.007***	-0.007***	-0.011***	-0.010***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Foreign born	0.020***	0.019***	0.012*	0.011	0.009
	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
Science	0.006**	0.007**	0.008***	0.008***	0.010***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Temporarily employed	0.234***	0.234***	0.204***	0.213***	0.226***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
R-squared	0.134	0.135	0.144	0.144	0.138
Observations	129,275	129,275	129,275	114,874	114,874

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). All regressions include county-by-year fixed effects. In addition to the displayed variables in column (1) I control for teacher's age, marital status, university education, vocational and special education indicator variables and workload. In column (2) on top of column (1) I control for student-teacher ratio in full time equivalence, share of female students, mean parental income, second order polynomial in school size and indicator for schools with less than 100 students. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7th that teachers work with. Sample reduced to municipalities, which do not experience reductions in teacher stock of more than 5% over the studied period.

Table A3. Baseline estimates restricted to the sample of teachers present in the first year of the analysis. The dependent variable is equal to unity if the teacher changes job.

VARIABLES	(1) Mobility	(2) Mobility	(3) Mobility
Log yearly earnings (1000SEK)			-0.040***
			(0.003)
Upper-secondary school	-0.000	0.007*	0.010***
	(0.002)	(0.004)	(0.004)
Private school	0.020***	0.030***	0.024***
	(0.007)	(0.007)	(0.007)
Share of immigrant students		0.011	0.016
		(0.016)	(0.016)
GPA		-0.001***	-0.001***
		(0.000)	(0.000)
Female	-0.007***	-0.007***	-0.009***
	(0.001)	(0.001)	(0.001)
Foreign born	0.009***	0.008**	0.005
	(0.003)	(0.003)	(0.003)
Science	0.008***	0.009***	0.009***
	(0.002)	(0.002)	(0.002)
Temporarily employed	0.177***	0.176***	0.161***
	(0.004)	(0.004)	(0.004)
R-squared	0.074	0.076	0.079
Observations	275,723	275,723	275,723

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). All regressions include county-by-year fixed effects. In addition to the displayed variables in column (1) I control for teacher's age, marital status, university education, vocational and special education indicator variables and workload. In column (2) on top of column (1) I control for student-teacher ratio in full time equivalence, share of female students, mean parental income, second order polynomial in school size and indicator for schools with less than 100 students. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7th that teachers work with. Sample reduced to teachers observed in the first year of data.

Details of sample construction

I construct the sample of lower and upper secondary school teachers for the school years 1996/1997 to 2006/2007. The information about teachers comes from the teacher registry and the analysis focuses on teachers working in grades 7-9 (lower secondary school) of compulsory education and in grades 1-3 (upper secondary school) of secondary education. The reason for restricting the analysis to these grade levels, is that I lack information on student characteristics for lower levels. Teachers who are on unpaid leave of absence or whose workloads are zero hours (i.e., they do not perform any pedagogical duties) are excluded from the analysis. Such teachers are treated neutrally in terms of mobility if they come back after the absence period to the same school. Similarly, I exclude teachers who are employed as principals, study counselors etc. In each year if a teacher has multiple entries in the registry, the observation with the highest workload is selected irrespectively whether it is at the same or at different schools.[16] The teacher registry is a high quality data set, that allows recovering information on school location (unique identifier), school ownership and type, teacher certification, workload, employment type (temporary vs. permanent), education and position.

Teachers are grouped into either lower or upper secondary education and teachers working in grades 7-9 are recovered by merging the teacher registry to the pupil registry via unique school identifier. There exist schools with more grades covered under the same school identifier (i.e. 1-9 or 4-9) and one possible source of bias would be, for instance, relating teachers who work with students in grades 1-3 to school characteristics measured for students in grades 7-9. Since I have information about the grades in which teachers work I address this issue by excluding teachers coded as primary (grades 1-3) and middle (grades 4-6) school teachers. Such a procedure does not solve the problem completely as some teachers (arts or music) are not necessarily coded by grades. Thus, I may still include some miscoded teachers, however, the share of miscoded teachers is likely low. Nonetheless, each included school serves grades 7-9 and only turnover between such schools is considered at lower secondary level.

Teachers are then linked (using unique identifier) to population registry, which covers all individuals living in Sweden. The population registry is a high quality data set that allows recovering information on gender, marital status, age, family composition (using unique family identifier), immigration history, education and income. Income is measured as a gross salary plus income from business and self-employment plus any work-related allowances. Investment losses are not included, and thus, income is lower-bounded at zero. The analysis is restricted to teachers aged 25-58 years, to abstract from mobility driven by educational attainment and retirement decisions.

[16] The workload of teachers having multiple positions at the same school is not summed and the highest workload position is selected.

The earnings registry often contains multiple entries per individual, which reflect different sources of labor compensations but are uniquely identifiable based on establishment identifier. This poses linking problem for individuals with multiple entries as I may miss-assign earnings from different establishment to a particular school code. Since there is no direct link between unique school code and establishment identifiers, I create such a link using a mode rule. In particular, based on the individuals with only one record I define most often occurring establishment identifier for each school code. I then use this data to resolve matching of individuals with multiple earnings entries.

The students' characteristics are based on "school in" and "school out" pupil registries. The lower secondary school composition is based on outgoing students. The quality of students in lower secondary school is measured based on their 9th grade outgoing grades. The measure is calculated for year t as a mean percentiled GPA from cohorts graduating in years t+1, t+2 and t+3.

The upper secondary school composition is based on all the students that are in a given school in a particular year. The quality of students in upper secondary school is measured based on their 9th grade grades. The main advantage of using lower secondary school grades as a measure of upper secondary school quality is that it is largely exogenous to upper secondary school teachers. I match these students to their parents using unique family identifier and obtain the family level socioeconomic indicator i.e. mean parental income.

Finally, having data with teachers and students I match the two using a unique school identifier. Naturally since the mobility itself is a lagged variable school year 2006/2007 is dropped from the analysis. The final sample includes 136 100 teachers and 622 453 person–year observations. I exclude the following observations from the sample: very small schools with number of teachers in full time equivalence less than 3 (5 170 observations), teachers that are below 25 years old (8 370 observations), teachers that are above 58 years old (82 298 observations), and schools with the number of students less than 15 (1 539 observations). The final sample consists of 121 580 teachers, 2703 unique schools and 525 076 person-years. Applying the monthly wages sample restriction further reduces the sample to 109 541 teachers, 2172 unique schools and 475 505 teacher-years.

Essay 2

Job mobility among high-skilled and low-skilled teachers

1 Introduction

Teachers are important for student achievement (Rockoff, 2004; Rivkin et al., 2005; Aaronson et al., 2007). Even though it has proved hard to pin-point exactly what makes a good teacher, a number of studies suggest that teacher effects can be related to observed measures of teacher skills. Experienced teachers have been shown to provide more skills to students than teachers that are new to the profession (Rockoff, 2004; Harris and Sass, 2011; Clotfelter et al., 2007). Few studies have found any effect of teacher education on student outcomes, but there is some evidence that more detailed information on teacher quality may be important in the production of skills. Clotfelter et al. (2007) find that teacher test scores and regular licensure have positive effects on student achievement. Rockoff and Speroni (2011) document that subjective evaluations of teacher effectiveness have a predictive power for the achievement gains of their students. On the other hand, Grönqvist and Vlachos (2008) find no overall relationship between teachers' cognitive and non-cognitive assessments and student outcomes.[1]

The quality of teachers may be of particular importance for disadvantaged students. For example, Grönqvist and Vlachos (2008) find that the effect of a teacher with high non-cognitive skills is stronger among low-performing students. At the same time, they also find that high cognitive skills' teachers might actually harm low-aptitude students. In the US Aaronson et al. (2007) find that teacher quality, measured by value added, is particularly important for lower-ability students and that a one standard deviation in teacher quality is worth as much as 24% of average test score gain between eight and ninth grade for students from the bottom tertile of the quality distribution. At the same time, this effect for the top tertile is only 6%. The heterogeneity in teacher effects found in the econometric analyses has also been confirmed using a random assignment of teachers to classrooms. Nye et al. (2004), using data from Tennessee STAR experiment, show that the variance of teacher effects is much larger in low than in high socioeconomic status schools. This means that in low SES schools, it matters more which teacher a child receives than it does in high SES schools. Furthermore, it suggests that interventions to replace less effective teachers with more effective teachers may be more promising for disadvantaged than for privileged children.

School principals can try to enhance the quality of the teacher stock, either by hiring good teachers or by firing bad ones (Böhlmark et al., 2012). However, the success of such employment policies depends, in part, on the skills of available teachers to hire. Many studies have documented falling quality of new entrants into the profession over the past decades, leading to

[1] Their empirical strategy relates within-school variation in teacher quality to within-student variation in performance between subjects, and may not be appropriate for identifying main effects of teacher quality.

deterioration of the skills in the pool of potential teachers to hire.[2] Thus, reducing turnover among high-quality teachers must probably be crucial for any principal wishing to sustain the competence level in their school.

A growing number of studies document turnover among teachers of different quality. One strand of the literature makes use of input-based measures of teacher quality, such as certification, education and experience (Boyd et al., 2005; Feng, 2010; Barbieri et al., 2011; Clotfelter et al., 2011). However, most of these studies use quite crude quality measures, which have shown to be only weakly related to student performance. Another strand of the literature exploits output measures of teacher quality, such as the estimated value-added of different teachers. This approach is not limited to observed determinants of student performance, but the validity and stability of teacher fixed effects models have been questioned in the literature (Rothstein, 2010).

In addition to the standard input-based teacher quality measures used in the literature (education and experience), this study makes use of a population-wide data on both cognitive and non-cognitive skills among male teachers (born 1951 or later) to study teacher turnover in Swedish secondary schools. In particular, I study differences in teacher mobility, to other schools or out of the profession, among high-quality and low-quality teachers. Further, I relate any differences in teacher turnover to a number of job attributes, such as student quality, teacher wages and type of contract.

This paper should also be of interest due to the uniqueness of the Swedish institutional setup. Unlike most countries (Falch and Strøm, 2005; Jackson, 2009; Falch, 2010), the Swedish labor market for teachers does not differ much from other white-collar job markets and is an excellent example of monopsonistic competition with individual wage bargaining and a growing private sector (Manning, 2011; Karbownik, 2013). Similar to other countries, Sweden also struggles with attracting high skilled individuals into the teaching profession and experiences teacher shortages, yet has introduced utterly different institutions.[3]

I show that university educated and experienced teachers are less likely to both leave their current school and the profession. Furthermore, using the unique enlistment records I document that teachers with high non-cognitive skills are less likely to change employers. At the same time, I do not find

[2] Grönqvist and Vlachos (2008) document a close to 20 percentile ranks decline in the average cognitive ability of Swedish teachers since the early 1990s and also a substantial decrease in social abilities and GPAs. Fredriksson and Öckert (2007) present evidence of a deterioration of returns to teacher education and experience among Swedish teachers. Similarly, Nickell and Quintini (2002) report severe declines in investment in teachers in Britain, while Leigh and Ryan (2008) find about 10 percentile rank declines in Australian teacher quality. Both Bacolod (2007) and Corcoran et al. (2004) document that contemporary teachers in the US are less qualified than their counterparts in the 1960s and 1970s.

[3] Björklund et al. (2006) or National Agency for Education (2003) provide details about teacher shortages.

robust correlations for cognitive skills when I control for standard teacher quality measures like education or experience. Moreover, I do not find any support for the common view that schools serving minority students experience higher turnover rates of high-quality pedagogues. Finally, I present robust negative correlation between teacher turnover and monetary compensations.

The paper is organized as follows: section two offers a short literature review, section three briefly presents the institutional background, data sources, and the econometric model, section four presents descriptive evidence, section five contains the main results, section six includes heterogeneity analyses, and finally section seven concludes.

2 Literature review

It is important from the education policy stand point to understand if disadvantaged schools experience outflow of high quality teachers or attract particularly bad teachers, and thus there is a growing number of studies that document turnover among teachers of different quality. Using the data from New York State, Lankford et al. (2002) show that urban (low-income, low-achieving and non-white) schools deter high quality teachers and that salary variation rarely compensates for the difficulties of teaching in these disadvantaged schools. Their measures of teacher quality are based on experience, formal education and its quality as well as certification. Furthermore, using the same dataset and quality measures Boyd et al. (2005) show that there is a significant heterogeneity in teacher responses when exposed to low-quality pupils. For example, when considering probability of separation, the top 75 percent of teachers, as measured by general knowledge certification exam, reacts much more strongly to low-aptitude students than does the bottom 25 percent. The differences in teacher turnover by experience are also found in Feng (2010), who uses Florida school teachers and explores an assignment to tough classrooms. She documents that it is rather inexperienced teachers that are most likely to exit the profession when facing low-achieving and misbehaving students. Clotfelter et al. (2011) using data from yet another State in the US, North Carolina, show that teachers with stronger qualifications are both more responsive to racial and socioeconomic mix of school's students and less responsive to salary changes. The authors use four measures of teacher qualifications: teachers' average licensure test scores, alma mater competitiveness, experience and certification. In the European context, Barbieri et al. (2011) document that experienced teachers are driven away from the most difficult schools and that the major discouraging factors include high shares of disabled and foreign students, as well as students who had to repeat a grade.

Research also suggests that teachers react to changes in their working environment. Studies show that teachers are responsive to even small variation in wages (Baugh and Stone, 1982; Murnane and Olsen, 1990; Figlio, 1997; Figlio, 2002; Feng, 2009; Falch, 2011; Karbownik, 2013). Another factor affecting teachers' turnover and compensations is the competition between publicly and privately run schools (Jackson, 2012; Hensvik, 2012). It is also important to understand the differences between the wages offered to teachers in education and in other sectors of the economy (Dolton and van der Klaauw, 1995, 1999; Brewer, 1996; Fredriksson and Öckert, 2007; Dolton and Marcenaro-Gutierrez, 2011). Non-pecuniary characteristics play an important role alike and sometimes they even dominate monetary compensations (Hanushek et al., 2004). As the literature suggests, teachers are generally discouraged by high fractions of poor, minority and low-achieving students (Falch and Strøm, 2005; Scafidi et al., 2007; Barbieri et al., 2011; Bonhomme et al., 2011).[4] Finally, there is evidence that the quality of match between a school and a teacher is an important issue (Jackson, 2013).

3 Institutional setting, data and empirical set-up

3.1 Institutions

The Swedish schooling system starts with voluntary pre-school and continues with nine years of compulsory education. Lower secondary school covers the grades 7-9. The 9th grade grades determine student's chances to advance to upper secondary school. Swedish municipalities are obliged by law to provide upper secondary schooling to all students who successfully completed compulsory education. Upper secondary school consists of different programs, lasts three years and provides eligibility for post-secondary education.

Private schooling is growing in Sweden and is encouraged by the government. In 1992, Sweden introduced a school voucher reform that allowed for both non-profit and for-profit independent schools. The municipality is obliged to pay the independent schools for each student they can attract, with an amount corresponding roughly to the average per student cost in the public schools.[5] Since the reform the fraction of private schools has risen, in particular at the upper secondary level. In the school year 2005/2006 there

[4] More recent literature relying on quasi-experimental methods (Jackson, 2009) and based on administrative data (Karbownik, 2013) finds rather heterogeneous impact of minorities on teacher turnover.

[5] An independent school receives around 85-95% of the average per student cost in public schools and this varies from year to year. Some municipalities also have a socioeconomic gradient for the school voucher. The private schooling was effectively introduced at lower secondary level in 1992, and at upper secondary level in 1994 (Böhlmark and Lindahl, 2007, 2008).

were 220 private upper secondary schools, which constituted 33.1% of all upper secondary schools in Sweden, a rise from 8.1% in 1996/1997. At the same time, the number of private lower secondary schools constituted only 15.8% of all schools at this level starting from 3.2% in 1996/1997.[6]

Teaching profession in Sweden is regulated and different qualifications are required depending on the subject taught and on the type of school. Teaching at the secondary school level requires completing special coursework beyond what is required from a compulsory school teacher. Individuals from other professions who want to become teachers need to supplement their professional degrees with a minimum of 1.5 years of preparation in pedagogy, didactics and teaching practice.

Municipalities are the primary employers of teachers in Sweden, and thus, handle the responsibility of recruiting them.[7] In practice, however, the decisions regarding recruitment, selection and employment of a teacher are made at the school level by a principal. Finally, teacher wages are determined at local level through individual bargaining between teacher and principal given the collective bargaining outcome set at the national level.[8]

3.2 Data

This paper utilizes Swedish population-wide registries. The main data source is the teacher registry that covers all teaches employed in Swedish schools in years 1996/1997 to 2006/2007. It contains information on teachers' education, specialization, experience, certification, place of work, type of contract (permanent vs. temporary) and workload. To these data I have matched background information on age, gender, immigration histories, employment and income for all teachers in the registers. The pupil registries for lower and upper secondary schools are used to obtain information on students in a given school. These allow linking children and their parents to schools, as well as obtaining the average percentile ranked GPA of the students. Administrative records on earnings provide information on teachers' monetary compensations. The details of the sample construction are discussed in the appendix.

Since, the core focus of this paper is on teacher quality, for the subsample of male teachers born 1951 or later, I use military enlistment data to obtain information on cognitive and non-cognitive test scores. Until the 1st of July 2010 the military service in Sweden was mandatory for all males aged 18-

[6] This information is based on registry data.

[7] For more information on the reform that shifted responsibility for schooling from the central government to municipalities see: Fredriksson and Öckert (2008). There is still a small fraction of schools run by county or state, however, those employ around 1% of all the teachers between 1996/1997 and 2005/2006. Those schools are excluded from the analysis since they have different sources of funding and their role is diminishing.

[8] Individualized pay was introduced in 1996 and is discussed in detail by Hensvik (2012) and in survey by Lindholm (2006).

47.[9] The enlistment procedure lasts two days and comprises of medical and physical assessments, cognitive ability tests and 20 minutes semi-structured interview with a trained, and often very experienced, psychologist (Mood et al., 2012). It was not possible to avoid military service by obtaining a low score on cognitive or non-cognitive ability assessments but about 5-10 percent of enlisted men did not attend the enlistment because of the mental or physical handicaps. The data is also restricted to the natives, since only Swedish citizens were allowed and obliged to attend the enlistment.

The cognitive assessment of Swedish conscripts has been conducted since the mid-1940s. The tests have changed somewhat over the years, but they have always been intended to measure the same four underlying cognitive traits: logic-inductive ability, verbal comprehension, spatial ability and technical comprehension.[10] Each of these tests was graded on a scale from 1 to 9, where 1 is the lowest possible and 9 is the highest possible score. These scores were then transformed to a discrete variable of general cognitive ability ranging from 1 to 9. In the analyses, I use the final score which is comparable across all years.

Similarly to cognitive assessment, the personality tests were introduced at the military enlistment in the early 1940s. All the men in the data had their psychological profiles evaluated according to a procedure that was adopted in 1969 and kept unchanged up to 1995 when it was subject to minor revisions. The personality assessment which is based mostly on behavioral questions can be categorized into four parts: social maturity (extroversion, having friends, taking responsibility, independence), psychological energy (perseverance, ability to fulfill plans, ability to remain focused), intensity (the capacity to activate oneself without external pressure, the intensity and frequency of free-time activities) and emotional stability (ability to control and channel nervousness, tolerance of stress, and disposition of anxiety). The general objective of the interview was to assess the conscript's ability to cope with the psychological requirements of the military service, and in the extreme case, war. As the final outcome of the interview the psychologists assign each man military aptitude score from 1 to 9, which is comparable over years.

I am able to recover information on cognitive and non-cognitive test scores for 89% of Swedish male teachers born in 1951 or later.[11] Since most

[9] At the end of 2000s not the whole population was drafted and thus the data are reliable only until 2006. The enlistment usually takes place right after upper secondary school graduation i.e., when man turns 18 or 19 years old. Among the teachers for whom I have data 96.3% did the enlistment when they were 18 or 19, 2.3% when they were 20, 0.3% when they were below 18 and the remaining 1.1% when they were older than 20 years old.

[10] Carlstedt (2000) describes the history of psychometric testing in the Swedish military. Unlike AFQT, the Swedish cognitive assessment is a good measure of a general intelligence.

[11] The first draft year I use is 1970 and the last one is 1999. Most of the data for individuals tested in 1978 are lost, and thus only 15 412 observations are recorded for this year. This loss is not systematically related to individual characteristics other than year of birth.

of the missing individuals were exempted from the draft due to mental and physical disabilities, there are differences in observables between them and those for whom the scores are available. More details regarding the construction of final scores used in the analyses are provided in the appendix. For details regarding the testing procedure itself and various applications of Swedish military enlistment registries see: Lindqvist and Vestman (2011).

3.3 Econometric modeling

The main analysis is done using a series of binary choice models that attempt to capture the manifestation of teachers' job preferences conditional on teacher quality. Since this paper is only descriptive I am not able to identify teacher's preferences in an econometric sense. Nonetheless, it should be intuitive that leaving employment j in favor of an alternative opportunity k is related to individual preferences with respect to employers j and k. Thus, I specify the following linear model. The dependent variable is equal to unity if a teacher leaves their current employer from year to year, and such a decision is regressed on teachers' working environment and their own characteristics. In particular, these binary models show whether teachers who remain in their appointments have, on average, different quality than those who leave their jobs.

From the policy point-of-view, one should also investigate what are the factors that drive high quality teachers to seek a better employment match as such sorting of teachers may indicate permanent quality drop of particular institutions, and thus, have adverse influence on student achievement.[12] Therefore, the heterogeneity analyses based on the differences in school characteristics shed light on what job characteristics are important for low and high quality teachers. Using the main specification, I also run separate regressions depending on teacher's destination. In particular, I specify two distinct variables of transition. These are: switching schools within a teaching profession and leaving teaching in favor of a different occupation.[13] This analysis could be of interest for policy makers, as losing highly educated pedagogues in favor of other sectors of the economy may lead to worsening productivity of the whole educational system in the future.

[12] High quality teachers are these with university education (Ehrenberg and Brewer, 1994; Harris and Saas, 2011), longer experience (Rockoff, 2004), above median cognitive and non-cognitive test scores (Hanushek, 1971; Harbison and Hanushek, 1992; Grönqvist and Vlachos, 2008). A university graduate is defined as an individual graduating three, four or five year long university (hogskoleutbildning) education or an individual with a research degree. Note that other forms of post-secondary (eftergymnasial) education are not treated as university graduates.

[13] Switching to primary education or adult education is treated as school-to-school mobility. Switching to kindergarten, pre-K or university education is treated as quit. The results are robust to various definitions of school-to-school mobility and quits and are available upon request.

In order to maintain simplicity of the interpretation of the results, the estimation strategy is based on the least squares using linear probability model.[14] The following econometric model is estimated:

$$y_{ijt} = \alpha_0 + \alpha_1 Q_{ijt} + \alpha_2 W_{ijt} + \alpha_3 X_{jt} + \alpha_4 P_{ijt} + \delta t \cdot c + \varepsilon_{ijt} \tag{1}$$

where y_{ijt} is equal to unity if teacher i leaves the current employer j at a period following t, W_{ijt} is teacher i earnings at school j and time t, X_{ijt} is a vector of observable school characteristics at institution j at time t (polynomial of school size, share of girls, student-teacher ratio in full time equivalence as a proxy for school resources, share of non-Nordic students, student's precentiled GPA and mean parental income), P_{ijt} is a vector of personal characteristics of teacher i at school j and time t (gender, non-Nordic teacher indicator, marital status indicator, three specialization indicators, workload, type of school, school ownership indicator and type of employment), and ε_{ijt} is an error term that represents unobserved characteristics, which is clustered at school level. I use four types of quality indicators, Q_{ijt}. In the full sample of teachers, the quality of teacher i at school j and time t is measured using experience and education. In the sample of younger males, which is of interest due to the unique data, I use cognitive and non-cognitive military assessment of a teacher as quality measures.[15] Vector of δs captures county-by-year fixed effects.

4 Descriptive evidence

This paper focuses on four measures of teacher quality: university education, teaching experience, cognitive and non-cognitive test scores. In order to better understand how these measures relate to particular school characteristics, Figures 1 and 2 plot their means against the deciles of student's GPA and share of minorities.[16] In particular, the figures illustrate the distribution of teacher quality across schools with different pool of students, which should help understanding what type of teachers in terms of quality cluster in a given type of schools.

[14] This method yields very similar estimates to the non–linear models. The regressions using logit and multinomial logit models with marginal effects evaluated at mean are available upon request. Majority of correlations between explanatory variables are below 0.1 and the correlogram is available upon request.

[15] The correlation coefficient between cognitive and non-cognitive assessment in the studied sample is 0.15, which is lower than the one reported by Grönqvist and Vlachos (2008) for the whole population (0.36).

[16] Lower secondary school GPA is the percentiled GPA in 9^{th} grade. See the appendix for details.

Teacher education correlates positively with student achievement measured by GPA, and the worst performing students are taught by a lower number of university educated teachers. At the same time, both low and high achievers are taught by rather less experienced teachers. Similar u-shaped pattern can be found in the relationship between the proportion of immigrant students and teachers experience. Finally, the slope of the relationship between the level of immigrants in school and share of teachers with university education is smaller than the one for student quality measured with GPA.

As far as intellectual assessment is concerned the patterns are mostly stable with a significant gap between cognitive and non-cognitive scores. First, teachers are more positively selected on cognitive than on non-cognitive scores. Second, teachers with higher cognitive abilities are matched to high-performing students. Third, teachers with higher cognitive skills are matched to schools with more minorities, while the opposite is true for teachers with high non-cognitive skills.

Figure 3 shows the evolvement of teacher mobility with different destinations for teachers with and without university degree. Figure 4 depicts the three mobility measures split by teacher experience, cognitive skills, and non-cognitive skills, respectively. Teacher turnover differs systematically by teachers' educational attainment. This difference is almost entirely driven by a higher probability for teachers without a university degree to leave the profession, while there is more or less no difference in the school-to-school mobility. Teacher turnover decreases with teacher experience. Less experienced teachers are also more likely to leave the profession than to move to another school, but the mobility to different occupations converges to school-to-school mobility for teachers with more than 8 years of experience. As far as intellectual assessment is concerned, turnover rate is stable across deciles of cognitive and non-cognitive scores, except for the bottom of the distributions, where it is larger for quits. Thus, it is the teachers with the very worst abilities that are more likely to leave the profession, partly offsetting the documented by others decline in skills among inflowing teachers.

Figure 1. Teacher quality by student GPA.

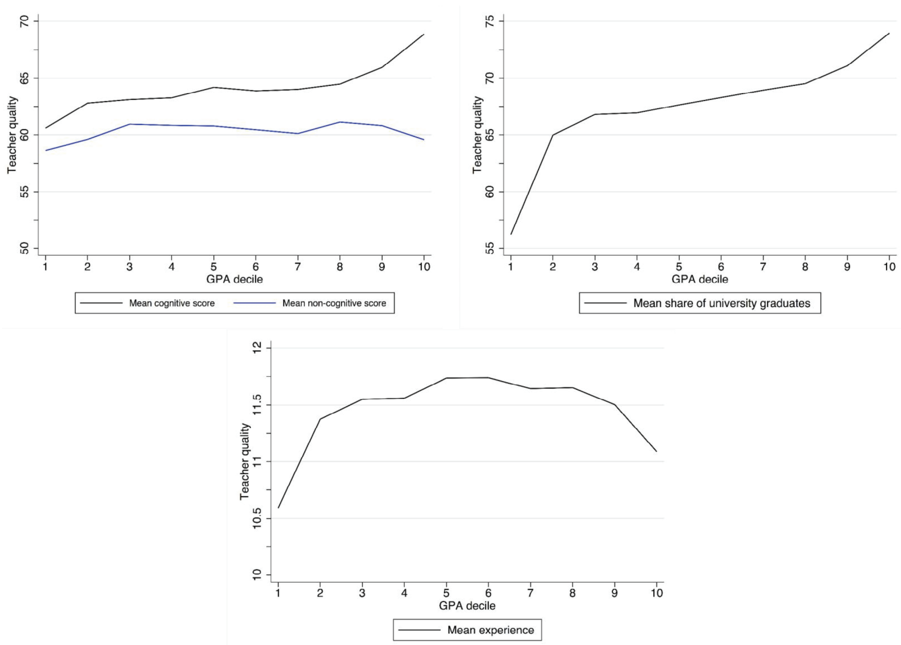


Figure 2. Teacher quality by share of minorities.

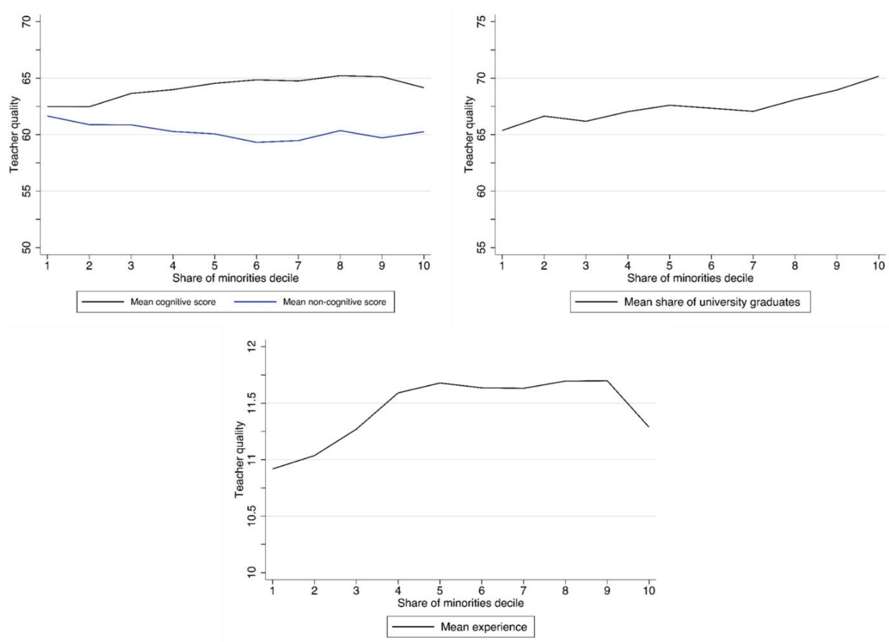


Figure 3. Turnover over time for teachers with and without university degree.

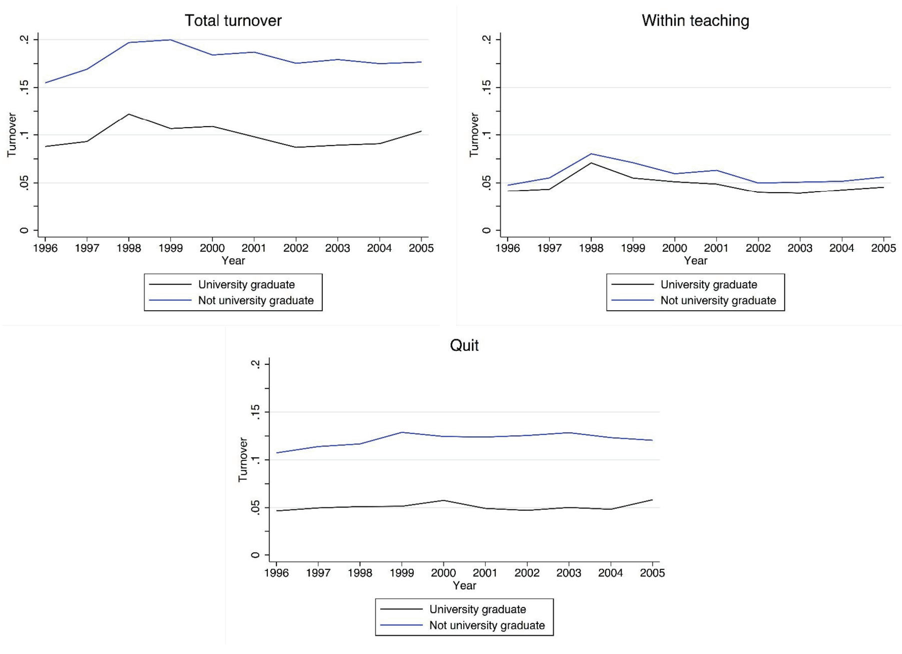
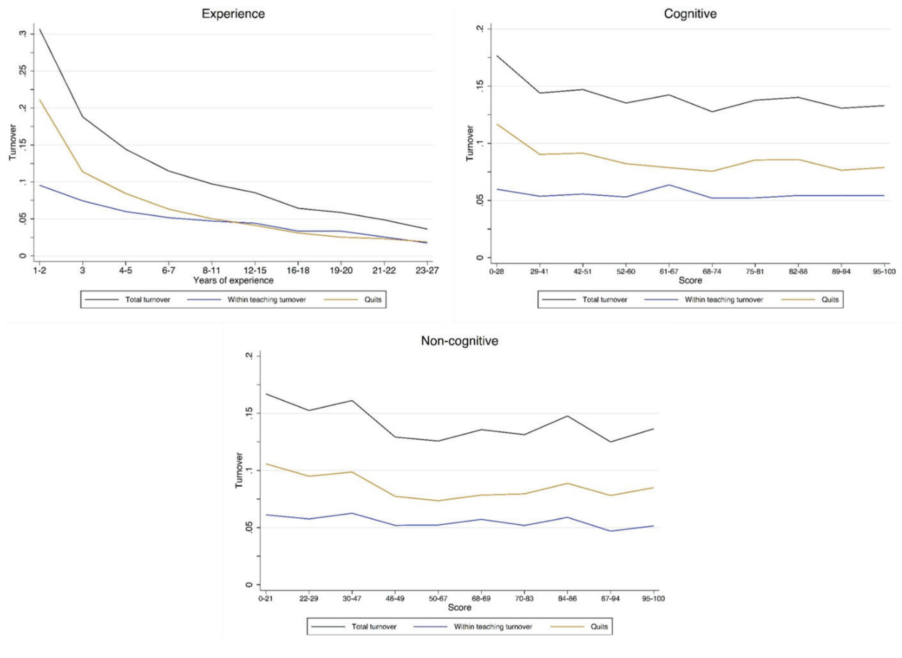


Figure 4. Experience, intellectual assessment and teacher turnover.



In sum, the descriptive evidence suggests that there is a substantial heterogeneity in teacher turnover with respect to teacher quality, irrespectively whether it is measured as formal education, tenure in teaching or intellectual and behavioral assessments. There are also differences in teacher quality across school characteristics i.e., teachers of different quality tend to cluster at schools with particular observable characteristics.

Table 1 presents the descriptive statistics of variables used in the econometric analysis. Panel A presents three turnover measures, panel B presents teacher quality measures, panel C presents personal and pecuniary characteristics, while panel D presents average school-level characteristics. Total turnover rate, is at 12.5%, which is lower than the overall turnover rate in all the occupations in Sweden (Edin et al., 2009). This could be driven by the fact that people who invest heavily in occupation-specific human capital (teaching) may have lower turnover rates in general. Although the quit rate in Sweden is larger than in Norway, these two countries share a common feature that the outflow from teaching is larger than the mobility within the profession. In the US registry data from Texas, Hanushek et al. (2004) find the opposite pattern (i.e., there is higher mobility within teaching rather than out of the profession). In panel B experience and university indicator are based on the whole sample of 525 076 observations from 2703 schools.[17] However, the intellectual assessment measures are based on the sample of native males, born prior to 1951 and drafted prior to 1970 with available data contributing 115 350 observations from 2628 schools. In the analyzed schools 67% of teachers are university graduates with an average experience of 11.5 years and scores of 64 and 60 points on a standardized 0-100 scale for cognitive and non-cognitive assessments, respectively.

In Swedish schools 56% of teachers are women, 6.8% come from non-Nordic countries, 20.8% are employed on temporary contracts and their average yearly earnings equals to 221 866 SEK. There is 15.6% science, 13.8% vocational and 6.6% remedial education teachers.[18] The fraction of teachers employed in private schools during the studied period rose from around 2% in 1996 to 10.5% in 2005. The student-teacher ratio in full time equivalence, which can be seen as proxy for school resources, is 11.5% and the average number of pupils is 574.[19] There is on average 8.3% non-Nordic immigrants

[17] Teacher experience is not available for all years, and thus, I use the predicted experience in the analysis. In particular, since the teacher registries date back to 1979 I explore this feature to construct the "in teaching predicted experience" variable. I create a panel of all teachers between 1979 and 2006 and link it to population enlistment data between 1985-2006 in order to obtain teacher's birth date. I then use all this information and tenure data provided in the later registries (since 1999 onwards) to construct the predicted measure of experience.

[18] Remedial education teacher (Speciallarare) works with students in need of special assistance concerning learning and development. Special education teachers focus on either language or mathematics.

[19] Number of students in lower-secondary school is measured as the sum of pupils attending grades 7 to 9 and it is provided in compulsory school registry by Statistics Sweden. Number

in Swedish schools. This number is larger than the one reported for Norway (Falch and Strøm, 2005).

Table 1. Descriptive statistics.

Variable	Mean	Standard deviation
Panel A: Teacher turnover		
Total mobility	0.125	(0.331)
Within profession mobility	0.051	(0.220)
Out of profession mobility	0.074	(0.262)
Panel B: Teacher quality		
Experience	11.445	(7.778)
University graduate	0.674	(0.469)
Cognitive test score*	0.641	(0.244)
Non-cognitive test score*	0.603	(0.273)
Panel C: Personal and pecuniary characteristics		
Female	0.562	(0.496)
Foreign	0.068	(0.251)
Married	0.573	(0.495)
Upper secondary	0.437	(0.496)
Private	0.056	(0.230)
Science	0.156	(0.363)
Vocational	0.138	(0.345)
Remedial	0.066	(0.249)
Temporary	0.208	(0.406)
Workload	86.488	(23.273)
Log yearly earnings (1000SEK)	5.290	(0.586)
Panel D: School characteristics		
Share of girls	0.482	(0.100)
Share of foreign students	0.083	(0.086)
Student-teacher ratio in full time equivalence	11.511	(3.241)
Number of students/100	5.739	(4.574)
Students' parents income in 1000SEK	380.201	(96.397)
Student's percentiled GPA	48.175	(6.708)
N	525 076	

Note: mean values, standard errors in parentheses.
*N = 115 350

5 Main results

The estimates presented in this section correspond to the model outlined in section 3.3. I estimate a binary linear regression model with county-by-year fixed effects where the dependent variable equals to unity if the teacher leaves a particular school from year t to year t+1, and zero otherwise.[20] The results are presented in table 2. Column (1) shows the raw correlation between the total turnover and teacher quality measured by university graduation and experience. Column (2) adds individual characteristics to the estimates from column (1). Column (3) provides estimates, including both individual and school level covariates. Column (4) adds yearly earnings to the specification from column (3). This allows me to understand if the differ-

of students in upper secondary school is measured based on the registry of students enrolled in grades 1 to 3 in upper secondary schools.

[20] Specifications with only year, or only county, or only year and county, or using municipality instead of county fixed effects has also been estimated and yield similar results. Including school fixed effects removes some of the variation that is of interest in the heterogeneity analyses presented in this paper.

ences in mobility by education and experience are driven by differences in earnings. At the same time, earnings are also a teacher quality measure, and thus the contribution of having higher education or more experience holding earnings constant is not trivial to interpret. Therefore, in the heterogeneity analysis I do not condition on earnings and use the specification from column (3).

Table 2. Main results using university education and experience. The dependent variable is equal to unity if the teacher changes job.

VARIABLES	(1) Mobility	(2) Mobility	(3) Mobility	(4) Mobility
University graduate	-0.040***	-0.005***	-0.004**	-0.002
	(0.002)	(0.001)	(0.001)	(0.001)
Experience	-0.009***	-0.004***	-0.004***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
R-squared	0.060	0.132	0.133	0.140
Observations	525,076	525,076	525,076	525,076
Personal characteristics		X	X	X
School characteristics			X	X
Log-earnings				X

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). All regressions include county-by-year fixed effects. Personal characteristics include: gender, immigration status, marital status, indicators for science, vocational and remedial specialization, indicator for temporarily employed, workload, indicators for upper secondary and private school teachers. School characteristics include: student-teacher ratio in full time equivalence, number of students and its square, indicator for schools with less than 100 students, share of girls and immigrants at school, mean percentiled student GPA and mean parental income. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7th that teachers work with.

The results from columns (1) to (3) suggest that schools in Sweden do not lose university educated and experienced teachers, as both of the coefficients are negative and significant. An additional year of experience is associated with 0.3 to 0.9 percentage points (pp) lower mobility while holding university diploma is associated with 0.4 to 4.0 pp lower mobility depending on the specification. In columns (2) and (3) the coefficient on university education decreases about 10 folds in comparison to column (1) and these changes are virtually entirely driven by accounting for temporary employment status. This likely is due to the Swedish institutional setting, where permanent contracts typically are given only to teachers who have a university degree. When the earnings are added (column (4)) both coefficients decrease even more and the coefficient on university education becomes insignificant. This could mean that principals may have a scope for changing the mobility behavior of teachers of different quality through manipulation of monetary compensations and type of employment, but since this paper documents only descriptive associations there may well be other explanations to the observed pattern. If these job attributes can help retain experienced and educated teachers, then one would expect the estimates of teacher quality to be weaker when the controls are added into the model.[21]

[21] I have also estimated models for public schools only using monthly wages. The results are similar to these reported in column (4) i.e., they yield an insignificant and close to zero coefficient on university education and a negative 0.3 pp estimate on experience. Since the information on monthly wages is available only for public school teachers and the main results using both compensation measures are similar then the analyses in this paper use earnings. I

It is a question of general interest, how individuals' intellectual capacities affect their decisions to change jobs. Table 3, in columns (1) to (4), re-estimates the specifications from table 2, while substituting education and experience by cognitive and non-cognitive test scores. In columns (5) and (6) I re-estimate specifications from columns (3) and (4) while controlling for education and experience. This allows me to understand the value-added from using the intellectual skills measures as compared to measures typically used in the literature. Panel A estimates association between non-cognitive skills and turnover unconditional on cognitive skills while panel B uses cognitive skills unconditional on non-cognitive skills. In panel C I include both measures simultaneously.

Results in columns (1) to (3) suggest a negative relationship between a propensity to leave current employment and both cognitive and non-cognitive skills with larger estimates for the former measure. This association is estimated to be between 4.7 and 0.9 pp depending on the specification and measure. In column (4) where I include earnings both coefficients decrease and only the estimate on non-cognitive skills remains significant. Estimates from column (5) suggest that controlling for teacher education and experience captures well the set of skills related to cognition, however, these measures are not so effective in terms of non-cognitive capabilities. Even when controlling for education, experience and cognitive skills I still find a significant and negative correlation of 1.0 pp between teacher mobility and non-cognitive skills. This is filtered out in column (6) where I also control for earnings but as it has been noted in the first paragraph of section 5 these results are not trivial to interpret given the fact that earnings itself is a measure of teacher quality.

Finally, I can only observe mobility if teachers leave their school from one year to another, however, it may be questioned whether this mobility is voluntary or not. In particular, there can be reshuffling of teachers between schools in municipality due to the fact that employment protection is based on an employment in municipality and not at the school (this naturally does not apply to privately owned institutions). It could also be the case that if one school has an opening for a teacher and there are other schools in the same municipality laying off teachers, there might be bargaining and reshuffling of teachers within the municipality. To address this issue I restrict the analysis to the sample of municipalities that never experienced reductions in the teacher stock by more than 5% over the studied period.

Tables A1 and A2 in the appendix present the estimation results using the sample described above and the specifications from tables 2 and 3. The sample size is reduced around four-fold, however, the majority of the results using education and experience remain unchanged. Unlike in table 3, how-

have also estimated models where the two quality measures are included separately, and the conclusions do not change. These are available upon request.

ever, the estimates on cognitive and non-cognitive skills in this restricted sample become mostly insignificant but very similar quantitatively. Thus, the lack of significance should rather be associated with increased standard errors due to reduction in sample size than with sample selection and changes in point estimates. Overall, these estimates suggest that the differences in mobility for teachers of different quality should not be driven by selective lay-offs when schools are down-sizing.

Table 3. Main results using cognitive and non-cognitive assessment. The dependent variable is equal to unity if the teacher changes job.

VARIABLES	(1) Mobility	(2) Mobility	(3) Mobility	(4) Mobility	(5) Mobility	(6) Mobility
Panel A: Non-cognitive assessment unconditional on cognitive score						
Non-cognitive score	-0.039***	-0.014***	-0.014***	-0.008**	-0.011***	-0.007
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
R-squared	0.011	0.116	0.117	0.129	0.122	0.131
Panel B: Cognitive assessment unconditional on non-cognitive score						
Cognitive score	-0.047***	-0.012**	-0.011**	-0.007	-0.003	-0.003
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
R-squared	0.011	0.116	0.117	0.129	0.122	0.131
Panel C: Both scores included.						
Non-cognitive score	-0.033***	-0.013***	-0.013***	-0.008*	-0.010**	-0.006
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Cognitive score	-0.041***	-0.010**	-0.009*	-0.006	-0.002	-0.002
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
R-squared	0.012	0.116	0.117	0.129	0.122	0.131
Observations	115,350	115,350	115,350	115,350	115,350	115,350
Personal characteristics		X	X	X	X	X
School characteristics			X	X	X	X
Log-earnings				X		X
University and experience					X	X

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). All regressions include county-by-year fixed effects. Personal characteristics include: marital status, indicators for science, vocational and remedial specialization, indicator for temporarily employed, workload, indicators for upper secondary and private school teachers. School characteristics include: student-teacher ratio in full time equivalence, number of students and its square, indicator for schools with less than 100 students, share of girls and immigrants at school, mean percentiled student GPA and mean parental income. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7th that teachers work with.

6 Heterogeneity analysis

So far the evidence suggests that schools in Sweden experience lower turnover rates of high skilled teachers, which is true both for the whole population and for the sample of schools where we shut down the potential for selective lay-offs. In the heterogeneity analyses, I give insights on how the high quality teachers match to the most disadvantaged schools. In particular, I analyze if teachers of different quality differ in the probability to leave schools with certain characteristics. For instance, high quality teachers may be more prone to leave schools with higher shares of minorities or schools with limited financial resources. The quality in table 4 is measured by education and experience, while in table 5 by cognitive and non-cognitive assessments.

University educated teachers tend to leave the private sector with higher likelihood, which works against the common perception that private schools cream skim the best teachers from the market (p-value: 0.028). It is also the highly educated for whom I find association between mobility and student quality (p-value: 0.000). Identical conclusion holds for school resources measured by student-teacher ratio in full-time equivalence. As far as experience is concerned, there is no significant positive correlation between mobility and working in private sector only among the least experienced teachers. The coefficient on student quality is insignificant also only for the least experienced individuals. Finally, unlike Hanushek et al. (2004) I do not find any relationship between the share of minorities at school and teacher mobility for teachers with different education or experience. This supports findings from Karbownik (2013) who found only scarce and heterogeneous evidence of increased teacher turnover in schools with high minority enrollment.

Table 4. Heterogeneity analyses in education and experience. The dependent variable is equal to unity if the teacher changes job.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	University graduate				Years of experience			
VARIABLES	Yes	No	0-2	3-5	6-10	11-15	16-20	20+
Share of	-0.007	0.031	-0.009	0.031	0.029	0.037	-0.004	0.007
immigrants	(0.015)	(0.025)	(0.029)	(0.026)	(0.023)	(0.026)	(0.019)	(0.016)
GPA	-0.001***	-0.000	-0.001	-0.001***	-0.001**	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mean parental	-0.000	-0.000	-0.000*	-0.000	-0.000	0.000	0.000	-0.000
income	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Private school	0.040***	0.024***	-0.000	0.021**	0.022***	0.039***	0.031***	0.038***
teacher	(0.006)	(0.008)	(0.010)	(0.009)	(0.007)	(0.009)	(0.010)	(0.009)
Student-teacher	-0.001***	0.000	0.000	0.001*	-0.001	-0.001	-0.002***	-0.001
ratio FTE	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R-squared	0.101	0.137	0.097	0.084	0.079	0.064	0.050	0.049
Observations	354,121	170,955	82,691	92,260	85,799	65,023	117,402	81,901

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). All regressions include county-by-year fixed effects. Personal characteristics include: gender, immigration status, marital status, indicators for science, vocational and remedial specialization, indicator for temporarily employed, workload, indicators for upper secondary and private school teachers. School characteristics include: student-teacher ratio in full time equivalence, number of students and its square, indicator for schools with less than 100 students, share of girls and immigrants at school, mean percentiled student GPA and mean parental income. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7th that teachers work with.

Table 5. Heterogeneity analyses in cognitive and non-cognitive assessment. The dependent variable is equal to unity if the teacher changes job.

VARIABLES	(1) Cognitive score below median	(2) Cognitive score above median	(3) Non-cognitive score below median	(4) Non-cognitive score above median
Share of immigrants	0.060	0.031	0.053	0.027
	(0.039)	(0.028)	(0.033)	(0.029)
GPA	-0.000	-0.001***	-0.001	-0.001**
	(0.001)	(0.000)	(0.000)	(0.000)
Mean parental income	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Private school teacher	0.026**	0.027***	0.014	0.038***
	(0.012)	(0.008)	(0.010)	(0.009)
Student-teacher ratio FTE	0.000	-0.001	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
R-squared	0.125	0.115	0.135	0.106
Observations	34,071	81,279	50,091	65,259

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). All regressions include county-by-year fixed effects. Personal characteristics include: marital status, indicators for science, vocational and remedial specialization, indicator for temporarily employed, workload, indicators for upper secondary and private school teachers. School characteristics include: student-teacher ratio in full time equivalence, number of students and its square, indicator for schools with less than 100 students, share of girls and immigrants at school, mean percentiled student GPA and mean parental income. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7th that teachers work with.

In table 5 I focus on the relationship between job characteristics and teacher turnover for teachers from different parts of intellectual assessment distribution. In particular, columns (1) and (3) report results for individuals below or equal to the median, while columns (2) and (4) report results for individuals above the median. Working in the private sector is equally associated with higher mobility among high and low cognitive abilities teachers, yet it is only positively correlated with mobility of high non-cognitive ability teachers (p-value: 0.021). Furthermore, table 5 again suggests no relationship between share of minorities and turnover. This is reassuring, as the disadvantaged schools in Sweden do not seem to lose their highly educated, experienced and skilled teachers. Finally, the negative association between mobility and student quality is confirmed for both above median cognitive and non-cognitive aptitude.

Table 6. Teacher quality and school types. The dependent variable is equal to unity if the teacher changes job.

VARIABLES	(1) Lower secondary	(2) Upper secondary	(3) Public	(4) Private
	Panel A: University education and experience measures			
University graduate	-0.005***	0.001	-0.003**	0.003
	(0.002)	(0.002)	(0.001)	(0.006)
Experience	-0.004***	-0.003***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
R-squared	0.144	0.128	0.137	0.096
Observations	295,454	229,622	495,556	29,520
	Panel B: Non-cognitive and cognitive measures			
Non-cognitive score	-0.023***	-0.001	-0.014***	0.012
	(0.006)	(0.006)	(0.004)	(0.017)
Cognitive score	-0.006	-0.010	-0.008	-0.019
	(0.007)	(0.007)	(0.005)	(0.019)
R-squared	0.132	0.113	0.122	0.103
Observations	58,567	56,783	107,020	8,330

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). All regressions include county-by-year fixed effects. Personal characteristics include: gender, immigration status, marital status, indicators for science, vocational and remedial specialization, indicator for temporarily employed, workload, indicators for upper secondary and private school teachers. School characteristics include: student-teacher ratio in full time equivalence, number of students and its square, indicator for schools with less than 100 students, share of girls and immigrants at school, mean percentiled student GPA and mean parental income. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7th that teachers work with. Regressions in panel B exclude gender and immigrant indicator as intellectual assessment is available only for native males. Columns (1) and (2) exclude high school indicator from regressions while columns (3) and (4) exclude private school indicator from regressions.

Since the 1990s there has been a heated discussion in the public debate and among researchers regarding allowing private sector to the public schooling system. Researchers assessed the influence of such changes on student (Ladd, 2002; Sandström and Bergström, 2005; Hsieh and Urquiola, 2006) and teacher (Hoxby, 2002; Hensvik, 2012; Jackson 2012) outcomes. Karbownik (2013) documents differences in turnover rates between private and public school teachers in Sweden. Furthermore, both tables 4 and 5 suggest that teachers of different quality experience differences in mobility depending on whether they work in private or public institution. Table 6 studies differences in mobility for different measures of teacher quality and different types of schools. In particular, columns (1) and (2) present differences between lower and upper secondary schools and columns (2) and (4) illustrate differences between public and private sector.[22]

Significant negative coefficient on university educated teachers is present only in lower secondary and public schools. In the former case, I can statistically reject the difference between lower and upper secondary school estimates (p-value: 0.055), however, in the latter case I am unable to reject the equality of estimates for public and private institutions (p-value: 0.306). These results might be driven by the fact that university graduates in public schools are different from those in private schools. Statistical investigation confirms that among university graduates those teaching in private schools differ significantly from those working in public schools as far as observable socio-economic characteristics are concerned. Nevertheless, even if these correlations are driven by selection into different sectors, they still should draw an attention of policy makers. The estimates on experience are similar for both school types and levels. Finally, I find strong and significant negative relationship between non-cognitive aptitude and mobility for lower secondary schools (p-value: 0.007). Similarly, I find a negative 1.4 pp association for public schools and I can reject that it is equal to insignificant but positive estimate for private schools (p-value: 0.086). Table 6 shows no differences in terms of cognitive skills.

The models used so far pool all destinations of teachers leaving the school together, however, there is research indicating that the correlations with teacher characteristics differ depending on the destination (Lankford et al., 2002). To investigate whether the relationship between teacher quality and teacher turnover depends on destination, I estimate the baseline specifications from tables 2 and 3 separately for mobility within teaching profession as well as out-of-teaching to a different occupation. Panel A reports estimates based on the specification from column (3) in table 2, while panel B reports estimates based on the specification from column (3) in panel C in table 3.

University educated teachers are more likely to change jobs within teaching, than leave for alternative occupations. Similar pattern can be observed as far as experience is concerned, however, here both coefficients in the within and out-of-profession mobility regressions are negative with the one on quits being significantly larger. There is no significant relationship between cognitive skills and either type of mobility. I do, however, find a significant and negative association between non-cognitive skills and both mobility measures. These coefficients are of similar magnitude and cannot be statistically distinguished from one another. Thus, I conclude that better

[22] This distinction is of interest as Karbownik (2013) shows that although there is no relationship between minorities and turnover rate in lower secondary and public schools, it is significant and positive at the upper secondary level and in private schools.

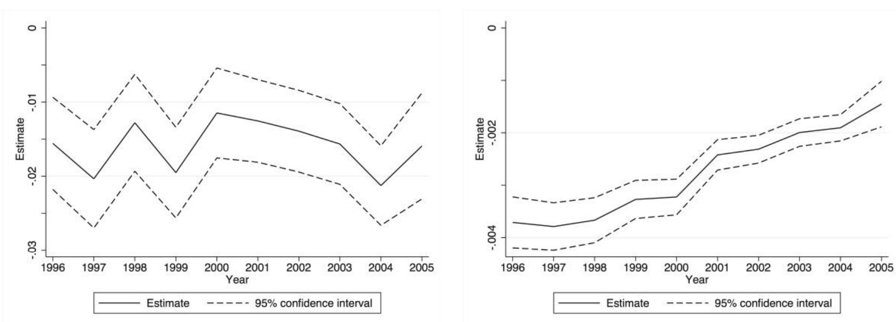
teachers are less likely to leave the profession and this is most pronounced in terms of formal education.

Table 7. Analyses by different destinations.

VARIABLES	(1) Within	(2) Quit
	Panel A: University education and experience measures	
University graduate	0.013***	-0.017***
	(0.001)	(0.001)
Experience	-0.001***	-0.003***
	(0.000)	(0.000)
R-squared	0.032	0.108
Observations	525,076	525,076
	Panel B: Non-cognitive and cognitive measures	
Non-cognitive score	-0.007***	-0.006*
	(0.003)	(0.003)
Cognitive score	-0.003	-0.006
	(0.003)	(0.004)
	0.027	0.099
	115,350	115,350

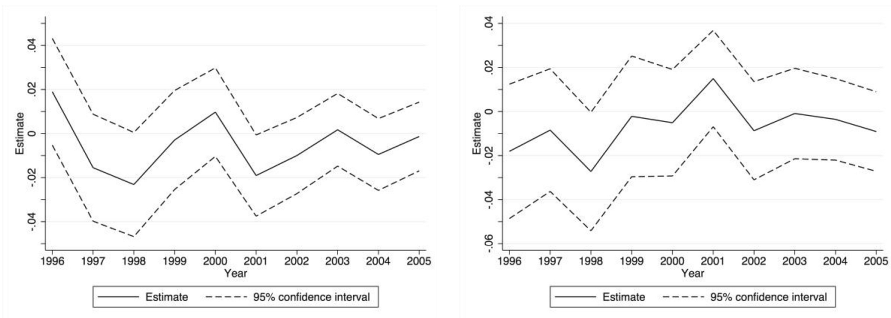
Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). All regressions include county-by-year fixed effects. Personal characteristics include: gender, immigration status, marital status, indicators for science, vocational and remedial specialization, indicator for temporarily employed, workload, indicators for upper secondary and private school teachers. School characteristics include: student-teacher ratio in full time equivalence, number of students and its square, indicator for schools with less than 100 students, share of girls and immigrants at school, mean percentiled student GPA and mean parental income. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7th that teachers work with. Regressions in panel B exclude gender and immigrant indicator as intellectual assessment is available only for native males.

Figure 5. Leaving the profession and teacher quality - estimates over time. University education and experience.



In their paper, Grönqvist and Vlachos (2008) document a declining quality of new teachers entering the profession in Sweden. In this research, I show that this decline in teacher quality is partly offset by a lower tendency for high-quality teachers (educated and experienced) to quit teaching. This tendency stayed roughly constant over time for university education and intellectual measures, with the former one being consistently negative and the latter two bouncing around zero (figures 5 and 6). It is, however, becoming less and less negative in terms of experience. In fact, between 1996 and 2005 the association between quitting teaching and experience decreased by a half. Finally in this paper, I focus on teacher quality conditional on the selection into teaching and show that among the pool of teachers who decide to pursue a teaching career it is the lower skilled ones who exit. My results do not give any insight about the total population of potential teachers, and in that sense cannot be directly compared to Fredriksson and Öckert (2007) who show that individuals with higher abilities do not enter teaching profession after teacher's training.

Figure 6. Leaving the profession and teacher quality - estimates over time. Cognitive and non-cognitive skills.



7 Conclusions

The contemporary literature on the teacher mobility lacked a detailed study relating turnover rates to teacher quality. Although high turnover rates may state a problem to some schools, principals (and students) are probably more concerned about the quality of teachers leaving the school. In particular, losing skilled teachers may be especially problematic for schools with many disadvantaged students. This paper attempts to fill in this gap in the literature on teacher turnover using unusually rich data on teacher skills for Swedish secondary school teachers covering years 1996/1997 to 2006/2007.

The results indicate that in Sweden schools do not seem to lose university educated and experienced teachers, and such teachers also do not leave the profession. In particular, teachers with high non-cognitive skills are less likely to change employers. This suggests that the drop in teacher quality documented by others is partly offset by lower tendency for high-quality teachers to leave the profession. I do not find any support for the common view that schools serving minority students experience high turnover rates and outflow of high quality pedagogues. There is no evidence that a higher share of minority enrollment correlates positively with quits of high quality teachers. Finally, a somewhat speculative interpretation of the results is that it may be possible to influence teacher's mobility decision through changes in their monetary compensations or type of employment.

References

Aaronson, D., L. Barrow, and W. Sander (2007), “Teachers and student achievement in the Chicago public high schools”, *Journal of Labor Economics*, 25:1, 95-135.

Bacolod, P.M., (2007), “Do alternative opportunities matter? The role of female labor markets in the decline of teacher quality”, *Review of Economics and Statistics*, 89:4, 737-751.

Barbieri, G., C. Rossetti, and P. Sestito (2011), “The determinants of teacher mobility: Evidence using Italian teachers’ transfer applications”, *Economics of Education Review*, 30:6, 1430-1444.

Baugh, H.W., and J.A. Stone (1982), “Mobility and wage equilibrium in the educator labor market”, *Economics of Education Review*, 2:3, 253–274.

Björklund, A., M.A. Clark, P-A. Edin, P. Fredriksson, and A. Krueger (2006), “The market comes to education in Sweden: An Evaluation of Sweden’s surprising school reforms”, Russel Sage Foundation, New York.

Bonhomme, S., G. Jolivet, and E. Leuven (2011), “Estimation of preferences for job attributes using job–to–job mobility: teacher turnover and school characteristics”, mimeo CREST.

Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff (2005), “Explaining the short careers of high – achieving teachers in school with low performing students” *American Economic Review*, 95:2, 166–171.

Böhlmark, A., and M. Lindahl (2007), “The impact of school choice on pupil achievement, segregation and costs: Swedish evidence”, IZA Working Paper 2786.

Böhlmark, A., and M. Lindahl (2008), “Does school privatization improve educational achievement? Evidence from Sweden’s voucher reform”, IZA Working Paper 3691.

Böhlmark, A., E. Grönqvist, and J. Vlachos (2012), “The headmaster ritual: The importance of management for school outcomes”, IFAU Working Paper 2012:16.

Brewer, J.D., (1996), “Career paths and quit decisions: evidence from teaching”, *Journal of Labor Economics*, 14:2, 313–339.

Carlstedt, B., (2000), “Cognitive abilities: Aspects of structure, process and measurement”, University of Gothenburg Doctoral Dissertation.

Clotfelter, T.C., H.F. Ladd, and J.L. Vigdor (2007), “Teacher credentials and student achievement: longitudinal analysis with student fixed effects”, *Economics of Education Review*, 26:6, 673-682.

Clotfelter, T.C., H.F. Ladd, and J.L. Vigdor (2011), “Teacher mobility, school segregation, and pay-based policies to level the playing field”, *Education Finance and Policy*, 6:3, 399-438.

Corcoran, P.S., W.N. Evans, and R.M. Schwab (2004), “Changing labor-market opportunities for women and the quality of teachers, 1957-2000”, *American Economic Review*, 94:2, 230-235.

Dolton, P., and W. van der Klaauw (1995), “Leaving teaching in the UK: a duration analysis”, *Economic Journal*, 105:429, 431–444.

Dolton, P., and W. van der Klaauw (1999), “The turnover of teachers: a competing risk explanation”, *Review of Economics and Statistics*, 81:3, 543–552.

Dolton, P., and O.D. Marcenaro-Gutierrez (2011), “If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance”, *Economic Policy*, 26:65, 5-55.

Edin, P-A., B. Holmlund, and O. Nordström Skans (2009), “Wage Dispersion Between and Within Plants: Sweden 1985-2000”, in E. Lazear and K. Shaw (eds.)

The Structure of Wages: an International Comparison, University of Chicago Press.
Ehrenberg, G.R., and D.J. Brewer (1994), "Do school teacher characteristics matter? Evidence from High School and Beyond", *Economics of Education Review*, 13:1, 1-17.
Falch, T., and B. Strøm (2005), "Teacher turnover and non–pecuniary factors", *Economics of Education Review*, 24:6, 611–631.
Falch, T., (2010), "The elasticity of labor supply at the establishment level", *Journal of Labor Economics*, 28:2, 237-266.
Falch, T., (2011), "Teacher mobility responses to wage changes: Evidence from a quasi – natural experiment" *American Economic Review*, 101:3, 460-465.
Feng, L., (2009), "Opportunity wages, classroom characteristics, and teacher mobility", *Southern Economic Journal*, 75:4, 1165-1190.
Feng, L., (2010), "Hire today, gone tomorrow: New teacher classroom assignment and teacher mobility", *Education Finance and Policy*, 5:3, 278-316.
Figlio, D., (1997), "Teacher salaries and teacher quality", *Economics Letters*, 55:2, 267-271.
Figlio, D., (2002), "Can public schools buy better-qualified teachers?", *Industrial and Labor Relations Review*, 55:4, 686-699.
Fredriksson, P., and B. Öckert (2007), "The supply of skills to the teacher profession", Mimeo Uppsala University.
Fredriksson, P., and B. Öckert (2008), "Resources and student achievement – evidence from a Swedish policy reform", *Scandinavian Journal of Economics*, 110:2, 277-296.
Grönqvist, E., and J. Vlachos (2008), "One size fits all? The effects of teacher cognitive and non-cognitive abilities on student achievement", IFAU Working Paper 25.
Harbison, W.R., and E.A. Hanushek (1992), "Educational performance of the poor. Lessons from rural northeast Brazil", Oxford University Press.
Hanushek, E.A., (1971), "Teacher characteristics and gains in student achievement: Estimation using micro data", *American Economic Review*, 61:2, 280-288.
Hanushek, E.A., F.J. Kain, and G.S. Rivkin (2004), "Why public schools loose teachers?", *Journal of Human Resources*, 39:2, 326–354.
Harris, N.D., and T.R. Sass (2011), "Teacher training, teacher quality and student achievement", *Journal of Public Economics*, 95:7, 798-812.
Hensvik, L., (2012), "Competition, wages and teacher sorting: four lessons from a voucher reform", *Economic Journal*, 122:561, 799-824.
Hsieh, T-C., and M. Urquiola (2006), "The effects of generalized school choice on achievement and stratification: Evidence from Chile's school voucher program", *Journal of Public Economics*, 90:8, 1477-1503.
Hoxby, M.C., (2002), "Would school choice change the teaching profession?", *Journal of Human Resources*, 37:4, 846-891.
Jackson, C.K., (2009), "Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation", *Journal of Labor Economics*, 27:2, 213-256.
Jackson, C.K., (2012), "School competition and teacher labor markets: Evidence from charter school entry in North Carolina", *Journal of Public Economics*, 96:5-6, 431-438.
Jackson, C.K., (2013), "Match quality, worker productivity and worker mobility: Direct evidence from teachers", *Review of Economics and Statistics*, forthcoming.

Karbownik, K., (2013), "The determinants of teacher mobility in Sweden", Mimeo Uppsala University.

Ladd, F.H., (2002), "School vouchers: A critical view", *Journal of Economic Perspectives*, 16:4, 3-24.

Lankford, H., S. Loeb, and J. Wyckoff (2002), "Teacher sorting and the plight of urban schools: the importance of alternative labor market opportunities and non – pecuniary variation", *Educational Evaluation and Policy Analysis*, 24:1, 37-62.

Leigh, A., and C. Ryan (2008), "How and why has teacher quality changed in Australia?", *Australian Economic Review*, 41:2, 141-159.

Lindqvist, E., and R. Vestman (2011), "The labor market returns to cognitive and noncognitive ability: evidence from the Swedish enlistment", *American Economic Journal: Applied Economics*, 3:1, 101-128.

Lindholm, A., (2006), "Lärarförbundet och lärarnas riksförbund medlemsundersökning om lön och arbetstid".

Manning, A., (2011), "Imperfect competition in labor markets", in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, 4:B, 973-1041.

Mood, C., J.O. Jonsson, and E.Bihagen (2012), "Socioeconomic persistence across generations: The role of cognitive and non-cognitive processes", in J. Ermisch, M. Jantti, and T. Smeeding (eds.) *Inequality from childhood to adulthood. A cross-national perspective of the transmission of advantage*, New York: Russell Sage.

Murnane, J.R., and R.J. Olsen (1990), "The effects of salaries and opportunity costs on length of stay in teaching: evidence from North Carolina", *Journal of Human Resources*, 25:1, 106–124.

National Agency for Education, (2003), "Behov av och tillgång på pedagogisk personal – Skolverkets bedömning 2003".

Nickell, S., and G. Quintini (2002), "The consequences of the decline in public sector pay in Britain: a little bit of evidence", *Economic Journal*, 112:477, F107-F118.

Rivkin, S.G., E.A. Hanushek, and J.F. Kain (2005), "Teachers, schools, and academic achievement", *Econometrica*, 73:2, 417–458.

Rockoff, E.J., (2004), "The impact of individual teachers on student achievement: evidence from panel data", *American Economic Review*, 94:2, 247-252.

Rockoff, E.J., and C. Speroni (2011), "Subjective and objective evaluations of teacher effectiveness: evidence from New York City", *Labour Economics*, 18:5, 687-696.

Rothstein, J., (2010), "Teacher quality in educational production: tracking, decay and student achievement", *Quarterly Journal of Economics*, 125:1, 175-214.

Sandström, F.M., and F. Bergström (2005), "School vouchers in practice: Competition will not hurt you", *Journal of Public Economics*, 89:2, 351-380.

Scafidi, B., D.L. Sjoquist, and T.R. Stinebrickner (2007), "Race, poverty, and teacher mobility", *Economics of Education Review*, 26:2, 145–159.

Appendix

Tables

Table A1. Estimation results on a sample of municipalities with limited reductions in teacher stock. The dependent variable is equal to unity if the teacher changes job.

VARIABLES	(1) Mobility	(2) Mobility	(3) Mobility	(4) Mobility
University graduate	-0.041***	-0.010***	-0.009***	-0.007**
	(0.004)	(0.003)	(0.003)	(0.003)
Experience	-0.009***	-0.004***	-0.004***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
R-squared	0.060	0.137	0.138	0.146
Observations	129,275	129,275	129,275	129,275
Personal characteristics		X	X	X
School characteristics			X	X
Log-earnings				X

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). All regressions include county-by-year fixed effects. Personal characteristics include: gender, immigration status, marital status, indicators for science, vocational and remedial specialization, indicator for temporarily employed, workload, indicators for upper secondary and private school teachers. School characteristics include: student-teacher ratio in full time equivalence, number of students and its square, indicator for schools with less than 100 students, share of girls and immigrants at school, mean percentiled student GPA and mean parental income. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7^{th} that teachers work with. Sample reduced to municipalities, which do not experience reductions in teacher stock of more than 5% over the studied period.

Table A2. Estimation results on a sample of municipalities with limited reductions in teacher stock. The dependent variable is equal to unity if the teacher changes job.

VARIABLES	(1) Mobility	(2) Mobility	(3) Mobility	(4) Mobility	(5) Mobility	(6) Mobility
Panel A: Non-cognitive assessment unconditional on cognitive score						
Non-cognitive score	-0.040***	-0.010	-0.011	-0.006	-0.008	-0.004
	(0.010)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
R-squared	0.013	0.125	0.127	0.140	0.131	0.142
Panel B: Cognitive assessment unconditional on non-cognitive score						
Cognitive score	-0.039***	-0.008	-0.007	-0.002	0.002	0.002
	(0.012)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
R-squared	0.013	0.125	0.127	0.140	0.131	0.142
Panel C: Both scores included.						
Non-cognitive score	-0.035***	-0.010	-0.011	-0.006	-0.009	-0.005
	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Cognitive score	-0.033***	-0.007	-0.005	-0.001	0.003	0.003
	(0.012)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
R-squared	0.014	0.125	0.127	0.140	0.131	0.142
Observations	28,874	28,874	28,874	28,874	28,874	28,874
Personal characteristics		X	X	X	X	X
School characteristics			X	X	X	X
Log-earnings				X		X
University and experience					X	X

Note: Standard errors clustered at school level (*** p<0.01, ** p<0.05, * p<0.1). All regressions include county-by-year fixed effects. Personal characteristics include: marital status, indicators for science, vocational and remedial specialization, indicator for temporarily employed, workload, indicators for upper secondary and private school teachers. School characteristics include: student-teacher ratio in full time equivalence, number of students and its square, indicator for schools with less than 100 students, share of girls and immigrants at school, mean percentiled student GPA and mean parental income. All regressions corrected for school mergers and dissolutions as well as for mobility in grades below 7^{th} that teachers work with. Sample reduced to municipalities, which do not experience reductions in teacher stock of more than 5% over the studied period.

Details of sample construction

I construct the sample of lower and upper secondary school teachers for the school years 1996/1997 to 2006/2007. The information about teachers comes from the teacher registry and the analysis focuses on teachers working in grades 7-9 (lower secondary school) of compulsory education and in grades

1-3 (upper secondary school) of secondary education. The reason for restricting the analysis to these grade levels, is that I lack information on student characteristics for lower levels. Teachers who are on unpaid leave of absence or whose workloads are zero hours (i.e., they do not perform any pedagogical duties) are excluded from the analysis. Such teachers are treated neutrally in terms of mobility if they come back after the absence period to the same school. Similarly, I exclude teachers who are employed as principals, study counselors etc. In each year if a teacher has multiple entries in the registry, the observation with the highest workload is selected irrespectively whether it is at the same or at different schools.[23] The teacher registry is a high quality data set, that allows recovering information on school location (unique identifier), school ownership and type, teacher certification, workload, employment type (temporary vs. permanent), education and position. The construction of teaching experience is presented in the descriptive statistics section.

Teachers are grouped into either lower or upper secondary education and teachers working in grades 7-9 are recovered by merging the teacher registry to the pupil registry via unique school identifier. There exist schools with more grades covered under the same school identifier (i.e. 1-9 or 4-9) and one possible source of bias would be, for instance, relating teachers who work with students in grades 1-3 to school characteristics measured for students in grades 7-9. Since I have information about the grades in which teachers work I address this issue by excluding teachers coded as primary (grades 1-3) and middle (grades 4-6) school teachers. Such a procedure does not solve the problem completely as some teachers (arts or music) are not necessarily coded by grades. Thus, I may still include some miscoded teachers, however, the share of miscoded teachers is likely low. Nonetheless, each included school serves grades 7-9 and only turnover between such schools is considered at lower secondary level.

Teachers are then linked (using unique identifier) to population registry, which covers all individuals living in Sweden. The population registry is a high quality data set that allows recovering information on gender, marital status, age, family composition (using unique family identifier), immigration history, education and income. Income is measured as a gross salary plus income from business and self-employment plus any work-related allowances. Investment losses are not included, and thus, income is lower-bounded at zero. The analysis is restricted to teachers aged 25-58 years, to abstract from mobility driven by educational attainment and retirement decisions.

The earnings registry often contains multiple entries per individual, which reflect different sources of labor compensations but are uniquely identifiable based on establishment identifier. This poses linking problem for individuals with multiple entries as I may miss-assign earnings from different establish-

[23] The workload of teachers having multiple positions at the same school is not summed and the highest workload position is selected.

ment to a particular school code. Since there is no direct link between unique school code and establishment identifiers, I create such a link using a mode rule. In particular, based on the individuals with only one record I define most often occurring establishment identifier for each school code. I then use this data to resolve matching of individuals with multiple earnings entries.

The students' characteristics are based on "school in" and "school out" pupil registries. The lower secondary school composition is based on outgoing students. The quality of students in lower secondary school is measured based on their 9th grade outgoing grades. The measure is calculated for year t as a mean percentiled GPA from cohorts graduating in years t+1, t+2 and t+3.

The upper secondary school composition is based on all the students that are in a given school in a particular year. The quality of students in upper secondary school is measured based on their 9th grade grades. The main advantage of using lower secondary school grades as a measure of upper secondary school quality is that it is largely exogenous to upper secondary school teachers. I match these students to their parents using unique family identifier and obtain the family level socioeconomic indicator i.e. mean parental income.

The enlistment registry covers period 1969 to 2006 and provides information on cognitive and non-cognitive assessments. Each of the parts that contribute to a final cognitive score is graded on 1 to 9 scale, and the final score is given in the same format. To make the variable more continuous and utilize all the information I predict the final score using its separate components. I obtain a variable with mean 97.4 and standard deviation of 23.7. The non-cognitive score is based on 1 to 9 scale and each of the four contributing personality traits is rated on 1 to 5 scale. Here again, I utilize all this information and I predict the final score using separate components. Then, I percentile rank all the male, native individuals by type of assessment and year of draft. This procedure yields ranking of individuals in every test in every draft year for the whole tested population. The data is linked to teacher registry via unique personal identifier and scores are assigned to native, male teachers for whom the data is available.

Finally, having data with teachers and students I match the two using a unique school identifier. Naturally since the mobility itself is a lagged variable school year 2006/2007 is dropped from the analysis. The final sample includes 136 100 teachers and 622 453 person–year observations. I exclude the following observations from the sample: very small schools with number of teachers in full time equivalence less than 3 (5 170 observations), teachers that are below 25 years old (8 370 observations), teachers that are above 58 years old (82 298 observations), and schools with the number of students less than 15 (1 539 observations). The final sample consists of 121 580 teachers, 2703 unique schools and 525 076 person-years. Applying the intellectual assessment sample restriction further reduces the sample to 26 235 teachers, 2628 unique schools and 115 350 teacher-years.

Essay 3

Do changes in student quality affect teacher mobility? Evidence from an admission reform

1 Introduction

Many educational interventions such as student busing, school choice or changes in admission policies impact the composition of students in schools. The interventions have been motivated by the idea that it could be beneficial for certain groups of students to meet better peers while keeping constant other inputs of the education production function.[1] This hypothesis relies heavily on the assumption that inputs of the educational production function are exogenous to student characteristics. It is quite possible, however, that low-performing students impose a heavier burden on teaching. Thus, if changes in student composition affect other factors of input such as teacher quality or school resources (Hanushek, 1986), then polices aimed at changing the peer group composition in schools may have unintended consequences.

In this paper, I study how exogenous changes in student composition affect teacher mobility. In particular, I investigate whether teachers who experience an inflow of high quality students are less likely to quit their jobs in comparison to teachers who face an inflow of lower quality students. Multiple correlational studies suggest that teacher mobility is negatively related to pupil quality.[2] At the same time, we know relatively little about whether this descriptive relationship can be given a causal interpretation, with the exception of a busing policy study by Jackson (2009). However, due to the nature of the policy, he focuses primarily on racial sorting, which only has a secondary relationship with student quality.[3]

Uncovering the causal relationship between student quality and teacher mobility should be a central priority for policy makers for two reasons. First, if worse quality students induce teachers, particularly of high quality, to leave their schools, then the problem with an inflow of less able students may be reinforced by higher teacher turnover and by unfavorable sorting of teachers. Second, the potential positive effects of policies aimed at reshuffling students between schools may be dwarfed by teacher mobility if high quality teachers leave in response to an inflow of low quality pupils. Hoxby (2002) shows that school choice creates a more high-powered incentive environment within the teaching profession and requires teachers to have higher levels of human capital and effort in return for higher marginal wages for such characteristics. In other words, it is possible that, on the margin,

[1] Examples of policies that lead to reshuffling of peers are: increased freedom in school choice (Cullen et al, 2006); school voucher programs (Hsieh and Urquiola, 2006); student busing (Jackson, 2009); increased competition from the private sector (Jackson, 2012; Hensvik, 2012); changes in school admission policies (Söderström and Uusitalo, 2010); and court-ordered desegregation (Reber, 2005).

[2] For example: Hanushek et al. (2004) for Texas; Falch and Strøm (2005) for Norway; Scafidi et al. (2007) for Georgia (US); Karbownik (2013) for Sweden.

[3] A third quasi-experimental study in the relevant literature is Feng et al. (2010), who study the effects of changes in school resources on teacher mobility.

schools enrolling more high quality students face pressure to retain high quality and fire low quality teachers. At the same time, schools that are adversely shocked may not be able to keep their best teachers, thus lowering the school quality even more.

I explore a major reshuffling of students induced by an admission reform introduced in the municipality of Stockholm, Sweden, in the fall of 2000. Prior to the reform, students applied only for a program and their grades from lower secondary school determined admission. Students could state their preferences for the school they would like to attend, but those living closest to a school had priority. Thus, although the program choice included an element of school choice, it essentially limited the choice of students living in less affluent neighborhoods as they never had a chance to be admitted to permanently oversubscribed programs in prestigious downtown schools.[4] The 2000 reform abolished all residence-based admission criteria and introduced a system that is based solely on lower secondary school performance. The reform was intended to undo the effects of residential segregation and to give the option of attending the most prestigious schools in downtown Stockholm to all students, irrespectively of where they lived.

I make use of rich registry data and a difference-in-differences strategy to identify the effect of student quality on teachers' decisions to leave their current employment. Since the composition of students changed exogenously and teachers faced students of utterly different quality before versus after the reform, the estimate can be treated as teacher preference for student quality, under certain theoretical assumptions on the teacher's utility function. In Section 7, I also consider a broader school-level perspective of the reform. In particular, I investigate whether the reform affected schools' hiring policies and if it changed an individual teacher's monetary compensation.

I find that a 10-percentile-point decline in average incoming student credentials increases 4-year separation rates by up to 9 percentage points (pp). The effect is driven primarily by teachers switching schools rather than teachers leaving the profession, and it is concentrated at the bottom half of the student quality distribution. The estimated effect is statistically and economically significant and similar across groups of teachers whose baseline mobility is very different. Furthermore, teachers seem to react to the direct measures of student quality. Once student credentials are taken into account, other characteristics like immigration background become unrelated to teacher mobility. Finally, I do not find any significant effects of changes in student quality on an individual teacher's earnings or school hiring policies.

[4] Although, Stockholm has a very well developed public transportation system, its housing market is highly regulated. It is much easier to buy or rent a flat in a low quality neighborhood and commute within the city than it is to get housing in an affluent location and cut down on transportation costs and time. This feature becomes even more important if the school admission system is, for the most part, residence based.

The reminder of the paper is organized as follows. Section 2 gives details regarding educational institutions in Sweden: the reform, data used and identifying variation. Section 3 presents a simple theoretical framework for teacher mobility and sets up the empirical analysis while Section 4 contains main results. Section 5 presents sensitivity analyses, while Section 6 includes heterogeneity analyses and Section 7 extends the analysis to school-level responses to the change in student quality. Finally, Section 8 concludes.

2 Institutions, reform, data and identification

2.1 Educational institutions in Sweden

The Swedish schooling system starts with voluntary pre-school and continues with nine years of compulsory education. Lower secondary school covers grades 7 to 9. The grades received in 9^{th} grade determine a student's chances to advance to upper secondary (high) school. Swedish municipalities are obliged by law to provide upper secondary schooling to all students who successfully completed compulsory education. Upper secondary school consists of different programs, lasts three years and typically provides eligibility for post-secondary education.

Private schooling is growing in Sweden and is encouraged by the government.[5] In 1992, Sweden introduced a school voucher reform that allowed for both non-profit and for-profit independent schools. The municipality is obliged to pay the independent schools for each student they can attract, with an amount corresponding roughly to the average per-student cost in the public schools.[6]

The teaching profession in Sweden is regulated and different qualifications are required depending on the subject taught and on the type of school. Teaching at the secondary school level requires completing special coursework beyond what is required from a compulsory school teacher. Individuals from other professions who want to become teachers need to supplement their professional degrees with a minimum of 1.5 years of preparation in pedagogy, didactics and teaching practice. However, uncertified teachers could also be hired on short-term contracts.

[5] The fraction of independent high schools has risen from 7.5% in the 1994/1995 school year to 32.0% in the 2004/2005 school year.

[6] An independent school receives around 85-95% of the average per-student cost in public schools and this amount varies from year to year. Some municipalities also have a socioeconomic gradient for the school voucher. Private schooling was effectively introduced at the lower secondary level in 1992, and at the upper secondary level in 1994 (Böhlmark and Lindahl, 2007).

Municipalities are the primary employers of teachers in Sweden, and thus, handle the responsibility of recruiting them.[7] In practice, however, the decisions regarding recruitment, selection and employment of a teacher are made at the school level by a principal. Finally, teacher wages are determined at the local level through individual bargaining between a teacher and a principal, given the collective bargaining outcome set at the national level.[8]

2.2 The admission reform

In the fall of 1999 the municipality of Stockholm passed legislation that changed the high school admission rules. Up to the 1999/2000 school year, students applied only for a program and their grades from lower secondary school determined admission. Students could state their preferences for which school they would like to attend, but those living closest to a school had priority. In practice, the educational administration first counted the number of places per program in any given municipality and then ranked the student choices according to grades, and accepted students to a certain program. Subsequently, they assigned the students to the specific schools based on their residence, and thus, assuming competitive grades, it was possible to get accepted into a better program in a school further away, but only if it was not oversubscribed with students residing in the neighborhood. For example, if school A, located in downtown Stockholm, excelled in a science program and there were enough students living nearby who subscribed to the program, then students with better grades residing in Tensta (a relatively poor and disadvantaged district in Stockholm) would be unable to gain admission to the program.[9] In particular, the restriction was binding for the two most popular and broadest programs: social sciences (samhällskunskap) and natural sciences (naturvetenskap). Generally, those from low-income, disadvantaged districts had virtually no chance of attending the most popular inner-city schools, even if they had competitive grades.

The cohort applying to high school in May 2000 for the 2000/2001 school year faced utterly different admission criteria. In line with the legislation, all residence-based school allocation within the municipality of Stockholm was abolished and replaced by a system based exclusively on grades from the 9th grade in lower-secondary school. In this paper, the grades of incoming high school students is the variable of interest.[10] In the new system, students apply for a specific program in a specific school and applicants are ranked by

---

7 For more information on the reform that shifted responsibility for schooling from the central government to municipalities see Fredriksson and Öckert (2008).

8 Individualized pay was introduced in 1996 and is discussed in detail by Hensvik (2012), in a survey by Lindholm (2006) and in a report by Skolverket (2009).

9 Independent high schools were allowed to select students on the basis of GPA also before the reform and there were no geographical restrictions in applying to these schools.

10 From here on, I refer to grades as students' credentials, student quality or student GPA.

schools and programs. If a student's first choice is not accepted, the second choice is considered, and so on. Importantly, this reform was introduced only in the municipality of Stockholm, and thus, the rest of Stockholm County was not affected.

It is important to note that most municipalities surrounding Stockholm do not offer all of the programs, and a student has the right to attend their chosen program in another municipality, financed by the municipality in which they reside. Cross-municipality commuting is relatively common in Sweden, and if increased school choice incentivizes more students from out-of-Stockholm to apply to schools in Stockholm, then they may crowd out students residing in Stockholm. Furthermore, Stockholm schools may decide to change the number of admitted students in response to higher demand for quality, which would in turn lead to either lower student-teacher ratio, and thus, impoverishment of school resources, or to the need for additional hires. I address the latter issue in Section 4. Finally, my calculations show that the fraction of students living outside of Stockholm municipality but attending Stockholm schools is stable at around 20% over the analyzed period.

Söderström and Uusitalo (2010) found clear evidence that the Stockholm admission reform affected both student mobility and the sorting of students by quality. In particular, the grade-based admission system increased the sorting of students to schools according to their ability, as well as ethnic and socio-economic background. However, the segregation between immigrants and natives increased more than one would expect as a result of increased sorting by ability. Edin et al. (2011) used the same strategy to evaluate the effects on student outcomes. They find either zero or negative effects on student performance. The authors conclude that their results do not support the idea that choice and competition improve performance. One possible mechanism behind this finding could be that schools that face inflowing students of poorer quality may also lose their best teachers.[11] Thus, this study evaluates how the resorting of students between schools in Stockholm affected teacher turnover rates.

2.3 Data and descriptive statistics

This paper utilizes Swedish population-wide registries. The main data source is the teacher registry that covers all teaches employed in Swedish schools during the 1991/1992 through 2004/2005 school years. It contains information on teachers' education, specialization, experience, certification, place of work, type of contract (permanent vs. temporary) and workload. I have matched background information on age, gender, immigration histories, education, employment and income to these data. The pupil registers for

[11] This assumes that there is a positive interaction effect between student quality and teacher quality in the production of student skills.

lower and upper secondary schools are used to obtain information on students in a given upper secondary school and their credentials from lower secondary school. All students have also been matched to their parents to obtain measures of family background. Administrative records on earnings provide information on teachers' monetary compensations. The details of the sample construction are discussed in the appendix.

Given the timing and the geographical implementation of the reform, I focus on secondary schools that have been in operation in Stockholm for all school years from 1991/1992 to 2004/2005. This avoids potential composition effects related to school openings and closures. However, all the results carry over if I use repeated cross-section of schools. Due to the reform implementation date there are no independent high schools in the 1991/1992 to 2004/2005 panel sample. I can observe, however, if a teacher leaves their current school in favor of a privately run institution. In the pooled sample of all secondary schools in Stockholm prior to the 1999/2000 school year there are 8 private schools out of 29 schools in total.

Since the reform was only implemented in the municipality of Stockholm, it is important for potential generalizations of the results to gauge how comparable the Stockholm population is to the overall population of teachers and schools in Sweden. Table A1 compares basic descriptive statistics for Stockholm and non-Stockholm schools for the last pre-reform (1999/2000) and first post-reform school year (2000/2001). It is clear that Stockholm is more affluent in many dimensions than the rest of Sweden. Schools in Stockholm admit students with higher credentials, who come from richer and better educated families, and whose fathers obtain higher cognitive and non-cognitive scores during military assessment. At the same time, these schools admit more minority students, which is not surprising given that Stockholm has a major concentration of immigrants to Sweden. Stockholm schools also have the advantage of employing more teachers with university diplomas; however, the teachers are on average less experienced.

The reform was implemented in the 2000/2001 school year, and thus, as a starting point, I present descriptive evidence for the 1999/2000 school year as the last pre-reform year and the 2000/2001 school year as the first post-reform year.[12] This paper focuses on the exogenous reshuffling of students within the municipality of Stockholm and responses of teachers when they face a different set of pupils. Therefore, in Table 1, I present descriptive statistics from the 1991/1992 to 2004/2005 panel of Stockholm schools for the immediate pre- and post-reform periods, separated by changes in their student composition. In particular, for each school j in the panel, I calculate the difference between mean-incoming-student credentials in the first post-reform year, 2000/2001, and the last pre-reform year, 1999/2000. Then, I order these differences from the schools most negatively affected to those

[12] Later in the paper I discuss, test and account for possible anticipation effects.

most positively affected and divide the ranking into tertiles. I call these schools downward, middle and upward shocked schools. The bottom of the table reports the number of schools and teachers in each group.

Table 1. Descriptive statistics – panel of Stockholm schools. Comparison across treatments.

Variables	Pre-reform = 1999			Post-reform = 2000		
	Change in student credentials					
	(1) 1/3 downward	(2) 1/3 middle	(3) 1/3 upward	(4) 1/3 downward	(5) 1/3 middle	(6) 1/3 upward
Outcome variable						
One-year mobility	0.10	0.16	0.11	0.10	0.13	0.08
	(0.30)	(0.37)	(0.32)	(0.30)	(0.33)	(0.27)
Treatment variable						
Incoming students' credentials	50.45	54.84	62.31	45.70	57.36	71.92
	(11.82)	(20.06)	(10.36)	(10.40)	(20.09)	(14.19)
Teacher characteristics						
Fraction of female teachers	0.56	0.53	0.45	0.53	0.56	0.46
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Mean teacher experience	13.00	11.13	11.79	13.01	11.21	11.46
	(6.92)	(7.04)	(7.59)	(7.49)	(7.03)	(7.73)
Fraction of teachers with university diploma	0.77	0.66	0.74	0.75	0.68	0.77
	(0.42)	(0.47)	(0.44)	(0.43)	(0.47)	(0.42)
Fraction of teachers employed on temporary contracts	0.18	0.26	0.24	0.22	0.27	0.23
	(0.38)	(0.44)	(0.43)	(0.42)	(0.45)	(0.42)
Mean yearly teacher earnings in 1000 SEK	245	218	216	248	231	223
	(84)	(76)	(80)	(88)	(77)	(88)
Student characteristics (alternative treatment variables)						
Share of immigrants	0.18	0.12	0.10	0.23	0.11	0.09
	(0.08)	(0.05)	(0.05)	(0.09)	(0.03)	(0.04)
Mean yearly parental income in 1000 SEK	346	404	422	330	438	486
	(44)	(100)	(58)	(48)	(81)	(106)
Mean parental education	12.34	13.12	13.66	12.37	13.22	13.77
	(0.55)	(1.33)	(0.90)	(0.43)	(1.28)	(0.83)
Mean paternal draft score	54.81	57.39	58.08	53.83	55.96	58.34
	(6.12)	(7.13)	(4.20)	(8.10)	(6.93)	(3.84)
Number of schools	5	5	5	5	5	5
Number of teachers	266	238	274	260	240	312

Note: Means and standard deviations. Columns (1) to (3) present descriptive statistics for the last pre-reform year while columns (4) to (6) present descriptive statistics for the first post-reform year. All descriptive statistics are based on the panel sample of Stockholm schools in operation between 1991 and 2004 and refer to incoming first year students as far as aggregate school characteristics are concerned. For each characteristic I report descriptive statistics for teachers and schools affected differently by the reform. In particular, columns (1) and (4) describe a third of most downward shocked schools. Columns (3) and (6) describe a third of most upward shocked schools. Columns (2) and (5) describe a third of middle tertile schools. Shock is defined as a difference between mean students' credentials measured by primary school 9th grade GPA (only first-grade students who applied to school in the same year) in high school j in the first post-reform year 2000 and mean students' credentials in the last pre-reform year 1999 in these same schools.

As is evident from Table 1, the reform indeed reshuffled incoming first-grade pupils between schools in Stockholm. In particular, student GPA in the top schools increased from 62.3 to 71.9 percentile points while it decreased at the bottom schools from 50.5 to 45.7 percentile points, widening the gap between best and worst schools from less than 12 to over 26 percentile points. This is equivalent to over two-thirds of a standard deviation change in student quality.

At the same time, other student characteristics correlated with student quality, such as parental income or share of minorities, also changed. For example, the gap between the best and worst schools in terms of mean parental income doubled, while in terms of the share of minority students increased by 75 percent. As a result of teacher turnover and school hiring decisions, the reform also affected the composition of the teacher stock. For example, there were on average more teachers with university diplomas in the upward shocked schools and more teachers on temporary contracts in downward shocked schools in the post-reform period in comparison to the pre-reform period. The gap in teacher compensation did not seem to widen, as it actually decreased from 29000 to 25000 Swedish Kronor. Interestingly, teachers in schools with better students earned less than those in schools with low-quality students, suggesting the presence of compensating wage differentials in a system with fairly flexible teacher pay scheme.

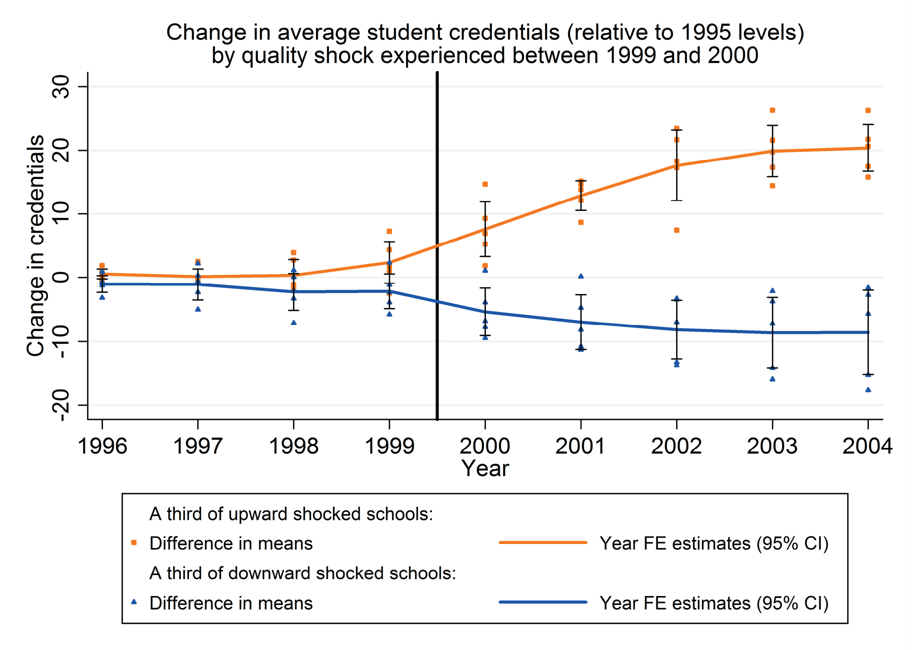
In summary, the descriptive evidence in Table 1 suggests that downward shocked schools attracted lower quality students even prior to the reform, but this gap increased after the reshuffling. Quite the opposite, however, is the relationship between school shock and teacher separation rates. Prior to the reform, the upward shocked schools experienced more one-year separations, but the fraction of separations is higher in downward shocked schools in the school year 2000/2001. Since these two facts are crucial for the identification in this paper I explore them further in Section 2.4.

2.4 Identifying variation

The implementation of the reform lead to abrupt changes in the sorting of students over schools in Stockholm. From one year to another, the same set of teachers experienced radical changes in the quality of the incoming students. In particular, some teachers ended up with lower quality pupils and some other teachers ended up with higher quality pupils than in the pre-reform period. The aim of this paper is to study how teacher mobility changed in response to this unexpected change in student quality. In this section I probe deeper into the changes in students' credentials and the changes in teacher mobility.

Figure 1 shows the differences in average student credentials for every year (1996 to 2004) and for upward and downward shocked schools relative to average student credentials in the same schools during the 1995/1996 school year. In Figure 1, these differences are plotted as points, while the vertical lines at each year show 95-percent confidence intervals from a linear regressions with the difference in the average first-year student credentials compared to 1996 as the dependent variable and year dummies (one for each year between 1996 and 2004) as independent variables. Figure 1 clearly shows that the reform caused a differential change in average student quality. Prior to the reform there are no significant differences in average students' credentials in upward and downward shocked schools, yet post-reform, the average credentials for these two groups of schools clearly diverge from one another. For the most part, I do not explore the changes in average characteristics in this paper, but rather, I focus on the changes in incoming student credentials since it is the margin for which the shock induced by the reform was the most pronounced. Naturally, the two measures are highly correlated, and Figure 2 confirms that the largest shock in incoming students quality occurred between the 1999/2000 and 2000/2001 school years, while the subsequently admitted cohorts mimicked the quality of the first graders from the 2000/2001 school year.[13]

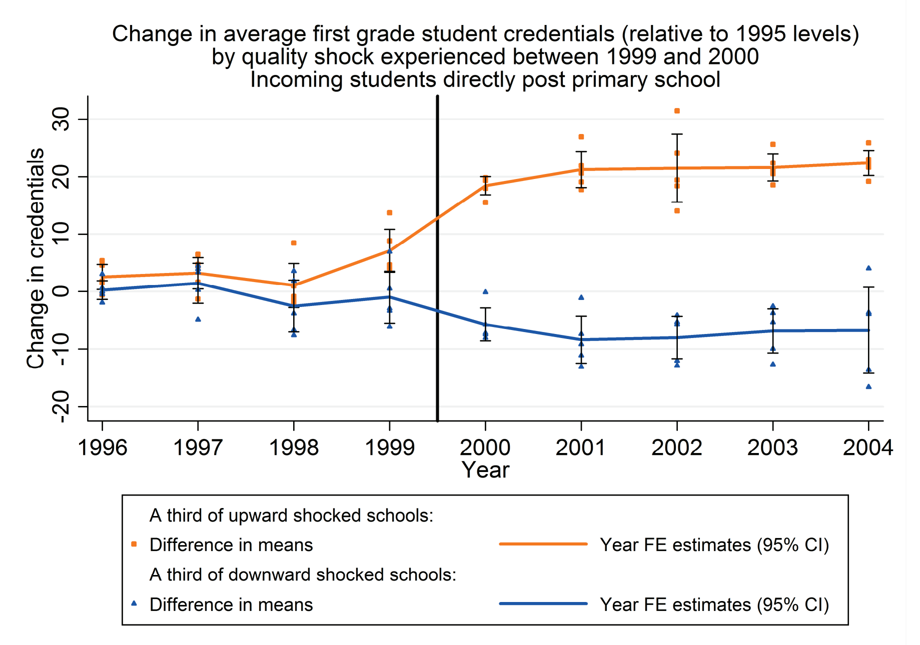
Figure 1. Variation in treatment: Correlation in mean school quality between 1995 and subsequent years. All grades and all students.



Note: Shock is defined as a difference between mean incoming students' credentials measured by primary school 9th grade GPA in high school j in the first post-reform year 2000 and mean incoming students' credentials in the last pre-reform year 1999 in these same schools. Based on the shock schools are divided into these that experience the most positive change (one-third upward shocked schools) and these that experience the least positive change (one-third downward shocked schools). Each point represents a difference between average all-grades credentials in these schools in a given year (1996 to 2004) and average all-grades credentials in these same schools in 1995. Each dot is related to a single difference for a single school. Lines plot coefficients and 95% confidence intervals from regressing these differences on year dummies (one for each year between 1996 and 2004). Robust standard errors. Black solid vertical line depicts reform implementation. Only schools that are present in the data in each year between 1991 and 2004 are included in the analysis.

[13] Throughout the paper I use the incoming students' credentials (Figure 2) as the main treatment variable, however, one might also think about using the average student quality from all grades (Figure 1). In fact, if we compare average student characteristics between school years 1999/2000 and 2000/2001 in a regression framework with year and school fixed effects then we are effectively comparing 3rd grade students in pre-reform period to 1st grade students in post-reform period. If the reform is truly exogenous then this should not make much of a difference because the correlation between 1st and 3rd graders in the pre-reform period will be high, while the correlation between 3rd graders in the pre-reform and 1st graders in the post-reform period will be low. The results are qualitatively similar irrespectively of the measure used and in fact they are larger quantitatively if I use all-grades GPA as student quality measure.

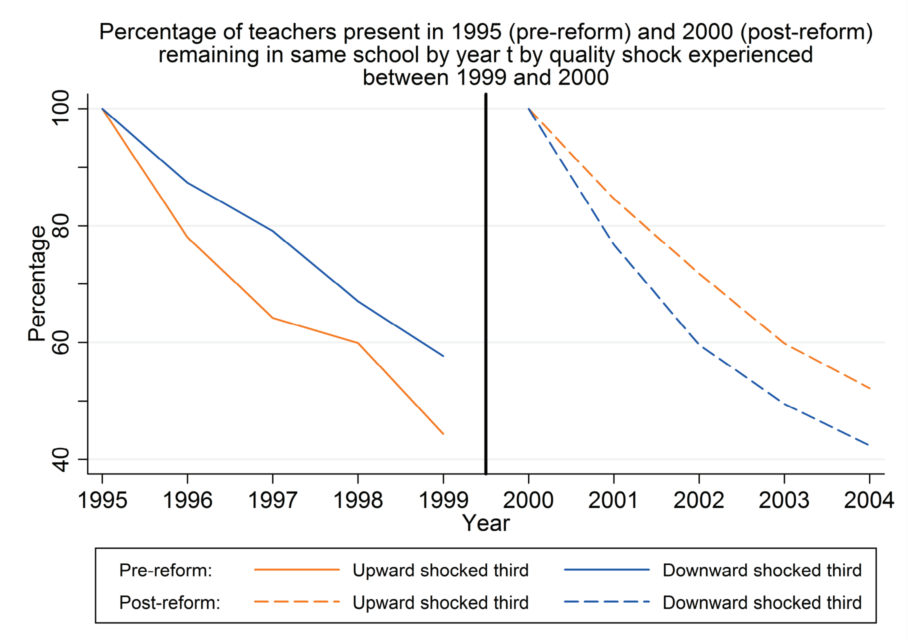
Figure 2. Variation in treatment: Correlation in mean school quality between 1996 and subsequent years. First grade students who applied to high school in the same year.



Note: Shock is defined as a difference between mean students' credentials measured by primary school 9th grade GPA (only students who applied to school in the same year) in first grade of high school j in the first post-reform year 2000 and alike defined mean students' credentials in the last pre-reform year 1999 in these same schools. Based on the shock schools are divided into these that experience the most positive change (one-third upward shocked schools) and these that experience the least positive change (one-third downward shocked schools). Each point represents a difference between incoming students' credentials in these schools in a given year (1996 to 2004) and incoming students' credentials in these same schools in 1995. Each dot is related to a single difference for a single school. Lines plot coefficients and 95% confidence intervals from regressing these differences on year dummies (one for each year between 1996 and 2004). Robust standard errors. Black solid vertical line depicts reform implementation. Only schools that are present in the data in each year between 1991 and 2004 are included in the analysis.

Figures 1 and 2 documented that the reform abruptly reshuffled students across schools in the municipality of Stockholm. In Figure 3, I provide some first evidence on how this reshuffling affected the probability that a teacher left their current employment. In particular, I start with the pool of teachers in 1995 (pre-reform) and in 2000 (post-reform), and I plot the fraction of teachers that remained employed from one up to four years. I plot these percentages separately for upward and downward shocked schools defined in the same manner as in Figures 1 and 2. Although the figure is uninformative about the pre-reform trends in teacher mobility and thus potential bias, it shows the mobility differences in levels before and after the reform for the two types of schools. For example, it depicts that upward shocked schools had higher levels of turnover before the reform, and that these same schools switched to having lower turnover rates in comparison to downward shocked schools post-reform. This is of importance as one might be worried that finding a negative effect of increased student quality on teacher turnover is driven by the fact that upward shocked teachers had lower mobility rates even prior to the reform. Figure 3 clearly shows that this is not the case and, if anything, the opposite is true. Thus, the empirical strategy should provide a lower bound estimate for the rate of mobility.

Figure 3. Variation in dependent variable: Teachers leaving their 1995 or 2000 employment.



Note: Shock is defined as a difference between mean students' credentials measured by primary school 9th grade GPA (only students who applied to school in the same year) in first grade of high school j in the first post-reform year 2000 and alike defined mean students' credentials in the last pre-reform year 1999 in these same schools. Based on the shock schools are divided into these that experience the most positive change (one-third upward shocked schools) and these that experience the least positive change (one-third downward shocked schools). Each point represents percentage of teachers who were teaching in school j in year 1995 (2000) and remain in this same school in year t. Black solid vertical line depicts reform implementation. Only schools that are present in the data in each year between 1991 and 2004 are included in the analysis.

3 Theoretical framework and empirical specification

3.1 Teachers' decision making process

The decision making process of teachers choosing whether to stay with their current employer or search for a new job can be framed within a turnover theory proposed by Jovanovic (1979). In the first period of time, I observe a teacher employed by a certain school and I assume that the employment decision was made so that it maximizes their utility with respect to the job characteristics (Jackson, 2013). For simplicity, let us assume that teachers only value the quality of their students and the monetary compensation they obtain from employment, and that they weakly prefer higher compensation and better students. Thus, the quality of the match between an individual teacher and school can potentially be altered either by changes in student composition or by changes in wages.

Since the admission system did not change over time, the expected quality of incoming students was roughly constant prior to the reform. Therefore, teachers did not expect that their match quality with respect to student quality would rapidly change and teachers with good matches were less likely to separate from their schools. Naturally, even without a policy change, teacher mobility is not zero. There are several reasons for this phenomenon. First, since at any point in time there are poor matches between schools and teachers – formed due to imperfect information or uncertainty about student composition – there are teachers switching schools in between school years. Second, there are teachers employed on fixed contracts (for example, as substitutes for permanent teachers who are on leaves) who leave their position once it can be filled again. Third, teachers retire or pass away, and thus, they drop out of the sample and new teachers need to be hired as replacements.[14] Having a poor match, however, is specific for a given school but not teaching as a profession, and thus, teachers with low quality matches should rather switch schools than leave the profession. On the other hand, retired or deceased teachers will naturally leave the profession. Finally, it is not clear a priori if teachers employed on fixed contracts are more likely to leave for a different occupation or switch schools within the profession.

So far I have discussed an individual teacher's separation decision - supply side. However, the decision naturally interacts with their employer's demand for new or existing teachers. Although firing teachers is relatively hard in Swedish schools, quitting is not. Thus, the principal's role in this optimization problem is related to either manipulating teacher compensation, or hiring new teachers when they face a teacher shortage, possibly as a result of increased mobility following the reform. However, the decisions made by principals regarding hires will only be observed after teachers decide whether to stay with his or her current school or separate. Therefore, the reform should not have an immediate influence on hiring policies but rather a delayed effect.

The framework discussed above generates two predictions that can be tested empirically: first, since teachers value working with high quality students, they will be less likely to leave schools experiencing inflow of students with better credentials; second, if monetary and student quality inputs to a teacher's utility function are jointly determined then non-switchers who experience an inflow of students with worse credentials should expect a rise in monetary compensation.

[14] I shut down the retirement channel by limiting the sample to teacher no older than 58 years of age, however, I cannot exclude any disability pensions.

3.2 Empirical specification

The reform can be described in two stages. First, it generated a change in the composition of incoming students in different schools, but it did not alter the average quality of students in the municipality of Stockholm. Figures 1 and 2 show that the reform indeed altered the student quality in different schools. Second, the change in student composition caused teachers to face a different set of students from one year to the next, and generated a reshuffling of teachers whose match quality had been exogenously altered. For this second stage to be due to changes in student quality only, I require that students did not select schools based on the underlying trends in teacher turnover – I discuss this possibility in the main results.

Since the reform was implemented in the school year 2000/2001, it is natural to first compare schools before and after this date which experienced different changes in student quality. Such a comparison yields a difference-in-differences estimator in which schools are treated to different extents, depending on the change in student quality. Thus, I compare teacher turnover in schools that experienced a sharp increase (or fall) in student quality to teacher turnover in schools where the student composition did not change that much.

Furthermore, since high school education in Sweden consists of three grades, it took up to three years for the reform to be fully implemented. Thus, in school year 2000/2001 only a third of the student stock had been admitted under the new rules and it was not until the school year 2002/2003 that the reform came into effect for the full student stock. Because of this feature of the reform, I study how teacher mobility changes up to three years after the reform. For the pre-treatment period not to overlap with the post-treatment period, I lag the pre-treatment measure of student quality one year for every additional year that I follow teacher mobility. In other words, a one-year teacher mobility analysis compares students in school year 1999/2000 to students in 2000/2001. A two-year mobility analysis compares students in school year 1998/1999 to students in 2000/2001, while a three-year mobility analysis compares students in school year 1997/1998 to students in 2000/2001.

Given the nature of the outcome variable I need at least two years to construct a single observation of the outcome variable, that is, I need to observe a teacher in periods t and $t+1$ to construct a mobility indicator. Since it took up to three years for the reform to be fully implemented, I construct three mobility measures. In each measure teacher is observed in school j in period t, and then separately in period $t+1$ (one-year mobility), period $t+2$ (two-year mobility) or period $t+3$ (three-year mobility). Thus, if I want to study the full effect of the reform, I compare the probability that teacher i in school j in 1997/1998 had left the school by 2000/2001 with the probability that teacher i in school j in 2000/2001 had left the school by 2003/2004. The treatment is set to the first year in the mobility window and, thus, compares the difference in incoming student quality in school year 1997/1998 to incoming student quality in school year 2000/2001. This can be written as:

$$(Y_{ij}^{2000} - Y_{ij}^{2000+k}) - (Y_{ij}^{2000-k} - Y_{ij}^{2000}) = \alpha + \beta(T_j^{2000} - T_j^{2000-k}) + \gamma(X_{ij}^{2000} - X_{ij}^{2000-k}) + \delta_j + \varphi_{2000} + \varepsilon_{ij} \qquad (1)$$

where i denotes individual teachers, j denotes schools and k denotes exposure length. The variable Y equals unity if teacher i is observed in school j in a given year and zero otherwise; T represents student quality or any alternative student characteristic measured at school j in a given year; X denotes individual teacher covariates including gender, marital status, immigration status, specialization (science, vocational, special education), university education indicator and experience; the parameters δ and φ are school and time fixed effects; and ε is a heteroskedasticity-robust standard error. The coefficient of interest in this paper is β and it identifies the effect of student quality on teacher mobility.

Equation (1) estimates the causal effect of student quality on the probability that a teacher separates from his or her current school, assuming that changes in student composition are not correlated with changes in teacher mobility in an absence of the reform. One testable implication of the identifying assumption is that post-reform changes in student quality in different schools are not correlated with pre-reform changes in teacher mobility in these schools. This examines if the assumption about common underlying trends in teacher turnover in the absence of the reform is plausible. For the placebo analysis to be meaningful, however, the placebo treatment period must not overlap with the true treatment period. Thus, studying pre-reform teacher mobility over a 3-year period requires lagging the outcome variable by three years. This can be written as:

$$(Y_{ij}^{2000-k} - Y_{ij}^{2000}) - (Y_{ij}^{2000-2*k} - Y_{ij}^{2000-k}) = \alpha + \beta(T_j^{2000} - T_j^{2000-k}) + \gamma(X_{ij}^{2000-k} - X_{ij}^{2000-2*k}) + \delta_j + \varphi_{2000-k} + \varepsilon_{ij} \qquad (2)$$

where Y, T, X, δ, φ and ε are defined as in Equation (1).

Equation (2) directly estimates the possibility of an anticipation effect. However, finding insignificant results in placebo estimates does not prove that the effect is not present as failing to reject a hypothesis does not imply it is true. Furthermore, one should focus not only on the second moment, which could be uninformative in the case of low precision in the estimates, but also on the point estimate which should be as close to zero as possible. Therefore, in order to be on the safe side, and since it is possible to directly account for an anticipation effect, I lag the dependent variable by one period in Equation (3). Such a procedure mechanically purges the possibility of a reaction to student quality in advance of the policy implementation. It requires, however, following teachers for four years for the reform to be fully implemented. In other words, the point estimates for one-, two- and three-year mobility estimated by Equation (1) should be compared to point estimates for two-, three- and four-year mobility estimated by Equation (3).

In specification described by Equation (3) I define the outcome variable as a comparison between the probability that teacher i in school j in 1995/1996 had left the school by 1999/2000 and the probability that teacher i in school j in 1999/2000 had left the school by 2003/2004. At the same time, the treatment compares the difference in incoming student quality between school year 1996/1997 and incoming student quality in school year 2000/2001. If there is no anticipation effect and the placebo regression specified in Equation (2) does not yield any large or significant results, then we should observe close to zero estimates in a one-period window in this specification. More formally, I can write:

$$(Y_{ij}^{1999} - Y_{ij}^{1999+k}) - (Y_{ij}^{1999-k} - Y_{ij}^{1999}) = \alpha + \beta(T_j^{2000} - T_j^{2000-k}) + \gamma(X_{ij}^{1999} - X_{ij}^{1999-k}) + \delta_j + \varphi_{1999} + \varepsilon_{ij} \qquad (3)$$

where Y, T, X, δ, φ and ε are defined as in Equation (1). Details about specific school years that I use for outcome and treatment variables of different exposure lengths in regressions defined by Equations (1) and (3) can be found in Tables A2 and A3. Note that in each regression I use only one pre- and one post-reform period, although I use multiple years to construct the outcome variables.

In order to illustrate the logic behind the difference-in-differences strategy used in this paper, Table 2 presents changes in teacher mobility over time for schools that experienced positive or negative changes in student quality, respectively.[15] I divide schools into two groups based on their changes in incoming student credentials between school years 1999/2000 (pre-reform) and 2000/2001 (post-reform). In the first column, I show data for one-third of schools with the most positive changes in incoming student credentials (one-third upward) while in the second column I show data for one-third of schools with the least positive (or negative) changes in incoming student credentials (one-third downward). On average, student quality increased by 15.79 percentile points in upward shocked schools and it decreased by 6.78 percentile points in downward shocked schools.[16] Concurrently, teacher mo-

[15] In order to provide better intuition about the timing of the reform and the reshuffling of students I start off with the model that does not account for the anticipation effect and does not require a lagged dependent variable. In Table A4, however, I also present the results for the Wald estimator accounting for the anticipation effects. Thus, Table A4 compares four-year mobility and four-year changes in student quality before and after the reshuffling started. The results are remarkably similar.

[16] This does not indicate that the average student quality in Stockholm increased due to the reform as the comparison excludes the middle quality schools. However, comparing the quality of incoming students between 1997 and 2000 indeed suggests that student quality increased

bility decreased by 20 pp in upward shocked schools and there was virtually no change in mobility in downward shocked schools.

Table 2. Effects of changes in students' credentials and probability of leaving school within 3-years. Wald estimator without accounting for an anticipation effect.

Effects of 3-year changes in student quality on 3-year teacher mobility			
	Schools		
	1/3 upward shocked	1/3 downward shocked	Difference
Treatment: Student quality - percentile ranked GPA from 9th grade in primary school. Incoming students graduating 9th grade in the same year.			
Year 2000	71.92	45.70	26.22***
	(14.19)	(10.40)	(1.06)
Year 1997	56.13	52.48	3.65***
	(13.41)	(12.68)	(1.03)
Difference	15.79***	-6.78***	**22.57*** **
	(1.09)	(0.99)	**(1.46)**
Dependent variable: Leaving school j from year 1997 to year 2000 (3-year mobility)			
Year 2000	0.17	0.22	-0.06*
	(0.37)	(0.42)	(0.03)
Year 1997	0.36	0.26	0.10***
	(0.48)	(0.44)	(0.04)
Difference	-0.20***	-0.04	**-0.16*** **
	(0.03)	(0.04)	**(0.05)**
Wald estimate			
-0.007* **			
(0.002)			

Note: Shock is defined as a difference between mean students' credentials measured by primary school 9th grade GPA (only students who applied to school in the same year) in first grade of high school j in the first post-reform year 2000 and alike defined mean students' credentials in the last pre-reform year 1999 in these same schools. Based on the shock schools are divided into these that experience the most positive change (one-third upward shocked schools) and these that experience the least positive change (one-third downward shocked schools). Only schools that are present in the data in each year between 1991 and 2004 are included in the analysis. It results in a sample of 15 schools. Dependent variable is defined as probability of leaving school j from school year 1997/1998 to school year 2000/2001 pre-reform and probability of leaving school j from school year 2000/2001 to 2002/2003 post-reform. Independent (treatment) variable is defined as difference in mean incoming students' credentials between 1997 in pre-period and 2000 in post-period. Differences report the interaction coefficients from regression of students' credentials or mobility on year dummy, upward shock dummy and their interaction. Wald estimate reports coefficient from instrumental variables regression of probability that teacher leaves school j on students' credentials, year dummy and upward shock dummy. Students' credentials are instrumented by interaction between year and shock. Robust standard errors and differences rounded to second decimal.

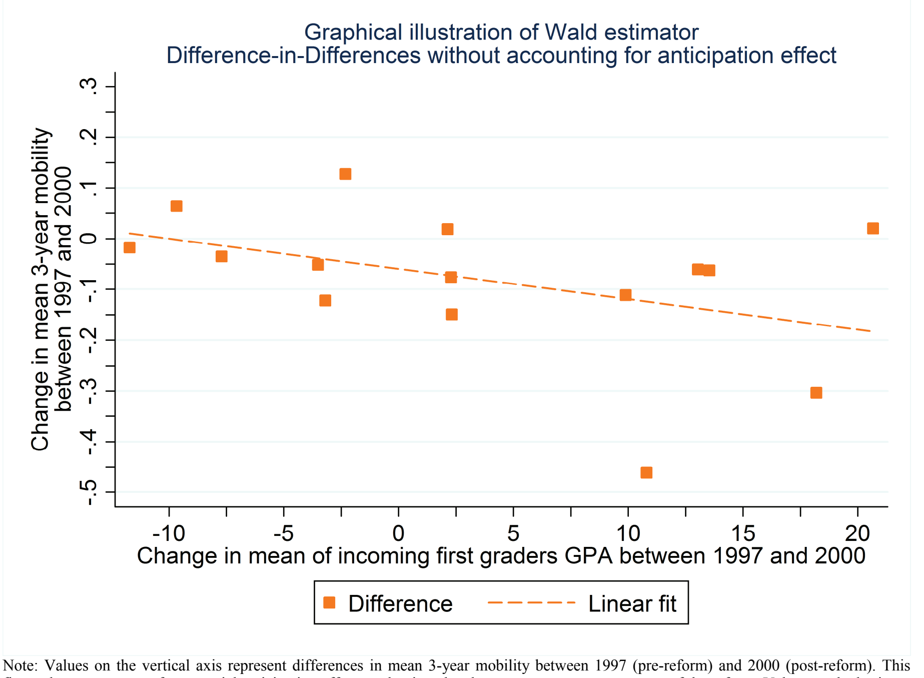
By calculating the ratio of the two changes (-16 pp divided by 22.57 percentile points) I obtain the Wald estimate of 3-year teacher mobility on incoming students quality. It implies that increasing incoming student credentials by 10 percentile points reduces teacher mobility by 7 pp. In the reminder of the paper I investigate whether these results hold up in a more formal regression analysis where the dummy variable for school shock is replaced with a continuous measure of incoming student credentials.

Finally, to illustrate how I exploit all of the variation in the changes in student quality, Figure 4 plots the differences in mobility for each school

by 6 percentile points. This fact can be driven by multiple factors: focusing on a panel of more stable schools, differential inflow of high-quality students from outside-of-Stockholm; or differential grade inflation. When analyzing all schools in Stockholm the average incoming students GPA is 54 in 1997, 58 in 1999, and 58 in 2000. Furthermore, my calculations show that there is no differential inflow of students residing outside of Stockholm. Thus, given that the averages in 1999 and 2000 are very similar but the average in 1997 is lower I conclude that over time there is some grade inflation at the upper end of the grade distribution. It is, however, small in comparison to the magnitude of the shock and should be purged by school and time fixed effects.

against the differences in the GPA of incoming students. This figure suggests that, on average, the negatively shocked schools experienced either small increases in teacher mobility or no changes at all. On the other hand, schools that were positively shocked were more likely to have experienced relatively large reductions in mobility. The dashed line in the figure shows a linear fit of the individual school observations and clearly points towards a negative relationship between changes in student quality and changes in teacher mobility.[17]

Figure 4. Difference-in-Differences. Probability of leaving school j in 3-years.



Note: Values on the vertical axis represent differences in mean 3-year mobility between 1997 (pre-reform) and 2000 (post-reform). This figure does not account for potential anticipation effects and quits related to rumours or announcement of the reform. Values on the horizontal axis represent changes in mean students' credentials between 2000 and 1997. Student credentials are based on first grade students who applied to high schools in the same year. Student credentials are measured using primary school 9th grade GPA. Line represents linear regression fit. Only schools that are present in the data in each year between 1991 and 2004 are included in the analysis.

4 Main results

I start by presenting the results for the regressions specified in Equation (1). I assume that teachers did not anticipate the changes in student quality. In Table 3, I report the estimates for the effects of changes in student quality on one-year mobility (row 1), two-year mobility (row 2) and three-year mobility (row 3). In column (1) I present correlations between GPA and mobility, in

[17] Figure A1 presents the same graph but for fully implemented reform specified by Equation (3). It points to the same conclusion as Figure 4.

column (2) I present difference-in-differences estimates without controlling for any observable teacher characteristics, while in column (3) I condition on a set of teacher controls. The estimates do not change much when I control for teacher characteristics, thus supporting the quasi-experimental nature of student resorting.[18] Since the reform gradually changed the student composition in schools, it is interesting to note that teachers' responses seem stronger the larger the share of students that gained admission under the new rules. The point estimate in row (3) in column (3) indicates that a 10-percentile-point increase in student quality reduces the probability of teacher turnover within three years by 7 pp.

Table 3. The effects of student credentials (first grade) on probability of leaving school j. No anticipation effects.

Dependent variable: probability of leaving school j within k years	(1) OLS	(2) DD	(3) DD
1-year mobility	0.000	-0.001	-0.002
	(0.001)	(0.002)	(0.002)
Observations		1,590	
2-year mobility	-0.002***	-0.002	-0.004*
	(0.001)	(0.002)	(0.002)
Observations		1,770	
3-year mobility	-0.001*	-0.006***	-0.007***
	(0.001)	(0.002)	(0.002)
Observations		1,710	
School and year fixed effects		X	X
Individual controls	X		X

Note: Teacher level regressions. Each estimate comes from a separate regression. Column (1) presents correlations conditional on individual teacher observable characteristics. Column (2) presents difference-in-differences estimates without controlling for any observable teacher characteristics. Column (3) adds individual level controls to column (2). Individual controls include: gender, marital status, immigration status, specialization (science, vocational, special education), university education indicators and experience. This table does not account for potential anticipation effect and quits related to rumours or announcement of the reform. The dependent variables are defined according to rows (1) to (3) in Table A2. The independent variables of interest measuring students' credentials are defined according to columns (2) and (4) in Table A2. Students' credentials measured by primary school 9th grade GPA (only students who applied to school in the same year) in first grade of high school j. Only schools that are observed in each year between 1991 and 2004 are included in the regressions. Robust standard errors.

The model underlying the estimates in Table 3 assumes that teachers did not anticipate the changes in student composition that followed the announcement of the reform in the fall of 1999. However, the GPA of incoming students at the school was made public around May 2000 and teachers could have left the school until October 2000.[19] The crucial question is whether teachers react to information about the quality of incoming students or the realization of the quality of incoming students. Thus, in an attempt to detect any potential anticipation effects, I estimate Equation (2), which is a placebo test of the difference-in-differences specification in Equation (1). The results are presented in Table 4. They clearly support the fact that teachers did not seem to respond to the information on future student quality. The estimates are insignificant and relatively small.

[18] Individual control variables do not include teacher earnings or type of contract as these might be an outcome of the reform. The estimates are identical whether I condition on earnings and type of contract or not.
[19] Although teachers could have left within a school year, such situations are rare, and this type of mobility would be captured by comparing two adjacent registers.

Table 4. Placebo analysis for regressions in table three. Effects of post reform changes in students' credentials on pre-reform changes in probability of leaving school j.

Variable of interest/Difference	(1) 1-year	(2) 2-years	(3) 3-years
1st graders quality	-0.001	-0.002	-0.001
	(0.002)	(0.002)	(0.002)
Observations	1,736	1,847	1,839

Note: Teacher level regressions. Each estimate comes from a separate regression. All point estimates come from difference-in-differences regressions including school and year fixed effects as well as individual controls (see column (3) in Table 3). The independent variables of interest measuring students' credentials are defined according to columns (2) and (4) in Table A2. Students' credentials measured by primary school 9th grade GPA (only students who applied to school in the same year) in first grade of high school j. The dependent variables are lagged by one exposure-period in comparison to these described in Table A2. That is in column (1) I compare one-year mobility in 1998/1999 to one-year mobility in 1999/2000. In column (2) I compare two-year mobility in 1996/1997 to two-years mobility in 1998/1999. In column (3) I compare three-year mobility in 1994/1995 to three-years mobility in 1997/1998. Only schools that are observed in each year between 1991 and 2004 are included in the regressions. Robust standard errors.

Thus far I have focused on the supply side of the teacher's labor market. Teachers who faced abrupt positive changes in the quality of their incoming students became less likely to separate from their current school. In this section, I analyze school responses to changes in student quality. In particular, I test whether the reform affected the number of enrolled students, which could mechanically lead to changes in teacher turnover. For example, if schools that experienced a positive shock to student quality also admitted more students after the reform, then it is possible that these schools attempted to retain or hire more teachers. Conversely, if unpopular schools both lost pupils and admitted students of lower quality after the reform they may have been forced to let some teachers go.[20] In addition, I analyze to what extent changes in student quality affected the number of teachers in the school as well as the student-teacher ratio.

Table 5 presents the effects of the reform on changes in the number of students, the number of teachers and the student-teacher ratio. Contrary to the mobility analysis, these regressions are based on a static model in which the outcome is determined at a given point in time, similar to the treatment, but not over multiple time periods as in the case of mobility. Furthermore,

[20] Note that the funding of schools in Sweden is tied to the number of enrolled students. The reform could also force some students to change schools as a response to changes in peer composition. I address this issues by estimating a model in which I define the outcome as the probability that I do not observe currently enrolled student i in school j in the next school year, and construct the mean probability at the school level. The regression framework is identical to Table 5 with mean probabilities as outcomes, and I lag the last pre-reform period by one (to 1998) in order to account for potential anticipation effects by the students. For each exposure length I find small but highly significant results on student mobility. In the anticipation year the point estimate is 0.001 and in subsequent differences these are -0.001, -0.0008 and -0.0005 for one, two and three year windows, respectively. Given that roughly 12% of students change schools from year to year, these effects are tiny in terms of magnitude, although they are statistically significant. Given an average school size of 825 pupils, the estimate suggests that when a school is shocked by a one-standard-deviation decrease in school quality, for each 91 pupils that would normally leave the school, an extra student will leave due to the changes in student quality. Since this estimate is tiny and I do not find any effects on the average school size, I conclude that general equilibrium effects are unlikely to play a major role in a teacher's decision making process.

since school composition was determined during the pre-period of September 1999, and the reform was not voted into power until later in 1999, there is no need to account for an anticipation effect in this setting. Thus, the outcome variable takes the form of a comparison between the last year prior to the reform and the first three years after the reform, but I set up the treatment as in all other regressions. The results in Table 5 show that neither the number of students, the number of teachers nor the student-teacher ratio responded to changes in student quality. Thus, it is unlikely that the changes in teacher mobility following the reform were a mechanical consequence of changes in school size or school resources.

Table 5. Difference-in-Differences: Effects of the reform on school size and resources.

VARIABLES	(1) 1-year	(2) 2-year	(3) 3-year
	Panel A: Number of students		
1st graders quality	0.143	1.272	2.749
	(2.742)	(3.298)	(3.608)
	Panel B: Number of teachers		
1st graders quality	0.182	0.195	0.230
	(0.367)	(0.466)	(0.755)
	Panel C: Student-teacher ratio		
1st graders quality	-0.018	-0.002	-0.002
	(0.068)	(0.062)	(0.089)
Observations	30	30	30

Note: School level difference-in-differences. Regressing number of students attending school (panel A), number of teachers at school (panel B) and student-teacher ratio at school (panel C) on students' credentials and school and time fixed effects. Students' credentials measured by primary school 9th grade GPA (only students who applied to school in the same year) in first grade of high school j. The independent variables of interest measuring student credentials are defined according to columns (2) and (4) in Table A3. The dependent variables are measured in 1999 in the pre-reform period and in 2000, 2001, 2002 in the post-reform period for 1, 2, and 3-year exposure, respectively. Only schools that are observed in each year between 1991 and 2004 are included in the regressions. Robust standard errors.

Table 6. The effects of student credentials (first grade) on probability of leaving school j. Accounting for anticipation effects.

Dependent variable: probability of leaving school j within k years	(1) OLS	(2) DD	(3) DD
1-year	-0.001**	-0.000	-0.001
	(0.001)	(0.002)	(0.002)
Observations		1,736	
2-years	-0.001	-0.003	-0.004**
	(0.001)	(0.002)	(0.002)
Observations		1,676	
3-years	-0.000	-0.006***	-0.007***
	(0.001)	(0.002)	(0.002)
Observations		1,667	
4-years	-0.001	-0.008***	-0.009***
	(0.001)	(0.002)	(0.002)
Observations		1,657	
School and year fixed effects		X	X
Individual controls	X		X

Note: Teacher level regressions. Each estimate comes from a separate regression. Column (1) presents correlations conditional on individual teacher observable characteristics. Column (2) presents difference-in-differences estimates without controlling for any observable teacher characteristics. Column (3) adds individual level controls to column (2). Individual controls include: gender, marital status, immigration status, specialization (science, vocational, special education), university education indicators and experience. This table through one-year lag in outcome variable (with respect to reform timing) accounts for potential anticipation effect and quits related to rumours or announcement of the reform. The dependent variables are defined according to rows (1) to (3) in Table A3. The independent variables of interest measuring students' credentials are defined according to columns (2) and (4) in Table A3. Students' credentials measured by primary school 9th grade GPA (only students who applied to school in the same year) in first grade of high school j. Only schools that are observed in each year between 1991 and 2004 are included in the regressions. Robust standard errors.

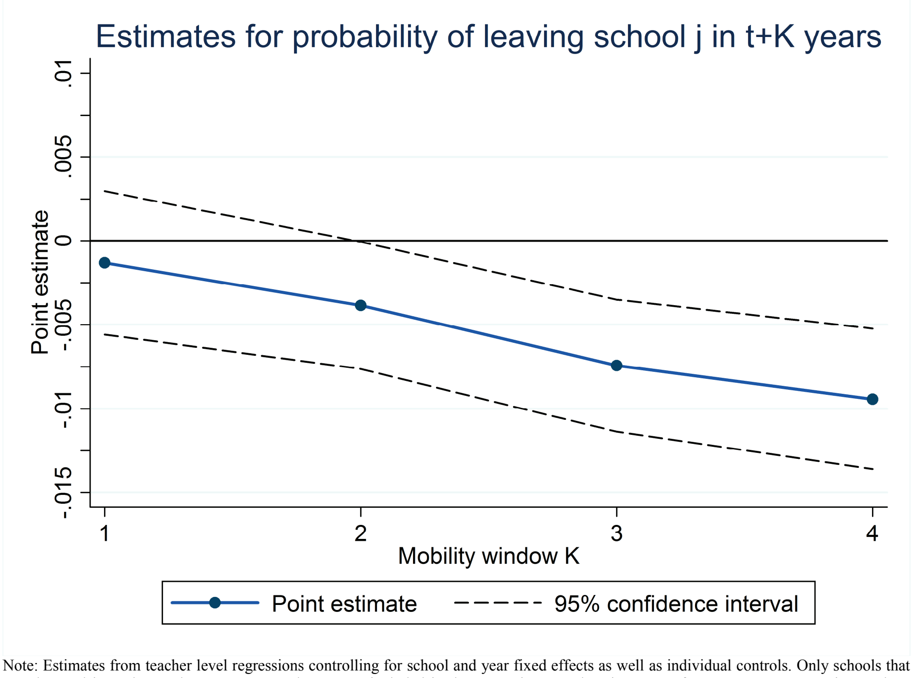
In Table 4 I showed placebo estimates suggesting that there is no significant anticipation effect in teacher turnover decisions. However, in order to further rule out the possibility of a biased estimate, I show estimates of the effects using the specification from Equation (3) in Table 6. It implies that teachers could not possibly anticipate changes in student quality because I lag the dependent variable by one year and the reform was not even announced yet early in the fall of school year 1999/2000. The first row of Table 6 can also be treated as a test for the anticipation effects as it compares one-year mobility in 1998/1999 to one-year mobility in 1999/2000. If there are no anticipation effects, then I should find an effect that is insignificant and close to zero. This is indeed the case, which should reassure the readers that teachers in Stockholm did not react significantly to the information and expectations, but rather they reacted to realized changes in student quality. Although the OLS point estimate in column (3) is significantly different from zero, it is very small in magnitude and does not point quantitatively towards any substantial bias. The estimates in column (3) for rows (2) through (4) in Table 6 that correspond to estimates from column (3) in Table 3 are slightly larger but in the same general ballpark. Thus far I have shown that there is no anticipation effects on average, yet it may well be the case that some teachers, such as better educated ones, are better at anticipating the effects of the reform than other teachers. Therefore, instead of presenting placebo estimates for all possible settings studied in this paper, I lag the outcome variables of interest in each case, and thus, mechanically purge the possibility of an anticipation effect.

The point estimate in row (4) of column (3) in Table 6 is the most important and most conservative estimate to be taken away from this paper. It suggests that when the reform was fully implemented a 10-percentile-point increase in student quality reduced the probability of a teacher leaving his or her school by 9 pp. Alternatively, a one standard deviation (17.3) increase in incoming student credentials decreased the probability of a separation within four years by 16 pp.[21] Given that the average four-year separation rate in this sample is 32%, the result implies a 50% reduction in mobility. As can be seen in Figure A1, there are two schools in Stockholm in which students improved by more than a standard deviation, three schools that improved by roughly three-quarters of a standard deviation, two schools that depreciated by a half of a standard deviation, and three schools that depreciated by roughly a third of a standard deviation. In summary, my findings are not only statistically significant, but also economically large and policy relevant.

[21] When I include the quadratic in students' credentials in the equation the coefficient on linear part remains negative and significant while the coefficient on quadratic term is positive and significant. Thus, the relationship between student quality and teacher mobility is estimated to be convex i.e., the higher inflow of good students has marginally diminishing effect on teacher separation rates.

Finally, in order to visualize how the effect of the changes in student quality evolved over time as the reform progressed, Figure 5 shows point estimates from column (3) of Table 6 with 95% confidence intervals. The line is clearly downward sloping, starting close to zero as there are virtually no anticipation effects. The F-test rejects the hypothesis that all four estimates are identical (p=0.030).

Figure 5. Difference-in-Differences estimates for different exposure lengths.



Note: Estimates from teacher level regressions controlling for school and year fixed effects as well as individual controls. Only schools that are observed in each year between 1991 and 2004 are included in the regressions. Each point comes from a separate regression. Robust standard errors. The dependent variables are defined according to rows (1) to (3) in Table A3. The independent variables of interest measuring students' credentials are defined according to columns (2) and (4) in Table A3. Students' credentials measured by primary school 9th grade GPA (only students who applied to school in the same year) in first grade of high school j. Individual controls include: gender, marital status, immigration status, specialization (science, vocational, special education), university education indicators and experience.

5 Is it really about student quality?

Thus far I have presented evidence that higher student quality reduces the probability that teachers leave their current employment. To the best of my knowledge this is the first paper that estimates causal effects of changes in student quality, as measured by academic credentials, on teacher labor supply decisions. Student quality is, however, correlated with other observable variables such as the fraction of minority students or parental wealth. For instance, Jackson (2009) used a similar identification strategy to gauge the causal effect of the reshuffling of minority students on teacher mobility. It is therefore relevant to ask whether it is direct measures of student quality or variables correlated with student quality that drive teachers' decisions.

Table 7. Probability of leaving school j. Alternative measures of student composition.

Dependent variable: probability of leaving school j within k years	(1) Unconditional	(2) Conditional on credentials
Share of immigrant students	1.199***	0.209
	(0.447)	(0.532)
GPA		-0.009***
		(0.002)
Mean parental income in 1000 SEK	-0.001**	0.003***
	(0.000)	(0.001)
GPA		-0.026***
		(0.005)
Mean parental education	-0.190***	-0.119**
	(0.049)	(0.053)
GPA		-0.007***
		(0.002)
Mean combined cognitive and non-cognitive paternal IQ	-0.017***	-0.008
	(0.004)	(0.005)
GPA		-0.007***
		(0.003)
Observations	1,657	

Note: Teacher level regressions controlling for school and year fixed effects as well as individual controls. Each row and column reports estimates from a separate regression. All regressions based on specification from Table 6, row (4) and column (3). In column (1) I substitute students' credentials with other mean school-level first grade characteristics, mainly, fraction of immigrants (row (1)), parental income (row (2)), parental education (row (3)) and paternal cognitive and non-cognitive military assessments (row (4)). These are correlated with first grader GPA (only students who applied to school in the same year) at the level of 0.39, 0.81, 0.91, 0.79, respectively. In column (2) I keep these alternative measures but also include first graders GPA (only students who applied to school in the same year). Individual controls include: gender, marital status, immigration status, specialization (science, vocational, special education), university education indicators and experience. Only schools that are observed in each year between 1991 and 2004 are included in the regressions. Robust standard errors.

My data include a number of background characteristics that may proxy for student quality such as whether a pupil is a first generation immigrant. I also have information on the yearly income and education of a pupil's parents, for which I compute school-level averages. Finally, I use military draft data with information on the cognitive and non-cognitive assessment of fathers. The results are presented in Table 7 where I focus on the specification of interest based on column (3) and row (4) from Table 6 with the estimate showing the effect for the fully implemented reform and accounting for an anticipation effect. The first row of Table 7 presents estimates in which the treatment is defined as a fraction of first generation immigrants (a correlation of 0.39 with GPA), the second row presents estimates for mean parental income (a correlation of 0.81 with GPA), the third row presents estimates for mean parental education (a correlation of 0.91 with GPA), and the fourth row presents estimates for mean combined cognitive and non-cognitive assessment of fathers (a correlation of 0.79 with GPA).[22] Column (1) presents the effects of the characteristics from rows (1) to (4) while column (2) adds a

[22] These data are available only for some fathers, and the coverage at school level increases from 24 to 51% over the time period used in this analysis. On average, I have information about fathers of 40% of pupils. This limitation is driven by the fact that the registries are not available for individuals tested before 1970 and immigrants. Nonetheless, I calculate the mean for all fathers with assessment information available in a given school.

student quality measure in a horse race between direct and indirect measures of student quality.

First, I focus on the unconditional effects. The estimate in row (1) confirms what other researchers have previously found in descriptive analyses, in particular, that the fraction of minorities at a school correlates positively with the probability of job separation (Hanushek et al., 2004; Falch and Strøm, 2005; Barbieri et al., 2011; Karbownik, 2013). Furthermore, unlike other researchers I do not find any evidence for the clustering of immigrant teachers and minority students in either specification. In row (2), the coefficient on mean yearly income in 100 000 SEK is -0.074 with a standard error of 0.035. This is a small estimate given that the mean yearly parental income in the studied group of schools is 377 696 SEK and in standard deviation terms it is roughly half of the effect estimated for student GPA. Similarly, rows (3) and (4) indicate significant and robust negative effects of increased parental quality on the probability of job separation. The intergenerational transmission of education has been well documented in the literature so it is not surprising that parental education is a good measure of student quality (Björklund et al., 2006). The last estimate is in line with Black et al. (2009), which documents an intergenerational mechanism of cognitive skills transmission, and thus, it is not surprising that teachers favor working with students whose fathers obtained relatively higher cognitive and non-cognitive scores in the military assessment.

In column (2), the estimates for the fraction of minorities and paternal military assessments become insignificant and decrease in size after controlling for students' credentials. The coefficient on mean parental income actually turns positive. On the other hand, the two coefficients in row (3) are negative when I include both parental education and student GPA. Overall, the estimates in column (2) suggest that teachers value primarily student quality, but that some of the response to changes in student quality is driven by changes in the students' socio-economic backgrounds.[23] In particular, teachers may prefer working with poorer students conditional on their high quality.

[23] Since direct (student quality) and indirect (share of immigrants, parental income and education, paternal military test scores) are highly correlated, one might be worried that models in column (2) pick up non-linear measures of student quality. When I add the square of student GPA to the estimates in column (2), however, it turns out to be positive and significant in all estimations (similar to the main specification). At the same time, the linear term in student quality remains highly significant and negative in all cases. Finally, the significant negative coefficient on parental education becomes insignificant suggesting that indeed it was picking up some non-linearity in student quality, however, the coefficient on parental income remains positive and significant with an identical point estimate.

6 Heterogeneity analysis

The richness and completeness of Swedish registry data allows me to investigate heterogeneity in the effects of student quality. It is important from a policy point of view to learn whether the effects of student quality vary by teacher characteristics. In particular, the consequences of the admission reform could be very different if high-quality teachers are more likely to leave the most disadvantaged schools. Therefore, I analyze how the response to changes in student composition differs by teacher quality, as measured by their formal education and experience. For male teachers born after 1951 I also have information about their cognitive and non-cognitive skills as measured at the military draft. Furthermore, I study how the response to changes in student credentials differs by the teacher's gender, specialization and type of contract. Then, I also divide teacher mobility by their destination. Finally, I split schools into quartiles of student quality distribution measured in the pre-reform period and by changes in the distribution of student quality.

Table 8 presents a range of heterogeneity findings. The table has the following structure: the first column reports the fraction of teachers in each group, while the second column reports the mean and a standard deviation of 4-year mobility for the group. The third column reports the point estimate and standard error of the effect of student quality on 4-year teacher mobility for the group. Finally, the fourth column presents a joint significance test for whether the point estimates for different sub-groups of teachers are different from one another.

I first consider the standard teacher quality measures. The first and second panels of Table 8 stratify teachers by their education and experience, which are important predictors of student achievement (Boyd et al., 2005; Harris and Sass, 2011). More than one-quarter of secondary school teachers in Stockholm do not have a formal university degree. Although these teachers have substantially higher turnover rates (42% vs. 29%), the estimated relationship between student quality and probability of leaving the current employment is remarkably similar (-0.007 vs. -0.009).

The same observation applies to teacher experience. Even though there are large differences in average turnover rates between groups (for example 54% for the least experienced teachers and only 20% for the most experienced ones), the point estimates are virtually identical, suggesting that the effects of student quality are similar across the distribution of teacher quality. It is worth noting, however, that although similar in percentage points and not significantly different, the point estimates suggest different relative reductions in mobility percent wise due to large differences in average mobility levels between the groups.

In rows (3) and (4) I further explore the uniqueness of the Swedish registry data and split teachers by their cognitive and non-cognitive skills that are available for all native males born in 1951 or later in Sweden. The sample size in this analysis is reduced dramatically to only 260 observations. The estimated responses to changes in student composition are somewhat different for teachers of different skills, but due to the relatively few observations I fail to reject that the estimates are different from one another.

Table 8. Heterogeneous effects by teachers' characteristics.

	Characteristic	Group	(1) Fraction [%]	(2) Mean mobility	(3) Estimate	(4) p-value difference
(1)	University education	Yes	72	0.286	-0.009***	
				(0.452)	(0.002)	
		No	28	0.422	-0.007*	0.706
				(0.494)	(0.004)	
(2)	Experience	0-5	24	0.536	-0.011**	
				(0.500)	(0.005)	
		6-15	36	0.323	-0.008**	
				(0.468)	(0.004)	0.917
		16+	40	0.195	-0.009***	
				(0.396)	(0.003)	
(3)	Cognitive assessment	High	73	0.381	0.001	
				(0.487)	(0.007)	
		Low	27	0.352	-0.017	0.207
				(0.481)	(0.012)	
(4)	Non-cognitive assessment	High	50	0.392	-0.013	
				(0.490)	(0.008)	
		Low	50	0.354	-0.001	0.323
				(0.480)	(0.008)	
(5)	Gender	Male	48	0.347	-0.011***	
				(0.476)	(0.003)	
		Female	52	0.303	-0.008***	0.492
				(0.460)	(0.003)	
(6)	Subject taught	Science	10	0.405	-0.011	
				(0.492)	(0.009)	
		Other	90	0.315	-0.009***	0.865
				(0.465)	(0.002)	
(7)	Type of contract	Permanent	81	0.270	-0.008***	
				(0.444)	(0.002)	0.619
		Temporary	19	0.554	-0.011**	
				(0.498)	(0.005)	

Note: Teacher level regressions controlling for school and year fixed effects as well as individual controls. Each row reports estimates from a separate regression. Column (1) reports fraction of individuals in each group while column (2) reports mean and standard deviation of a dependent variable (4-year mobility) in each group. Column (3) reports point estimates from regression specified as in Table 6, row (4) and column (3) for each group separately. Column (4) presents the joint significance test for the analysed groups in difference-in-differences model from column (3). Individual controls include: gender, marital status, immigration status, specialization (science, vocational, special education), university education indicators and experience. In row (1) a university graduate is defined as an individual graduating three, four or five year-long university education or individual with a research degree. Other forms of post-secondary education are not treated as university graduates. In row (6) science teachers include: mathematics, physics, chemistry, biology and computer science subjects. In rows (3) and (4) sample is restricted to native, males for whom both cognitive and non-cognitive assessment is observed. Cognitive and non-cognitive test scores are available for 89% of Swedish male population born 1951 or later. Low score is defined as below or equal to median in population percentiled draft-year distribution, while high score is defined as above median in population percentiled draft-year distribution. Sample size 1657 based on 1995 and 1999 comparison. Sample size for cognitive and non-cognitive skills regression is 260 based on 1995 and 1999 comparison. Only schools that are observed in each year between 1991 and 2004 are included in the regressions. Robust standard errors.

I also consider whether the estimated effect of student quality varies by teacher gender. While female teachers are somewhat less mobile than male teachers, the difference-in-differences estimates for both groups are virtually identical (-0.011 and -0.008, respectively). Another important group of teachers which often gets a lot of attention in media and research are science teachers (Edmark and Nordström Skans, 2010). On the one hand, providing these skills to students may be important for their chances on the labor market. On the other hand, teachers with this specialization may have favorable outside options. Thus, it is worth learning how changes in student quality affect teachers in mathematics, physics, chemistry, biology, and computer science in comparison to other teachers (row (6)). Even though science teachers have higher mobility rates I fail to find any significant evidence that they respond stronger to changes in student composition than other teachers.

Finally, I present estimates separately for teachers on permanent and temporary contracts. The latter teachers are typically employed on fixed-term contracts, often as replacements for teachers on extended leave, and are exposed to higher probabilities of job separation. Nearly 20% of teachers in Stockholm are employed on a temporary basis and they have more than twice as high turnover rates as permanently employed teachers. The estimated coefficients indicate, however, that the effects of student quality are virtually identical irrespectively of the type of employment. The evidence suggests that most teachers are affected to the same extent by changes in student quality. This may indicate that schools that end up with lower quality students are likely to lose all types of teachers and not only the best (or the worst) ones.[24]

The models used so far pool all of teacher mobility into one destination. However, previous research indicates that the correlations with teacher characteristics differ depending on the destination (Lankford et al., 2002). In Table 9, I investigate whether the effects of changes in student credentials are stronger along some mobility margins than others. In particular, I estimate the effect of student quality on teacher mobility within high schools (row (1)), to all levels of education (row (2)), to private schools (row (3)), out of the profession (row (4)) and to high schools with a higher quality of students (row (5)). Since it should be of particular interest to policy makers if highly educated teachers tend to leave the profession in response to such a reform, I also estimate the above specifications separately for the whole population (column (2)) and for teachers with university degree (column (4)).

[24] This statement might not be completely accurate as the groups presented in Table 8 overlap. To purge this confounding factor I use teacher's individual characteristics to predict their 4-year mobility and divide it into 10 mutually exclusive groups. I then run heterogeneity analysis using these groups. Even though the groups range in mean predicted mobility from 12.7% to 68.5%, the estimated effects are very similar and in a range between -0.025 to 0.005, and the slope of the line for 10 estimates is insignificant -0.0002. Thus, I conclude that the estimated effects of student quality are indeed identical for all teachers.

Table 9. Heterogeneity analysis: Effects by teachers' destination.

	(1)	(2)	(3)	(4)
Sample	All teachers		Teachers with university degree	
Dependent variable: probability of leaving school j within k years	Mean	Estimate	Mean	Estimate
Mobility within high schools	0.093	-0.006***	0.083	-0.005***
	(0.290)	(0.001)	(0.276)	(0.002)
Mobility within schooling	0.158	-0.007***	0.153	-0.006***
	(0.365)	(0.002)	(0.360)	(0.002)
Mobility to private school	0.011	-0.000	0.008	-0.001*
	(0.104)	(0.001)	(0.091)	(0.001)
Out-of-teaching	0.166	-0.003	0.133	-0.003
	(0.372)	(0.002)	(0.340)	(0.002)
To a higher quality high school	0.048	-0.003***	0.051	-0.002**
	(0.213)	(0.001)	(0.221)	(0.001)
Observations	1,657		1,192	

Note: Teacher level regressions controlling for school and year fixed effects as well as individual controls. Each row in columns (2) and (4) reports estimates from a separate regression. Columns (1) and (3) present means and standard deviations of dependent variables. Column (2) presents estimates for all teachers while column (4) presents estimates for teachers with university diploma. Estimates in columns (2) and (4) are based on specification from Table 6, row (4) and column (3). Dependent variable in row (1) equals unity if teacher leaves for another teaching position in high school. Dependent variable in row (2) equals unity if teacher leaves for another teaching position within primary or secondary schooling. Dependent variable in row (3) equals unity if teacher leaves for another teaching position in a primary or secondary private school. Dependent variable in row (4) equals unity if teacher leaves for another occupation outside of teaching. Dependent variable in row (5) equals unity if teacher leaves for high school with higher student quality than their initial allocation. Rows (2) and (4) add up to total mobility measure used in previous specifications. A university graduate is defined as an individual graduating three, four or five year-long university education or individual with a research degree. Other forms of post-secondary education are not treated as university graduates. Individual controls include: gender, marital status, immigration status, specialization (science, vocational, special education), university education indicators and experience. Only schools that are observed in each year between 1991 and 2004 are included in the regressions. Robust standard errors.

The estimates in the first two rows in column (2) show that increases in student credentials reduce the probability that a teacher will leave their current school for either a different high school or for any other primary or secondary school to the same extent. These estimates are also virtually identical for teachers with university education. Furthermore, I do not find any significant relationship between student quality and moving to a private school for all teachers, while the estimate is small and significant for teachers with a university diploma. On the contrary, I find significant and negative estimates on the probability of leaving to a school with a higher quality of students for all teachers. This potentially provides meaningful information about the direct manifestation of teacher preferences. Teachers seem to value the quality of their students, as they flee the adversely shocked schools in favor of schools with higher student quality. Finally, I do not find any effects of student quality on the probability that teachers leave the profession. This is in accordance with Jackson (2013), who argues that teachers will adjust their match quality within the profession rather than through outflow from the profession.

The last aspect of the heterogeneity analysis investigates the distributional effects of changes in student quality. These might be especially important as adversely shocked teachers tend to move to schools that have a better pool of students. I investigate this phenomenon in two ways. First, in Table 10, I study how teachers employed initially in schools from different parts of the student quality distribution respond to changes in their pupils' composition. Second, in Table 11, I study how teachers react to changes in the fraction of students from different parts of the quality distribution. For every school and year, I calculate the fraction of students admitted from each quartile of the quality distribution. Then, I use these four variables in separate regressions as a substitute for the average of student credentials. Thus, Table 10 reports heterogeneous responses to the same treatment, while Table 11 documents reactions to heterogeneous treatments.

Table 10. Heterogeneity analysis: Effects by pre-reform school quality.

Quartile of student quality	(1) Fraction [%]	(2) Mean mobility	(3) Estimate	(4) p-value difference
Bottom	27	0.206	-0.012***	
		(0.405)	(0.005)	
Lower middle	22	0.306	-0.028***	
		(0.462)	(0.006)	0.007
Higher middle	29	0.240	-0.003	
		(0.428)	(0.005)	
Top	22	0.137	-0.003	
		(0.344)	(0.005)	

Note: Teacher level regressions controlling for school and year fixed effects as well as individual controls. Each row in column (3) reports estimates from a separate regression. Column (1) reports fraction of individuals in each group while column (2) reports mean and standard deviation of a dependent variable (4-year mobility) in each group. Column (3) reports point estimates from regression specified as in Table 6, row (4) and column (3) for each group separately. Column (4) presents the joint significance test for the analysed groups in difference-in-differences model from column (3). Individual controls include: gender, marital status, immigration status, specialization (science, vocational, special education), university education indicators and experience. Sample sizes based on 1995 and 1999 comparison are 315, 258, 333 and 256 for rows (1) through (4), respectively. Student quality is divided into four quartiles based on the quality in school year 1996/1997 i.e., baseline student quality. Only schools that are observed in each year between 1991 and 2004 are included in the regressions. Robust standard errors.

The results in Table 10 indicate that only teachers employed in the bottom half of the distribution respond to changes in student quality.[25] The coefficients on the upper half are identical for both quartiles at -0.003 with a standard error of 0.005. It should be of interest that most of the turnover occurs in the second quartile of the distribution, suggesting that teachers on the margin that experience a mixture of high and low quality students on a daily basis react most strongly to the reshuffling. In fact, teachers at the bottom of the student quality distribution are the only ones who are significantly more likely to leave the profession in favor of a different occupation. Among the worst performing schools, the point estimate of -0.008 suggests that a 10-percentile-point decrease in student quality increases the probability that a teacher leaves his or her school for a job in a different profession by 8 pp. At the same time, I do not find any significant results for within-teaching mobility for the lowest quality schools, with an estimate of -0.005 and a standard error of 0.004, yet I find strong negative estimates for within-profession mobility for the second lowest quartile – an estimate of -0.023 with a standard error of 0.004. This last piece of evidence suggests that teachers at the

[25] When I split the sample into halves I only find significant estimate for the bottom half. It is -0.018 with SE of 0.004, while the estimate for the top half is -0.003 with SE of 0.003. The two coefficients are different at 1% level. When I split the sample into tertiles I find significant estimates for bottom and middle tertile. These are both -0.014 with SEs of 0.005 and 0.004, respectively. The coefficient for the top tertile is -0.002 with SE of 0.004 and the three coefficients are significantly different from one another at 10% level.

bottom of the distribution prefer to leave the profession when facing an adverse shock to student quality. However, their colleagues who are at the margin and who experience a mix of good and bad students seek a higher quality match within the occupation.

Table 11. Heterogeneity analysis: Effects by changes in fraction of students in quartiles of quality distribution.

Quartile	(1) Bottom	(2) Lower middle	(3) Higher middle	(4) Top
Mean fraction	0.183	0.235	0.254	0.328
	(0.163)	(0.120)	(0.095)	(0.242)
Fraction of students	1.511***	0.609***	-0.766***	-0.149
in k-th quartile	(0.255)	(0.190)	(0.162)	(0.108)

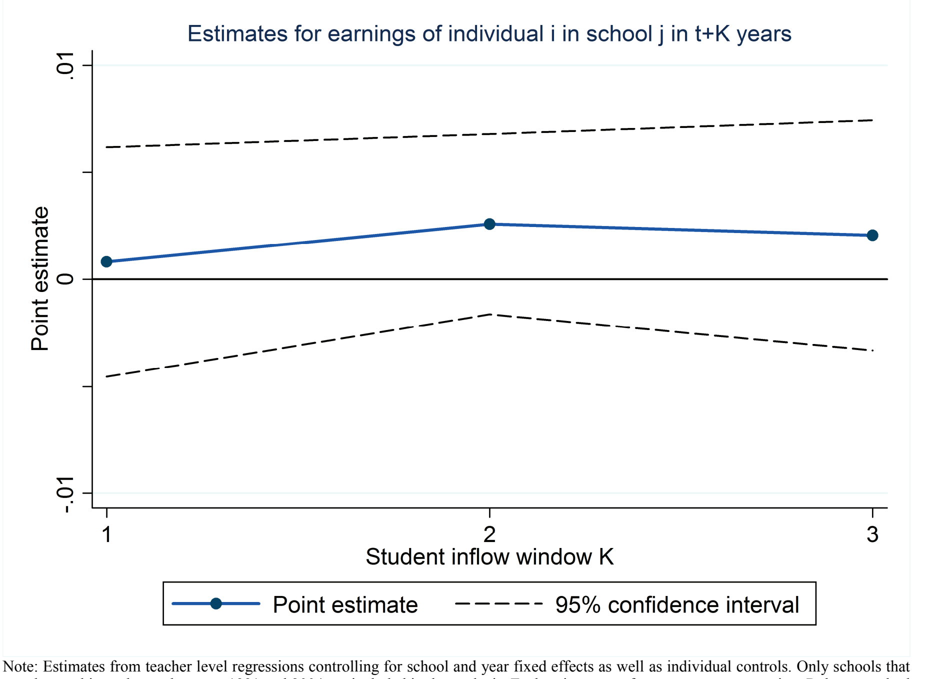
Note: Teacher level regressions controlling for school and year fixed effects as well as individual controls. Each column in the second row reports estimate from a separate regression. First row presents mean and standard deviation of the share of students in a given quartile of the quality distribution based on all first grades that applied to schools in the year of graduation. Point estimates based on regression specified as in Table 6, row (4) and column (3) for each group separately. Individual controls include: gender, marital status, immigration status, specialization (science, vocational, special education), university education indicators and experience. Sample size based on 1995 and 1999 comparison is 1657 observations. Only schools that are observed in each year between 1991 and 2004 are included in the regressions. Robust standard errors.

Finally, Table 11 leads to similar conclusions regarding the quality of students. Teachers who experience an inflow of students coming from the bottom half of the quality distribution are more likely to leave their current employment. The point estimate for the lowest quartile is also more than twice the size of the coefficient for the second lowest quartile suggesting that the really bad students have a significantly higher pushing out effect than their moderately more able peers. Here, it is also the case that only an increase in the fraction of students from the bottom quartile induces teachers to leave the profession. One the other hand, an increase in the fraction of students that come from the upper half of the quality distribution actually decreases the probability that teachers separate from their school. It is also interesting to note that this effects is driven by the mediocre students scoring above the median rather than by the very top students.

7 School responses: teacher earnings and hiring policy

Swedish teachers' wages are determined at a national level with some room for individual wage bargaining. Since the reform only affected the admission system in the municipality of Stockholm, any effects on wage bargaining at the national level were likely small. It is thus interesting to investigate whether the principals at Stockholm schools used teacher wages as a way to compensate for the changes in the other attributes of schools. Figure 6 shows the point estimates together with 95-percent confidence intervals. Similarly to regressions in Table 5, this analysis is based on the static model in which the earnings are determined at a given point in time and do not require using multiple time periods to construct a single dependent variable. It is plausible however, that if teachers expected changes in student quality they could have renegotiated their monetary compensations in the school year 1999/2000 as an insurance against a potential shock. In fact, wage renegotiation is probably more plausible in this setting than changes in employment. Therefore, when analyzing earnings I account for the anticipation effects and compare the second-to-last year prior to the reform to the three years post-reform. All point estimates are positive, but they never reach statistical significance. If anything, the results indicate that schools with a positive shock to student quality raise wages in an attempt to retain old teachers and attract new ones.

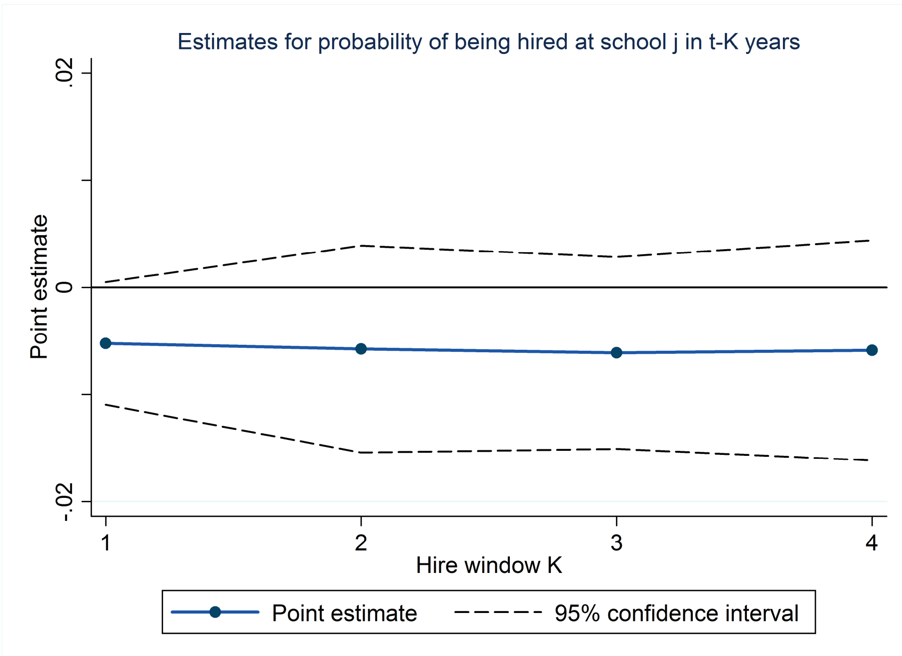
Figure 6. Difference-in-Differences estimates for individual teacher's earnings.



Note: Estimates from teacher level regressions controlling for school and year fixed effects as well as individual controls. Only schools that are observed in each year between 1991 and 2004 are included in the analysis. Each point comes from a separate regression. Robust standard errors. The dependent variables are earnings in 1998 in pre-period and earnings in 2000, 2001 and 2002 in post-period for one, two and three year differences, respectively. The independent variables of interest measuring students' credentials are defined in year 1999 in pre-period and in 2000 in post-period. Students' credentials measured by primary school 9th grade GPA (only students who applied to school in the same year) in first grade of high school j. Individual controls include: gender, marital status, immigration status, specialization (science, vocational, special education), university education indicators and experience.

In Figure 7, I present results on teacher hiring. This analysis, akin to the analysis of mobility, is based on a dynamic treatment of the outcome variables. In particular, I need two time periods to define a single outcome variable. The dependent variable is defined as the probability of being hired a year, two, three or four years prior to baseline in pre- and post-treatment period. Here again I need to use four years in order to account for the anticipation effects. Even though I do not find any effects of the reform on changes in the number of teachers, it is plausible that principals might have attempted to contract some extra teachers if they expected their schools to be adversely shocked, resulting in an outflow of their current staff. Although statistically insignificant, all the estimates are negative, which suggests that schools that experience an increase in student quality retain their current teachers and, thus, reduce new hires.

Figure 7. Difference-in-Differences estimates for probability of being a new hire.



Note: Estimates from teacher level regressions controlling for school and year fixed effects as well as individual controls. Only schools that are observed in each year between 1991 and 2004 are included in the regressions. Each point comes from a separate regression. Robust standard errors. The dependent variable in pre-period ends in school year 1999/2000 in each case. That is for one year window I code hired teacher as the one that is present in school j in school year 1999/2000 but was not present in school year 1998/1999. Identical logic applies for longer (2, 3 and 4) exposure lengths, thus for 4-year hire window in the pre-period I code teachers as hired in school year 1999/2000 if they were not present in school j in school year 1995/1996. In the post-reform period I define hires for school years 2000/2001 (1-year), 2001/2002 (2-year), 2002/2003 (3-year) and 2003/2004 (4-year). They correspond to being hired in these years and not being present in school j in school year 1999/2000. The independent variables of interest measuring students' credentials are defined in year 1999 in pre-period and in 2000 in post-period. Students' credentials measured by primary school 9th grade GPA (only students who applied to school in the same year) in first grade of high school j. Individual controls include: gender, marital status, immigration status, specialization (science, vocational, special education), university education indicators and experience.

8 Conclusions

A number of educational policies involve placing certain groups of students in a more favorable school environment, in hopes that interacting with better peers would boost their school performance. However, the success of such policies relies on, among other things, how teachers respond to changes in student quality. This paper provides evidence on the causal effect of student quality on teacher mobility, using abrupt changes in the credentials of the incoming students following an admission reform in Stockholm. I use data on teachers, students and their parents for Swedish high schools covering years 1991/1992 to 2004/2005.

The results show that an increases in student quality leads to lower teacher mobility and that the effect is increasing as the reform progresses. A 10-percentile-point increase in incoming student credentials decreases the probability that a teacher will leave their school by up to 9 pp. I show that this effect is robust to different model specifications and I account for the fact that the change in student quality in different schools might be related to pre-existing trends in teacher mobility. The effect is very similar across all types of teachers and is found mostly for mobility between schools rather than out of the profession. It is also present only in the lower half of the student quality distribution. Furthermore, teachers seem to react mostly to direct measures of student quality (credentials) rather than to characteristics that are correlated with student quality (immigrant status, parental income and schooling, paternal cognitive and non-cognitive skills). Finally, I do not find any significant effects of changes in student quality on teacher's earnings or school hiring policies.

References

Barbieri, G., C. Rossetti, and P. Sestito (2011), "The determinants of teacher mobility: Evidence using Italian teachers' transfer applications", *Economics of Education Review*, 30:6, 1430-1444.

Björklund, A., M. Lindahl, and E. Plug (2006), "The origins of intergenerational associations: Lessons from Swedish adoption data", *Quarterly Journal of Economics*, 121:3, 999-1028.

Black, S., P. Devereux, and K. Salvanes (2009), "Like father, like son? A note on the intergenerational transmission of IQ scores", *Economics Letters*, 105:1, 138-140.

Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff (2005), "Explaining the short careers of high – achieving teachers in school with low performing students", *American Economic Review*, 95:2, 166–171.

Böhlmark, A., and M. Lindahl (2007), "The impact of school choice on pupil achievement, segregation and costs: Swedish evidence", IZA Working Paper 2786.

Cullen, B.J., B.A. Jacob, and S.D. Levitt (2006), "The effects of school choice on participants: Evidence from randomized lotteries", *Econometrica*, 74:5, 1191-1230.

Edin P-A., B. Holmlund, S. Sandgren Massih and R. Uusitalo (2011), "School choice and student outcomes: Evidence from a Swedish reform", Mimeo Uppsala University.

Edmark, K., and O. Nordström Skans (2010), "Science of success: the causal effect of high school math and science on labor market outcomes", mimeo IFAU.

Falch, T., and B. Strøm (2005), "Teacher turnover and non–pecuniary factors", *Economics of Education Review*, 24:6, 611–631.

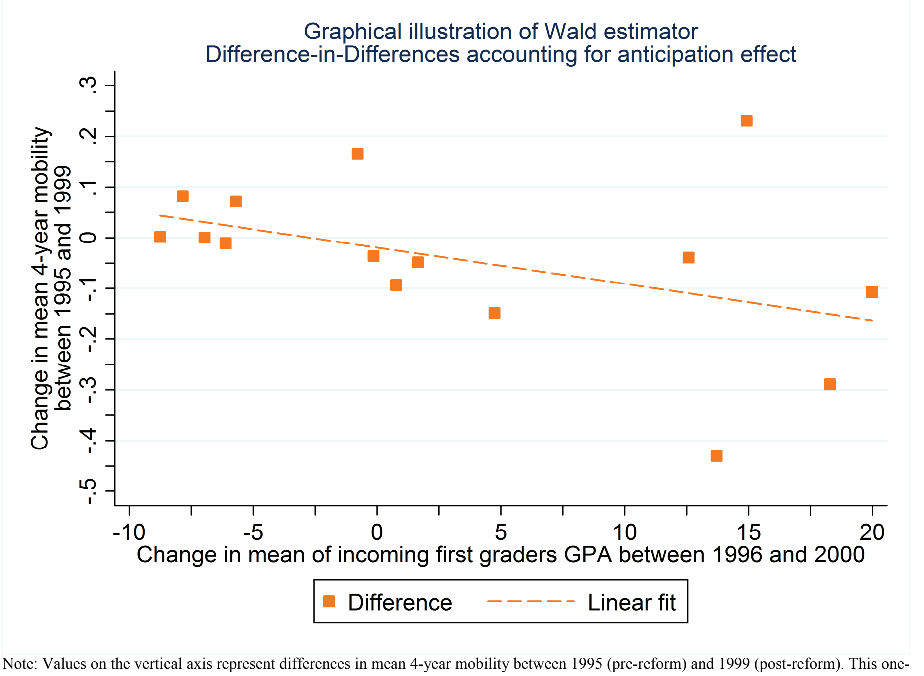
Feng, L., D. Figlio, and T. Sass (2010), "School accountability and teacher mobility", NBER Working Paper 16 070.

Fredriksson, P., and B. Öckert (2008), “Resources and student achievement – evidence from a Swedish policy reform”, *Scandinavian Journal of Economics*, 110:2, 277-296.
Hanushek, E.A., (1986), “The economics of schooling: Production and efficiency in public schools”, *Journal of Economic Literature*, 24:3, 1141-1177.
Hanushek, E.A., F.J. Kain, and G.S. Rivkin (2004), “Why public schools loose teachers?”, *Journal of Human Resources*, 39:2, 326–354.
Harris, N.D., and T.R. Sass (2011), “Teacher training, teacher quality and student achievement”, *Journal of Public Economics*, 95:7, 798-812.
Hensvik, L., (2012), “Competition, wages and teacher sorting: four lessons from a voucher reform”, *Economic Journal*, 122:561, 799-824.
Hoxby, C., (2002), “Would school choice change the teaching profession?”, *Journal of Human Resources*, 38:4, 846-891.
Hsieh, C-T., and M. Urquiola (2006), “The effects of generalized school choice on achievement and stratification: Evidence from Chile’s voucher program”, *Journal of Public Economics*, 90:8-9, 1477-1503.
Jackson, C.K., (2009), “Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation”, *Journal of Labor Economics*, 27:2, 213-256.
Jackson, C.K., (2012), “School competition and teacher labor markets: Evidence from charter school entry in North Carolina”, *Journal of Public Economics*, 96:5-6, 431-438.
Jackson, C.K., (2013), “Match quality, worker productivity, and worker mobility: Direct evidence from teachers”, *Review of Economics and Statistics*, forthcoming.
Jovanovic, B., (1979), “Job matching and the theory of turnover”, *Journal of Political Economy*, 87:5, 972-990.
Karbownik, K., (2013), “The determinants of teacher mobility in Sweden”, Mimeo Uppsala University.
Lankford, H., S. Loeb, and J. Wyckoff (2002), “Teacher sorting and the plight of urban schools: the importance of alternative labor market opportunities and non – pecuniary variation”, *Educational Evaluation and Policy Analysis*, 24:1, 37-62.
Lindholm, A., (2006), “Lärarförbundet och lärarnas riksförbund medlemsundersökning om lön och arbetstid”.
Reber, J.S., (2005), “Court-ordered desegregation: Successes and failures integrating American schools since Brown versus Board of Education”, *Journal of Human Resources*, 40:3, 559-590.
Scafidi, B., D.L. Sjoquist, and T.R. Stinebrickner (2007), “Race, poverty, and teacher mobility”, *Economics of Education Review*, 26:2, 145–159.
Skolverket (2009), “Resursfördelning utifrån förutsättningar och behov”, Skolverket Rapport 330.
Söderström, M., and R. Uusitalo (2010), “School Choice and Segregation: Evidence from an Admission Reform”, *Scandinavian Journal of Economics*, 112:1, 55-76.

Appendix

Figures

Figure A1. Difference-in-Differences. Probability of leaving school j in 4-years.



Note: Values on the vertical axis represent differences in mean 4-year mobility between 1995 (pre-reform) and 1999 (post-reform). This one-year lag in outcome variable (with respect to the reform timing) accounts for potential anticipation effects and quits related to rumours or announcement of the reform. Values on the horizontal axis represent changes in mean students' credentials between 2000 and 1996. Student credentials are based on first grade students who applied to high schools in the same year. Students' credentials are measured using primary school 9th grade GPA. Line represents linear regression fit. Only schools that are present in the data in each year between 1991 and 2004 are included in the analysis.

Tables

Table A1. Descriptive statistics – comparison of Sweden and Stockholm.

Variable	(1)	(2)	(3)	(4)
	Pre-period = 1999		Post-period = 2000	
	Sweden	Stockholm	Sweden	Stockholm
One-year mobility	0.13 (0.34)	0.14 (0.35)	0.11 (0.31)	0.15 (0.36)
Fraction of female teachers	0.49 (0.50)	0.55 (0.50)	0.48 (0.50)	0.53 (0.50)
Mean teacher experience	12.00 (7.27)	11.05 (7.24)	11.94 (7.61)	10.96 (7.48)
Fraction of teachers with university diploma	0.65 (0.48)	0.72 (0.45)	0.65 (0.48)	0.72 (0.45)
Fraction of teachers employed on temporary contracts	0.20 (0.40)	0.21 (0.41)	0.20 (0.40)	0.22 (0.41)
Mean yearly teacher earnings in 1000 SEK	224 (78)	217 (84)	226 (79)	226 (87)
Students' credentials	48.95 (11.12)	56.73 (16.48)	48.99 (11.58)	57.35 (18.34)
Share of immigrants	0.08 (0.06)	0.13 (0.07)	0.09 (0.06)	0.13 (0.07)
Mean yearly parental income in 1000 SEK	338 (62)	396 (117)	356 (65)	423 (128)
Mean parental education	11.99 (0.82)	13.05 (1.15)	12.16 (0.79)	13.20 (1.17)
Mean paternal draft score	51.60 (5.86)	58.10 (7.86)	51.88 (5.78)	56.92 (8.91)
Number of teachers	20 795	1304	21 675	1364

Note: Means and standard deviations. Columns (1) and (3) present statistics for all high school teachers in Sweden (excluding Stockholm municipality) in years 1999 and 2000 from schools that were in operation prior to school year 1999/2000. Columns (2) and (4) present statistics for all high school teachers in Stockholm municipality in years 1999 and 2000 from schools that were in operation prior to school year 1999/2000.

Table A2. Definitions of mobility and students' credentials variables. No anticipation effects.

	(1)	(2)	(3)	(4)
Mobility	Post-period mobility	Post-period GPA	Pre-period mobility	Pre-period GPA
1-year	00/01 to 01/02	2000	99/00 to 00/01	1999
2-year	00/01 to 02/03	2000	98/99 to 00/01	1998
3-year	00/01 to 03/04	2000	97/98 to 00/01	1997

Note: Table presents length of mobility in rows. First row defines mobility for teachers leaving in period t+1, second row in t+2 and third row in t+3. Column (1) defines the post-reform period dependent variables while column (2) defines post-reform treatment variables. Column (3) defines the pre-reform period dependent variables while column (4) defines pre-reform treatment variables.

Table A3. Definitions of mobility and students' credentials variables. Anticipation effects present.

	(1)	(2)	(3)	(4)
Mobility	Post-period mobility	Post-period GPA	Pre-period mobility	Pre-period GPA
1-year	99/00 to 00/01	2000	98/99 to 99/00	1999
2-year	99/00 to 01/02	2000	97/98 to 99/00	1998
3-year	99/00 to 02/03	2000	96/97 to 99/00	1997
4-year	99/00 to 03/04	2000	95/96 to 99/00	1996

Note: Table presents length of mobility in rows. First row defines mobility for teachers leaving in period t+1, second row in t+2, third row in t+3 and fourth row in t+4. Column (1) defines the post-reform period dependent variables while column (2) defines post-reform treatment variables. Column (3) defines the pre-reform period dependent variables while column (4) defines pre-reform treatment variables.

Table A4. Effects of changes in students' credentials and probability of leaving school within 4-years. Wald estimator accounting for an anticipation effect.

Effects of 4-year changes in student quality on 4-year teacher mobility			
	Schools		
	1/3 upward shocked	1/3 downward shocked	Difference
Treatment: Student quality - percentile ranked GPA from 9th grade in primary school. Incoming students graduating 9th grade in the same year.			
Year 2000	73.90	46.04	27.87***
	(13.23)	(10.19)	(1.02)
Year 1996	56.35	50.05	6.29***
	(13.90)	(10.51)	(1.00)
Difference	17.56***	-4.02***	**21.58*** **
	(1.10)	(0.88)	**(1.41)**
Dependent variable: Leaving school j from year 1996 to year 2000 (4-year mobility)			
Year 1999	0.24	0.30	-0.06
	(0.43)	(0.46)	(0.04)
Year 1995	0.41	0.29	0.11***
	(0.49)	(0.46)	(0.04)
Difference	-0.16***	-0.01	**-0.17*** **
	(0.04)	(0.04)	**(0.05)**
Wald estimate			
	-0.008* **		
	(0.003)		

Note: Shock is defined as a difference between mean students' credentials measured by primary school 9th grade GPA (only students who applied to school in the same year) in first grade of high school j in the first post-reform year 2000 and alike defined mean students' credentials in the last pre-reform year 1999 in these same schools. Based on the shock schools are divided into these that experience the most positive change (one-third upward shocked schools) and these that experience the least positive change (one-third downward shocked schools). Only schools that are present in the data in each year between 1991 and 2004 are included in the analysis. It results in a sample of 15 schools. Dependent variable is defined as probability of leaving school j from school year 1995/1996 to school year 1999/2000 pre-reform and probability of leaving school j from school year 1999/2000 to 2003/2004 post-reform. Independent (treatment) variable is defined as difference in mean incoming students' credentials between 1996 in pre-period and 2000 in post-period. Differences report the interaction coefficients from regression of students' credentials or mobility on year dummy, upward shock dummy and their interaction. Wald estimate reports coefficient from instrumental variables regression of probability that teacher leaves school j on students' credentials, year dummy and upward shock dummy. Students' credentials are instrumented by interaction between year and shock. Robust standard errors and differences rounded to second decimal.

Details of sample construction

I construct the sample of high school teachers for the school years 1991/1992 to 2004/2005. The information about teachers comes from the teacher registry and the analysis focuses on teachers working in grades 1 to 3 of secondary education (high school) that were in operation in Stockholm municipality prior to school year 1999/2000. Teachers who are on unpaid leave of absence or whose workloads are zero hours (i.e., they do not perform any pedagogical duties) are excluded from the analysis. Such teachers are treated neutrally in terms of mobility if they come back after the absence period to the same school. Similarly, I exclude teachers who are employed as principals, study counselors etc. In each year if a teacher has multiple entries in the registry, the observation with the highest workload is selected irrespectively of whether it is at the same or at different schools.[26] The teacher registry is a high quality data set, that allows recovering information on school location (unique identifier), school ownership and type, teacher certification, workload, employment type (temporary vs. permanent), education and position.

[26] The workload of teachers having multiple positions at the same school is not summed and the highest workload position is selected.

Teacher experience is not available for all years, and therefore, I use predicted experience in the analysis. In particular, since the teacher registries date back to 1979 I explore this feature to construct the “in teaching predicted experience” variable. I create a panel of all teachers between 1979 and 2006 and link it to population enlistment data between 1985 and 2006 in order to obtain teacher’s birth date. I then use all this information and tenure data provided in the later registries (since 1999 onwards) to construct the predicted measure of experience.

Teachers are then linked (using unique identifier) to population registers, which covers all individuals living in Sweden. The registers include information on gender, marital status, age, family composition (using unique family identifier), immigration history, education and income. Income is measured as a gross salary plus income from business and self-employment plus any work-related allowances. Investment losses are not included, and thus, income is lower-bounded at zero. The analysis is restricted to teachers aged 25-58 years, to abstract from mobility driven by educational attainment and retirement decisions.

The students’ characteristics are based on “school in” and “school out” pupil registries. The secondary school composition is based on all the students that are in a school in a given year. The quality of students in secondary school is measured based on their 9th grade grades. I percentile rank students for each subject and take the average across all subjects. The average GPA is then percentile ranked again. I match students to their parents using unique family identifier and obtain the family level socioeconomic indicators i.e. mean parental income, mean parental education and the cognitive and non-cognitive skill of the fathers from the military enlistment.

The enlistment registry covers period 1969 to 2006 and provides information on cognitive and non-cognitive assessments. All skill measures are percentile ranked by year of draft. The data is linked to teachers and students’ fathers using the unique personal identifier.

Finally, having a dataset with teachers and students I match the two using the unique school identifier. I exclude schools with less than three employed teachers (in full time equivalence) and schools with less than 15 students. I also restrict the analysis to teacher aged 25-58 years. I then select schools that operate within the municipality of Stockholm and were in operation prior to school year 1999/2000. This results in a sample of 15765 teacher-year observations, which is based on 3621 unique teachers from 29 schools. In this paper I focus on a balanced panel of schools, i.e. I restrict the sample to schools present in the data for all years between 1991/1992 and 2004/2005. I also drop teachers from Skärholmens Gymnasium because this school did not admit any new students in school year 1998/1999. The final sample consists of 2758 teachers, 15 unique schools and 12226 person-years.

Essay 4

Who gets to look nice and who gets to play? Effects of child gender on household expenditure

1 Introduction

Children's gender has been demonstrated to influence family stability (Dahl and Moretti, 2008), fertility (Ben-Porath and Welch, 1976; Das, 1987), abortion rates (Sen, 1990; Jha et al., 2006) and labor market activity (Lundberg and Rose, 2002; Ichino et al., 2011).[1] While a single causal mechanism behind these findings has not been identified, the possible explanations include strict preferences of parents for a specific gender of their child, differences in costs of bringing up boys and girls and the importance of gender specific roles in the upbringing process. Thus, while behavior associated with the birth of a child of a specific gender can be viewed as a reflection of gender bias in preferences, there are also explanations for different behavior of parents of boys and girls, once they are born, under unbiased preferences. Differential behavior of mothers on the labor market, for example, has been interpreted as strategic response to the increased probability of divorce following the birth of a girl (Ichino et al., 2011). In such a case higher divorce risk is considered a reflection of biased preferences and the higher labor supply of girls' mothers interpreted as its further consequence.

In developing countries children's gender has also been shown to determine the level of child related expenditure on schooling, nutrition and care (Jayachandran and Kuziemko, 2011; Barcellos et al., 2012). In these cases parental behavior is again explained by either higher returns to investment in boys or by strict gender bias against girls. Given these findings and the evidence on the role of children's gender in determining family and labor market outcomes in the developed world, it is surprising that there has so far been no study linking child gender to household expenditure. Studying expenditure patterns may provide further clues to understand the mechanisms behind the already identified effects on parental outcomes. Moreover, if there is differential treatment of boys and girls by their parents it should be reflected in the way households allocate their resources. The latter case is of particular importance in the light of the growing evidence on the role of early interventions (Blau and Currie, 2006; Carneiro and Ginja, 2008; Cascio, 2009; Almond and Currie, 2011), suggesting that differential treatment of boys and girls might have significant consequences for adult outcomes of men and women. Investments in the form of prenatal care, vaccinations or medical care have also been shown to be vital for children's development in the developed world (Aizer, 2003; Figlio et al., 2009; Levine and Schanzenbach, 2009). Along these lines, we would also expect that different levels of expenditure on goods that can be thought of as investment in the child's human capital might play an important role for children's future outcomes.

[1] For an excellent review of economic, sociological and psychological studies see Raley and Bianchi (2006).

The reason for these gaps in the literature relates primarily to data availability. From the above perspective the Polish Households' Budgets Survey (PHBS), which we use in this paper, offers a unique chance to study detailed patterns of household expenditures differentiated by gender and age. To our best knowledge this is the first such an analysis in the developed world context and the only other study relating gender to family expenditure though focusing on food comes from India (Behrman 1988), where treating the first-born child's gender as random may be problematic due to sex-selective abortions and infanticide (Jha et al., 2006). Apart from information on the main parental outcomes studied earlier in the literature, the PHBS data contains also expenditure information on over 400 specific items. From among these expenditures we can distinguish several items, which on the one hand can be assigned to adults by gender, and on the other, to children below the age of 13. This feature of the data provides a unique opportunity to study the so far unexamined effects of children's gender.[2]

Using data for the years 2003-2010 we compare 15,000 families with first-born girls and 16,198 families with first-born boys to study the differential patterns of household expenditure. Given the design of the data 35.7% of the first group of families and 35.4% of the latter are observed in two consecutive years, which in total gives us a sample of 48,397 observations. Given previous findings in the literature we first discuss three main potential confounding factors, namely marital stability, fertility and labor supply, which have all been found to correlate with the gender of the first child and may also affect household expenditure. We then move directly to examine several broad expenditure categories, such as food, clothing, health, education or transport. Secondly, we analyze expenditure levels on adult clothing items distinguished by gender thanks to the fact that the PHBS separately collects expenditure on (adult) men's and women's shoes and clothing. Finally, we can identify several child related expenditure items in the data. For example, we can distinguish spending on shoes and clothing for children aged below 13. Additionally, we extend the child related categories to include expenditure on kindergarten, private schooling and tutoring, and examine two specific categories covering expenditure, which is likely to include

[2] In the United States the Consumer Expenditure Survey also contains some information on child specific expenditure. Among others these are clothing and education. In the US data children goods are classified up to the age of 17 that is way into teenagehood. It might cause a problem because the older the children the more they take active part in consumption decisions. Therefore, in the US we often cannot distinguish between the decision of a teenager and a parent. Another problem in the case of the US is the much more lenient abortion legislation and broader access to IVF. We know that IVF treatment affects the probability of having multiple births (Schieve et al., 1999) and there is also evidence that immigrants from Asia keep their skewed towards boys gender preferences even long after immigration to Northern America (Almond et al., 2013). From this perspective Poland offers higher quality data, homogenous population, strict abortion legislation and only limited access to IVF treatment. It is thus unlikely that our estimates will be biased due to lack of randomness in gender of a child.

mainly child related goods ("Games, toys and hobby" and "Educational books and materials").

We find that first-born girl increases overall spending on clothing and shoes by around 3.5%. This effect, in the sample of all households, correspond to 7.0% increase in woman's spending, 6.4% decrease in man's spending and 9.7% increase in child's spending. We examine the findings for all families with children and for the subsamples of married and non-married parents. While most of the results do not differ by marital status, we find three statistically significant differences. These differences in the behavior of married and non-married parents can reflect different mechanisms including potential endogeneity of marital status with respect to the gender of the first child. Given that we show strong effects of first child's gender on partnership stability (Table 3) we conduct the heterogeneity analysis using the subsample of married parents. While this in itself limits the validity of our conclusions to a subsample of the data, it allows us to interpret them without regard to the indirect effect of partnership stability on family resources. Since the subsample of married couples is clearly a selected one, in the appendix, we also present heterogeneity estimates including all households as suggested by Ichino et al. (2011).

The results related to parental expenditure are particularly interesting from the point of view of the evidence on the effects of children's gender to date. While, since we only have information on expenditure on household level, we cannot distinguish who takes the specific spending decisions, this is in our view the first evidence which is suggestive of the fact that the gender of the first child may affect parental expenditure preferences. The fact that households with first-born girls spend 7.3% more on adult female clothing and 4.8% on adult female shoes is difficult to square with any other mechanism. One could hypothesize some complex form of strategic behavior, along the lines of Ichino et al. (2011), but any such a story would have problems with reconciling it with diminished expenditure on the adult male items. This would suggest that having a girl, which has been earlier shown to affect even voting preferences (Oswald and Powdthavee, 2010), may also influence preferences related to household expenditure in the developed world.

Along these lines the estimates on child related expenditure suggest another important result concerning early determination of social roles of boys and girls, including differential investments in early human capital developments reflected in kindergarten expenditure. Such an early "assignment" of gender roles could suggest a so far unexamined channel of gender bias in child investment, expressed through the type of expenditure on boys and girls. We are unable to say what causes the observed differences and directly identify them as discrimination in expenditure patterns. We also do not know what happens to these children later in the adolescence and adulthood and what consequences the differential treatment might have on them. We speculate, however, that these consequences could carry over into adult life and contribute to sustaining gender inequality.

2 Data and sample statistics

We use a dataset from the Polish Household Budget Survey (PHBS) for years 2003–2010. It is a nationally representative dataset collected annually by the Central Statistical Office in Poland.[3] The data includes information on household demographic composition, labor market activity, as well as income and expenditure data. In total, we have information on 286 379 households and 857 843 individuals over the eight years from 2003 to 2010. Since the dataset does not contain retrospective fertility information we rely only on the contemporaneous family composition. Individuals in every household are matched into families, which we define as a single adult or a couple (married or cohabiting) with any dependent children. This is done using available information on the relationship to the head of household and detailed pairing in the data using information on the unique identifiers of mothers, fathers and partners of each individual. Following other studies in the literature we limit the analysis to mothers aged between 18 and 40, who had their first child at the earliest at age of 16. The limit for the age of the oldest child is set at 12 years old, which on the one hand follows the practice of other studies (Dahl and Moretti, 2008), and on the other hand corresponds to the grouping of expenditure information.[4] Because expenditure data is collected at household and not family level we additionally limit the sample to households where there is only one family with children below 13 years of age. This does not preclude the possibility of there being more than one family in the household (for example parents living with children and their grandparents). In fact such complex households are relatively common in Poland (Haan and Myck, 2012). In the full PHBS sample 71.1% of households contain only one family, 22.2% include two and 6.7% three or more. In the sample used for the analysis 67.7% are single-family households.[5] We further restrict the sample to families with a mother present in the household,

---

[3] For more information on the methodology used by the Polish Central Statistical Office see Barlik and Siwiak (2011). The methodology is approved and monitored by the EUROSTAT. A summary of the survey methodology is given in the Appendix.

[4] Sample selection bias is likely to be very small as schooling in Poland is compulsory until the age of 18 and most children live with their parents until at least that age. We limit the age of the mother at 40 so we may also erroneously treat the second child as the first one. Such cases are not very likely, as they apply only to mothers: (a) who had a child aged 21 or less (40-21 = 19), (b) whose first child, aged 19+ is no longer in education and does not count as a dependent child, (c) had a second child at least four years later. Initial results of parental outcomes using 15 years old cutoff are presented in Karbownik and Myck (2011).

[5] Results are robust to limiting sample to households with a single family. These are available from the authors upon request.

and where the child-mother relationship is clearly specified in the data. We also exclude twins and triplets at first birth, widowed mothers and lone fathers.[6]

The analysis is conducted for the full sample of families and then separately for the sample of married couples and non-married families. This differentiation is dictated by the relatively strong effects of child gender on partnership stability which we present in Table 3. If the material condition of families is affected by the marital status of parents, and the latter driven to some extent by the gender of children, then any identified effect of gender on expenditure in the full sample could be a consequence of different partnership arrangements of girls' and boys' families, rather than directly of different expenditure behavior of parents of boys and girls. Section 3.1 also presents analysis related to other potential sources of bias, namely the indirect effects of gender through fertility and maternal labor supply.

Descriptive sample statistics are presented in Table 1 separately for all families and for married couples. The sample size for all families, used in the main analysis, is 48 397, and for the married couples it is 42 102. Among all families, 9.0% of children live without a father. This number can be decomposed into 4.6% of mothers who never marry and 4.4% of mother who are divorced or separated. There is also 237 widows which we exclude from the main analysis. The average number of children in the main sample is 1.63 and it is lower than among married couples. Furthermore, married couples have larger families, as the probabilities of having two children and three or more children are higher for this group. Importantly, however, the share of first-born girls is virtually identical in both samples. Finally, the demographic and socio-economic characteristics of the mother are similar in both examined samples.

The Polish Household Budget Survey contains detailed information on over 400 specific household expenditure items collected over a period of a month. These items are aggregated into 11 basic broad categories of expenditures such as for example food, clothing, housing and energy, health, education and transport. Table 2 provides the full list of the categories and information on mean expenditure levels in the two samples we consider. The households in the full sample spend on average 612 PLN ($192) on food and non-alcoholic drinks and 163 PLN ($51) on clothing and shoes. These values are only slightly different for the married sample (respectively 621 PLN and 170 PLN).[7]

---

[6] Lone fathers are defined as families in which mothers do not live with their children in the household. Since paternal custody is extremely rare in Poland any gender-bias in these situations could not be confirmed.

[7] All absolute values are given in Polish zloty (PLN) in June 2006 prices. The exchange rate between the US dollar and the PLN on 14.06.2006 was: $1=3.194 (National Bank of Poland, exchange rate Table A).

Table 1. Descriptive statistics – demographics and labor market.

	All families		Married couples	
	Mean	Standard deviation	Mean	Standard deviation
Sample means on family level:				
Living without a father	0.090	0.287	-	-
Never married	0.046	0.209	-	-
Separated or divorced	0.044	0.206	-	-
Married	0.870	0.336	1.000	-
Number of children	1.633	0.766	1.665	0.770
- one child	0.511	0.500	0.479	0.500
- two children	0.383	0.486	0.407	0.491
- three or more children	0.107	0.309	0.114	0.317
First born girl	0.482	0.500	0.480	0.500
Age of mother	30.242	4.553	30.429	4.429
Age of mother at first birth	24.104	3.845	24.235	3.778
Mother's education:*				
- Basic	0.352	0.478	0.341	0.474
- Secondary	0.363	0.481	0.364	0.481
- Higher	0.285	0.451	0.295	0.456
Mother works	0.522	0.500	0.533	0.499
Mother's income	669.876	1000.445	674.709	1009.328
Observations	48397		42102	
Families	31198		27132	

Notes: The samples include families in which the mother is younger than 41 and older than 17 and had the first child at the earliest at the age of 16; children's age 0-12; expenditure information for households with at most one family with children aged 0-12.
* Education categories cover: basic – no formal education, primary education, gymnasium and vocational education; secondary – secondary academic and secondary vocational education; higher education – education degree higher than secondary;
Source: authors' own calculations based on the PHBS data (2003-2010).

Additionally the data separates spending on such items as shoes and clothing into male and female adult (aged 13+) expenditure and child (aged <13) expenditures.[8] Moreover, the detailed categories allow us to identify the following items (see Table 2):

- games, toys and hobby (labeled as "Games and toys");
- educational books, educational stationary and other materials ("Educational materials");
- kindergarten and pre-kindergarten care expenditure ("Kindergarten and pre-K");
- schooling expenditure ("Schooling");
- private tutoring ("Tutoring").

While the first two of the five above categories could include spending on adult goods (e.g. on sports or fishing equipment and on training or education books unrelated to children's education), they are most likely to cover child related expenses. The last three categories are directly related to expenditure on children, although given that we focus on households with children below the age of 13 the incidence of both schooling and private tutoring is very low

[8] The total clothing and shoes category contains adult (male and female) and child clothing and shoes as well as several smaller items such as coloring or cleaning.

(see Table 2). This is because the majority of children in this age group in Poland attend state schools and at this level of schooling private tutoring is not very common. 15% of households in both the overall sample and married couples declare expenditure on kindergarten and pre-kindergarten care. As we can see in Table 2, 58% of families used in the analysis declare positive expenditure on child clothing, with the average expenditure of about 41 PLN ($13). Positive spending on games and toys is recorded in about 40% of the households, and about 35% declare positive expenditure on educational books and materials, with the average amounts spent on each of these items equal to about 20 PLN ($6) in the full sample and 19 PLN ($6) in the married sample. All of the five above categories can be treated as directly related to expenditures on children and they could also be treated as items of expenditure reflecting "human capital investment" in the children.

Table 2. Descriptive statistics – expenditure information.

	All families		Married couples	
	Mean	Standard deviation	Mean	Standard deviation
Broad expenditure items (average amounts)				
Food and non-alcoholic drinks	611.8	267.5	620.9	266.5
Alcohol tobacco and drugs	66.4	95.0	66.8	94.0
Clothing and shoes	163.3	253.7	169.5	259.9
Housing costs and energy	491.1	561.5	496.6	573.5
Housing equipment	151.0	395.3	157.2	413.5
Health	101.3	157.1	103.6	157.7
Transport	290.0	1003.1	311.3	1045.5
Communication	117.6	101.1	120.4	101.9
Recreation and culture	206.9	359.3	214.5	370.0
Education	50.3	168.7	52.2	173.5
Restaurants and hotels	48.7	197.9	50.7	187.1
Gender-specific adult expenditure (average amounts):				
Male shoes	9.1	39.0	9.9	40.7
Male clothing	27.0	99.4	29.4	104.2
Female shoes	18.3	52.4	18.2	52.4
Female clothing	46.4	109.0	47.1	109.4
Child-related expenditure (average amounts):				
Games and toys	20.3	55.3	21.3	57.1
Educational materials	19.4	62.2	19.9	62.9
Clothing	41.2	72.1	42.8	73.6
Shoes	20.3	40.5	20.8	40.9
Kindergarten and pre-K	30.1	94.3	31.1	96.1
Schooling	5.8	46.9	6.0	45.6
Tutoring	0.9	12.6	0.9	12.9
Child-related expenditure (any positive expenditure):				
Games and toys	0.40	0.49	0.41	0.49
Educational materials	0.35	0.48	0.36	0.48
Clothing	0.58	0.49	0.59	0.49
Shoes	0.37	0.48	0.38	0.49
Kindergarten and pre-K	0.15	0.36	0.15	0.36
Schooling	0.10	0.30	0.10	0.30
Tutoring	0.01	0.10	0.01	0.10
Total declared expenditure:	2523.5	1953.1	2594.3	2000.3
Observations	48397		42102	
Families	31198		27132	

Notes: The samples include families in which the mother is younger than 41 and older than 17 and had the first child at the earliest at the age of 16; children's age 0-12; expenditure information for households with at most one family with children aged 0-12. Values in June 2006 prices.
Source: authors' own calculations based on the PHBS data (2003-2010).

3 Modeling the effect of children's gender on household expenditures

Our identification strategy, as in the case of most studies quoted above, relies on treating the child's sex at birth as randomly determined. While some doubts have been raised with respect to the randomness of this outcome (Das Gupta, 2005; Hesketh et al., 2005) virtually all studies rely on this approach. The assumption of gender randomness implies that any differences that we observe in terms of household expenditure can be attributed to the gender of the child. Since the higher parity fertility might be endogenous (as showed in table 4), the most common approach in the literature is to focus on the gender of the first child, in which case the estimated model for each of the expenditure items takes the following form:

$$E_i^j = (\text{First child girl}_i)'\alpha_1 + X_{1i}'\alpha_2 + X_{2i}'\alpha_3 + \varepsilon_i \tag{1}$$

where E_i^j is the expenditure of household *i* in expenditure group or item *j*, vector X_1 contains mothers' socio-demographic characteristics (mother's age at first birth, cubic polynomial in age, educational attainment indicators), while X_2 includes town size indicators, regional and year dummies.[9] The *First child girl* indicator takes value 1 if the first-born child was a girl, and zero if it was a boy and ε_i is the residual, which is clustered at household level because some households are observed twice in our data. Since we are interested in estimating the differences between a single girl-birth and a single boy-birth we exclude twin and triplet births at first pregnancy from the sample. Equation (1) is estimated by OLS in levels. To be able to compare the results across papers in each case we also report the percent effect, which is the odds ratio minus one.[10]

Our approach loses its validity in case of sex selective abortion, but this can be safely assumed away given the strict anti-abortion legislation in Poland and important cultural factors. Furthermore, as we pointed out in the Introduction, if we want to interpret α_1 as a reflection of expenditure differ-

---

[9] One might be worried that maternal education and town size are endogenous from the perspective of first child gender. First, when we do not control for these the results do not change. Second, in tables A1 and A2 in the appendix we directly show that gender of a child is not related to these controls.

[10] One can imagine comparing twin-girl births to twin-boy girls. This idea although interesting mixes the effects of gender and the effects of fertility. Furthermore, in our data we do not have enough power to credibly conduct such an analysis. Results using log expenditure are available upon request. Given that we observe some households multiple times we have also estimated random effects models and models where we only keep the first interview for each household. Most of the results survive these robustness checks both qualitatively and quantitatively. Some coefficients, however, become insignificant because of the reduced sample size in the latter method. These results are available upon request.

entiation with respect to the gender of children we ought to examine the role of any potential intermediate confounding factors. These include the effect of child's gender on partnership stability, fertility and mothers' labor supply which have been shown in other studies to be related to the gender of the first child (Dahl and Moretti, 2008; Ichino et al., 2011).

How much families spend on different types of goods depends to a large extent on their available resources which can be affected by the three noted channels, and these in turn could be influenced by the gender of children. For example, if a first-born boy increases the probability of partnership stability, and this has a positive effect on family resources, expenditure levels in such families could be higher. This would show up in the estimations as the effect of a first-born boy, but could reflect only the indirect effect of higher resources among families with a first-born boy, and not the effect of a different expenditure pattern directly resulting from the gender of the first child. The validity of our conclusions drawn from the analysis of the relationship between family expenditure and gender of children will thus hang heavily on the role of these indirect channels. Detailed estimates of the effects of first child's gender on these three outcomes are presented below. As we demonstrate there is a significant relationship between the gender of the first child and partnership stability as well as fertility, but no significant effects on maternal labor supply. We argue that partnership stability is the principal channel which reflects differentiation in behavior of parents by the gender of their first child. There are no effects of child gender on maternal labor supply in Poland and the fertility channel cannot be strictly assigned to parental gender preferences and may just be a consequence of the partnership stability effect (more stable couples decide to have more children).[11]

3.1 Potential confounding factors: partnership stability, fertility and labor supply

Marital stability

Table 3. First child's gender and family status.

VARIABLES	(1) Living without father	(2) Mother never married	(3) Mother separated or divorced
First child a girl	0.006*	0.006**	0.001
	(0.003)	(0.002)	(0.002)
% effect	6.9	13.2	1.3
Observations	48,397	48,397	44,967

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Families with children living at home aged between 0 and 12, mothers aged <41 and >17, mother's age at first birth at least 16 years old.
Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

[11] Expenditure estimates have also been conducted with controls for fertility (controlling for having more than one child or more than two children) and they do not change qualitatively. These results are available upon request.

In Table 3 we present regression results for the model specified in equation (1) for the probabilities of living without a father, mother being never married and mother being divorced or separated conditional on being ever married.[12] Significant coefficient on the "*first child girl*" variable has usually been interpreted in the literature as a reflection of parents' gender preferences through its effect on the stability of parental partnership. Our results confirm strong and statistically significant influence of the gender of the first-born child on family structure. A *first child girl* increases the probability of children living without a father and mother never marrying. Probability of living without a father increases by 6.9% when comparing a family whose first child was a boy to a family whose first child was a girl. We need to remember though that this estimate provides the total effect including all the possible indirect effects that operate through subsequent fertility choices. The probability of the mother never marrying is also higher if a first-born child is a girl, with a statistically significant percent effect of 13.2%. Unlike in the previous research, however, we do not find any significant effects of child gender on probability of divorcing or separating conditional on being ever married. The results suggest that gender of a child can have a detrimental effect on family stability, and thus it may also influence family resources and the expenditure decisions. Since the first child girl reduces the stability of the family, it might well also bias downwards the consumption estimates.

Fertility

Another important channel which can indirectly affect family expenditure is the consequence of the gender of the first child on fertility, and its subsequent effect on total and per capita resources. For example, if parents have preferences for having a boy (or a girl) the gender of the first child may lead to different subsequent fertility decisions. Furthermore, child gender could influence not only fertility per se but also the spacing between the first and subsequent children, and closely spaced siblings might impose larger financial burden on the household's budget. These decisions could also be affected indirectly through the effect on partnership stability. For example, more stable relationships might result in higher fertility. As we show below once we take this indirect channel into account, and impose some specific assumptions about parental gender preferences our data is inconclusive as to the overall role of the gender of the first child on fertility. Main fertility estimates are presented in Table 4. Columns (1) to (3) present the results for the sample of all families, while columns (4) to (6) present the results for married couples.

[12] Following the literature we present OLS results. Probit estimates, available from the authors upon request, are qualitatively identical and quantitatively very close to results based on the linear probability models.

Table 4. Effects of first child's gender on fertility.

	All families			Married couples		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Total number of children	Two or more children	Time since first birth	Total number of children	Two or more children	Time since first birth
First child a girl	-0.008	-0.011**	-0.026	-0.006	-0.013**	-0.050
	(0.007)	(0.005)	(0.040)	(0.008)	(0.005)	(0.043)
% effect	-0.5	-2.2	-0.0	-0.4	-2.4	-0.1
Observations	48,397	48,397	48,397	42,102	42,102	42,102

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Families with children living at home aged between 0 and 12, mothers aged <41 and >17, mother's age at first birth at least 16 years old.
Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

In both samples the results are virtually identical and suggest that although a first-born's gender does not have an effect on family size or spacing between first and second birth, it does affect negatively the fertility decisions at parity two.[13] Furthermore, the magnitude of the effects is much smaller than in the stability analysis suggesting that fertility should not be a major driver in the expenditure analysis. It is notable though that the negative coefficient on *first child girl* in the fertility equation points towards girl preferences which seems to contradict our family stability findings. We need to remember though that higher fertility will be driven by sample selection bias related to partnership stability.[14] The reason for the potential bias is the fact that unlike in the case of married parents, for non-married mothers, if their fertility is affected by the separation, we do not observe their child preference as reflected in the number of children conditional on the gender of the first-born. To examine the potential extent of this bias, we estimate the role of gender of the first child with *assumed* gender preferences of parents who are no longer living together. This allows us to construct bounds for extreme behavior using assumptions concerning child's gender preferences.

Table 5. Probability of two or more children under alternative assumptions of preferences of separated parents.

	(1)	(2)	(3)
VARIABLES	All separated parents have boy preferences	All separated parents have girl preferences	Separated parents have either boy or girl preferences
First child a girl	0.029***	-0.047***	-0.010**
	(0.005)	(0.005)	(0.005)
% effect	5.8	-8.9	-2.0
Observations	48,397	48,397	48,397

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Families with children living at home aged between 0 and 12; mothers aged <41 and >17, mother's age at first birth at least 16 years old. Imputations of children for separated families adjusted for the probability of having more than one child. In column (3) - separated parents with a first-born girl are assumed to have "boy preferences", and those with a first-born boy are assumed to have "girl preferences".
Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

[13] We do not find any statistically significant evidence that gender affects fertility at any other parity margin. Results are available upon request.
[14] Although we can observe the current marital status, we do not observe the entire partnership history. It is possible that some of these women have been previously divorced, and since women whose first child is a girl are more likely to get divorced and have lower fertility, the relationship between child gender and fertility will be biased towards finding a negative relationship in the fertility analysis. In our case the coefficients for the married and the total sample are similar so any bias should not be severe.

Table 5 presents the estimates of the effect of *first child girl* on the probability of having two or more children under such assumptions.[15] First, we assume that all parents who no longer live together have "boy preferences", i.e. we assume that parents with a first-born girl would have had another child had they stayed together (column 1). Further, we assume that all separated parents have "girl preferences", i.e. that those with a *first-born boy* would have another child had they not separated (column 2). Finally, we assume that separated parents with a *first-born girl* have "boy preferences", and those with a *first-born boy* have "girl preferences" (column 3).

Our results confirm the existence of the potential sample selection bias, and demonstrate that the negative effect of gender of the first child may turn positive (and statistically significant) under the assumption that all separated parents had "boy preferences". The assumption that generates these results is arguably very strong, and under the more natural one (column 3) the estimated coefficients still suggest overall "preferences" in favor of girls. Thus, it is not implausible to argue that there is a negative and significant effect of first-born girl on fertility, although it seems that the larger-in-magnitude channel which could affect our expenditure results is through family stability.

Maternal labor supply

Results presented in Table 6 show the effect of children's gender on mothers' employment and labor market income. The sample focuses on the one hand on all families (columns 1 and 2) and on the other hand on widows (column 3).[16] The first column provides the total effect while the latter two columns intend to uncover the direct effect of gender on labor supply decisions that are independent from fertility and marital stability. First, we study a sample of mothers whose first child is no more than two years old (column 2). Arguably in this case the majority of the women decide not to have another child, at least temporarily. Secondly, we analyze a sample of widowed mothers whose marriage ended through an exogenous shock.[17] Unlike Ichino et al. (2011) we do not find any evidence that gender of a first-born child affects significantly any of the labor market outcomes.[18] Even considering

[15] The estimations are carried out so that families without a father are imputed additional children conditional on the gender of their first child. Imputations are adjusted by the probability of having another child, i.e. only parents with higher than average probability of having another child (estimated on the sample of non-separated parents) are imputed an additional child. Details of these estimations are available from the authors upon request.

[16] Although we do not use widows in the consumption analysis in table 6 we document that we cannot confirm findings from Ichino et al. (2011) in the widowed sample.

[17] Given the sample size we treat the widowhood results with some caution, however in order to obtain more reliable estimates we used the whole sample and interacted widowhood with first-born child gender. The results are available upon request and we do not find any significant effects of either the gender itself or the interaction term.

[18] In their paper Ichino et al. (2011) do not present any results for paternal labor supply decisions. Since these might also influence household consumption we also run specifications

the size of the coefficients the effects would be relatively small. Thus, in the case of Poland, unlike in the advanced economies studied by Ichino et al. (2011), we reject the hypothesis that the gender of a first-born child matters for maternal labor supply. Furthermore, this suggests that child's gender should not affect expenditure patterns through the indirect effect on mothers' labor supply decisions.

As we showed in the above analysis of the three indirect channels which may affect our estimates of the relationship between the gender of a first child and expenditure patterns it is only the first one which could play a significant role. Because of the importance of the partnership stability channel, analysis of expenditure patterns is conducted on the entire sample as well as on subsamples differentiated by marital status. Heterogeneity analysis presented in Section 5 shows results for the full sample and for the sample of married couples. While there are differences in some of the estimates, few of these are statistically significant and the general pattern of the results holds in all samples. We thus argue that the identified relationship between expenditure patterns and children's gender reflects its direct effect and is not driven by the effects of the three explored channels on family resources.

Table 6. Effect of first child's gender on mother's labor supply.

	(1)	(2)	(3)
	All families		Widows
VARIABLES	0-12	0-2	0-12
Panel A: Probability of working			
First child a girl	-0.000	0.002	-0.067
	(0.005)	(0.009)	(0.074)
% effect	-0.0	0.5	-11.1
Panel B: Monthly labor income			
First child a girl	-9.633	-3.587	10.814
	(9.100)	(20.699)	(123.912)
% effect	-1.4	-0.6	1.7
Observations	48,397	10,647	237

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Families with children living at home aged between 0 and 12, mothers aged <41 and >17, mother's age at first birth at least 16 years old.
Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

4 Differential expenditure by gender of the first child. Baseline results

Below we present results of the estimation of the model outlined in Section 3 for various expenditure categories. The results of the baseline estimations are presented in Tables 7-11. Expenditure items are first analyzed in the 11 broad categories (Table 7). We then look at specific types of adult clothing

from columns (1) and (2) for fathers in married families. We do not find any effects of a first-born girl on the extensive margin of paternal labor supply and we find only marginally significant small, negative 3.3% effect of having a first-born girl on paternal income in the sample of children aged 0-2. Note that this sign works in the opposite direction than the effect found in Ichino et al. (2011) for females. These results are available from the authors upon request.

expenditure differentiated by gender (Tables 8 and 9) and finally at child related expenditure (Tables 10 and 11). From among the broad expenditure categories we find a statistically significant effects of the gender of the first child for "Food and non-alcoholic drinks", "Clothing and shoes" and "Communication", which are respectively 1.1% lower, 3.5% higher and 2.1% higher among households with first-born girls compared to those with first-born boys. All these results hold in the sample of married couples, however, only the result on "Communication" expenditure is present in the non-married sample and it is barely significant. The difference between the married and non-married sample, however, is not statistically significant in any category.

Going deeper into the more detailed expenditure categories we find stable pattern by first child's gender on gender-specific expenditure on adult clothing and shoes (Tables 8 and 9). In the full sample we find that households spend 7.3% more on female clothing and 4.8% more on female shoes and that on the extensive margin households with a first-born girl are 2.8% more likely to spend money on female clothing and shoes and 3.7% less likely to declare expenditure on male clothing and shoes items (3.2% among married). In the sample of married couples the estimates on the intensive margins for female clothing and shoes are even higher at 8.0% and 6.8%, respectively. The probabilities for this combined category of female "Clothing and shoes" are positive 2.8% and 3.2% for all and married women, respectively. On the extensive margin, we could not identify corresponding effects for the non-married sample (except for barely significant estimate for male shoes), but with the exception of female shoes the signs of the estimated coefficients are consistent with the estimates for the married sample.

The results we find for adult clothing and shoes correspond in an interesting way with the findings for the respective child categories (Tables 10 and 11). Households with first-born girls spend 9.7% more on child clothing compared to those with a first-born boy (10.3% among married couples). Result from Tables 8 and 10 may thus suggest a degree of complementarity of expenditures between mothers' and daughters' clothing consumption, and a positive effect of a first-born girl on both. It is notable, though that the differences by gender are driven by the married sample (effects are twice as large) and in this case the differences between the coefficients estimated for the married and non-married samples are only marginally statistically insignificant. The pattern of spending found on the intensive margin is reflected also on the extensive margin (Table 11) but here the differences between estimates in married and non-married sample are not so pronounced. We find that households with first-born girls are 4.3% more likely to declare spending on children's clothing. This effect is similar in the full sample and in the two subsamples, although it is statistically insignificant among the households with non-married parents.

Among the analyzed child related expenditures we find statistically significant effects of the gender of a first-born child on several other items. On the intensive margin, these include "Kindergarten and pre-kindergarten care" and "Games and toys" expenditure, which in the full sample are lower among households with first-born girls by 6.0% and 12.8% respectively. On the extensive margin, households with first-born girl are less likely to declare expenditure within these categories by 3.2% (insignificant) and 5.9%. The results on the intensive margin hold for the married sample, and while kindergarten expenditure among non-married parents does not seem to differ by the gender of a first-born child (coefficient larger but insignificant), the effect on expenditure on "Games and toys" for this subsample (-27.6%) is almost triple the size of the effect estimated for married households. This effect on the extensive margin is also three times as strong for the non-married compared to the married sample (14.5% vs. 4.7%). On the extensive margin (Table 11) we also find that parents of first-born girls more frequently declare expenditure on "Educational books and other materials". This applies both to married (4.4%) and non-married (8.6%) households.[19] Finally, although expenditure on private schooling in Poland is relatively low the data suggest opposite effects of child gender on this type of expenditure depending on the marital status. Married parents of first-born girls are 5.7% more likely to declare expenditure on schooling while non-married parents are 16.1% less likely to do so.

[19] In our view, this does not contradict findings for the kindergarten expenditure on the extensive margin. When we split this category into educational books, educational magazines, educational prints and educational stationary then we find positive and significant coefficients on the last two categories, 0.001 (p-value 0.093) and 0.018 (p-value 0.000), respectively. This suggests that families have a lot of tiny expenditure on educational stationary (like pencils or crayons) that are skewed towards girls, however, these are so minor that we do not observe any differences on the intensive margin in the expenditure on educational materials.

Table 7. First child gender and broad expenditure categories.

VARIABLES	(1) Food and non-alcoholic drinks	(2) Alcohol, tobacco and drugs	(3) Clothing and shoes	(4) Housing costs and energy	(5) Housing equipment	(6) Health	(7) Transport	(8) Communica-tion	(9) Recreation and culture	(10) Education	(11) Restaurants and hotels
Panel A: All households (N=48,397)											
First child a girl	-6.442**	0.772	5.675**	-2.822	-2.631	0.700	8.641	2.443**	-1.957	0.384	-0.989
	(2.770)	(1.006)	(2.460)	(5.270)	(3.618)	(1.501)	(9.206)	(0.993)	(3.362)	(1.630)	(1.847)
% effect	-1.1	1.2	3.5	-0.6	-1.7	0.7	3.0	2.1	-0.9	0.8	-2.0
Panel B: Households with married partners (N=42,102)											
First child a girl	-5.072*	1.315	6.948**	-5.146	-2.019	0.945	8.993	2.332**	-1.454	0.882	-0.071
	(2.944)	(1.071)	(2.706)	(5.743)	(4.075)	(1.614)	(10.321)	(1.074)	(3.698)	(1.785)	(1.874)
% effect	-0.8	2.0	4.2	-1.0	-1.3	0.9	2.9	2.0	-0.7	1.7	-0.1
Panel C: Households with non-married partners (N=6,295)											
First child a girl	-12.577	-2.408	-1.957	15.265	-5.097	-1.209	10.894	4.097*	-4.452	-3.152	-8.212
	(7.655)	(2.877)	(5.342)	(12.370)	(5.978)	(4.051)	(15.993)	(2.478)	(7.285)	(3.706)	(7.718)
% effect	-2.3	-3.7	-1.6	3.4	-4.5	-1.4	7.7	4.2	-2.8	-8.0	-20.7
p-value B=C	0.358	0.223	0.135	0.131	0.670	0.620	0.920	0.512	0.712	0.325	0.304

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Children living at home aged between 0 and 12; mothers aged <41 and >17; mother's age at first birth at least 16 years old. Samples restricted to households with only one family with dependent children.

Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

Table 8. First-child gender and adult clothing expenditure. Intensive margin.

	(1)	(2)	(3)	(4)	(5)	(6)
	Clothing and shoes		Clothing		Shoes	
VARIABLES	Man	Woman	Man	Woman	Man	Woman
Panel A: All households (N=48,397)						
First child a girl	-2.284**	4.173***	-1.798**	3.269***	-0.699*	0.862*
	(1.004)	(1.245)	(0.900)	(1.046)	(0.359)	(0.490)
% effect	-6.4	7.0	-6.5	7.3	-7.4	4.8
Panel B: Households with married partners (N=42,102)						
First child a girl	-2.114*	4.800***	-1.670*	3.613***	-0.684*	1.202**
	(1.125)	(1.341)	(1.014)	(1.130)	(0.402)	(0.526)
% effect	-5.4	7.9	-5.5	8.0	-6.7	6.8
Panel C: Households with non-married partners (N=6,295)						
First child a girl	-2.674*	-0.229	-2.221	0.927	-0.528	-1.546
	(1.625)	(3.400)	(1.354)	(2.827)	(0.593)	(1.341)
% effect	-18.0	-0.4	-19.1	2.3	-13.2	-8.0
p-value B=C	0.776	0.167	0.744	0.376	0.827	0.056

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Children living at home aged between 0 and 12; mothers aged <41 and >17; mother's age at first birth at least 16 years old. Samples restricted to households with only one family with dependent children.

Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

Table 9. First-child gender and adult clothing expenditure. Extensive margin.

	(1)	(2)	(3)	(4)	(5)	(6)
	Clothing and shoes		Clothing		Shoes	
VARIABLES	Man	Woman	Man	Woman	Man	Woman
Panel A: All households (N=48,397)						
First child a girl	-0.015***	0.016***	-0.013***	0.016***	-0.008***	0.009**
	(0.005)	(0.005)	(0.004)	(0.005)	(0.003)	(0.004)
% effect	-3.7	2.8	-3.7	3.0	-7.2	4.1
Panel B: Households with married partners (N=42,102)						
First child a girl	-0.013***	0.018***	-0.012**	0.018***	-0.008**	0.013***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.003)	(0.004)
% effect	-3.2	3.2	-3.2	3.4	-6.0	5.6
Panel C: Households with non-married partners (N=6,295)						
First child a girl	-0.011	0.000	-0.008	0.002	-0.009*	-0.014
	(0.010)	(0.013)	(0.009)	(0.013)	(0.006)	(0.011)
% effect	-6.1	0.1	-5.2	0.5	-17.3	-5.7
p-value B=C	0.806	0.196	0.707	0.270	0.778	0.025

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Children living at home aged between 0 and 12; mothers aged <41 and >17; mother's age at first birth at least 16 years old. Samples restricted to households with only one family with dependent children.
Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

Table 10. First-child gender and child expenditure. Intensive margin.

VARIABLES	(1) Games and toys	(2) Educational materials	(3) Clothing	(4) Shoes	(5) Clothing and shoes	(6) Kindergarten and pre-K	(7) Schooling	(8) Tutoring
Panel A: All households (N=48,397)								
First child a girl	-2.773***	0.524	3.816***	-0.204	3.531***	-1.859**	-0.038	0.057
	(0.548)	(0.623)	(0.700)	(0.385)	(0.877)	(0.925)	(0.471)	(0.117)
% effect	-12.8	2.7	9.7	-1.0	6.1	-6.0	-0.7	6.6
Panel B:Households with married partners (N=42,102)								
First child a girl	-2.522***	0.272	4.212***	-0.228	3.913***	-1.803*	0.361	0.035
	(0.607)	(0.678)	(0.767)	(0.418)	(0.960)	(1.008)	(0.474)	(0.128)
% effect	-11.2	1.4	10.3	-1.1	6.6	-5.6	6.2	3.8
Panel C: Households with non-married partners (N=6,295)								
First child a girl	-4.415***	2.512	1.388	0.082	1.323	-1.989	-2.912	0.261
	(1.060)	(1.543)	(1.560)	(0.970)	(2.000)	(2.175)	(1.915)	(0.254)
% effect	-27.6	17.5	4.7	0.5	2.9	-8.2	-46.6	51.4
p-value B=C	0.119	0.182	0.103	0.769	0.241	0.938	0.096	0.427

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Children living at home aged between 0 and 12; mothers aged <41 and >17; mother's age at first birth at least 16 years old. Samples restricted to households with only one family with dependent children.

Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

Table 11. First-child gender and child expenditure. Extensive margin.

VARIABLES	(1) Games and toys	(2) Educational materials	(3) Clothing	(4) Shoes	(5) Clothing and shoes	(6) Kindergarten and pre-K	(7) Schooling	(8) Tutoring
Panel A: All households (N=48,397)								
First child a girl	-0.024***	0.017***	0.024***	-0.000	0.012**	-0.005	0.003	0.000
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.003)	(0.001)
% effect	-5.9	4.8	4.3	-0.1	1.7	-3.2	3.3	3.8
Panel B: Households with married partners (N=42,102)								
First child a girl	-0.020***	0.016***	0.026***	-0.001	0.012**	-0.005	0.006*	0.000
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.003)	(0.001)
% effect	-4.7	4.4	4.5	-0.2	1.8	-3.5	5.7	1.1
Panel C: Households with non-married partners (N=6,295)								
First child a girl	-0.051***	0.024**	0.017	0.005	0.013	-0.000	-0.014**	0.003
	(0.012)	(0.011)	(0.013)	(0.012)	(0.013)	(0.009)	(0.007)	(0.002)
% effect	-14.5	8.6	3.5	1.6	2.2	-0.3	-16.1	56.1
p-value B=C	0.016	0.504	0.535	0.648	0.943	0.595	0.011	0.265

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Children living at home aged between 0 and 12; mothers aged <41 and >17; mother's age at first birth at least 16 years old. Samples restricted to households with only one family with dependent children.

Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

5 Heterogeneity analysis

To gain a better understanding of the factors driving the baseline results, in this section we look at the effect of child's gender on adult clothing and child related expenditure in several subsamples of married couples (the results for all families are presented in Tables A3 to A5 in the Appendix). The aim of this exercise, is on the one hand to examine the role of parental education in driving expenditure differentiation by gender of children, and on the other to analyze if there is any consistent pattern of this differentiation by cohort and maternal age at first birth. Other studies have found that various reflections of the gender bias and discrimination are negatively correlated with the level of education (Desai and Alva, 1998), and one could also expect that with the growing concerns and awareness about gender equality we should find stronger expenditure differentiation by child's gender among earlier cohorts. Finally, there is a large body of literature relating teenage childbearing and limited resources to adverse outcomes (Geronimus and Korenman, 1992; Akerlof et al., 1996).

The subsamples for heterogeneity analysis are split by maternal education (above secondary, secondary and below secondary), maternal cohorts divided in such a way that we construct three subsamples of comparable size (born prior to 1975, 1975-1978, and after 1978) and maternal age at first birth (first child born before age of 21, between 21 and 26, and after 26). The results are presented in Tables 12 to 14, respectively. These show the role of a first-born child's gender on expenditure on adult gender specific clothing and shoes, and the six categories of child specific expenditure that we analyzed in Section 4 (Games and toys, Educational materials, Clothing and shoes, Kindergarten and pre-kindergarten, Schooling and Tutoring).

As we can see in Tables 12 to 14 there is a clear pattern in the differentiation by child's gender of expenditure on adult clothing and shoes (columns 1 and 2) with respect to education, cohort or age at first birth. In the case of male clothing and shoes we find evidence of expenditure differentiation by the gender of the first child (with the exception of heterogeneity by maternal education). We show negative effects of similar magnitude for youngest cohorts and mothers who had their first child between the age of 21 and 26. Quite opposite we find barely significant but positive 13.1% increase in the spending on male adult clothing among youngest mothers. In the same group we also find 21.8% increase in spending on women's clothing. This is more than three times the effect found for mothers that had their first child later in life. Similarly we find large positive effects on adult female clothing for low educated mothers. If clothing (or at least some part of it) could be seen as a form of conspicuous consumption, then these findings may be related to the results found in Charles et al. (2009) who identify higher propensity for conspicuous consumption among lower socio-economic groups. The split by cohort suggests that results are driven by the earlier cohorts of parents.

Among those born prior to 1975, households with first-born girls spend as much as 16.7% more on female clothing and shoes compared to those with first-born boys. The results for younger cohorts (born in 1975 and later) are much smaller and statistically insignificant and coefficient equality for the three subsamples is in this case strongly rejected (see bottom row of Table 13).

Such a consistent pattern by education and age at first birth can also be found in Tables 12 to 14 with respect to the results on child related expenditure (columns 3-8). For example, lowest educated mothers and those who had their first child while teenagers reduce spending on kindergarten and pre-kindergarten by 12.5% and 16.9% if the first-born child is a girl in comparison to a boy, respectively. On the contrary, however, these mothers do not seem to spend significantly more on children clothing if the first-born child is a girl. In the "Games and toys" category expenditure differentiation by gender of the first child is almost equally strong among highest and lowest educated mothers and equals -12.9% and -14.1%, respectively. Similar pattern is observed for mothers who have their first child when they were relatively younger versus relatively older.

Gender driven differentiation of child related expenditure by cohort shows no stable pattern (Table 13, columns 3-8). Expenditure differentiation in the "Games and toys" category is strongest in the middle cohort (born between 1975 and 1978) in which households with first-born girls spend 14.5% less compared to those with first-born boys. The effect in this category is -11.0% among the oldest cohort, and -7.4% among the youngest cohort. Spending on children's "Clothing and shoes", on the other hand, is almost as strongly differentiated by gender among households with youngest and oldest mothers, and favors first-born girls by 9.9% and 8.2%, respectively. Similar to the analysis by maternal education, in all cohorts we find negative effects of first-born girls on "Kindergarten and pre-K" expenditure. These, however, are only statistically significant for the youngest cohort of mothers where the spending on childcare is 11.5% lower if a first-born child is a girl.

Table 12. First-child gender and expenditure. Households with married partners. Intensive margin. Maternal education categories.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Adult clothing and shoes		Children expenditure					
VARIABLES	Man	Woman	Games and toys	Educational materials	Clothing and shoes	Kindergarten and pre-K	Schooling	Tutoring
	Panel A: Above secondary (N=12,438)							
First child a girl	-4.753	4.346	-4.764***	0.915	6.366***	-2.952	0.602	0.047
	(3.026)	(3.386)	(1.527)	(1.088)	(2.134)	(2.770)	(1.376)	(0.340)
% effect	-7.5	4.3	-12.9	5.3	8.3	-4.9	6.2	3.0
	Panel B: Secondary (N=15,323)							
First child a girl	-1.793	5.597***	-1.388	0.382	4.042***	-1.158	0.187	0.063
	(1.545)	(2.130)	(0.912)	(1.095)	(1.545)	(1.362)	(0.600)	(0.173)
% effect	-5.0	10.3	-6.9	2.0	7.0	-4.2	3.8	7.5
	Panel C: Below secondary (N=14,341)							
First child a girl	0.199	4.493***	-1.752**	-0.496	1.740	-1.488*	0.193	-0.035
	(0.999)	(1.263)	(0.687)	(1.288)	(1.332)	(0.868)	(0.273)	(0.144)
% effect	1.0	14.5	-14.1	-2.2	3.8	-12.5	5.8	-8.1
p-value identical	0.212	0.899	0.145	0.704	0.160	0.844	0.958	0.903

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Children living at home aged between 0 and 12; mothers aged <41 and >17; mother's age at first birth at least 16 years old. Samples restricted to households with only one family with dependent children.
Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

Table 13. First-child gender and expenditure. Households with married partners. Intensive margin. Maternal cohorts.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Adult clothing and shoes		Children expenditure					
VARIABLES	Man	Woman	Games and toys	Educational materials	Clothing and shoes	Kindergarten and pre-K	Schooling	Tutoring
	Panel A: Born prior to 1975 (N=17,288)							
First child a girl	0.067	9.927***	-2.291**	-0.013	5.359***	-0.931	1.156	0.158
	(1.707)	(2.173)	(0.956)	(1.226)	(1.637)	(1.671)	(0.992)	(0.280)
% effect	0.2	16.7	-11.0	-0.1	8.2	-2.8	12.3	9.8
	Panel B: Born between 1975 and 1978 (N=12,958)							
First child a girl	-2.950	2.821	-3.393***	1.062	1.309	-1.807	-0.062	-0.012
	(1.848)	(2.253)	(1.064)	(1.299)	(1.646)	(1.840)	(0.719)	(0.160)
% effect	-7.4	4.5	-14.5	5.8	2.2	-5.2	-1.3	-1.9
	Panel C: Born after 1978 (N=11,856)							
First child a girl	-4.208*	-0.038	-1.760	-0.135	4.864***	-3.048*	-0.185	-0.044
	(2.453)	(2.603)	(1.158)	(0.872)	(1.651)	(1.664)	(0.287)	(0.094)
% effect	-10.8	-0.1	-7.4	-1.6	9.9	-11.5	-11.8	-21.3
p-value identical	0.281	0.008	0.560	0.735	0.164	0.666	0.431	0.790

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Children living at home aged between 0 and 12; mothers aged <41 and >17; mother's age at first birth at least 16 years old. Samples restricted to households with only one family with dependent children.
Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

Table 14. First-child gender and expenditure. Households with married partners. Intensive margin. Maternal age at first birth.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Adult clothing and shoes		Children expenditure					
VARIABLES	Man	Woman	Games and toys	Educational materials	Clothing and shoes	Kindergarten and pre-K	Schooling	Tutoring
Panel A: First child born before the age of 21 (N=6,916)								
First child a girl	3.453*	8.837***	-2.249*	-1.377	3.640	-2.812*	1.218	-0.317
	(1.934)	(3.213)	(1.181)	(1.912)	(2.239)	(1.449)	(0.783)	(0.250)
% effect	13.1	21.8	-14.1	-5.8	7.2	-16.9	33.8	-36.7
Panel B: First child born between 21 and 26 (N=24,202)								
First child a girl	-3.716**	3.439**	-1.425*	1.440	3.971***	-2.065*	-0.258	0.108
	(1.473)	(1.585)	(0.736)	(0.931)	(1.227)	(1.110)	(0.626)	(0.154)
% effect	-10.1	6.1	7.2	7.0	6.9	-7.3	-4.0	11.7
Panel C: First child born after 26 (N=10,984)								
First child a girl	-1.984	5.492*	-5.038***	-0.661	4.394**	-0.732	1.252	0.127
	(2.635)	(3.242)	(1.491)	(1.081)	(2.071)	(2.777)	(1.080)	(0.328)
% effect	-3.9	6.6	-15.4	-4.3	6.4	-1.5	21.5	14.0
p-value identical	0.012	0.308	0.094	0.217	0.969	0.789	0.248	0.324

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Children living at home aged between 0 and 12; mothers aged <41 and >17; mother's age at first birth at least 16 years old. Samples restricted to households with only one family with dependent children.
Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

6 Conclusions

Gender of children has been shown to influence decisions of their parents in several important dimensions including partnership stability, fertility and labor market activity (Dahl and Moretti, 2008; Lundberg and Rose, 2002; Ichino et al., 2011). There is also ample evidence from the developing countries that parents treat boys and girls differently when it comes to decisions on human capital investment (e.g. Jayachandran and Kuziemko, 2011; Barcellos et al., 2012). In both cases the mechanisms believed to be responsible for parental decisions involve either biased preferences against one gender or optimization mechanism reflecting different costs of investment in boys and girls or different returns from these investments.

Some of the findings presented in this paper can also be explained within these frameworks. Parents may be biased against girls when it comes to expenditure on ”Games and toys” (on average 12.8%), and against boys in expenditure on clothing (9.7%). They may also differentiate expenditure on boys and girls believing in different returns from such an ”investment” reflecting the strongly stereotypical approach to gender roles. In our view however, some of our results are difficult to square with the standard explanations for differentiated outcomes by the gender of children. As we showed in Section 4 parents in households with a first-born girl spend more on female clothing and female shoes (by 7.3% and 4.8% in the full sample) and less on the corresponding goods for males (by 6.5% and 7.4%) compared to households with first-born boys. These differences in expenditure patterns are robust to the indirect effects of children's gender on partnership stability, fertility and maternal labor market behavior. In our view it is impossible to reconcile these findings with any of the standard approaches and without resorting to the direct effect of children's gender on parental preferences.

Some evidence on such effects with regard to voting preferences already exists in the literature (Oswald and Powdthavee, 2010) and in our view this paper sheds new light on the effect of children's gender on their parents. Since expenditure data is collected at household level, we cannot identify who makes the expenditure decisions and so we can only speculate whose preferences are affected and how. Assuming, however, that parents have some autonomy in deciding on how much money to spend on their clothes and shoes, our findings point towards the effect of having a girl on higher expenditure of their mothers on these items. Perceived in this light the explanation of higher spending on children's clothing among households with first-born girls could be explained by a form of complementarity between spending of mothers and daughters, and not necessarily by resorting to differentiated investment of parents in their daughters “looks”.

A similar story could apply to the case of expenditure on "Games and toys", in particular that this is an overall household level category and covers not only spending on children's goods. If we allow for the direct effect of having a boy on parental preferences then the higher spending within this category (11.2% and 27.6% among households with married and non-married parents, respectively) could be explained from this perspective.

The fact that the gender of a first-born child affects parental preferences sheds new light on the results found so far in the literature in particular with respect to the effect of gender on labor market behavior (Ichino et al., 2011). At the same time, however, the explanations behind our findings should not overshadow the potential consequences of parental behavior on their off-spring's future outcomes. While it is uncertain how the amount of spending on clothing and toys affects children's development and their prospects, it is well documented that pre-school formal childcare has positive long-term effects (Blau and Currie, 2006). Households of married parents in Poland spend 6.0% less on kindergarten and pre-kindergarten care if their first-born child is a girl than otherwise. While these effects are relatively small they may translate into important disadvantages in school age and adult life. At the same time, we find that parents with first-born girls are more likely to buy any educational materials, which is driven by negligible expenditures on items like pencils or crayons that do not affect the average expenditure level on educational materials.

The overall differentiation of expenditure on child related goods by gender of the first child, regardless of the mechanism which drives it, may be seen as a reflection of early assignment to social role models. From this point of view, it is notable that there is no clear pattern in the degree of the effects by the education of mothers or their cohort. Such an early assignment of social roles, with implications that girls are expected to look nice and boys are freer to play, may act to slow down the process towards greater gender equality. Such a perspective on the findings of our paper would call for the need for better understanding of parental behavior and the reasons behind the observed differential treatment of boys and girls, and potentially for policy interventions in terms of access to childcare and awareness of gender issues.

References

Aizer, A., (2003), "Low take-up in Medicaid: Does outreach matter and for whom?", *American Economic Review Papers and Proceedings*, 93(2), 238-241.

Akerlof, G.A., J.L. Yellen and M.L. Katz (1996), "An analysis of out-of-wedlock childbearing in the United States", *Quarterly Journal of Economics*, 111:2, 277-317.

Almond, D., and J. Currie (2011), "Human capital development before age five", *Handbook of Labor Economics*, 4:4b, 1315-1486.

Almond, D., L. Edlund and K. Milligan (2013), "Son preference and the persistence of culture: Evidence from South and East Asian immigrants to Canada", *Population and Development Review*, 39:1, 75-95.
Barcellos, S., L. Carvalho, and A. Lleras-Muney (2012), "Child gender and parental investments in India: Are boys and girls treated differently?", NBER WP 17781.
Barlik, M., and K. Siwiak (2011), "Metodologia badania budżetów gospodarstw domowych." Zeszyty Metodologiczne GUS, Zakład Wydawnictw Statystycznych, Warszawa.
Behrman, J., (1988), "Intrahousehold allocation of nutrients in rural India: Are boys favored? Do parents exhibit inequality aversion?", *Oxford Economic Papers*, 40:1, 32-54.
Ben-Porath, Y., and F. Welch (1976), "Do sex preferences really matter?", *Quarterly Journal of Economics*, 90, 285-307.
Blau, D., and J. Currie (2006), "Pre-school, day care, and after-school care: Who's minding the kids?", *Handbook of the Economics of Education,* 2:2, 1163-1278.
Carneiro, P., and R. Ginja (2008), "Preventing behavior problems in childhood and adolescence: evidence from Head Start", Mimeo.
Cascio, E., (2009), "Do investments in universal early education pay off? Long-term effects of introducing kindergartens into public schools", NBER WP 14951.
Charles, K.K., E. Hurst and N. Roussanov (2009), "Conspicuous consumption and race", *Quarterly Journal of Economics*, 124:2, 425-467.
Dahl, G.B., and E. Moretti (2008), "The demand for sons", *Review of Economic Studies*, 75:4, 1085-1120.
Das, N., (1987), "Sex preferences and fertility behavior: A study of recent Indian data", *Demography*, 24, 517-530.
Das Gupta, M., (2005), "Explaining Asia's missing women: A new look at the data", *Population and Development Review*, 31:3, 529-535.
Desai, S. and S. Alva (1998), "Maternal education and child health: Is there a strong causal relationship?", *Demography*, 35:1, 71-81.
Figlio, D., Hamersma S., and J. Roth (2009), "Does prenatal WIC participation improve birth outcomes? New evidence from Florida", *Journal of Public Economics*, 93:1-2, 235-245.
Geronimus, A.T., and S. Kornman (1992), "The socioeconomic consequences of teen childbearing reconsidered", *Quarterly Journal of Economics*, 107:4, 1187-1214.
Haan, P., and M. Myck (2012), "Multi-family households in a labour supply model: a calibration method with application to Poland", *Applied Economics*, 44:22, 2907 - 2919.
Hesketh, T., Lu L. and Z.W. Xing (2005), "The effect of China's one-child family policy after 25 years", *New England Journal of Medicine*, 353:11, 1171-1176.
Ichino, A., E-A. Lindström, and E. Viviano (2011), "Hidden consequences of a first-born boy for mothers", IZA Discussion Papers, No. 5649.
Jayachandran, S, and I. Kuziemko (2011), "Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India." *Quarterly Journal of Economics*, 126:3, 485-1538.
Jha, P., R. Kumar, P. Vasa, N. Dhingra, D. Thiruchelvam, and R. Moineddin (2006), "Low male-to-female sex ratio of children born in India: national survey of 1.1 million households", *Lancet*, 367, 211-218.
Karbownik, K., and M. Myck (2011), "Mommies' girls get dresses, daddies' boys get toys: Gender preferences in Poland and their implications", IZA Discussion Papers 6232.

Levine, P.B., and D.W. Schanzenbach (2009), "The impact of children's public health insurance expansions on educational outcomes", NBER WP 14671.
Lundberg, S., and E. Rose (2002), "The effects of sons and daughters on men's labor supply and wages" *Review of Economics and Statistics*, 84:2, 251-268.
Oswald, A.J., and N. Powdthavee (2010), "Daughters and left-wing voting" *Review of Economics and Statistics*, 92:2, 213-227.
Raley, S., and S. Bianchi (2006), "Sons, daughters, and family processes: Does gender of children matter?", *Annual Review of Sociology*, 32, 401-421.
Schieve, L.A., H.B. Peterson, S.F. Meikle, G.Jeng, I.Danel, N.M. Burnett and L.S. Wilcox (1999), "Live-birth rates and multiple-birth risk using in vitro fertilization", *Journal of American Medical Association*, 282:19, 1832-1838.
Sen, A., (1990), "More than 100 million women are missing", *New York Review of Books*, 37:20, 61-66.

Appendix

Tables

Table A1. Validity of using education controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Education categories		Education grouped into three groups				Groups			
VARIABLES	Raw	Controls	Raw	Controls	High raw	High controls	Mid raw	Mid controls	Low raw	Low controls
First child a girl	-0.003	-0.006	-0.003	-0.005	0.003	0.004	-0.002	-0.003	-0.001	-0.001
	(0.020)	(0.018)	(0.009)	(0.008)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)
Observations	48,397	48,397	48,397	48,397	48,397	48,397	48,397	48,397	48,397	48,397

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). The dependent variables are all education categories (columns (1) and (2)), three education categories (columns (3) and (4)), three education categories estimated in separate regressions as dummy variables (columns (5) to (10)). No control variables included in odd columns. Even columns control for age at first birth, third order polynomial in maternal age, year and regional fixed effects. Children living at home aged between 0 and 12; mothers aged <41 and >17; mother's age at first birth at least 16 years old. Samples restricted to households with only one family with dependent children.
Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

Table A2. Validity of using town controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Town categories		Raw						Controls					
VARIABLES	Raw	Controls	500+	200-500	100-200	20-100	0-20	Village	500+	200-500	100-200	20-100	0-20	Village
First child a girl	0.005	-0.007	0.004	-0.002	-0.000	-0.005	-0.004	0.008	0.003	0.001	0.000	-0.005	-0.004	0.005
	(0.020)	(0.019)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.006)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.006)
Observations	48,397	48,397	48,397	48,397	48,397	48,397	48,397	48,397	48,397	48,397	48,397	48,397	48,397	48,397

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). The dependent variables are all town categories (columns (1) and (2)), six town categories estimated in separate regressions as dummy variables without (columns (3) to (8)) and with (columns (9) to (14)) controls. Controls include age at first birth, third order polynomial in maternal age, year and regional fixed effects. Children living at home aged between 0 and 12; mothers aged <41 and >17; mother's age at first birth at least 16 years old. Samples restricted to households with only one family with dependent children.
Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

Table A3. First-child gender and expenditure. Households with all mothers. Intensive margin. Maternal education categories.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Adult clothing and shoes		Children expenditure					
VARIABLES	Man	Woman	Games and toys	Educational materials	Clothing and shoes	Kindergarten and pre-K	Schooling	Tutoring
			Panel A: Above secondary (N= 13,789)					
First child a girl	-5.111*	3.817	-5.084***	1.171	5.659***	-3.184	-0.460	0.067
	(2.766)	(3.190)	(1.410)	(1.041)	(1.990)	(2.613)	(1.460)	(0.318)
% effect	-8.6	3.8	-14.1	6.8	7.5	-5.3	-4.4	4.3
			Panel B: Secondary (N= 17,582)					
First child a girl	-2.146	4.883**	-1.667**	0.622	4.394***	-1.145	0.156	0.134
	(1.439)	(2.045)	(0.832)	(1.007)	(1.419)	(1.269)	(0.539)	(0.161)
% effect	-6.4	8.8	-8.5	3.4	7.9	-4.2	3.3	17.8
			Panel C: Below secondary (N= 17,026)					
First child a girl	0.052	3.785***	-2.074***	-0.165	0.947	-1.679**	0.018	-0.040
	(0.882)	(1.151)	(0.613)	(1.148)	(1.200)	(0.788)	(0.244)	(0.126)
% effect	0.3	12.3	-17.1	-0.8	2.1	-14.2	0.6	-9.5
p-value identical	0.121	0.894	0.099	0.689	0.058	0.777	0.919	0.692

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Children living at home aged between 0 and 12; mothers aged <41 and >17; mother's age at first birth at least 16 years old. Samples restricted to households with only one family with dependent children.
Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

Table A4. First-child gender and expenditure. Households with all mothers. Intensive margin. Maternal cohorts.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Adult clothing and shoes		Children expenditure					
VARIABLES	Man	Woman	Games and toys	Educational materials	Clothing and shoes	Kindergarten and pre-K	Schooling	Tutoring
Panel A: Born prior to 1975 (N= 19,225)								
First child a girl	-0.167	9.476***	-2.358***	0.203	5.459***	-0.850	0.135	0.241
	(1.559)	(2.040)	(0.880)	(1.162)	(1.531)	(1.550)	(1.051)	(0.264)
% effect	-0.5	16.0	-11.7	0.7	8.5	-2.7	1.4	15.7
Panel B: Born between 1975 and 1978 (N= 14,655)								
First child a girl	-2.899*	2.049	-3.367***	1.294	0.771	-1.744	-0.004	-0.033
	(1.663)	(2.125)	(0.981)	(1.205)	(1.530)	(1.739)	(0.644)	(0.149)
% effect	-8.0	3.3	-14.9	7.2	1.3	-5.1	-0.1	-5.3
Panel C: Born after 1978 (N= 14,517)								
First child a girl	-4.308**	-0.265	-2.661***	0.357	4.132***	-3.287**	-0.134	-0.037
	(2.066)	(2.323)	(0.994)	(0.769)	(1.445)	(1.477)	(0.270)	(0.079)
% effect	-12.0	-0.5	-11.7	4.5	8.8	-12.5	-8.6	-19.7
p-value identical	0.234	0.003	0.740	0.766	0.082	0.514	0.957	0.598

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Children living at home aged between 0 and 12; mothers aged <41 and >17; mother's age at first birth at least 16 years old. Samples restricted to households with only one family with dependent children.
Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

Table A5. First-child gender and expenditure. Households with all mothers. Intensive margin. Maternal age at first birth.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Adult clothing and shoes		Children expenditure					
VARIABLES	Man	Woman	Games and toys	Educational materials	Clothing and shoes	Kindergarten and pre-K	Schooling	Tutoring
Panel A: First child born before the age of 21 (N= 8,763)								
First child a girl	3.011*	7.632***	-3.006***	-1.216	2.240	-2.692**	0.611	-0.198
	(1.615)	(2.822)	(1.018)	(1.619)	(1.917)	(1.278)	(0.689)	(0.211)
% effect	12.8	18.2	-19.1	-5.5	4.6	-16.5	16.4	-26.1
Panel B: First child born between 21 and 26 (N= 27,358)								
First child a girl	-3.897***	3.198**	-1.779***	1.594*	3.705***	-2.176**	-0.160	0.168
	(1.336)	(1.493)	(0.671)	(0.868)	(1.138)	(1.036)	(0.559)	(0.145)
% effect	-11.3	5.7	-9.3	8.0	6.6	-7.8	-2.7	19.2
Panel C: First child born after 26 (N= 12,276)								
First child a girl	-2.345	4.030	-4.682***	-0.279	4.474**	-0.389	-0.266	0.053
	(2.391)	(3.038)	(1.370)	(1.015)	(1.924)	(2.601)	(1.299)	(0.299)
% effect	-4.9	4.9	-14.9	-1.9	6.8	-0.8	-3.9	6.0
p-value identical	0.004	0.380	0.139	0.192	0.699	0.728	0.655	0.362

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Control variables include: mother's age at first birth, cubic polynomial in age, educational attainment indicators, town size indicators, regional and year dummies. Children living at home aged between 0 and 12; mothers aged <41 and >17; mother's age at first birth at least 16 years old. Samples restricted to households with only one family with dependent children.
Source: authors' own calculations based on the Polish Household Budget Survey data, 2003-2010.

Polish Household Budget Survey – summary of the methodology

The Polish Household Budget Survey is a representative survey of Polish households surveying over 37000 households per year. The survey is conducted every year and is spread over the entire calendar year with each household surveyed over a period of a month during which households record their expenditures and incomes. This information is complemented with an additional interview which is conducted at the end of each quarter of data collection (so called quarterly interview). Each year since 2005, when the most recent sampling procedure was introduced, the target sample is 37584 households.

In a case of refusal to participate among households from the principal gross sample, households are replaced by another household from a reserve list of randomly chosen households. This reserve list is prepared separately for each sampling unit. Households which drop out of the survey in the first half of their survey month are also replaced by households from the reserve list. Those who drop out in the second half of the month are not replaced. Households from the principal gross sample which agree to participate are re-interviewed in the same month of the following year. Households from the reserve list are not re-interviewed. The survey methodology has been developed in accordance with the EUROSTAT guidelines.

The overall response rate in the survey in 2010 was 50.2%. Survey non-response was either due to refusal to participate (48.1%), survey drop out during its duration (1.6%) or refusal to complete the final quarterly interview (0.1%). From among households which were approached to complete the survey for the first time in 2010 (either from the principal gross sample or from the reserve list) 59.5% did not participate in the survey, and from among those who participated in the previous year 14.9% did not complete the survey for the second time.

Essay 5

For some mothers more than others: How children matter for labour market outcomes when both fertility and female employment are low

1 Introduction

Wide spread entry of women into the labour force has been one of the most pronounced socio-economic developments in the 20th century, and high levels of female employment are crucial from the point of view of continued economic growth and financial stability of many welfare systems (Galor and Weil, 1996; Lagerlof, 2003; Klasen and Lamanna, 2009). At the same time, demographic changes determined by the current and future fertility levels will play a vital role in shaping these developments and will affect the costs of social programs. Given the potentially strong link between female employment and family size it seems that understanding the relationship between the two ought to be at the heart of policy discussions, especially in countries that are characterised by both low fertility and low female employment. In particular, in the light of rising unemployment in low-fertility countries, which have been most severely affected by the economic crisis such as Greece, Spain or Latvia, our findings may serve as a guide with respect to the relationship between fertility and labour supply in an environment, which will likely become more common in Europe in the near future.[1]

Employment rates of women with children, in particular those with young kids, are generally lower in comparison to women who either never had children or whose children are older or no longer live with their parents (Gronau, 1973; Schultz, 1990; Leibowitz et al., 1992; Ahn and Mira, 2002; Adsera, 2005). On the one hand, the presence of children induces various obvious constraints on labour market activity and may affect individual preferences over consumption and leisure.[2] On the other hand, women who have lower inclination to work may decide to have more children than those who are more strongly attached to the labour market, implying self-selection into larger families among women with weaker labour market attachment. This would result in lower rates of labour market participation among these mothers even without the causal link running from family size to lower employment. Such a potential selection means that a simple cross-sectional correlation between employment and the number of children could be biased (Killingsworth and Heckman, 1986; Blundell and Macurdy, 1999), and OLS estimates would overstate the negative effect of family size on maternal labour market outcomes. On the other hand, if stable labour market position of women implies having more children, then the OLS would understate the negative effect of family size on employment. The latter is more likely to

[1] Exogenous changes in policies have been used to identify changes in female labour supply. These include reforms that affect individuals' work incentives (Blundell et al., 2008) or tax reforms (Blundell et al., 1998). Blundell et al. (2005) provide theoretical collective framework for analyzing the labour supply in households with children.

[2] The presence of children in the family significantly affects female preferences for leisure and thus women's labour supply elasticities (Heckman, 1993; Joshi, 1998; Blau and Kahn, 2005).

occur in countries with less mature labour markets for which we have little evidence on the causal relationship between family size and employment.

In this paper, we follow a classical approach to identify the causal effects of children on labour market outcomes i.e., we use exogenous variation in the number of children driven by multiple births (Rosenzweig and Wolpin, 1980a; Rosenzweig and Wolpin, 1980b; Bronars and Grogger, 1994; Angrist and Evans, 1998; Jacobsen et al., 1999; Caceres-Delpiano, 2006; Vere, 2011).[3] Since parents expect to have a single offspring as a result of a pregnancy while in turn they get two (or more) kids, there is an exogenous variation in the size of the family that is independent from preferences related to the labour market.[4]

The distinguishing feature of this study is that the analysis is conducted on data from a regime characterised by a combination of relatively low levels of female employment and a low fertility rate. For the purpose of the analysis we use the Household Budget Survey data from Poland, a country with one of the lowest fertility and female employment rates in Europe. Partly as a result of this, the country faces one of the most severe demographic changes in the coming decades with old-age dependency ratio in 2050 at about 53.0. With fertility at 1.4 in 2009 Poland lags far behind countries such as Ireland (2.1), France (2.0), the UK (1.9) or Sweden (1.9).[5] In addition to low fertility levels, Poland has one of the lowest rates of female employment in the European Union, far below those of such countries as the Netherlands, Germany or Sweden. At the same time, during the 2000s Poland was on the stable path of economic growth with the annual GDP growth in the range of 1.6 to 6.8%.

These stylised facts make Poland an extremely valuable reference case for the analysis of the causal relationship between family size and employment in a low fertility – low female employment context, which to our knowledge has never been studied before.[6] The combination of low female employment

---

[3] In an earlier version of this paper we have also used gender preferences (Angrist and Evans, 1998; Chun and Oh, 2002; Angrist et al., 2006; Cruces and Galiani, 2007; Daouli et al., 2009 or Hirvonen, 2009) but this instrument turned out to be quite weak and unreliable in the case of Poland. We refer readers interested in these results to Karbownik and Myck (2012).

[4] Note that twining rates may not be purely random. For example, women with family history of twining have higher incidence of subsequent multiple births. Furthermore, twining rates increase with maternal age, being a twin, use of fertility drugs and specific nutritional aspects (Waterhouse, 1950; Bulmer, 1970; Lichtenstein et al., 1996; Westergaard et al., 1997). In the analysis we control for maternal age as well as age at first birth, and treat the instrument as exogenous. The incidence of in-vitro fertilization is still very low in Poland. Although the official statistics are not maintained, NGOs reports from late 2000s suggest that around 1.5% of live births is due to IVF procedures.

[5] Data on fertility rates and old-age dependency ratios are taken from EUROSTAT.

[6] There is a number of studies linking family size (fertility) and female employment based on data from the former Soviet bloc countries, yet to our knowledge these do not include a single causal study: Hungary (Saget, 1999), Romania (Fong and Lokshin, 2000), Poland (Matysiak, 2009; Bardasi and Monfardini, 2009) and the former East Germany (Bonin and Euwals, 2002).

and low fertility is particularly challenging from the policy-making point of view when a strong negative causal relationship between family size and female employment exists. In such a case any potential increases in fertility would reduce the effects of policies aimed at higher female labour market participation. On the other hand, if the relationship between family size and employment is weak, the policies aimed at gains in both of these domains could operate without significant negative spillover effects. Since this relationship may differ by individual and family characteristics in the paper we present detailed heterogeneity analysis. The findings for some of the groups suggest a positive bias in the OLS estimates of the effects of family size on employment, which are robust to several specifications. These results, while at first sight counterintuitive shed a new light on the analysed relationship and offer additional insights from the policy perspective. The conclusions are in our view valid not only for the Polish case, but could have broader application to other countries characterised by the combination of low fertility and low female employment.

Our results confirm the overall negative relationship between number of children and female labour supply. In line with the endogeneity hypothesis, the simple OLS estimates overstate the negative effect of childbearing on female labour force participation, but in the overall sample this bias is small. In the sample of all mothers with at least one child, we find that an additional child reduces the mother's probability of employment by 6.7 percentage points and it averages over all the subsequent children above the first one. Thus, the marginal effect of going from first to second child is larger in reality. The corresponding effect estimated for OLS is -8.3pp. The negative causal effect of additional children in the sample of all mothers with at least two children is much smaller (-2.9pp) and statistically insignificant, while the OLS suggests a statistically significant correlation of -6.8pp. This suggests endogeneity between fertility and labour market choices among families with more than two children. Naturally, given the estimation strategies we take, we can only examine the relationship between family size and labour market outcomes for families with at least one child and this limitation should be kept in mind throughout the discussion, i.e. we cannot explore the difference between having versus not having any children.[7]

Heterogeneity analysis shows significant variation in the nature of the family size – the labour market attachment relationship in Poland. We find that the negative causal effect established in the full sample is driven primarily by women who are highly educated, who come from the younger cohorts and who had their first child later in life. Of a particular interest should be the fact that in all of these subsamples we find strong negative effects of family size and additionally a positive bias of the OLS estimates relative to

[7] Aguero and Marks (2011) study the effects of going from zero to one child on maternal labour supply using infertility shocks in developing countries.

the 2SLS coefficients. The latter finding suggests that it is women with the strongest labour market attachment and/or with most secure labour market position who have larger families. We attribute that to the fact that in low fertility and low employment societies only families with secured labour market positions can afford to have children, and in particular more than one child (Brewster and Rindfuss, 2000). For women with less than higher education, who had first child in their teenage years and for those from earlier cohorts (born before 1978) we find no statistically significant causal effects of additional children on employment. Thus, in these cases the negative OLS coefficients result from the fact that it is the women with weaker labour market attachment who choose to have larger families. Lack of causal effects of larger families on female employment in these cases may also be related to an easier access to informal childcare among less educated women and those who had their children at a younger age (who may be living closer to their families) as well as to formal childcare which is often subject to means testing in Poland. The findings for earlier cohorts on the other hand, may reflect long-run effects of larger families which is likely to be lower compared to short-run effects. What is novel about these findings is that they suggest that it would be naive to expect that lower employment among women might result in higher fertility. In fact, if anything, the reverse is more likely to hold. Poor economic prospects of families would in such cases further aggravate the long-term socio-economic consequences of economic downturns with significant implications for countries affected by the recent depression.

Finally, we could not identify any significant causal effects of the number of children on female employment in the sample in which we approximate complete fertility history by looking at women whose last birth was more than six years prior to the interview. For this sample, however, using the twinning instruments we find strong and significant negative effects of family size on maternal labour income, and - in the case of families with at least two children - also on the income of fathers.

The rest of the paper is organised as follows. In the next section we describe the data and provide a set of summary statistics. We then present and discuss the estimation strategy (Section 3), which is followed by the main results of the paper and the heterogeneity analysis in Section 4. Section 5 concludes the paper.

2 Data and descriptive statistics

Our analysis is based on a dataset from the Polish Household Budget Survey (PHBS) for the years 2003–2010.[8] The PHBS is a nationally representative dataset collected annually by the Polish Central Statistical Office. The data includes information on household demographic composition, labour market activity, as well as detailed income and expenditure data. In total, we have data on 286 379 households and 857 843 individuals over eight years. The dataset does not contain retrospective fertility information, and thus we can rely only on contemporaneous family composition. Individuals in the data are matched into families, defined as a single adult or a couple (married or cohabiting) with any dependent children, through available relationship information. Since we use contemporaneous family information we restrict the sample to families with a mother present in the household, and where the child-mother relationship is clearly specified in the data. Following similar studies in the literature we limit the analysis to mothers aged between 18 and 40, who had their first child at the earliest at the age of 16, and whose oldest child was at most 15 years old at the time of the interview.[9] Additionally we impose the restriction that the youngest child is at least six months old to avoid potential bias due to lower labour market activity of mothers during the initial months following childbirth.[10]

The descriptive statistics are presented in Table 1 where we show information separately for families with at least one and at least two children. Statistics for the subsample of married or cohabiting mothers (below referred to as “couples sample”) differ very little from the full sample of mothers and we present them separately in the Appendix (Table A1). The sample size for families with at least one child is 60256 (52 986 couples), and for families with two or more children is 33 010 (30 573 couples). Given the design of the survey approximately 36% of these families are observed in two consecutive years. Among families with at least one child the average number of children is 1.74. About 14% of mothers in the sample have three or more children. Among those with two or more children the number of children (at 2.35) and the proportion with three or more children (at about 26%) in the full and in the cohabiting sample are essentially the same. Both the number of children and the proportion of mothers with two or more kids are lower

---

[8] For a summary of the survey methodology see the Appendix. Details of the methodology are given in Barlik and Siwiak (2011).

[9] The dataset contains a very small number of families with children without a mother. We do not have precise information if the mother in the data is the biological mother, but the families we use are limited only to the cases where the mother-child relationship is specified in the data. There is a number of cases where the children fulfill our age criteria but where only the father is identified in the data – 235 families. Since these are very rare and special cases we exclude them from the analysis.

[10] We impose the restriction at the threshold corresponding to statutory maternal leave in Poland. This additional restriction does not have any substantial effect on the results.

when compared to other studies in the literature (e.g. Angrist and Evans, 1998; Vere, 2011; Cruces and Galiani, 2007), which reflects the distinguishing feature of our study.[11] About 54% of mothers in the sample are working, and employment rates are very similar for the sample with at least one and at least two children. While this is higher than the rates in studies using Greek or Latin American data (Cruces and Galiani, 2007; Daouli, 2009), the rates of maternal employment in Poland are lower compared to those in the US (Angrist and Evans, 1998; Vere, 2011) and much lower in comparison to other European countries.[12] Employment rates of husbands, or partners of mothers (Table A1) is also similar in the two samples at about 81%. In both samples the raw female employment rate falls for women with three or more children by about 4 percent compared to mothers with either one or two children.

We use the number of children as our – potentially endogenous – family size variable in the analysis.[13] We instrument it by twins at first birth (*twins-1*; e.g. Rosenzweig and Woplin, 1980a) for families with at least one child and by twins at second birth for families with at least two children (*twins-2*; e.g. Angrist and Evans, 1998). We take a multiple birth as an observed case of twins in the family identified by month of birth of the children (in the sample of all families with at least one child we have four triplet births, two of these born at first birth and two at the second birth; we drop these households from the analysis). The mean of the *twins-2* indicator (0.010) is slightly lower than the mean of the *twins-1* indicator (0.011), which might be related to the fact that the probability of having twins rises with mother's age at conception (Mittler, 1971). Since this could be an outcome of the mother's choice, and thus affects the exogenous nature of the instrument, we incorporate demographic characteristics of the mother in the analysis.

In Table A3 in the Appendix, we present evidence on correlations between maternal education, maternal birth cohort and maternal age at first birth, and family size. These regressions suggest little endogeneity concern in the case of maternal education, age at first birth and cohorts in the full sample. Therefore, our heterogeneity analysis presented in Section 4.2 focuses on these three dimensions.[14]

---

[11] The only causal study where we found even lower fraction of women having more than two kids is Greece, with about 21% (Daouli et al., 2009).

[12] Angrist and Evans (1998) report between 56 and 66%; Cruces and Galiani (2007) report between 22 and 32%; Daouli (2009) reports between 25 and 38%; Vere (2011) reports between 54 and 66%. According to OECD maternal employment in 2010 was at the level of 69%, 70% and 73% in the Netherlands, Sweden and Switzerland respectively.

[13] Results using indicator variables for more than one child or more than two children give similar conclusions. These results are available from the authors upon request.

[14] We experimented with various other interesting dimensions like, for example, paternal income but these turned out not to be statistically independent of family size.

Table 1. Descriptive statistics – all families with children.

	With at least one child		With at least two children	
	Mean	Standard deviation	Mean	Standard deviation
Number of children	1.740	(0.846)	2.351	(0.693)
- one child	0.452	(0.498)	-	-
- two children	0.404	(0.491)	0.737	(0.441)
- three or more children	0.144	(0.351)	0.263	(0.441)
Twins at first birth (twins-1)	0.011	(0.102)	-	-
Twins at second birth (twins-2)	-	-	0.010	(0.098)
Same sex of first two born children	-	-	0.509	(0.500)
Two first born girls	-	-	0.237	(0.425)
Two first born boys	-	-	0.271	(0.445)
Age of mother	31.352	(4.844)	32.761	(4.127)
Age of mother at first birth	23.563	(3.662)	22.905	(3.253)
Mother's education:*				
- basic	0.385	(0.487)	0.457	(0.498)
- secondary	0.364	(0.481)	0.346	(0.476)
- higher	0.252	(0.434)	0.197	(0.398)
Mother works:**	0.539	(0.499)	0.538	(0.499)
- one child	0.541	(0.498)	-	-
- two children	0.548	(0.498)	0.548	(0.498)
- three or more children	0.508	(0.500)	0.508	(0.500)
Mother's labour income:	677.54	(975.73)	603.85	(918.45)
- one child	766.82	(1033.91)	-	-
- two children	684.24	(959.18)	684.24	(959.18)
- three or more children	379.10	(749.14)	379.10	(749.14)
Number of families	38383		20802	
N	60256		33010	

Notes: The samples include families in which the mother is younger than 41 and older than 17 and had the first child at the earliest at the age of 16; children's age ranges from 6 months to 15 years; labour incomes are unconditional monthly net values indexed by CPI to June 2006.
* Education categories cover: "basic" – no formal education, primary education, gymnasium and vocational education; "secondary" – secondary academic and secondary vocational education; "higher education" – education degrees higher than secondary;
** The sample sizes are 27246 for mothers with one child; 24314 for mothers with two children and 8696 for mothers with three or more children. Same sample sizes apply to mother's labour income.
Source: authors' own calculations based on the PHBS data (2003-2010).

3 Estimation strategy

We use exogenous variation in family size in the form of twinning (*twins-1* and *twins-2*), and examine the effects of family size measured as the number of children on employment and labour income. We thus consider the following linear model:

$$Y_i = X_i^{'}\alpha_1 + C_i^{'}\alpha_2 + \varepsilon_i \qquad (1)$$

where Y_i is a measure of labour supply (employment or labour income) of mother or father *i*, X_i is a set of control variables with respect to fertility, such as age of the mother at first birth, a polynomial in mother's age at the time of interview, as well as time and regional (voivodship) effects; C_i is the endogenous family size variable and ε_i is the residual. The error term is clustered at household level since we observe the same households multiple times in our data. We also exclude from the estimation births at parity higher than two. The first-stage equation (2) describes a relationship for twinning at j^{th} parity:

$$C_i = X_i^{'}\beta_1 + (twins - j_i)^{'}\gamma_k + \upsilon_i \quad (2)$$

where $Cov(twins - j_i, \upsilon_i) = 0$; $j=1,2$ is the indicator of twin birth parity and γ_k (k=1,2) are the first stage effects of the instruments.

In order for the instruments to be robust, in addition to their exogeneity with respect to labour market outcomes, we also need a strong relationship between the instruments and the endogenous variables. Table 2 presents the first stage results for the full sample of families linking the instruments to our family size variable. Using twins at either first or second birth is strongly correlated with the number of children in the family. The effects are highly significant with large t- and F-statistics. First birth twinning effect is about 0.64, while twining at second birth naturally has a larger impact of around 0.84.[15]

Table 2. OLS estimates of first stage relationships - all families.

	(1)	(2)
	Dependent variable: number of children	
Instruments for family size:		
Twins at 1st birth	0.643***	
	(0.028)	
	[22.66]	
Twins at 2n birth		0.840***
		(0.037)
		[22.67]
Partial R-squared	0.008	0.015
F-statistics on excluded instruments	514	514
LM statistic on underidentification test	362	233
N	60,256	33,010

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1), t-statistics on the coefficients in square brackets. All regressions include year and region specific effects. The additional covariates include age of mother at first birth, and a polynomial of mother's current age. Sample of mothers aged <18; 40> with oldest child younger than 16 years, who gave the first birth at the age of 16 at the earliest and whose last birth was 6 months prior to survey at the latest.
Source: authors' calculations based on BBGD data 2003-2010.

The discussion so far assumes homogenous causal effects i.e., the effects are not differentiated across different groups of the population. The literature has also tried to study heterogeneous treatment effects in which one allows for distribution of causal effects across individuals (Imbens and Angrist, 1994; Angrist and Krueger, 2001; Imbens, 2009). These are important from the perspective of external validity of our results and require stricter assumptions. In particular in our approach we need to assume that the instrument is as good as random (independence assumption), that the instrument operates only through a single and known channel (exclusion restriction) and that the instrument has the same effect on everyone who is affected (monotonicity assumption). In this paper the three assumptions are equivalent to assuming

[15] The coefficients obtained for Poland are generally larger than those for the US reported in Vere (2011). This conforms with differences in family size/fertility between Poland and the US. If families on average decide to have fewer children, the effect of a twin birth on family size will be larger.

that twins occur randomly in population (or randomly conditional on observables), that twins do not affect maternal labor supply except through changes in family size, and that there is no one who got twins and did not increase their family size.

The first assumption is likely to hold in the case of Poland as IVF and other ARTs are rare because of legislative and cultural reasons. As has been demonstrated in earlier studies (Bronars and Grogger, 1994), the rate of twining increases with maternal age. However, we control for this directly in the regressions. In table A3 we further show that twinning is uncorrelated with other socioeconomic variables except for age, which partially speaks to the exclusion restriction as it decreases the probability that twins affect maternal labor supply, say, through lower educational attainment of mothers. The last assumption is satisfied by the nature of twinning. It seems therefore that, given that the three assumptions are likely to hold, we should be able to reliably estimate the causal effects of family size on female labor supply for those women who are affected by having multiple birth instead of a singleton, i.e. the local average treatment effect. Estimating LATE makes the direct comparison to the OLS estimates somehow problematic, but, we nonetheless follow the literature in this respect and try to assess the bias by comparing OLS to the IV-LATE estimates.

4 Results

Estimation results presented below are grouped into three sections. In Section 4.1 we show the baseline results estimated for the full and couples samples. Section 4.2 presents estimates of heterogeneity analysis using subsamples split by characteristics which have been established to be uncorrelated with our instruments (see Table A3 in the Appendix), namely mother's education, birth cohort and mother's age at first birth. Following this, we analyse the longer run effects of children on parental outcomes by focusing on samples that are likely to represent women with complete or close-to-complete fertility, which we take to be delineated by the time since the last birth to be higher than six years. While without either retrospective data on past or declarative data on future childbearing a strict completed fertility sample cannot be created, we take our definition to be its close approximation. The purpose of this analysis is, on the one hand, to look at a sample where future fertility considerations no longer affect current labour market situation, and, on the other, to examine if the number of children has longer run consequences on labour market outcomes for parents whose children are already of school age.

4.1 Baseline results

The baseline results are presented in Tables 3 and 4 for the full and the couples' samples, respectively. In the former, we show the effects of the number of children on probability of observing a working mother in the household and her labour income, while in the latter we include also mother's partners' labour market outcomes. Columns (1) and (2) of the tables show results for families with at least one child, while columns (3) and (4) for the sample with at least two children.

Table 3. OLS and 2SLS estimates of labour supply models – all families.

	(1)	(2)	(3)	(4)
	With at least one child		**With at least two children**	
	OLS	2SLS twins-1	OLS	2SLS twins-2
	Dependent variable: mother works			
Number of children	-0.083***	-0.067**	-0.068***	-0.029
	(0.003)	(0.033)	(0.004)	(0.035)
	Dependent variable: mother's labour income			
	-211.571***	-171.997***	-199.913***	-114.523
	(5.591)	(64.269)	(7.098)	(69.747)
N	60,256	60,256	33,010	33,010

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Sample of all families - for selection criteria see Table 1. Columns (1) and (2) – families with at least one child; columns (3) and (4) – families with at least two children. All regressions include following covariates: age of mother at first birth, a polynomial of mother's current age as well as year and region specific effects.
Source: authors' calculations based on BBGD data 2003-2010.

OLS estimates suggest a strong negative relationship between family size and maternal labour market outcomes. Mothers' probability of working is reduced with each child by 8.3 percentage points (pp) in the sample of all families with children, and by 6.8pp in the sample of families with two or more children. These results suggest lower correlations than those found in Rosenzweig and Wolpin (1980b) and Caceres-Delpiano (2006). 2SLS results for maternal employment hold in the sample of mothers with at least one child, however the values of coefficients are lower. Namely, each additional child (second and subsequent children) reduces maternal employment by about 6.7pp.

For families with at least two children, the estimated 2SLS coefficient is still negative but of a much lower magnitude compared to the OLS estimates (-2.9pp) and it is no longer statistically significant. No statistical significance in specification in column (4) results despite the acceptable strength of the twining instrument (see Table 2). All this suggests that family size in Poland reduces employment up to the second child, but the causal effects of the number of children disappear for higher parities. Thus, increasing the number of children from two to three has no significant causal effect on female employment, and the observed lower employment rates of mothers with more than two children are due to the endogenous nature of fertility choices.

Since we assumed that treatment is as good as random, that the twinning results in an increase in family size by a single child and that twining affects maternal labor supply only through increased family size, we can interpret the LATE, in addition to effect for compliers, as the average causal effect on women who are not treated i.e., have only one (two) child(ren). This is because all women who have a multiple first or second birth end up with two or three children, so there are no never takers in response to the twinning instrument. The group of compliers in the LATE is still different from the group estimated by OLS, but we believe that the special interpretation of the twinning instrument, including the effect for non-treated, and the heterogeneity analysis presented in the next section should address the heterogeneity in the treatment effects (Angrist and Pischke, 2009).

OLS estimates presented in Table 3 further indicate a negative relationship between maternal labour income and the number of children in the magnitude of between 200 PLN and 212 PLN per month per child. This negative relationship between the number of children and labour income holds and is statistically significant in the 2SLS regression using the twinning at first birth, and thus, it can be given a causal interpretation. The magnitudes in specifications in columns (2) and (4) are lower compared to the OLS estimates at -172 PLN in the sample with at least one child and at insignificant -115 PLN in the sample with at least two children, but they represent substantial reductions in income given the average incomes of 678 PLN and 604 PLN, and median incomes of 189 PLN and 0 PLN in the two investigated samples, respectively. The strong and statistically significant causal effect of the number of children on labour income suggests also "penalties" on the labour market for some women on the intensive margin.

Table 4. OLS and 2SLS estimates of labour supply models. Couples sample.

	(1)	(2)	(3)	(4)
	With at least one child		**With at least two children**	
	OLS	2SLS twins-1	OLS	2SLS twins-2
	Dependent variable: mother works			
Number of children	-0.082***	-0.063*	-0.065***	-0.029
	(0.003)	(0.037)	(0.005)	(0.036)
	Dependent variable: mother's labour income			
Number of children	-209.529***	-159.273**	-199.513***	-104.910
	(6.017)	(73.401)	(7.423)	(71.731)
	Dependent variable: father works			
Number of children	0.001	0.021	-0.005*	0.019
	(0.002)	(0.016)	(0.003)	(0.015)
	Dependent variable: father's labour income			
Number of children	-94.521***	218.908	-176.360***	-110.286
	(9.750)	(156.316)	(12.574)	(113.669)
N	52,986	52,986	30,573	30,573

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Sample of couples – for sample selection criteria see Table A1 in the Appendix. Columns (1) and (2) – families with at least one child; columns (3) and (4) – families with at least two children. All regressions include following covariates: age of mother at first birth, a polynomial of mother's current age as well as year and region specific effects.
Source: authors' calculations based on BBGD data 2003-2010.

The nature of family size decisions may be different among single mothers and those living in couples, and the investigation of couples enables us to estimate also the family size effects for fathers, or to be precise, for partners of mothers as in the case of couples we do not impose the restriction of the mother's partner to be identified in the data as the child's father.

In Table 4 we re-estimate the specifications from Table 3 for couples (summary of first stage equations are given in panel A in Table A2). Neither the OLS nor the 2SLS estimates for mothers in couples deviate much in magnitude from the results in the full sample of mothers. For paternal labour market outcomes, the OLS results indicate negative correlations between the number of children and labour income. The OLS estimates in the sample of families with at least two children also pick up a correlation between the number of children and father's labour supply on the extensive margin with a small statistically significant negative coefficient (-0.5pp). In the causal estimates, however, the negative effects on the intensive margin are no longer significant. This does not confirm earlier findings of the effect of children on paternal labour market outcomes using fixed effects models (Lundberg and Rose, 2002).[16] Thus, our results provide no significant causal evidence on the effect of number of children on fathers' extensive or intensive margin of labour supply decisions.[17]

4.2 Heterogeneity analysis

The relationship between labour supply and childbearing is likely to differ by women's education (Gronau, 1986), which affects labour market opportunities (Psacharopoulos, 1985; Altonji and Blank, 1999; Card, 1999) and marital matching (Becker, 1973; Becker, 1974; Chiappori et al., 2009), all of which in turn may affect household income, labour market activity and the family size. Furthermore, it seems crucial from the policy point of view to understand if and how the effects of the number of children on labour market outcomes differ in specific population subgroups, in particular in relation to characteristics correlated with income. If there are significant differences between groups then clear identification of those in most need of policy intervention could potentially help in the choice of a particular policy, e.g. between benefit increases and tax reductions for families. It also seems important to understand if the relationships are stable across different cohorts of families, and try to identify any observable trends as well as separate out potential short-run and long-run effects. Additionally, the effects of children, and the degree of endogeneity of the examined relationship, might differ by the age of mother at first birth given the strong effects of children on labour market careers of mothers. Therefore, in this Section, we present the analyses for the full sample of mothers, which is split conditional on:

[16] Angrist and Evans (1998) using the *twin-2* instrument find positive relationship but their coefficients are insignificant.

[17] Out of all the pairs (OLS and 2SLS) of regressions presented in Tables 3 and 4 only the estimates for father's labor income yield statistically significant Hausman-Wu test suggesting endogenity problems.

- mother's education (below high school, high school, above high school);
- mother's cohort (born before 1973, between 1973-1977 and after 1977)
- mother's age at first birth (before the age of 21, between 21-26 and after age of 26).[18]

The results of our heterogeneity analyses are presented in Tables 5, 6 and 7. As we can see in Table 5, the negative correlation between the number of children and mothers' work and income is most pronounced for the higher educated mothers. All OLS estimates suggest a negative relationship between the number of children and the two labour market outcomes. Once we look at the causal estimates, however, the strongest effects are found for the sample with at least one child among the most educated mothers. One child among these mothers reduces maternal employment by as much as 14.3pp and labour income by 299 PLN per month. Both of these are higher in magnitude than the OLS estimates for this sample, but the difference is only statistically significant on the extensive margin. It suggests an unexpected direction of the endogeneity bias, pointing towards the interpretation that in this group of mothers it is those with the highest labour market attachment who decide to have more children, which results in the downward bias of the OLS estimates.

The relationship between the probability of working and family size found in the OLS regression for low and middle educated women confirms the expected direction of the endogeneity. Namely, that the lower employment among those with higher number of children is, at least partially, driven by the fertility choices of women with lowest labour market attachment. All 2SLS estimates for the two lower educated groups are statistically insignificant, which suggests no significant causal effect of children on female employment but since the confidence intervals are wide we cannot rule out potentially large and economically meaningful effects. In particular, in the case of middle educated mothers the magnitude of the causal estimates is an insignificant +1.6pp and it changes from the statistically significant OLS estimate of -8.1pp per additional child. For both samples of mothers with lowest education and for those with at least two children in the middle education group we identify negative causal effects of children on labour income in the range of around 105-129 PLN per month.

[18] In Table A3 in the Appendix we demonstrate the validity of the choice of the three conditioning variables by which we split the sample.

Table 5. Heterogeneity analysis by mother's education. All families.

		(1) "1+ children" OLS	(2) "2+ children" OLS	(3) "1+ children" 2SLS	(4) "2+ children" 2SLS
Below high school		Dependent variable: mother works			
	Number of children	-0.057***	-0.054***	-0.073	0.018
		(0.004)	(0.006)	(0.066)	(0.055)
		Dependent variable: mother's labour income			
	Number of children	-113.613***	-104.013***	-104.937*	-117.846***
		(3.922)	(4.884)	(61.829)	(42.098)
	N	23,201	15,075	23,201	15,075
High school		Dependent variable: mother works			
	Number of children	-0.081***	-0.039***	0.016	-0.042
		(0.005)	(0.009)	(0.049)	(0.061)
		Dependent variable: mother's labour income			
	Number of children	-216.179***	-196.635***	-40.028	-128.578*
		(8.276)	(12.750)	(74.707)	(77.595)
	N	21,903	11,426	21,903	11,426
Above high school		Dependent variable: mother works			
	Number of children	-0.081***	-0.087***	-0.143**	-0.068
		(0.006)	(0.014)	(0.057)	(0.059)
		Dependent variable: mother's labour income			
	Number of children	-203.419***	-211.404***	-298.843*	57.967
		(23.185)	(46.208)	(159.814)	(238.045)
	N	15,152	6,509	15,152	6,509

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). For sample restrictions see Table 1. All regressions include the following covariates: age of mother at first birth, a polynomial of mother's current age as well as year and region specific effects. "1+ children" – families with at least one child; "2+ children" – families with at least two children.
Source: authors' calculations based on BBGD data 2003-2010.

We also confirm a degree of heterogeneity in the relationship between family size and labour market outcomes in the analysis by mothers' birth cohorts (Table 6). We set the cohort thresholds at birth years, which allow the division of the main sample of mothers with at least one child into three subsamples of similar size. This implies thresholds set at birth years before 1973, between 1973 and 1977, and after 1977. For the oldest cohorts and families with one or more children, the OLS coefficient on the number of children suggests a reduction in employment by 6.2pp for each additional child. This effect for the middle and latest cohort is -8.3pp and -4.0pp, respectively. We also find heterogeneity in the estimates for the sample of mothers with two or more children, where the coefficients for the two older groups are between -6.5pp and -6.2pp but for the youngest group it increases to -3.6pp per child.

In the sample of mothers with at least one child we cannot identify any statistically significant causal effects of the number of children on maternal employment for women in the two elder cohorts. For the youngest cohort, however, the causal effect of the number of children is strongly negative (-14.5pp) and statistically significant. This may reflect the important short-run effects of family size on employment, and it once again suggests selection into fertility among women with higher labour market attachment, and thus, a downward OLS bias in this group. For this cohorts, the causal negative effect of additional children on maternal employment is more than three times higher when compared to the OLS estimate, and this difference is statistically significant.

Table 6. Heterogeneity analysis by mothers' cohort. All families.

		(1) "1+ children" OLS	(2) "2+ children" OLS	(3) "1+ children" 2SLS	(4) "2+ children" 2SLS
Mothers born after 1977		Dependent variable: mother works			
	Number of children	-0.040***	-0.036***	-0.145***	-0.018
		(0.006)	(0.011)	(0.048)	(0.058)
		Dependent variable: mother's labour income			
	Number of children	-103.894***	-137.661***	-94.823	93.841
		(9.249)	(14.538)	(91.323)	(176.548)
	N	17,988	6,012	17,988	6,012
Mothers born between 1973 and 1977		Dependent variable: mother works			
	Number of children	-0.083***	-0.062***	-0.037	-0.012
		(0.005)	(0.008)	(0.057)	(0.061)
		Dependent variable: mother's labour income			
	Number of children	-202.781***	-179.555***	-106.463	-134.307
		(8.834)	(11.021)	(112.264)	(106.288)
	N	20,838	12,110	20,838	12,110
Mothers born before 1973		Dependent variable: mother works			
	Number of children	-0.062***	-0.065***	-0.017	-0.068
		(0.004)	(0.006)	(0.063)	(0.060)
		Dependent variable: mother's labour income			
	Number of children	-193.091***	-203.580***	-318.090***	-234.636**
		(8.638)	(10.262)	(123.171)	(112.724)
	N	21,430	14,888	21,430	14,888

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). For sample restrictions see Table 1. All regressions include the following covariates: age of mother at first birth, a polynomial of mother's current age as well as year and region specific effects. "1+ children" – families with at least one child; "2+ children" – families with at least two children.
Source: authors' calculations based on BBGD data 2003-2010.

It is also worth noting here the pattern of the results identified for the oldest cohort. Causal estimates for mothers born before 1973 suggest no effect of children on the probability of work, and large and statistically significant negative causal effects of the number of children on labour incomes. It points to a potentially important medium or long term consequence of children on the intensive margin of the female labour market outcomes, which we investigate further below (Section 4.3.) by looking at a sample of families with the last recorded birth at least six years prior to the survey. This, on the one hand, approximates a selection of families with close to, or complete, fertility histories and focuses the analysis on parents with children beyond pre-school. On the other hand, it also allows us to look at the nature of long-term effects of children on labour market outcomes.

The direction of the OLS bias is again positive once we look at the subsample of mothers who were older than 26 at first birth (Table 7). For this group of mothers while the OLS estimate suggests a negative correlation of 5.8pp per child in the sample with at least one child, the estimated causal effect is -12.1pp. Similarly to the lower educated mothers, the effects of the number of children on employment are insignificant among mothers who had their first child at the age of 26 or less. The correlation in these results is perhaps not surprising as the level of education is likely to correlate with the age at first birth, but the consistency of the positive OLS bias of the estimates, which we find in different subsamples, is noteworthy. Our results suggests that given the significant causal negative effect of another child among these groups of mothers (better educated, younger and who gave birth to their first child earlier), they will be more likely to decide to have another (second) child only if they have a strong attachment to the labour market and/or strong preferences for work.

Table 7. Heterogeneity analysis by mothers' age at first birth. All families.

		(1) **"1+ children" OLS**	(2) **"2+ children" OLS**	(3) **"1+ children" 2SLS**	(4) **"2+ children" 2SLS**
First birth after 26		Dependent variable: mother works			
	Number of children	-0.058***	-0.064***	-0.121**	0.030
		(0.008)	(0.018)	(0.055)	(0.082)
		Dependent variable: mother's labour income			
	Number of children	-229.453***	-373.862***	-230.536*	-187.377
		(22.900)	(47.022)	(140.125)	(245.595)
	N	11,898	4,396	11,898	4,396
First birth between 21 and 26		Dependent variable: mother works			
	Number of children	-0.079***	-0.063***	-0.023	-0.038
		(0.004)	(0.006)	(0.049)	(0.045)
		Dependent variable: mother's labour income			
	Number of children	-235.331***	-222.747***	-186.058**	-45.369
		(6.982)	(9.463)	(76.541)	(96.153)
	N	35,679	20,541	35,679	20,541
First birth before age 21		Dependent variable: mother works			
	Number of children	-0.084***	-0.068***	-0.110	-0.041
		(0.005)	(0.007)	(0.109)	(0.086)
		Dependent variable: mother's labour income			
	Number of children	-175.868***	-155.490***	-37.297	-181.859**
		(7.284)	(8.737)	(174.040)	(87.378)
	N	12,679	8,073	12,679	8,073

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Based on the full sample of families. For sample restrictions see Table 1. All regressions include the following covariates: age of mother at first birth, a polynomial of mother's current age as well as year and region specific effects. "1+ children" – families with at least one child; "2+ children" – families with at least two children.
Source: authors' calculations based on BBGD data 2003-2010.

4.3 Long-term effects of the number of children

Table 8. OLS and 2SLS estimates of labour supply models – all families.
Time since last birth more than 6 years.

	(1) **With at least one child** OLS	(2) 2SLS twins-1	(3) **With at least two children** OLS	(4) 2SLS twins-2
	Dependent variable: mother works			
Number of children	-0.044***	-0.037	-0.047***	-0.041
	(0.005)	(0.042)	(0.008)	(0.049)
	Dependent variable: mother's labour income			
Number of children	-215.077***	-147.158*	-223.343***	-202.877***
	(9.981)	(82.827)	(13.168)	(70.571)
N	24,623	24,623	13,793	13,793

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Families in which the mother is younger than 41 and older than 17 and had the first child at the earliest at the age of 16 and the last birth more than 6 years prior to interview; children's age from 6 to 15 years; Columns (1) and (2) – families with at least one child; columns (3) and (4) – families with at least two children. All regressions include following covariates: age of mother at first birth, a polynomial of mother's current age as well as year and region specific effects.
Source: authors' calculations based on BBGD data 2003-2010.

Results in this section focus on the sample of families in which the time since the birth of the youngest child is more than six years, meaning they naturally focus on a sample of older mothers (mean age of 34.1 and 34.2 in the two investigated samples) and approximate complete fertility histories, as well as examine the situation of mothers in families where all children are already of school age but still in the household. The results, presented in Tables 8 and 9 (first stages are presented in panels B and C in table A2) for the full and the couples' samples respectively, are broadly in line with those for the oldest cohort from Table 6. We still find negative correlations between female labour market outcomes in the OLS regressions. The causal nature of these effects holds, however, only for maternal labour incomes in the 2SLS estimates with the exception of the estimate for the couples' sample with at least one child. The estimates suggest that mothers' labour incomes are reduced by 147 PLN and 203.00 PLN per month for each child in the samples with at least one and at least two children, respectively. Like in the results in Table 4, the causal effect of children on paternal incomes in the sample with at least one child is positive but not statistically significant. The 2SLS estimates in the case of the sample with at least two children suggest a negative effect of children on the income of fathers/partners in the range of 256 PLN per month. This suggests that among larger families in the longer-run not only mother's but also father's income is reduced as a result of a higher number of children.

Table 9. OLS and 2SLS estimates of labour supply models. Couples sample. Time since last birth more than 6 years.

	(1)	(2)	(3)	(4)
	At least one child		**At least two children**	
	OLS	2SLS twins-1	OLS	2SLS twins-2
	Dependent variable: mother works			
Number of children	-0.043***	-0.035	-0.046***	-0.039
	(0.005)	(0.047)	(0.009)	(0.050)
	Dependent variable: mother's labour income			
	-213.474***	-122.208	-231.512***	-186.434**
	(10.850)	(97.961)	(13.411)	(73.330)
	Dependent variable: father works			
	0.003	0.033	0.001	-0.015
	(0.003)	(0.023)	(0.005)	(0.028)
	Dependent variable: father's labour income			
	-157.941***	191.659	-234.933***	-256.019*
	(17.157)	(192.814)	(23.444)	(143.979)
N	21,109	21,109	12,572	12,572

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). Sample of couples; families in which the mother is younger than 41 and older than 17 and had the first child at the earliest at the age of 16 and the last birth more than 6 years prior to interview; children's age from 6 to 15 years; Columns (1) and (2) – families with at least one child; columns (3) and (4) – families with at least two children. All regressions include following covariates: age of mother at first birth, a polynomial of mother's current age as well as year and region specific effects.
Source: authors' calculations based on BBGD data 2003-2010.

5 Conclusions

The combination of high levels of female employment and fertility is crucial from the point of view of sustained economic growth and future financial stability of welfare systems. Yet, if family size strongly limits mothers' labour market activity achieving these two objectives may prove difficult. The analysis in this paper focuses on the identification of causal estimates of the effects of family size on labour market outcomes using data from the Polish Household Budget Surveys for years 2003-2010. We applied 2SLS estimations using twining as the source of exogenous variation in the family size. To our knowledge this is the first set of causal estimates for a regime from a developed country where both fertility and female employment are low and for any of the countries of Central and Eastern Europe.

Overall results using the twinning instrument are consistent with the literature (Rosenzweig and Wolpin, 1980a; Vere, 2011) and confirm the negative effect of an additional child on female employment of about 6.7pp in the sample of mothers with at least one child. This is, however, only slightly less negative compared to the OLS estimates of about -8.3pp. These causal effects apply only up to the parity of two. While OLS estimates for families with at least two children are still negative and statistically significant (-6.8pp) we could not identify any causal effect of the number of children on female employment for families with two or more children. Thus, lower employment among mothers with more than two children seems to be a result of fertility choices among mothers with lower labour market attachment. Relative to other findings in the literature, our twinning results are generally larger for families with more than one child. Furthermore, these results seem to be similar irrespectively whether we use OLS or IV, whereas in the US studies the OLS were severely downward biased.

In most cases, OLS estimates exaggerate the negative effects of children on maternal labour supply on the extensive and the intensive margin but once we differentiate the analysis by maternal education, cohort and age at first birth we demonstrate that for some groups the effect of endogeneity may actually be reversed. Thus, the OLS may in some cases underestimate the negative causal effects of children. It is the case for mothers with higher education, those from the cohort born after 1977, and those who had their first child aged more than 26. In all of these cases we find the negative causal effect of an additional child to be in the range of negative 14.5pp to negative 12.1pp, compared to the OLS estimates of -8.1pp, -4.0 and -5.8pp, respectively. To our knowledge such an effect has not been found in the earlier studies, and it points towards the hypothesis that in these groups it is the stable employment and good career outlooks that determine choices concerning a higher number of children. Therefore, it is women with greater labour market attachment that decide to have a higher number of children. At the same time, for mothers with less than higher education, for those from earlier cohorts, and for those who had their first child aged up to 26 we find no evidence of the causal effect of children on employment. These estimates are generally lower when compared to the OLS results and statistically insignificant. Therefore, our heterogeneity analysis suggests that for some groups of women good labour market prospects may be key determinants of their fertility decisions.

The Polish policy context may give some clues to the identified effects. Poland is distinguished by low formal childcare enrollment rate for pre-school children (43% compared to EU average of 84% in 2011) and low financial support for families through child-related benefits (0.7% of the GDP compared to EU average of 2.3% in 2009). At the same time, childcare support from relatives and informal nannies is widespread and early retirement and easy access to disability pensions facilitates this form of childcare provision among women in retirement and pre-retirement age. Our heterogeneity analysis is consistent with this institutional framework. Given extremely low levels of financial support for low-income families (approximated here with low level of maternal education), mothers in these families simply cannot afford to stay out of work. These mothers, are also more likely to live close to their relatives and thus have better access to informal childcare. Such scenarios may be more likely in particular among mothers who had their first baby at an early age. All these factors provide explanation for the insignificant effects of family size on employment among these subgroups of mothers. On the other hand, mothers with higher education (and those who had their child aged 27 or older, which is strongly correlated with the education level), will be more likely to rely on formal childcare and live away from their relatives. Childcare constraints may be much more significant for these mothers, which is reflected in the high estimates of the causal effect of family size on employment.

The fact that it is women with strong work preferences or labour market attachment who chose to have more children in the presence of such constraints has important implications for policy in the current economic situation. If this were the case then poor employment outlook in many countries affected by the recent economic crisis, and already characterised by low fertility, may further aggravate their demographic situation.

In many cases where we find a negative causal effect of family size on employment of mothers we also confirm the negative influence of the number of children on female labour incomes. Such negative effects on incomes are additionally found for mothers with low and medium education, for those in the oldest cohort and for those who had first child earlier in their life, where we could not identify any causal effect on employment. Furthermore, we could find very little evidence on the negative effect of the number of children on fathers' labour outcomes. The only exception is the sample of families in which we approximate full fertility history by limiting the sample to mothers whose youngest child was born at least six years before the survey. For this sample using twinning instruments we identify negative effects of children on the intensive margin of labour supply in the case of mothers with at least one and at least two children, and for fathers with at least two children.

The findings suggest several important policy conclusions and new directions for further research. From the analysis it is clear that mothers, but not fathers, suffer the negative labour market consequences of childbearing in Poland. These effects are particularly strong for well-educated women, for women from younger cohorts, and those who had their first child later in life, and they apply principally up to parity two. While mothers with more than two children are less likely to work, it is due to the fertility choices of women with weaker labour market attachment rather than the causal effect of the higher number of children. In many subsamples of women, however, we find negative consequences of children in terms of lower labour incomes. These effects also extend beyond the time of early childhood.

References

Adsera, A., (2005), "Vanishing children: From high unemployment to low fertility in developed countries", *American Economic Review*, 95:2, 189-193.

Aguero, J., and M. Marks (2011), "Motherhood and female labor supply in the developing world: evidence from infertility shocks", *Journal of Human Resources*, 46:4, 800-826.

Ahn, N., and P. Mira (2002), "A note on the changing relationship between fertility and female employment rates in developed countries", *Journal of Population Economics*, 15:4, 667-682.

Altonji, G.J., and R.M. Blank (1999), "Race and gender in the labor market", *Handbook of labor economics*, 3:3, 3143-3259.

Angrist, D.J., and W.N. Evans (1998), "Children and their parent's labor supply: evidence from exogenous variation in family size", *American Economic Review*, 88:3, 450-477.

Angrist, D.J., and A.B. Krueger (2001), "Instrumental variables and the search for identification: From supply and demand to natural experiments", *Journal of Economic Perspectives*, 15:4, 69-85.

Angrist, D.J., V. Lavy, and A. Schlosser (2006) "New evidence on the causal link between the quantity and quality of children", IZA Discussion Paper 2075.

Angrist, D.J., and S.J. Pischke (2009), "Mostly harmless econometrics", *Princeton University Press.*

Bardasi, E., and C. Monfardini (2009), "Women's employment, children and transition. An empirical analysis for Poland", *Economics of Transition*, 17:1, 147-173.

Barlik, M., and K. Siwiak (2011), "Metodologia badania budżetów gospodarstw domowych", Zeszyty Metodologiczne GUS, Zakład Wydawnictw Statystycznych, Warszawa.

Becker, S.G., (1973), "A theory of marriage: part I", *Journal of Political Economy*, 81:4, 813-846.

Becker, S.G., (1974), "A theory of marriage: part II", *Journal of Political Economy*, 82:2, S11-S26.

Blau, D.F., and L.M. Kahn (2005), "Changes in the labor supply behavior of married women: 1980-2000", NBER Working Paper 11230.

Blundell, R., A.S. Duncan, and C. Meghir, (1998), "Estimating labor supply responses using tax reforms", *Econometrica*, 66:4, 827-862.

Blundell, R., and T. Macurdy (1999), "Labor supply: A review of alternative approaches", *Handbook of Labor Economics*, 3:1, 1559-1695.

Blundell, R., P. Chiappori and C. Meghir (2005), “Collective labor supply with children”, *Journal of Political Economy*, 113:6, 1277-1306.
Blundell, R., M. Brewer and M. Francesconi (2008), “Job changes and hours changes: Understanding the path of labor supply adjustment”, *Journal of Labor Economics*, 26:3, 421-453.
Bonin, H., and R. Euwals (2002), “Participation behavior of East German women after German unification”, IZA Discussion Papers 413.
Brewster, L.K., and R.R. Rindfuss (2000), “Fertility and women’s employment in industrialized nations”, *Annual Review of Sociology*, 26, 271-296.
Bronars, G.S., and J. Grogger (1994), “The economic consequences of unwed motherhood: using twin births as a natural experiment”, *American Economic Review*, 84:5, 1141-1156.
Bulmer, M.G, (1970), “The biology of twinning in man”, Oxford University Press
Caceres-Delpiano, J., (2006), “The impacts of family size on investment in child quality”, *Journal of Human Resources*, 41:4, 738-754.
Card, D., (1999), “The causal effect of education on earnings”, *Handbook of Labor Economics*, 3:1, 1801-1863.
Chiappori, P-A., M. Iyigun, and Y. Weiss (2009) “Investment in schooling and the marriage market”, *American Economic Review*, 99:5, 1689-1713.
Chun, H., and J. Oh (2002) “An instrumental variable estimate of the effect of fertility on the labour force participation of married women”, *Applied Economics Letters*, 9:10, 631-634.
Cruces, G., and S. Galiani (2007) “Fertility and female labor supply in Latin America: new causal evidence”, *Labour Economics*, 14:3, 565-573.
Daouli, J., M. Demoussis and N. Giannakopoulos (2009), “Sibling-sex composition and its effects on fertility and labor supply of Greek mothers”, *Economics Letters*, 102:3, 189-191.
Fong, M., and M. Lokshin (2000), “Child care and women’s labor force participation in Romania”, World Bank Policy Research Working Papers 2400.
Galor, O., and D.N. Weil (1996), “The gender gap, fertility, and growth”, *American Economic Review*, 86:3, 374-387.
Gronau, R., (1973), “The effect of children on the housewife’s value of time”, *Journal of Political Economy*, 81:2, S168-S199.
Gronau, R., (1986), “Home production – a survey”, *Handbook of labor economics*, 1:1, 273-304.
Heckman, J.J., (1993), “What has been learned about labor supply in the past twenty years?”, *American Economic Review*, 83:2, 116-121.
Hirvonen, L., (2009), “The effect of children on earnings using exogenous variation in family size: Swedish evidence”, SOFI Working Paper 2/2009.
Imbens, G., and J.D. Angrist (1994), “Identification and estimation of local average treatment effects”, *Econometrica* 62:2, 467-475.
Imbens, G., (2009), “Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009)”, Mimeo Harvard University.
Jacobsen, P.J., J. Wishart Pearce III and L.J Rosenbloom (1999), “The effects of childbearing on married women’s labor supply and earnings: using twin births as a natural experiment”, *Journal of Human Resources*, 34:3, 449-474.
Joshi, H., (1998), “The opportunity costs of childbearing: more than mothers’ business”, *Journal of Population Economics*, 11:2, 161-183.
Karbownik, K., and M. Myck (2012), “For some mothers more than others: how children matter for labour market outcomes when both fertility and female employment are low”, IZA Discussion Papers 6933.

Killingsworth, R.M., and J.J. Heckman (1986), "Female labor supply: A survey", *Handbook of Labor Economics*, 1:1, 103-204.
Klasen, S., and F. Lamanna (2009), "The impact of education inequality in education and employment on economic growth: New evidence from a panel of countries", *Feminist Economics*, 15:3, 91-132.
Lagerlof, P.N., (2003), "Gender equality and long-run growth", *Journal of Economic Growth*, 8(4), 403-426.
Leibowitz, A., J.A. Klerman and L.J. Waite (1992), "Employment of new mothers and child care choice: Differences by children's age", *Journal of Human Resources*, 27:1, 112-133.
Lichtenstein, P., P.O. Olausson and A.J. Kallen (1996), "Twin births to mothers who are twins: a registry based study", *British Medical Journal*, 312:70350, 879-881.
Lundberg, S., and E. Rose (2002), "Effects of sons and daughters on men's labor supply and wages", *Review of Economics and Statistics,* 84:2, 251-268.
Matysiak, A., (2009), "Employment first, then childbearing: women's strategy in post socialist Poland", *Population Studies*, 63:3, 253-276.
Mittler, J.P., (1971), "The study of twins", Harmondsworth Middlesex: Penguin.
Rosenzweig, R.M., and K.I. Wolpin (1980a), "Testing the quantity-quality fertility model: the use of twins as a natural experiment", *Econometrica*, 48:1, 227-240.
Rosenzweig, R.M., and K.I. Wolpin (1980b), "Life-cycle labor supply and fertility: causal inferences from household models", *Journal of Political Economy*, 88:2, 328-348.
Psacharopouslo, G., (1985), "Returns to education: a further international update and implications", *Journal of Human Resources*, 20:4, 583-604.
Saget, C., (1999), "The determinants of female labour supply in Hungary", *Economics of Transition*, 7:3, 575-591.
Schultz, P.T., (1990), "Testing the neoclassical model of family labor supply and fertility", *Journal of Human Resources*, 25:4, 599-634.
Vere, P.J., (2011), "Fertility and parents' labour supply: new evidence from US census data", *Oxford Economic Papers*, 63:2, 211-231.
Waterhouse, A.H.J., (1950), "Twinning in twin pedigrees", *British Journal of Social Medicine*, 4:4, 197-216.
Westergaard, T., J. Wohlfahrt, P. Aaby and M. Melbye (1997), "Population based study of rates of multiple pregnancies in Denmark, 1980-94", *British Medical Journal*, 314:7083, 775-779.

Appendix

Tables

Table A1. Descriptive statistics – couples sample.

	With at least one child		With at least two children	
	Mean	Standard deviation	Mean	Standard deviation
Number of children	1.780	(0.851)	2.352	(0.695)
- one child	0.423	(0.494)	-	-
- two children	0.425	(0.494)	0.736	(0.441)
- three or more children	0.153	(0.360)	0.264	(0.441)
Twins at first birth (twins-1)	0.011	(0.103)	-	-
Twins at second birth (twins-2)	-	-	0.010	(0.100)
Same sex of first two born children	-	-	0.508	(0.500)
Two first born girls	-	-	0.235	(0.424)
Two first born boys	-	-	0.273	(0.445)
Age of mother	31.517	(4.724)	32.797	(4.094)
Age of mother at first birth	23.682	(3.638)	22.982	(3.253)
Mother's education:*				
- basic	0.377	(0.485)	0.449	(0.497)
- secondary	0.363	(0.481)	0.348	(0.476)
- higher	0.260	(0.438)	0.203	(0.402)
Mother works:**	0.547	(0.498)	0.542	(0.498)
- one child	0.552	(0.497)	-	-
- two children	0.552	(0.497)	0.552	(0.497)
- three or more children	0.516	(0.500)	0.516	(0.500)
Mother's labour income:	681.82	(986.59)	605.51	(923.88)
- one child	785.91	(1057.38)	-	-
- two children	687.75	(966.10)	687.75	(966.10)
- three or more children	376.53	(748.49)	376.53	(748.49)
Father works:	0.806	(0.396)	0.812	(0.391)
- one child	0.797	(0.402)	-	-
- two children	0.810	(0.393)	0.810	(0.393)
- three or more children	0.818	(0.386)	0.818	(0.386)
Father's labour income:	1574.386	(1579.555)	1528.989	1541.926
- one child	1636.31	(1627.48)	-	-
- two children	1630.68	(1591.68)	1630.68	(1591.68)
- three or more children	1245.86	(1354.51)	1245.86	(1354.51)
Number of families	33732		19259	
N	52986		30573	

Notes: The samples include families in which the mother is younger than 41 and older than 17 and had the first child at the earliest at the age of 16; children's age from 0-15; labour incomes are unconditional monthly net values indexed by CPI to June 2006.
* Education categories cover: "basic" – no formal education, primary education, gymnasium and vocational education; "secondary" – secondary academic and secondary vocational education; "higher education" – education degree higher than secondary;
** The sample sizes are 22413 for mothers with one child; 22494 for mothers with two children and 8079 for mothers with three or more children. Same sample sizes apply to mother's labour income and paternal labour market supply variables.
Source: authors' own calculations based on the PHBS data (2003-2010).

Table A2. OLS first stage relationships and the strength of the instruments.

Dependent variable: number of children	(1) **twins at 1st birth**	(2) **twins at 2nd birth**
Panel A: Couples sample		
t-statistic on the instrument	19.86	22.10
Partial R-squared	0.008	0.016
F-statistic on excluded instruments	395	489
LM statistic on underidentification test	299	225
N	52986	30573
Panel B: All families, time since last birth more than 6 years		
t-statistic on the instrument	19.96	15.74
Partial R-squared	0.015	0.031
F-statistic on excluded instruments	359	248
LM statistic on underidentification test	207	105
N	24623	13793
Panel C: Couples sample, time since last birth more than 6 years		
t-statistic on the instrument	16.08	15.16
Partial R-squared	0.013	0.033
F-statistic on excluded instruments	258	230
LM statistic on underidentification test	169	99
N	21109	12572

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1). All regressions include year and region specific effects. The additional covariates include age of mother at first birth, and a polynomial of mother's current age. Sample of mothers aged <18; 40> with oldest child younger than 16 years old who gave the first birth at the age of 16 at the earliest and whose last birth was 6 months prior to survey at the earliest.

Source: authors' calculations based on BBGD data 2003-2010.

Table A3. Validity of heterogeneity analyses.

	(1) **With covariates** Twins-1	**(2)** Twins-2	**(3)** **Raw correlations** Twins-1	**(4)** Twins-2
Panel A: Maternal education, full sample				
Secondary school	-0.001	0.001	-0.001	0.001
	(0.001)	(0.002)	(0.001)	(0.002)
	0.338	0.596	0.329	0.672
Below secondary school	-0.001	0.001	-0.001	0.001
	(0.001)	(0.002)	(0.001)	(0.002)
	0.392	0.576	0.389	0.661
R-squared	0.001	0.001	0.000	0.000
N	60,256	33,010	60,256	33,010
Panel B: Maternal cohorts (no year fixed effects), full sample				
Middle age group	0.001	0.001	0.002	0.001
	(0.001)	(0.002)	(0.001)	(0.002)
	0.258	0.742	0.243	0.745
Youngest	-0.002	0.001	-0.002	0.001
	(0.001)	(0.002)	(0.001)	(0.002)
	0.177	0.703	0.189	0.699
R-squared	0.001	0.001	0.000	0.000
N	60,256	33,010	60,256	33,010
Panel C: Maternal age at first birth, full sample				
Birth between 21 and 26	0.002	-0.002	0.002	-0.002
	(0.001)	(0.002)	(0.001)	(0.002)
	0.116	0.164	0.131	0.190
Birth after 26	0.010***	-0.001	0.010***	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
	0.000	0.576	0.000	0.675
R-squared	0.002	0.001	0.001	0.000
N	60,256	33,010	60,256	33,010

Notes: Covariates in columns (1)-(2) include: year and region fixed effects. Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1).

Source: authors' calculations based on BBGD data 2003-2010.

Table A4. OLS estimates of first stage relationships by maternal education - all families.

	(1)	(2)	(3)
	Dependent variable: number of children		
Instruments for fertility	**Below high school**	**High school**	**Above high school**
	Twins at 1st birth		
Twins at 1st birth	0.528***	0.701***	0.660***
	(0.058)	(0.045)	(0.026)
	[9.07]	[15.50]	[25.46]
Partial R-squared	0.004	0.012	0.016
F-statistics on excluded instruments	82	240	648
LM statistic on underidentification test	83	145	153
N	23,201	21,903	15,152
	Twins at 2nd birth		
Twins at 2n birth	0.815***	0.832***	0.887***
	(0.069)	(0.041)	(0.038)
	[11.81]	[20.13]	[23.18]
Partial R-squared	0.011	0.022	0.037
F-statistics on excluded instruments	140	405	537
LM statistic on underidentification test	89	100	58
N	15,075	11,426	6,509

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1), t-statistics on the coefficients in square brackets. All regressions include year and region fixed effects. The additional covariates include age of mother at first birth, and a polynomial of mother's current age. Sample of mothers aged <18; 40> with oldest child younger than 16 years, who gave the first birth at the age of 16 at the earliest and whose last birth was 6 months prior to survey at the latest.
Source: authors' calculations based on BBGD data 2003-2010.

Table A5. OLS estimates of first stage relationships by maternal cohorts - all families.

	(1)	(2)	(3)
	Dependent variable: number of children		
Instruments for fertility	**Born after 1977**	**Born 1973 – 1977**	**Born before 1973**
	Twins at 1st birth		
Twins at 1st birth	0.813***	0.619***	0.579***
	(0.044)	(0.042)	(0.056)
	[18.65]	[14.68]	[10.30]
Partial R-squared	0.016	0.008	0.005
F-statistics on excluded instruments	348	216	106
LM statistic on underidentification test	127	153	95
N	17,988	20,838	21,430
	Twins at 2nd birth		
Twins at 2n birth	0.874***	0.811***	0.810***
	(0.041)	(0.047)	(0.073)
	[21.10]	[17.24]	[11.03]
Partial R-squared	0.030	0.015	0.011
F-statistics on excluded instruments	445	297	122
LM statistic on underidentification test	56	99	82
N	6,012	12,110	14,888

Notes: Standard errors clustered at household level (*** p<0.01, ** p<0.05, * p<0.1), t-statistics on the coefficients in square brackets. All regressions include year and region fixed effects. The additional covariates include age of mother at first birth, and a polynomial of mother's current age. Sample of mothers aged <18; 40> with oldest child younger than 16 years, who gave the first birth at the age of 16 at the earliest and whose last birth was 6 months prior to survey at the latest.
Source: authors' calculations based on BBGD data 2003-2010.

Table A6. OLS estimates of first stage relationships by maternal age at first birth - all families.

	(1)	(2)	(3)
	Dependent variable: number of children		
Instruments for fertility	**After 26**	**21 – 26**	**Before 21**
	Twins at 1st birth		
Twins at 1st birth	0.688***	0.587***	0.511***
	(0.034)	(0.044)	(0.084)
	[20.52]	[13.37]	[6.09]
Partial R-squared	0.023	0.006	0.003
F-statistics on excluded instruments	421	179	37
LM statistic on underidentification test	162	154	37
N	11,898	35,679	12,679
	Twins at 2nd birth		
Twins at 2n birth	0.913***	0.856***	0.693***
	(0.051)	(0.051)	(0.073)
	[17.76]	[16.76]	[9.52]
Partial R-squared	0.042	0.016	0.008
F-statistics on excluded instruments	316	281	91
LM statistic on underidentification test	43	139	49
N	4,396	20,541	8,073

Notes: Standard errors clustered at high school level (*** p<0.01, ** p<0.05, * p<0.1), t-statistics on the coefficients in square brackets. All regressions include year and region fixed effects. The additional covariates include age of mother at first birth, and a polynomial of mother's current age. Sample of mothers aged <18; 40> with oldest child younger than 16 years, who gave the first birth at the age of 16 at the earliest and whose last birth was 6 months prior to survey at the latest.
Source: authors' calculations based on BBGD data 2003-2010.

Polish Household Budget Survey – summary of the methodology

The Polish Household Budget Survey is a representative survey of Polish households surveying over 37000 households per year. The survey is conducted every year and is spread over the entire calendar year with each household surveyed over a period of a month during which households record their expenditures and incomes. This information is complemented with an additional interview which is conducted at the end of each quarter of data collection (so called quarterly interview). Each year since 2005, when the most recent sampling procedure was introduced, the target sample is 37584 households.

In a case of refusal to participate among households from the principal gross sample, households are replaced by another household from a reserve list of randomly chosen households. This reserve list is prepared separately for each sampling unit. Households which drop out of the survey in the first half of their survey month are also replaced by households from the reserve list. Those who drop out in the second half of the month are not replaced. Households from the principal gross sample which agree to participate are re-interviewed in the same month of the following year. Households from the reserve list are not re-interviewed. The survey methodology has been developed in accordance with the EUROSTAT guidelines.

The overall response rate in the survey in 2010 was 50.2%. Survey non-response was either due to refusal to participate (48.1%), survey drop out during its duration (1.6%) or refusal to complete the final quarterly interview

(0.1%). From among households which were approached to complete the survey for the first time in 2010 (either from the principal gross sample or from the reserve list) 59.5% did not participate in the survey, and from among those who participated in the previous year 14.9% did not complete the survey for the second time.