Economics of education: Policies and effects

Anders Björklund and Peter Fredriksson

Long-term effects of early childhood care and education
Christopher Ruhm and Jane Waldfogel

Recruiting, retaining, and creating quality teachers
C. Kirabo Jackson

On the margin of success? Effects of expanding higher education for marginal students
Björn Öckert

Gender differences in education
Tuomas Pekkarinen

Educating children of immigrants: Closing the gap in Norwegian schools
Bernt Bratsberg, Oddbjørn Raun and Knut Reed

The effects of education on health and mortality
Bhashkar Mazumder

Economics of education: Policies and effects

Anders Björklund and Peter Fredriksson

Long-term effects of early childhood care and education
Christopher Ruhm and Jane Waldfogel

Recruiting, retaining, and creating quality teachers
C. Kirabo Jackson

On the margin of success? Effects of expanding higher education for marginal students
Björn Öckert

Gender differences in education
Tuomas Pekkarinen

Educating children of immigrants: Closing the gap in Norwegian schools
Bernt Bratsberg, Oddbjørn Raun and Knut Reed

The effects of education on health and mortality
Bhashkar Mazumder
The *Nordic Economic Policy Review* is published by the Nordic Council of Ministers and addresses policy issues in a way that is useful for informed non-specialists as well as for professional economists. All articles are commissioned from leading professional economists and are subject to peer review prior to publication. The review appears twice a year. The journal is distributed free of charge to the members of the Nordic economic associations. The easiest way to subscribe to the NEPR is therefore to become a member of one of these associations, i.e.,

- Denmark: Nationaløkonomisk Forening
- Finland: Taloustieteellinen Yhdistys
- Norway: Samfunnsøkonomene
- Sweden: Nationalekonomiska Föreningen.

For institutional subscriptions, please contact nepr@ies.su.se.
This publication has been published with financial support by the Nordic Council of Ministers. However, the contents of this publication do not necessarily reflect the views, policies or recommendations of the Nordic Council of Ministers.

www.norden.org/en/publications

Nordic co-operation

Nordic co-operation is one of the world's most extensive forms of regional collaboration, involving Denmark, Finland, Iceland, Norway, Sweden, and the Faroe Islands, Greenland, and Åland.

Nordic co-operation has firm traditions in politics, the economy, and culture. It plays an important role in European and international collaboration, and aims at creating a strong Nordic community in a strong Europe.

Nordic co-operation seeks to safeguard Nordic and regional interests and principles in the global community. Common Nordic values help the region solidify its position as one of the world's most innovative and competitive.

Nordic Council of Ministers
Ved Stranden 18
DK-1061 Copenhagen K
Phone (+45) 3396 0200

www.norden.org
Content

Economics of education: Policies and effects
   Anders Björklund and Peter Fredriksson .................. 7

Long-term effects of early childhood care and education
   Christopher Ruhm and Jane Waldfogel .................... 23
   Comment by Tarjei Havnes .................................. 53

Recruiting, retaining, and creating quality teachers
   C. Kirabo Jackson ........................................... 61
   Comment by Torberg Falch .................................. 105

On the margin of success? Effects of expanding higher education for marginal students
   Björn Öckert .................................................. 111
   Comment by Torbjørn Hægeland .............................. 159

Gender differences in education
   Tuomas Pekkarinen ......................................... 165
   Comment by Anna Sjögren .................................. 197

Educating children of immigrants:
Closing the gap in Norwegian schools
   Bernt Bratsberg, Oddbjørn Raaum and Knut Røed ........ 211
   Comment by Lena Nekby .................................... 253

The effects of education on health and mortality
   Bhashkar Mazumder ......................................... 261
   Comment by Kjell G. Salvanes .............................. 303
Economics of education: Policies and effects

Anders Björklund* and Peter Fredriksson**

Education is important. It influences the growth prospects of a country (see Krueger and Lindahl, 2001) and causes differences in individual labor market success. The individual wage return to a year of schooling varies somewhat over countries and time periods but is typically estimated to be between 4 and 10 percent. Analogously, the effect of education and growth varies somewhat over studies but Krueger and Lindahl (2001) argue that it is of the same order of magnitude as the individual wage return. Moreover, some researchers argue that these wage returns underestimate the full return to schooling, since more educated individuals (it is argued) have more enjoyable lives along a variety of dimensions (see Oreopoulos and Salvanes, 2011). Thus, education has effects on efficiency as well as equity. Therefore, it is natural that school performance attracts a great deal of attention in the public discussion.

In our own home country, the reduction in the international test performance since the early or mid 1990’s has been widely discussed. Table 1 illustrates the trend decline.

These data suggest substantial declines in student achievement from the mid 1990’s. The decline at the upper-secondary level corresponds to 0.8-0.9 of a standard deviation and the decline in grade 8 mathematics to 0.4-0.5 of a standard deviation.

---

* Swedish Institute for Social Research, Stockholm University, anders.bjorklund@sofi.su.se.
** Department of Economics, Stockholm University, peter.fredriksson@ne.su.se.
Table 1. Results for Sweden in PIRLS and TIMSS 1991-2008

<table>
<thead>
<tr>
<th>Year</th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 8</th>
<th>Upper secondary</th>
<th>Grade 8</th>
<th>Upper secondary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991</td>
<td>513</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td></td>
<td>540</td>
<td>502</td>
<td>553</td>
<td>578</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>498</td>
<td>561</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td></td>
<td>499</td>
<td></td>
<td>524</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
<td>549</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td>491</td>
<td></td>
<td>511</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td>412</td>
<td></td>
<td>497</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change</td>
<td>-27</td>
<td>-49</td>
<td>-90</td>
<td>-42</td>
<td>-81</td>
<td></td>
</tr>
</tbody>
</table>

Source: Trend estimates in PIRLS (Progress in International Reading Literacy Study) and TIMSS (Trends in International Mathematics and Science Study)

Notes: The results are normalized such that the international average is 500 and the standard deviation among all participating students is 100. The row labeled “Change” shows the change between the latest observation and the earliest observation shown in the table. For reading comprehension, we have added the change between 2006 and 2001 in grade 4 to the change between 2001 and 1991 in grade 3.

The trend measures reported in Table 1 are only consistently available for Sweden. There is a trend estimate for Norway for science at the upper secondary level, however. In Norway, there was a decline of almost 0.5 of a standard deviation between 1995 and 2008.

Absent trend measures for the other Nordic countries, one can have a look at PISA, OECD’s study of 15-year-old students (see OECD, 2010). Results from PISA are available every third year between 2000 and 2009. Now, PISA is not designed to measure trends in achievement and the domains tested are slightly different than in PIRLS and TIMSS. Nevertheless, qualitatively, the trend changes for Sweden are consistent with Table 1. Across all subjects, there are declines in performance between 2000 and 2009 which amount to 0.16-0.19 of a standard deviation.

For the other Nordic countries, there are no declines which are equally systematic (and substantial) across subjects as in Sweden. In terms of the levels of performance in 2009, Finland is the shining star scoring among the top countries in all subjects. To find Denmark, Iceland, Norway, and Sweden, one has to move down the ranking; these four countries all score

---

1 For instance, in TIMSS, pure mathematics skills are tested to a greater extent than in the corresponding mathematics test in PISA, which requires reading comprehension in order to solve the problem.
around 500, implying that they are average performers in the PISA study (the overall international average in PISA is 500).

To summarize, education is important along a variety of dimensions. The recent experience in school achievement has been somewhat different across the Nordic countries and it seems that Swedish policy makers should be particularly concerned.

This issue of the *Nordic Economic Policy Review* (NEPR) deals with pertinent questions in the Economics of Education. We have chosen the questions with an eye towards what are policy-relevant questions and what is available elsewhere. On the last point, the predecessor of NEPR – the *Swedish Economic Policy Review* – published an issue on the topic: What can educational policies achieve? That issue included review articles on school competition and school choice (see Hoxby, 2003) as well as resources (see Gustafsson, 2003). Therefore, we have decided to leave these policy tools aside for this NEPR issue.

We have chosen to focus on three direct policy questions: What are the long-run effects of early childhood interventions? (see Ruhm and Waldfogel); How do you recruit, retain, and create quality teachers? (see Jackson); What is the return to increasing access to higher education? (see Öckert).

In addition, this issue includes three papers on education of general interest: What are the labor market consequences of increased female participation in education? (see Pekkarinen); How do immigrants fare in education? (see Bratsberg et al.); Does education improve health and increase longevity? (see Mazumder). We now motivate these six questions in more detail.

### 1. What are the long-run effects of early childhood interventions?

The recent literature in economics has seen a great deal of interest directed towards early intervention (Cunha and Heckman, 2007 is one example). The argument in favor of early intervention is the idea that skills acquired at an earlier stage influence the productivity of subsequent skill investments. Therefore, early interventions are more effective than later interventions since early investments influence the entire path of human
capital accumulation. In fact, targeting children in poor families early may improve both equity and efficiency.

The Nordic countries have in common that (high-quality) pre-school care is available for most families; they also have in common that parental leave is relatively extended and generous.

Publicly provided pre-school care was built up in Sweden during the 1970’s (see Figure 1). The 1970’s saw a similar expansion in Norway (see Havnes and Mogstad, 2011). By 2009 the participation rates among four-year olds exceed 90 percent in all Nordic countries except Finland (see Eurostat database).

Figure 1. Percent of 1-5 year-olds participating in pre-school activities, Sweden 1950-2009

Christopher Ruhm and Jane Waldfogel survey recent research on these topics. They distinguish between studies of parental leave policies and early childhood education. When it comes to parental leave policies, their overall conclusion is that there is only limited evidence that expansions of leave durations in Europe have led to long-run improvements in educational and labor market outcomes of the children whose parents are affected by them. The most notable exception is a study of a Norwegian expansion of paid leave from zero to four months, covering precisely the very young child ages where the gains may be particularly large.
In contrast, they find that expansions of early childhood education are found to yield benefits at school entry, in adolescence, and during adulthood. Furthermore, these gains are largest for those who are disadvantaged, e.g., those who come from low-income or immigrant families. The exceptions to these patterns in the results are found in studies of the US and Canada.

Ruhm and Waldfogel acknowledge the difficulties in estimating the causal effects of such programs, but also note that several recent studies for Nordic countries are more compelling than many previous studies. The fact that Nordic researchers have had access to population register-based data sets has helped improve research in the field. Ruhm and Waldfogel recommend more research of a similar type but with a focus on broader outcomes such as health, family functioning and criminality.

2. What measures improve the quality of teaching?

The discussion about the effects of different education policy measures is typically intense, among researchers as well as policy makers. Sometimes there is profound disagreement among researchers; see the interchange between Hanushek (2003) and Krueger (2003) on class size, for instance.

Everyone agrees, however, that teachers are important. The best estimates suggest that the identity of the teacher accounts for ten percent of the standard deviation in student outcomes (see, e.g., Rockoff 2004). Thus, there are substantial “teacher effects”.

But it is less clear how the best teachers are attracted to the profession. For one thing, most observed characteristics do not correlate with the teacher effects. In fact, teacher experience is the only observed characteristic that consistently has a positive effect on student outcomes. This implies that what makes a good teacher is mostly unobserved.

With one important exception, the attractiveness of the teacher profession has declined in the Nordic countries: In Sweden, the number of applicants per available slot in teacher education has hovered between one and two since the late 1980’s (see Björklund et al., 2005); the situation in Norway during the 2000’s was equally dismal (Falch and Strøm, 2009). The exception is Finland, where the situation is very different. In Finland, there were ten applicants per slot throughout the 2000’s (Sahlberg, 2011).
Lack of demand contributes to a worse selection of prospective teachers. Figure 2 shows different indicators of teacher qualifications measured prior to entry into teacher education. There has been a marked downward trend among graduates during the last two decades. From 1992 to 2007, there was a decline of the upper-secondary school grade-point-average (GPA) by roughly 15 percentile ranks for women and 10 percentile ranks for men. Analogous trend declines have been documented in several other countries (see Nickell and Quintini, 2002 for the UK and Corcoran et al., 2004 for the US).

The paper by Kirabo Jackson discusses different policy tools that potentially raise the quality of teaching. He discusses three approaches to improve the effectiveness of teaching: attract talented individuals into the profession, create incentives for exerting optimal effort and provide professional development so that teachers have the skills to be effective.

**Figure 2. Ability ranks of new subject teachers in Sweden (aged 25-30), 1980-2007**


Jackson’s review shows that all these three approaches can give meaningful improvements in student outcomes. There are large and persistent differences in teaching ability: having a teacher at the 85th percentile of the value-added distribution versus a teacher at the median would be associated with a 0.1 standard deviations higher student achievement. Similar improvements can be obtained if teachers exert more effort.
Moreover, well-designed mentoring, training, and evaluation programs have gains that are similar in magnitude.

Thus, there are examples of successful policies. But there is no consensus about best practices. In fact, there are few rigorous studies of evaluation programs, training programs, and pay-for-performance programs. Therefore, Jackson concludes that policy makers might wish to experiment with different versions of these programs before implementing them on a larger scale.

One common element of the different approaches to improving teacher quality is that they require that teaching effectiveness be measured in some way. According to Jackson, one of the most important policies that districts and schools should adopt is to collect and use data on teacher performance. Such data can be used as a diagnostic tool, as an outcome measure on which to base pay or personnel decisions, or as an evaluative tool.

3. What is the return to increasing access to higher education?

In the Nordic countries, there are numerous policies to improve access to university education. Examples of such policies include a generous system providing study grants and subsidized loans, and the regionalization of higher education.

The argument for introducing such policies has been tied to improving the access for talented youths in families without a tradition of university education. But since the policies are general in nature, a much wider set of the population can take advantage of such measures.

Despite the prevalence of policies to improve access to higher education, there are substantial differences in the transition rate to university by family background. Figure 3 provides an example for Sweden. It shows the odds of entering higher education for children with high-educated parents relative to children with low-educated parents.

Throughout the 20 years spanned by these birth cohorts, children with high-educated parents are more than twice as likely to enter university education. There is a downward trend in the odds ratio starting with children born in 1970, however. This might be interpreted as an effect of
policy. Whether this conjecture is correct and what the returns are to marginal expansions of higher education are questions addressed by Björn Öckert’s contribution to this volume.

Figure 3. Probability of entering university education for children with high-educated parents relative to children with low-educated parents

![Graph showing probability of entering university education for children with high-educated parents relative to children with low-educated parents]

Source: Björklund et al. (2010). We thank Björn Öckert for supplying this figure.

Note: Children with high-educated parents have (at least) one parent with at least three years of university-preparatory upper-secondary education.

Öckert uses detailed Swedish register data to analyze enrollment and marginal returns along three dimensions: parental income, distance to college, and GPA from upper-secondary school. The idea is that the parental income dimension sheds some light on policies that reduce the direct cost of attending university; the distance to college dimension sheds some light on regional expansions of university colleges; the GPA dimension, finally, sheds some light on policies that vary the admission standards.

He finds that when tuition is already low, and generous student grants are available for all, there is little leeway for increasing enrollment by reducing the direct costs. In Sweden, no students seem to be credit constrained in the literal meaning of the word. There are no differences in enrollment by family income conditional on observed characteristics (such as upper-secondary school GPA).

Still, there are substantial differences in earlier school performance and the choice of upper-secondary school programs between children from richer and poorer backgrounds. Thus, individuals may be more broadly constrained by their family background. These constraints cannot
be addressed by student grants and loans. Policies aimed at weakening the observed link between family income and college attendance should probably be implemented earlier rather than later.

If policy can attract individuals from poorer backgrounds, the evidence suggests that the returns to higher education for this group are relatively high. The estimated return to college is highest among children from the poorest backgrounds; then the return typically falls with family income.

Individuals living close to a college are more likely to enroll in college. Thus, expanding college education regionally may attract new students to college. Students living further away from college are more likely to graduate, probably because they need to complete their studies to collect a higher return to college. Thus, regional expansions of higher education in areas where the supply has been scarce may have relatively high returns.

The return to college increases sharply by upper-secondary school GPA. For the very top students, the returns in Sweden are on par with the returns to college reported in, e.g., the US. A lowering of the admissions standards only in oversubscribed programs could potentially attract more high-ability students to university. A limitation of this strategy, however, is that applicants who are rejected from popular programs often gain admission to other programs. Thus, it is not entirely clear that such a policy is efficient in attracting more high-ability students to college.

4. What are the labor market consequences of increased female participation in education?

One of the most striking trends over the past decades is the increase in female educational participation. This trend is visible in most countries of the world and the Nordic countries feature one of the largest gender differences in educational participation.

Figure 4 (taken from Tuomas Pekkarinen’s contribution to this volume) illustrates this claim for the Nordic countries and for the US which is included as a benchmark. In most Nordic countries (Denmark is the exception) there is female dominance starting with individuals born around 1950.
Figure 4. Mean years of schooling by birth cohort in a selection of countries

Source: See Pekkarinen, Figure 1, in this volume.

A closer look at the data reveals that overall female dominance comes from the fact that females are more likely to participate in tertiary education and they are less likely to drop out of secondary education. Judging from recent trends in international data, it is likely that female dominance in educational attainment will be stronger in the coming decades.

These aggregate trends should have implications for gender differences in the labor market. The purpose of Toumas Pekkarinen’s contribution to this volume is to discuss these implications. Moreover, he discusses whether policy action is needed to address the gender imbalance and, if so, whether some policies are more likely to be effective than others.

The widening gender gap in education partly reflects stalling or even falling male educational attainment. This may be a cause for concern for policymakers since recent trends in the labor market make education even more important for labor market outcomes. Low-educated men are increasingly employed in low-paid low-skill jobs. Furthermore, male labor force participation rates are declining across the whole OECD and this decline is particularly strong for low-educated men. Finally, Pekkarinen argues that the negative externalities associated with male failure in education are particularly large.
Pekkarinen then discusses various policies that address gender inequalities in education. Single-sex education is more likely to widen the gender gaps in education than to decrease them. There are indications that male students are more responsive to school resource investments such as reductions in class size. Early intervention policies, on the other hand, seem effective in improving both female and male long-term outcomes without reducing the gender gaps. There is also suggestive evidence that the structure of the educational system, and the timing of tracking in secondary school in particular, affects male and female students differently.

5. How do immigrants fare in education?

Another striking trend is the increasing prevalence of immigrants in the Nordic societies. Immigration to the Nordic countries started at different time points and has had a different character over time.

The first country to receive significant amounts of immigrants was Sweden. In the 1960’s, a substantial amount of labor immigrants arrived from Finland and Southern Europe. In the 1970’s, labor migration turned into refugee migration, however. Sweden received immigrants from Chile while Denmark received substantial inflows from Asia.

Refugee migration then increased during the 1980’s and 1990’s with wars in the Middle East and the Balkans as well as famine in Africa. By then, Norway had started to take on a substantial amount of refugees. Finland and Iceland were the last of the five Nordic countries to see immigration.

Figure 5 illustrates the immigration experiences of the Nordic countries between 1990 and 2011. The solid lines show the percentage of the overall population that is born abroad, while the dashed lined provides the same information for the population of secondary school ages (ages 15-19).

During the 2000’s, the share of the population born in another country continued to rise in the Nordic countries. Norway and Iceland, in particular, received a surge of labor migrants from Eastern Europe. This surge of immigration to Iceland stopped with the onset of the financial crisis in 2008. By 2011, the share of the population born abroad amounted to 9.3
percent in Denmark, 4.6 percent in Finland, 10.9 percent in Iceland, 11.6 percent in Norway, and 14.7 percent in Sweden.

**Figure 5. Percent of the population born abroad, 1990-2011**

Immigrant children typically do worse than native-born children in most countries. This is to be expected. Immigrants have to learn a new language, immigrant-sending countries are less affluent than immigrant-receiving countries, and there may, of course, be distress caused by the immigration experience itself (and by the events leading to immigration).

Bernt Bratsberg, Oddbjorn Raaum, and Knut Røed use Norway as a case to examine immigrant children’s educational attainment. By exploiting the country’s register data covering the whole population, they can provide a detailed comparison of the educational attainment of immigrant and native children. They follow all 15 cohorts of students who left compulsory school 1990-2004. Their key educational attainment indicator is whether the student has completed upper secondary education within five years of graduating from compulsory school. With this outcome variable, they compare natives with two immigrant groups: immigrant children born in Norway and immigrant children born abroad. Among native children, a quite stable share around 70 percent has completed upper second-
ary school as compared to 60-65 percent for immigrant children born in Norway and 50-55 percent for immigrant children born abroad. However, the authors stress a positive trend among immigrant children born in Norway who, by the end of the period, reach a completion rate of 65 percent, quite close to natives. Another finding is that immigrant children have a substantially higher educational attainment than their parents, suggesting that moving from a low- to a high-income country presents improved opportunities for education.

The authors also perform several regression analyses where they relate educational attainment to different background factors. These analyses yield several interesting insights. They find that parental resources such as earnings and education are more weakly associated with educational attainment among immigrants than among natives, probably reflecting that observed measures of such resources are less appropriate measures of long-run resources due to an ongoing integration process. They also replicate findings from other countries about the role of age at immigration with a particularly strong gradient after the age of seven. Finally, they find strong country-of-origin effects, which can be useful for the targeting of resources to municipalities, schools and students.

6. The effect of education on health and longevity

Across all countries, the more educated are in less hazardous jobs and tend to live longer. The question is whether this is a correlation – stemming from other traits that increase both education and longevity – or whether this should be viewed as a causal effect. If it is a causal effect, the full return to schooling is probably substantially higher than what is estimated from conventional earnings differentials.

Table 2 shows an example of the strength of the correlation between education and mortality in Sweden. We look at expected remaining lifetime at the age of 30 when most people have completed their education. We distinguish between three levels of education; compulsory, upper secondary, and post secondary school. The difference in longevity between the highest and lowest educational level is close to five years, it is constant over time and quite similar for men and women. We consider
this to be a substantive difference and yet the educational levels are defined in rather a crude way.

**Table 2. Expected remaining lifetime at the age of 30 by educational level. Sweden 2000 and 2010**

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Compulsory</td>
<td>Upper secondary</td>
</tr>
<tr>
<td>2000</td>
<td>46.6</td>
<td>49.0</td>
</tr>
<tr>
<td>2010</td>
<td>48.0</td>
<td>50.5</td>
</tr>
</tbody>
</table>


The contribution by Bhaskar Mazumder discusses whether descriptive evidence like that shown in Table 2 should be viewed as a result of education. Thus, the precise question is: Does education have a causal effect on health and longevity? Mazumder starts by noting that there are indeed some plausible mechanisms via which education may have such an effect. Financial resources via higher paying and more stable jobs is one, greater knowledge and improved ability to process information is another, and behavioral changes such as valuing the future more due to higher education is yet another mechanism.

Mazumder mainly focuses on a set of recent studies from seven countries – including Denmark and Sweden – that use the variation in education generated by compulsory schooling laws. He finds these studies most convincing because they use an outside (“exogenous”) force such as a reform to infer the impact of education on health and mortality. Mazumder argues that these studies point to little evidence of causal effects. The implication of this result is that the strong association between education and health and mortality is primarily due to common determinants of education on the one hand and health and mortality on the other.

Mazumder stresses that studies relying on reform-induced changes in education estimate “local” effects. The effects are valid for those who are directly affected by the reform, and may not be representative of the effects of education at other levels. Therefore, he also examines studies that use other methods. Some of these studies are suggestive of some causal effects of more education. He concludes that more research is needed on this topic.
References


Long-term effects of early childhood care and education*

Christopher Ruhm** and Jane Waldfogel***

Summary

This paper critically reviews what we know about the long-term effects of parental leave and early childhood education programs. We find only limited evidence that expansions of parental leave durations improved long-run educational or labor market outcomes of the children whose parents were affected by them, perhaps because benefits are hard to measure or confined to subgroups, or because leave entitlements were sufficiently long, even before recent extensions, to yield most potential benefits. In contrast, expansions of early education generally yield benefits at school entry, adolescence, and in adulthood, particularly for disadvantaged children; however, the gains may be less pronounced when high-quality subsidized child care was available prior to the program expansion or when subsidies increased the use of low-quality care.

Keywords: parental leave, early childhood care, education.
JEL classification numbers: I21, J13, J24.

* We gratefully acknowledge support from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) through R01HD047215 and R24HD058486, as well as the University of Virginia Bankard Fund. The content is solely the responsibility of the authors and does not necessarily represent the official views of NICHD or the National Institutes of Health.

** Frank Batten School of Leadership and Public Policy, University of Virginia and National Bureau of Economic Research, ruhm@virginia.edu.

*** Columbia University School of Social Work, jw205@columbia.edu.
Early childhood has been the focus of much research and policy in recent years. Particular attention has been directed to two types of policies – parental leave and early childhood education. Although distinct, both policy instruments alter the balance between parental care and non-parental care and education in early childhood, and may also influence the quality of the care and education that young children receive. As such, these policies are considered to not only improve child and family well-being in the all-important early childhood period but also to have the potential to improve long-term outcomes for children. The two policies can also substitute, to some extent, for each other. For example, some countries may choose to combine lengthy parental leaves with a relatively limited use of early non-parental care, while others may provide shorter leaves but earlier entry into non-parental care.

Until recently, evidence on the long-term effects of these policies has been lacking, and evidence from the new wave of studies is not always clear or consistent. Hence, in this paper, we provide a critical review of what we know, and do not know, about the long-term effects of parental leave and early childhood education, with particular focus on policies that are universal, or nearly so, for children of specified ages.

There are a variety of ways through which parental leave and universal early childhood education may affect short- and long-term developmental outcomes. This review will mostly not attempt to determine the contributions of these alternative mechanisms, or the interactions between them, in part because of limitations in existing research. However, two causal pathways deserve to be mentioned. First, as emphasized by James Heckman and co-authors (e.g. Cunha and Heckman, 2007; Heckman, 2008), the returns may be higher to investments made early in life than to those made at later ages. The key reasons are that early childhood may be a sensitive period, where the benefits are particularly high, and because early investments raise the productivity of future investments (dynamic complementarity). Kindergarten, pre-kindergarten and other educational related programs could thus be important because of the direct investments in human capital that they provide. These effects may also differ across types of children and depending on the nature of the investments. For instance, the language components of such programs may be particularly important for immigrant children whose native language differs from that of the country in which they reside. Second, there may be indi-
rect effects. Among the most important of these may be that expanded parental leave entitlements and universal or near universal early education may increase the employment of parents – particularly mothers – by reducing the opportunity costs of work. Such employment may have positive effects if it raises family incomes or negative impacts if parental time investments in children are more productive than those obtained in non-parental settings.

The structure of the paper is as follows. We first discuss challenges in estimating the effects of the policies of interest. Next, we provide a brief overview of parental leave policies and review the evidence on the effects of these policies on child outcomes. Section 3 provides an overview of recent expansions in large-scale or universal early childhood education and reviews the evidence on the effects of these programs on child outcomes. In the concluding section, we discuss policy implications as well as implications for future research.

1. Challenges in measuring policy impacts

At the onset, we should recognize several challenges in obtaining credible and accurate estimates of the long-term effects of these early childhood policies. One issue relates to the difficulty in obtaining data that are appropriate and sufficiently detailed for this task. We would like to look at the impacts of policies years or (ideally) decades in the future, but most longitudinal data sets (e.g. the National Longitudinal Survey of Youth and Panel Study of Income Dynamics in the United States or the British Household Panel Survey or various birth cohort studies in the United Kingdom) are simply not up to this task. They generally do not follow subjects for a sufficient length of time or include subjects who experienced different policy regimes. When they do, the analysis samples are typically too small to provide sufficient statistical power, particularly for interventions whose effects may be reinforced or offset by moderating or mediating factors occurring in later years. It is at least partially for this reason that estimates of short-term effects are far more voluminous and, frequently, more credible than estimates of long-term effects. However, a promising development is the increased use of registry (administrative) data by researchers, mainly in Europe. Registry data have several ad-
vantages. First, in many cases, information is available for long periods of
time (sometimes over the full lifecycle) for large representative samples
or even for the entire cohort in a country. Second, administrative data can
sometimes be combined from a variety of different sources including, for
example, the social insurance, education, employment, and health sys-
tems, to provide information over a variety of relevant dimensions of the
individual’s life. On the other hand, while rich in many ways, registry
data are often quite limited in others, simply because many types of per-
sonal information are not contained in the registries. For example, there is
generally no information on attitudes, many types of parental inputs (like
time spent investing in children) or, in some cases, basic demographic
variables such as family characteristics.

Even ignoring potential data problems, it is often difficult to identify
causal effects of parental leave or early childhood education policies be-
cause of potential omitted variables biases and endogenous enactment of
the policies. To provide one example, a country may put new early child-
hood education policies in place because it thinks that its children are
doing poorly, relative to those in other nations. Simple correlations are
then likely to understate the effectiveness of the policy. Conversely, the
benefits of the policies may be overstated if they tend to be enacted by
countries that already had relatively rich environments for child health
and development.

Tremendous progress has been made in recent years in developing
techniques to deal with these potential problems. Three are particularly
important. The first are “difference-in-difference” (or DD) estimates. In
DD models, researchers identify a “treatment” group expected to be af-
fected by the policy and a “control” group whom the policy should not
influence. For instance, when considering parental leave policies, it will
sometimes be appropriate to use children whose mothers worked during
pregnancy as the treatment group and those whose mothers did not as the
control group.¹ Differences in the outcomes of interest for children born
(or reaching a relevant age) before and after the implementation of a new
policy will be measured, and these will be compared for the treatment and
control groups. The basic assumption is that any differences observed for
the control group result from factors other than the policy and that these

¹ However, if the decision to work during pregnancy is influenced by parental leave policies,
this division into treatment and control groups may be inappropriate.
unaccounted for factors will have the same influence on the treatment group. Removing them then allows for the estimation of the causal effects of the policies. However, while helpful, DD methods are not a panacea because it is often difficult to cleanly separate individuals into treatment and control groups and because the assumption that unobserved factors influence the treatment and control groups identically may not hold. Specifically, the standard DD model includes a control for a policy main effect (a dummy variable taking the value of one for periods in which the policy operates and zero otherwise) and an interaction between the policy dummy variable and an indicator for belonging to the treatment group anticipated to be affected by the policy. The key assumption is that trends in (but not levels of) the outcome variable would be the same for the treatment and control groups in the absence of the policy being implemented. If trends are not the same, or if the treatment and control groups are not clearly defined, the interaction coefficient may understate or overestimate the true policy treatment effect.

A second approach is to use the method of instrumental variables (IV). Assume that we are interested in knowing how a variable $X$ affects an outcome $Y$ but a standard regression of $Y$ on $X$ will yield biased estimates because there are unobserved influences on $Y$ that are correlated with $X$. This bias can be eliminated if an “instrument” $Z$ can be found that is correlated with $X$ but does not independently affect $Y$, other than through its correlation with $X$ (i.e. the instrument is uncorrelated with the error term in the estimation equation). For instance, we might want to estimate the effect of child care – of a particular type or quality – on child outcomes, but the estimates may be biased because families do not randomly choose whether or not to use the specified type of child care. In this case, we might use child care policies (such as the availability of subsidies or the location of centers) as an instrument. IV estimates frequently encounter two problems. First, the assumption that the instrument has no independent effect on the outcome of interest may not hold. In the example just provided, it is possible that the subsidies or child care availability influence child outcomes in ways other than through changes in child care provision (e.g. through altering maternal employment or family income). Second, the effects of the instruments may be limited to a relatively small subset of the group in which we are interested and the impact of the policy on these subgroups may differ from those for the entire population of
interest. For instance, many child care subsidies primarily or exclusively target mothers of low-income children and the effects might be quite different than those that would occur if children from higher-income families were also included.

Third, and increasingly common in recent years, is the use of regression discontinuity (RD) techniques. The key aspect of these methods involves identifying treatment and control groups that are otherwise very similar but because of a feature of the way in which the specific policy was implemented are alternatively affected and not affected by it. Consider an early childhood education policy that affects children born on or after January 1 of a given year. Researchers using RD methods might then compare children born just before and just after this date. For instance, those born in January are the treatment group while those born in December of the previous year are the control group. The critical assumption is that children born in December would have the same outcomes as those born in January were it not for the implementation of the policy. Although this assumption might not hold, a variety of methods can be used to test whether it appears to be reasonable. RD methods provide a powerful way of providing causal estimates when the design conditions are met. However, as with instrumental variables, it will often be difficult to obtain treatment effects across the full range of policy variation that is of interest.

2. Parental leave

2.1 Origins and current situation

Parental leave policies have long been a feature of European welfare states. Beginning with Bismarck at the end of the nineteenth century, governments have enacted legislation to provide mothers with a period of

---

2 Thus, instrumental variables models obtain what is referred to as a “local average treatment effect”.

3 For instance, pre-treatment and post-treatment trends can be compared and “falsification tests” can be implemented whereby expected null effects are tested for when choosing to conduct an RD analysis using a cutoff date other than that of the actual policy implementation. RD and DD methods can also be combined by including, as an additional control, the difference-in-difference for groups just before and after the policy change compared to those during the same calendar time period in previous or subsequent years.
paid and job-protected leave around the time of childbirth. Such leaves are thought to be essential to protect the health of mothers and also to provide them with sufficient time to provide care to vulnerable newborns. Policies have been expanded in recent decades to supply longer periods of parental leave (much of which is available to both mothers and fathers) and frequently more flexibility in its use. Since mothers typically take the vast majority of parental leave, some countries have established small amounts of dedicated paternity leave or enacted bonus arrangements where the total duration of leave available is longer if both parents use some of it. All European nations offer at least 14 weeks of paid maternity leave, with a wage replacement rate of 70 to 100 percent, with at least six months of paid leave being the norm and job-protected partially paid leave of nine months or more being common (e.g. offered in Austria, Denmark, Finland, France, Germany, Greece, Iceland, Italy, Norway, Portugal, Sweden and the United Kingdom). Only the US among advanced industrialized countries does not provide paid maternity leave as a matter of right to all working mothers (nor does it guarantee paid paternity leave).

A large body of research has examined the effects of maternity leave reforms on women’s labor supply and fertility decisions, and a smaller number of analyses have studied consequences for infant health (e.g. see Ruhm, 1998, 2000; Tanaka, 2005). However, until recently, studies have not examined longer-term effects of maternity leave policies on child health and development.

The expected effects of leave policies on long-term child health and development are not completely clear a priori. First, the results of the reforms will depend on whether and how they affect maternal time at home, as well as their influence on mediating factors like family income. The reach of any laws will further depend on the share and characteristics of the population covered, as well as on the specific provisions of the legislation. Generously paid maternity leave extensions would presumably induce covered women to stay home for a longer period of time, without substantially decreasing the family income, while at the other extreme, unpaid leave would have less effect on women’s behavior and would, for those taking it up, reduce family income unless there are offsetting increases in spousal labor supply. Second, the effect of the reforms

---
4 See Ruhm (2011) for a more extensive discussion of the issues covered in this paragraph.
will depend on when in childhood they occur. The consequences could be very different for having a mother at home in the first weeks and months of the child’s life than later in infancy or toddlerhood. For instance, several studies have shown that maternity leave extensions early in life increase the likelihood and duration of breast-feeding (e.g. Baker and Milligan, 2008) which may, in turn, be beneficial for child health and development, but such effects are probably less important when considering reforms that extend leave later in the first year of life or in subsequent years. Third, the effects of any extensions will depend on the counterfactual. Where the quality of child care is high, children may benefit less from parental leave extensions than they do in situations where the quality of non-parental care is low or high quality care is difficult to afford. This discussion suggests that the effects of maternity leave reforms may not be consistent across countries or situations and that is indeed the case, as we shall see below.

2.2 The evidence

There are now four analyses of the effects of maternity leave on young adult or adolescent outcomes, all of which take advantage of policy changes in Europe that extended the period of leave and thus provide information about the effects of allowing mothers to stay home for a longer period of time in the first year or years of a child’s life.\(^5\) We begin our review with these studies and then turn to evidence from research on effects on younger children. For the most part, these latter studies examine the consequences of the recent Canadian maternity leave extensions, and differences in US state maternity leave policies.

\(^5\) Here, we focus on the effects of maternity leave. As discussed, several European countries have also instituted paid paternity leave or have passed laws reserving a portion of parental leave for fathers (so-called “daddy quotas”). While several studies have examined the effects of these laws on father involvement, employment, and earnings (see e.g. Ekberg et al., 2005; Johansson, 2010; Rege and Solli, 2011), there is little evidence to date on the effects of these programs on child outcomes. An exception is a recent Norwegian study that finds that setting aside a month of paid leave for fathers led to improved school performance for children when assessed at the age of 16, but only for those whose fathers were more highly educated than their mothers (Cools et al., 2011).
Studies assessing effects on young adults or adolescents

Denmark
Rasmussen (2010) examines the effects of a reform that extended parental leave benefits from 14 weeks to 20 weeks in 1984. She first presents data showing that the actual leave use increased by about six weeks following the reform, suggesting that the take-up of the extended benefits was high. She then uses a regression discontinuity design – where the treatment group includes children born during the two months immediately following the effective date of the leave expansion and the controls are those born in the two months before it – to analyze data on young adults’ high school enrolment and completion, and high school grade point average (GPA) measured at the age of 21.6 Rasmussen estimates that the reform had no effect on either outcome, with fairly precise estimates that reject the possibility of strong positive or negative impacts. Robustness tests using a difference-in-difference method and examining an alternative outcome (test scores from PISA for 15-year olds) yield similar conclusions. These results indicate that extending maternal time at home from 14 to 20 weeks did not affect later child outcomes, perhaps because the quality of the alternative to parental care, Danish child care, was high to start with and may even have improved with fewer children in care at young ages.

Germany
Dustmann and Schönb erg (2008) use regression discontinuity and difference-in-difference methods to analyze the effects of three reforms to maternity leave provisions in Germany: the first in 1979 extending paid leave from two to six months; the second in 1986 lengthening paid leave from six to ten months; and the third in 1992 increasing unpaid leave from 18 to 36 months.7 Each of these reforms reduced the share of wom-

---

6 Researchers have previously used RD designs to examine how parental leave reforms affect other outcomes. For instance, Lalive and Zweimüller (2009) use Austrian social security data to investigate how increases in parental leave durations from one to two years and then a subsequent decrease to eighteen months affected the fertility and return to work decisions of Austrian mothers.

7 Their primary RD approach involves comparing outcomes for children born one month before and after the reforms. They then further compare these differences to those for children born in the corresponding months in earlier or later years (where there was no policy change and so any differences will be unrelated to policy).
en working during the affected period, with the largest effects for the first reform (which the authors estimate affected some 10-15 percent of new mothers). In addition, the third reform reduced family incomes (since it extended leave but without pay). For the 1979 reform, they examine the effect on wages and unemployment for youth aged 25-26 and find no evidence of positive consequences, but with the caveat that they can only observe wages for those in regular employment and no longer in schooling. For the 1986 policy change, they study the impact on the type of high school the youth attended or had graduated from by the age of 18-20 (in the German system, only one type, gymnasium, provides direct access to university education) and again find no evidence of benefits of the extension. For the 1992 reform, they analyze the type of high school attended at age 13-14 as well as grade attendance and grade repetition; again they find no support for the hypothesis that longer leaves improved child outcomes, and the data reject the possibility of large benefits for children.

Norway

Carneiro et al. (2010) study a 1977 reform that increased paid maternity leave entitlements from zero to four months and unpaid entitlements from three to twelve months. They document that the reform led to an approximately four-month increase in maternal time at home (from an average of eight months pre-reform to twelve months post-reform). Combining regression discontinuity with difference-in-differences and examining outcomes for young adults as late as the age of 29, they find that the reform decreased high school dropout rates. The largest effects were observed for children of the least educated mothers and those whose mothers were in the bottom two quartiles of expected leave-taking prior to the reforms (i.e. mothers who pre-reform would have been anticipated to take 0.4 to 5.1 months of leave). They also find positive impacts of the leave expansions on male IQ and height (outcomes they observe only for men because they are obtained from military draft data). The authors suggest that they identify long-term benefits of maternity leave extensions, whereas other studies do not, because they focus on women who were

---

8 Their identification strategy is similar to that of Dustmann and Schönberg (2008). Specifically, they use an RD approach comparing results for children born just before versus just after the reforms and then compare these differences to those for three control groups – children born at the same time but whose mothers were ineligible for parental leave and those with eligible mothers but born in the same calendar months either two years earlier or two years later.
eligible under the reforms, rather than all new mothers, and also because
the policy change they study extended leave in the first year of life and
without reducing family income.9 They also note that the counter-factual
in Norway during this time period would have been informal child care
rather than the high-quality child care and education on offer in many
European countries in later years.

Sweden
Liu and Skans (2010) study a reform extending parental leave benefits
from 12 months to 15 months for children born beginning in October
1988.10 They use a difference-in-difference approach that relies on vari-a-
tions in leave entitlements available during 1988, with data from 1987
and 1989 also included to purge the model of season-of-birth effects.
They document a very high parental leave take-up, with the result that the
policy change effectively increased the amount of time new mothers
stayed home by up to three months, while also reducing the share of chil-
dren in non-parental care. Analyzing test scores and grades at the age of
16, the authors find no average effect of the reform on school perfor-
ance and are able to reject the possibility of large benefits. However,
the results vary by maternal education, with children of well-educated
mothers (those with at least some education beyond secondary school)
benefiting from the extension. These results suggest that children with
well-educated parents do gain from being home with their parents (and
delaying entry into child care) at 12-15 months. A likely reason is be-
cause the quality of parental care in these families exceeds the quality of
non-parental care, whereas this is not the case in less educated house-
holds.

Studies assessing effects on young children

Canada
leave reform enacted in 2000, extending the period of paid maternity

9 When they ignore eligibility, they cannot reject that the policy reforms have no effect.
10 Leave was extended by one month for children born in August 1988 and by two months
for those born in September.
leave from six months to twelve months for most of Canada. Children born to single parent households are omitted from the analysis because they would have been affected by other related policy changes. Their main results rely on an IV strategy, where the key independent variable is the number of months the mother spends at home with her child during the first year of life, and this is instrumented by whether the child was born before or after December 31, 2000, when the maternity leave expansion took effect. Baker and Milligan (2008) show that for treated mothers, the reform increased the time at home by three months, from an average of six months to nine months, and increased the duration of breastfeeding by one month. Focusing on health and developmental outcomes for children up to the age of 24 months, Baker and Milligan (2010) uncover no evidence of favorable effects of the reform. Analyzing health and developmental outcomes for children aged four to five years, Baker and Milligan (2011) again find little indication of positive impacts (they note some small possible benefits in terms of behavior, but also some small detriments in terms of cognitive development). Haeck (2011) also analyzes four to five year olds and finds that the effects vary depending on the method used. Applying standard difference-in-difference methods she finds no effects (in line with Baker and Milligan, 2011) but when applying matching difference-in-difference methods, the maternity leave extension is estimated to have positive effects on children’s health and cognitive development, and on related aspects of family functioning.

US

A recent study by Washbrook et al. (2011) investigates the effects of state maternity leave laws and two related policies (welfare related work extensions and child care subsidies) on children’s school readiness, using data from a large nationally representative birth cohort, the Early Childhood Longitudinal Study – Birth Cohort. The authors apply a DD method to the data available for a single cohort by considering differences between states with and without specified policies and for treatment (con-

11 The most important exception was that Quebec provided 70 weeks of leave both before and after the reform. Two provinices, Saskatchewan and Alberta, did not change the duration of job-protected leave until 2001.
12 They cannot easily use the RD approaches discussed for the European studies because they use survey, rather than administrative, data and thus, do not have large numbers of observations for children born immediately before and after the policy reforms.
3. Early childhood education

3.1 Origins and current situation

A good deal of attention has been focused recently on the potential role of early childhood education for improving child health and development, and narrowing gaps in outcomes between disadvantaged children and their more advantaged peers. Yet the evidence on which such claims are based has, until recently, been quite restricted, mainly deriving from a limited number of small-scale trials that delivered early childhood services to disadvantaged children in the US. These trials show convincingly that high-quality early childhood education can improve child outcomes, particularly for the disadvantaged, and can yield benefits exceeding costs.\(^{13}\) Evidence on Head Start, a compensatory early education program targeted to low-income children in the US, has also shown positive impacts for the disadvantaged children served by this program.\(^{14}\) But evi-

\(^{13}\) The results have been convincing, in part, because some of these trials involved experimental designs – where children were randomly placed in the treatment and control groups. Probably the most famous example is the Perry Preschool project, which randomized 123 children in Ypsilanti (Michigan). The treatment group received 2.5 hours a day of educationally-oriented preschool, along with weekly home visits by teachers. Long-term gains, now evaluated through the age of 40, have been observed for the treatment group along a variety of dimensions including educational attainment, employment and earnings, and social dimensions such as reduced criminality (see e.g. Karoly et al., 2005; Heckman et al., 2010).

\(^{14}\) The strongest evidence on Head Start comes from a recent experimental study, the Head Start Impact Study (HSIS) as well as several econometric studies that have used rigorous methods (e.g. family fixed effects, regression discontinuity) to control for selection bias. The HSIS
dence on large-scale or universal programs has been lacking. Yet this is exactly the evidence needed by policymakers in countries such as the US, which lack large-scale or universal programs, as they consider whether to move forward with such programs.

Fortunately, while the US has been studying its small-scale trial programs and its compensatory Head Start program, a number of countries around the world have implemented large-scale or universal programs. Indeed, it has become normative in many countries for the state to provide publicly funded and universal preschool for children beginning at the age of three or four, typically a year or two before the start of formal school entry (Organization for Economic Cooperation and Development, 2006). Do such programs improve child outcomes and help close achievement gaps between disadvantaged and advantaged children? Policymakers introducing and supporting such programs assume this to be the case, but until recently the effects of these universal programs on child outcomes had not been studied.

However, burgeoning recent research sheds some light on whether large-scale or universal provision of early childhood education improves child health and development, and narrows gaps in those domains. We will next focus on these new studies. We are particularly interested in analyses that shed some light on long-term effects but also review research providing evidence about medium- and short-term consequences, where these provide information about outcomes such as test scores or school progression that can plausibly be related to longer-term effects.

The expected effects of universal early childhood education policies on long-term outcomes for children will depend on several factors. In general, the results will depend on whether and how the policies affect children’s care arrangements, as well as their influence on mediating factors like parental employment and income. In particular, the effect of such policies will depend on the quality and type of early childhood care and education provided, the population to whom it is provided, and the age of the children when they are enrolled. In addition, the effects of such policies will depend on the counter-factual: early education reforms that move children into school or center based settings from low quality in-
formal care arrangements will have a different effect than those that move children from parental care. Thus, the effects of early childhood education reforms may not be uniform across countries or time periods.

3.2 The evidence

We begin with studies that have followed children exposed to universal preschool through adulthood, providing information about effects on long-term outcomes such as completed schooling, employment, and earnings. Then, we turn to research that has followed children to adolescence, supplying evidence on effects on medium-term outcomes such as middle school or secondary school test scores and grade retention. We conclude with more recent analyses of the effects of universal preschool programs on school readiness or primary school outcomes.

Studies assessing effects on adult outcomes

Denmark
Bingley and Westergaard-Nielsen (2012) analyze long-term effects of preschool expansions that occurred in the late 1970’s and early 1980’s, using OLS regression techniques. They estimate the effect of area preschool density when the child was aged zero to six on completed schooling and earnings at age 22-30. Preschool density is positively associated with completed schooling, and this effect tends to be larger for disadvantaged children, particularly daughters of less-educated mothers. There is also a favorable impact of preschool density on adult earnings, although the evidence is mixed as to whether this effect varies by family background.

France
Dumas and Lefranc (2012) use standard regression methods and IV techniques to evaluate long-term effects of the preschool expansions that occurred in the 1960’s and 1970’s, a period when the share of three-year-olds enrolled in preschool rose from 35 percent to 90 percent, while the share of four-year-olds enrolled grew from 60 percent to 100 percent.\footnote{Identification in the IV models relies on regional variations in the access to preschool in the 1970’s.}
They uncover positive impacts of preschool on grade repetition, test scores, high school graduation as well as on adult wages. These effects are particularly large for children from disadvantaged or intermediate (rather than advantaged) backgrounds.16

Norway
Several studies investigate the effects of preschool in Norway. Here we discuss examinations of long-term consequences. Havnes and Mogstad (2011) study the preschool expansions that occurred following the passage of the Kindergarten Act in 1975, after which federal preschool funding increased dramatically, especially in the late 1970’s, and the share of three- to six-year-olds for whom there was a subsidized child care slot rose almost threefold – from under ten percent in 1975 to over 28 percent in 1979. The authors use data on young adults aged 30-33 and apply a difference-in-difference methodology that is identified by municipality disparities in the amount by which the reform increased the availability of child care.17 Expanded preschool availability is found to raise children’s subsequent educational attainment (more years of schooling, higher rate of college attendance, and lower rate of high school dropout) and labor market participation, while reducing welfare receipts. The effects were largest for children of low-educated mothers. In terms of mechanisms, the expansion of preschool did not induce more maternal labor supply; rather, it led families to use more formal child care in place of informal care.

US
There are no studies estimating long-term effects of universal prekindergarten programs, because these have only recently been introduced in the US (studies examining impacts on medium- and short-term outcomes are discussed below). However, there is research investigating the consequences of introducing kindergarten programs (serving children aged five). Cascio (2009, 2010) exploits data on the differential timing of state funding, during the 1960’s and 1970’s, to examine the long-run effects of kindergarten.18 She finds that universal kindergarten reduced the share of

---

16 Two French studies (Caille, 2001; Goux and Maurin, 2010) examine the short-term effects of recent extensions of preschool to two-year-olds. We discuss these in a later section.

17 Municipalities with an above the median percentage point growth in child care coverage rates constitute the treatment group; those below the median are in the comparison group.

18 Nineteen states began funding kindergarten for the first time between 1966 and 1975.
21 to 35 year old adults who were high school dropouts or were incarcerated; however, these effects are only found for whites. The absence of comparable effects for blacks may be due to the fact that kindergarten substituted for enrollment in other early childhood education programs (in particular, Head Start) for many poor black children. No effect of kindergarten is found on public assistance receipt, employment, or earnings for either racial group. Using a similar empirical strategy, Dhuey (2011) analyzes long-term effects of the same kindergarten expansions on the educational and labor market outcomes for adult males aged 30-50. She finds marginally significant positive effects of kindergarten availability on the share with at least some college education, weeks worked, and wages for Hispanic men, but no corresponding effects for white or black males. She hypothesizes that the effects may be larger for Hispanics because, in the absence of kindergarten, they are less likely to attend educationally-oriented preschools and are more likely to be placed in informal non-center based care.

**Studies assessing effects on adolescents**

**Germany**

Spiess et al. (2003) use data from the German Socio-Economic Panel (GSOEP) for 1984-1994 to examine how kindergarten attendance is related to 7th grade school placement. Identification is based on a relatively limited set of controls and the authors acknowledge the possibility that kindergarten attendance differs based on unobserved characteristics.\(^{19}\) Noting this potential shortcoming, no effect of kindergarten is found for German citizens; however, immigrants attending kindergarten are more likely to be placed in the intermediate (Realschule) or university preparatory (Gymnasium) tracks and much less often placed in the less demanding Hauptschule track (the probability of attending the least demanding track is predicted to decrease from 72 to 46 percent for immigrant children who attended kindergarten, versus those who did not).

---

\(^{19}\) Additional caveats include the limitation of the study to West Germany and small sample sizes (206 German citizens and 110 residents).
India
Hazarika and Viren (2010) analyze how attending government sponsored early childhood developmental facilities (called Anganwadi), analogous facilities run by non-governmental organizations, or preschool classrooms, from the age of zero to six, is related to school enrollment and grade progression at ages 7-19. Since participation is potentially endogenous, they use an IV approach with enrollment in an early childhood developmental program instrumented by the presence in the village of an Anganwadi center or a school offering preschool classes. Previous participation in an early childhood program is estimated to raise the school enrollment of 7-19-year-olds by 31 percentage points – with strong positive effects estimated for both 7-12 and 13-19-year-olds – and also speeds the grade progression conditional on enrollment.

Norway
Black et al. (2010) examine the effects of government subsidized preschool on national exam grades for children at the age of 16, using variation in child care prices as an instrument for preschool attendance. Preschool attendance at ages three to five is found to have a positive effect on children’s future national exam grades, with the largest impacts for children from low-income families. (Results by the mother’s education level are mixed.) Drange and Telle (2010) use a difference-in-difference approach to study the 1988 introduction of free preschool for five-year-olds in two districts in Oslo (which eliminated co-payments otherwise required of parents).\(^{20}\) The program was intended to raise the school achievement of children of immigrants and the authors find that it did lead to higher grade point averages at the age of 16 for these children; however, this effect was only significant for girls.

Sweden
Fredriksson et al. (2010) analyze the effects of preschool attendance on test scores at the age of 13 for four cohorts of children born between 1967 and 1982, a period when preschool enrollment increased dramatically. They focus on the gap in test scores between children of immigrants and those with native-born parents, thus accounting for general cohort-

\(^{20}\) The comparison group consists of children in districts that did not make preschool free of charge at this point in time.
specific trends (but not those differing by immigration status of the parents). Preschool attendance is found to significantly close a portion of the language score disparity between children of immigrants, and their peers with native-born parents. However, preschool attendance is not observed to affect the gap in scores on another test, measuring inductive reasoning, nor whether children completed an academic secondary school degree.

Uruguay

Berlinski et al. (2008) analyze a government sponsored expansion in preschool provision that occurred during the late 1990’s/early 2000’s. Enrollment in public preschools increased by 76 percent between 1995 and 2004, with a particularly large growth for disadvantaged children (e.g. enrollment of children from households in the bottom quintile of the income distribution rose from 20 percent in 1991 to 60 percent in 2002). Using data on children aged 7-15 and controlling for family fixed effects to address possible selection into preschool, children who attended preschool were more likely to be enrolled in school and had completed more grades; both effects increased with child age.21 The estimated impacts were particularly large for children with low-educated parents or living outside the capital, Montevideo.

US

There are no studies that follow children exposed to universal prekindergarten programs to adolescence, because these programs are still quite recent in the United States. However, one of the studies discussed above in the context of effects on adults (Dhuey, 2011) finds that kindergarten expansions led to reduced grade retention among Hispanic children, non-English speakers, children of immigrants, and children from low SES households.

---

21 Thus, identification is based on differences in outcomes for children attending preschool compared to siblings who did not. This procedure controls for time-invariant household factors but not for potential confounding factors that differ across siblings.
Studies assessing effects on school readiness or primary school performance

Argentina

Berlinski et al. (2009) analyze how a government sponsored expansion in preschool provision affected children’s test scores and behavior in third grade. Subsequent to a 1993 law mandating universal preschool (between 1993 and 1999), the government added new preschool classrooms in local primary schools. The take-up was very high, with each new slot leading to approximately one additional child being enrolled in preschool. Using an identification strategy that exploits regional differences in the availability of new preschool spaces, the authors find that an additional year of preschool increased language and math test scores by around 0.23 of a standard deviation, and also led to improved attention, effort, class participation, and discipline. The test score effects were larger for children living in areas with relatively high poverty rates (differential effects on behavioral outcomes not reported).

Canada

Baker et al. (2008) study the introduction of a universal $5-a-day child care subsidy program for children aged zero to four in Quebec. The program did not explicitly provide preschools or prekindergarten – and thus differs from the universal preschool programs in other countries considered here – but instead supplied subsidies such that parents’ maximum child care expense would be $5 per day. The program led to large increases in maternal employment and in the use of non-parental child care. Using a DD strategy, changes in outcomes are compared between children in Quebec (the treatment group) and the rest of Canada (the controls). They find significant negative effects of the subsidy program on socio-emotional and health outcomes of children aged below five. The authors examine a range of possible mechanisms and highlight the likely roles of strong increases in maternal labor supply that are accompanied by worse parental health and relationship quality, and more hostile, less consistent parenting. In addition, a large part of the expansion of child care slots was for informal care which may have relatively poor average quality. Using a similar econometric strategy, Lefebvre et al. (2011) confirm that the subsidies significantly increased mothers’ employment and child
care use. They further demonstrate that the program had negative effects on children’s vocabulary (PPVT) scores at the age of five, which they attribute to the low average quality of child care available in Quebec during this period.

Denmark

Datta Gupta and Simonsen (2010a) examine the effect of universal preschool on child behavior at the age of seven, taking advantage of municipality-level differences in the availability of preschool slots (with children placed in family day care when preschool is unavailable). Children who attended preschool at the age of three (for less than 30 hours per week) do not differ in behavior from those exclusively cared for by parents, whereas those in family day care had more behavioral problems than either of the other groups. In a related paper, Datta Gupta and Simonsen (2010b) reported that by the age of eleven, children who attended family day care no longer had worse behavior than those who attended preschool (and they did not differ on cognitive outcomes), but they liked school less. Esping-Andersen et al. (2012) examine the effects of attending preschool, as compared to family day care or parental care, on children’s reading scores at the age of eleven. Using regression models that include fairly extensive controls (but do not account for unobserved heterogeneity between those attending different types of child care), they find that preschool attendees have significantly higher test scores, with the largest effects for children from the bottom of the income or reading score distributions.

France

Caille (2001) looks at how preschool enrollment at the age of two (rather than the usual age of three) is related to early school performance for children born at the beginning of the 1990’s. Using logistic regression models with controls for child and family characteristics, he finds that children starting preschool at two are slightly less likely to be retained than those beginning at three, with especially large effects for children of immigrants. Goux and Maurin (2010) take advantage of regional variations to examine how early (at the age of two) preschool enrollment is

22 The sample includes children born in September or October 1995 and followed through 2003.
related to subsequent school achievement, and find no evidence of any significant differences.

Germany
Felfe and Lalive (2011) investigate how center-based care provided to zero to three year olds affects social development, language skills and school grades measured at ages two through ten, exploiting differential post-unification changes in the availability of childcare slots between East and West Germany. Formal care is associated with small developmental benefits for the average child (estimated from OLS models) but large and lasting benefits for the marginal child (estimated from IV models). The difference occurs in part because the newly cared for children tend to be from low socio-economic status families where formal childcare, and the induced increases in maternal employment, are most beneficial.

US
There have been several studies of the recently introduced universal pre-kindergarten programs, which now cover more than 25 percent of the four-year-olds and a much smaller share (four percent) of the three-year-olds (Barnett et al., 2010; Ruhm, 2011). Most of this research focuses on school readiness, although some also examines outcomes later in primary school.

Magnuson et al. (2007a, 2007b) use data from the Early Childhood Longitudinal Study-Kindergarten Cohort of children in kindergarten in 1998 to examine the effects of attending pre-kindergarten as distinct from other types of childcare or parental care. Their estimates include an exceptionally detailed set of controls for child, family, background, school and neighborhood characteristics. IV estimates are also provided in Magnuson et al. (2007a), with access to state prekindergarten used as instruments. They show that children, particularly those who are disadvantaged, attending prekindergarten in the year before kindergarten enter school with better math and reading skills; however, some of these gains

23 There are also several recent investigations of child care subsidy programs for low-income families in the United States (see e.g. Herbst and Tekin, 2010a, 2010b, 2011; Johnson et al., 2011) but these targeted programs are not our focus here. Similarly, we do not discuss the many studies examining the effects of child care or preschool attendance on child outcomes (see the review in Waldfogel, 2006), as they do not examine universal or large-scale preschool programs.

24 Access is measured by state prekindergarten spending per poor child aged below six or the fraction of children aged below six in the state attending prekindergarten programs.
dissipate by fifth grade (around the age of eleven) or may be offset by later compensatory education that targets less prepared children. In addition, Magnuson et al. (2007a) find that children attending pre-kindergarten enter school with higher levels of behavioral problems, which persist to first grade; however, this does not occur for those attending a pre-kindergarten program at their current public school.

Figlio and Roth (2009) provide further evidence on the behavioral effects of pre-kindergarten, using detailed longitudinal administrative data for children attending public school in Florida. Using an IV strategy identified by the presence or absence of a public prekindergarten program in the student’s local zoned elementary school, they find that attending pre-kindergarten reduces behavioral problems, and suspension or grade retention rates during the first few years of primary school. The largest benefits are observed for children from families with low levels of education or from disadvantaged neighborhoods.

Fitzpatrick (2008) uses data on fourth graders from the National Assessment of Educational Progress (NAEP) and a difference-in-difference methodology to show that Georgia’s universal pre-kindergarten program increased the reading and math scores of disadvantaged children and those in small towns or rural areas, while reducing the probability of grade retention. The effects for other groups are less consistent. Fitzpatrick (2010) examines the effects of universal pre-kindergarten programs in Georgia and Oklahoma on children’s preschool enrollment and mothers’ employment. Universal pre-kindergarten is shown to raise preschool enrollment by at least 14 percent but with little effect on mothers’ employment, except in rural areas.

Gormley et al. (2005, 2008) have carried out several studies of Oklahoma’s universal pre-kindergarten program. Using a regression discontinuity method, taking advantage of the birthday cut-off for enrollment, they find positive effects on children’s literacy and math, as well as socio-emotional development, with the largest benefits for disadvantaged children. Wong et al. (2008) apply a similar RD approach to analyze the effects of five state pre-kindergarten programs (in Missouri, Oklahoma, New Jersey, South Carolina, and Wyoming). The results vary somewhat by state but, overall, are positive for language skills, literacy, and math.

---

25 The treatment group consists of Georgia children turning four years old on or after September 1, 1995. The comparison group includes Georgia children turning four before that date, as well as children in other states.

26 In related work, Fitzpatrick (2010) examines the effects of universal pre-kindergarten programs in Georgia and Oklahoma on children’s preschool enrollment and mothers’ employment. Universal pre-kindergarten is shown to raise preschool enrollment by at least 14 percent but with little effect on mothers’ employment, except in rural areas.

27 Studies of other individual state programs also report benefits of pre-kindergarten at school entry, as well as some positive effects sustained until later grades. These studies (of
4. Discussion

Research on the long-run effects of parental leave and early childhood education policies is growing rapidly, but we cannot yet say with confidence exactly what the consequences of these policies are. In part, this uncertainty results from the extensive data required to conduct such research and because the average effects may depend on the details of the policies implemented and the institutional environment (e.g. the quality and affordability of non-parental child care). Despite this uncertainty, some trends are beginning to emerge.

Expansions of parental leave policies may be desirable for a variety of reasons, including their role in improving the labor market outcomes of women, assisting families in balancing the competing demands of home and the workplace, and possibly improving maternal and child health. However, there is currently only limited evidence that expansions of leave durations in Europe have led to long-run improvements in educational or labor market outcomes of the children whose parents are affected by them. One possibility is that such benefits are hard to measure or are confined to specific sub-groups. This would be the case, for instance, where benefit extensions affected relatively few parents or where the effects varied among those covered. Consistent with this, we note that in Carneiro et al.’s (2010) study of the Norwegian expansions, no effects were uncovered when the analysis sample included parents likely to be ineligible for the benefit increases. Conversely, beneficial consequences were found when the estimates were focused on the women likely to be influenced by the expansions. A second potential explanation is that the leave durations in Europe were sufficiently long, even before recent extensions, so as to yield most of the possible benefits to children. Investments of parental time may have a particularly high return when they occur at the very beginning of the child’s life and European countries have long provided leave during this period. Consistent with this, we note that the strongest evidence of long-term gains is in Carneiro et al.’s (2010) study of the Norwegian expansion of paid leave from zero to four months, covering precisely the very young child ages where the gains

---

programs in California, Louisiana, Michigan, New Jersey, New Mexico, North Carolina, and Tennessee) are reviewed in Barnett et al. (2010); see also Weiland and Yoshikawa (2011) who study a pre-kindergarten program in Massachusetts.
may be particularly large. This distinction is also relevant when considering policy changes in the United States, where rights to paid leave are almost completely absent, and entitlements to short unpaid leave are available to some but not all new parents.

By contrast, the results are far more uniform for research examining early childhood education. Expansions of these programs, to cover either more children of a given age or younger children, are found to yield benefits at school entry, in adolescence, or for adults in virtually all studies. Generally, these gains are largest for those who are disadvantaged (e.g., those who come from low-income or immigrant households) and in many studies, the positive effects are largely restricted to these groups. Exceptions to this pattern are obtained in some US investigations, probably because high quality subsidized child care had been available to many low-income children through Head Start, even before the expansions. Similarly, the mixed or negative results obtained in Canada probably occurred because the expanded child care subsidies increased the use of relatively low-quality informal care (and simultaneously raised maternal employment).

These findings suggest that expansions of early child care and education yield benefits but with two caveats. First, obtaining these favorable consequences requires an institutional structure where high quality non-parental care is available and affordable. Second, the gains are generally largest for disadvantaged children. This suggests that the case for expansions of high-quality early education targeting disadvantaged children is stronger than that for universal programs. However, universality may have other advantages. For example, it may increase the political support for maintaining high program quality and foster a unity of national early childhood experiences. Moreover, most of the aforementioned benefits observed for disadvantaged children came from universal programs and we cannot conclude that narrowly targeted programs would necessarily have the same effects.

There is much scope for interesting future research. Our current understanding of long-term consequences remains quite limited, in part because many interesting policy innovations have not been in place for a sufficient period of time to observe any such effects with confidence. In addition, longer elapsed time periods between early childhood and the outcomes of interest would permit the study of potential mediating and
moderating factors, the effects of which are quite challenging to account for. Moreover, the set of dependent variables previously examined is relatively restricted, most often focusing on educational and labor market outcomes. For example, we know almost nothing about the medium- or longer-run effects of these programs for health, family functioning, criminality, social cohesion, or happiness, all of which represent important topics for subsequent investigation.

References


Datta Gupta, N. and Simonsen, M. (2010b), Where to put the kids? Effects of type of non-parental child care on pre-teen skills and risky behavior, manuscript, University of Aarhus.


Gormley, W., Gayer, T., Phillips, D. and Dawson, B. (2005), The effects of universal pre-K on cognitive development, Developmental Psychology 41, 872-884.


Heckman, J.J. (2008), Schools, skills, and synapses, Economic Inquiry 46, 289-324.


Weiland, C. and Yoshikawa, H. (2011), The impact of an urban universal public prekindergarten program on child language, literacy, and numeracy skills at kindergarten entry, paper presented at the Association for Public Policy and Management Conference, Boston, MA, November 4-6, 2011.

Comment on Ruhm and Waldfogel: Long-term effects of early childhood care and education

Tarjei Havnes

The empirical literature on early childhood interventions has grown rapidly in recent years, and such interventions have entered the policy agenda in many countries. While there is a number of recent reviews of this many-faceted literature (e.g. Blau and Currie, 2006; Almond and Currie, 2010; Baker, 2011), Ruhm and Waldfogel distinguish themselves by focusing on large-scale interventions available to the general population, rather than the targeted programs that received a great deal of attention in previous literature. Reviewing the rapidly evolving literature evaluating parental leave and early childhood care and education programs, they focus in particular on studies that provide information on the long-term effects on children of these programs, evidence that is just starting to become available. This nicely complements previous reviews and will be very helpful in providing a broad view into the literature, both across interventions and a wide range of countries.

In general, there are two main issues that need to be resolved when evaluating empirical studies. First, can we trust the estimates to be correct for the population under study? This is the question of internal validity, and it is arguably the central problem in empirical analysis. For instance,
children who attend and do not attend child care are likely to be different in dimensions that are both observable and unobservable to the researcher. Comparing outcomes directly, we then run the risk of mixing up the effects of child care with the underlying differences that are determining child care attendance itself. This is a fundamental problem that requires great care by the researcher. Ruhm and Waldfogel discuss important ways that have been used in the literature to alleviate this problem, and focus their review on studies that have taken care to deal with it.

The second issue is whether estimates are applicable to other populations than the one actually under study. That is, are they externally valid? In practice, policy makers will need to extrapolate from studies of a particular time or event, and often from countries or populations that are quite different from the intended recipients of the policy. Further, policy makers will often want to implement policies that differ in potentially important ways from the policies that have been studied in the literature. Therefore, it is of central importance to understand and be critical as to how the effects of a policy may depend on its particular features and the target population. This is the focus of the current paper.

1. The case for heterogeneous effects

There are several reasons to expect heterogeneity in the effects of early childhood interventions. First, there may be differences in the programs implemented. In particular, there may be important differences in the educational content of the program, the targeted population and the actual population affected by the program. Second, there may be differences in the effects of the same program in different populations. For instance, we might expect that the effect of introducing a child care program on children’s development is quite different depending on whether the program is voluntary or mandatory and depending on the age profile of the users.

In particular, the size and even the direction of the effect of a child care program may crucially depend on the counterfactual mode of care, i.e. the mode of care the child would have attended had the program not

---

1 I phrase in terms of child care, though only a slight rephrasing should be necessary to apply this to parental leave.
been introduced. Following Blau and Currie (2006), there are three main modes of care.

- Parental care
- Informal out-of-home care
- Formal out-of-home care

Depending on the particular program and context, a child care program that successfully offers or promotes formal out-of-home care may therefore shift children either from parental care to formal care or from informal care to formal care. Similarly, a program that successfully promotes parental leave-taking may shift children from formal care to parental care or from informal care to parental care. The impact on child development or maternal employment would be expected to differ substantially depending on which shift is the most salient. Which shift is the most salient is, in turn, likely to be highly specific to the policy under study, depending for instance on the age of the children targeted or affected, the nature of the program and the availability of formal and informal care prior to the implementation of the program.

An important path to heterogeneity in the impact of child care programs is therefore the parental choice of care prior to the reform. To be specific, say that parents invest in their children’s “quality” \( q \) by trading off a numeraire consumption good \( c \).\(^2\) For instance, parents could invest in their children by decreasing labor supply, by paying for higher quality out-of-home care, or by purchasing child goods.

Taking into account the sacrifice in terms of foregone consumption necessary to create an additional unit of child quality, the top left-hand panel of Figure 1 draws the set of feasible combinations of consumption and child quality. As for the standard production possibility set, this consumption frontier is decreasing and concave, reflecting both the fact that investing in additional units of child quality is harder as the quality level increases, and that the value of consumption is higher the less we are initially consuming due to child investments.

---

\(^2\) This framework is introduced in Havnes and Mogstad (2010). The model builds on the seminal writings of Becker (1964) who introduces the term child quality as an aggregate of the cognitive and non-cognitive determinants of a child’s measurable well-being as an adult.
Consider now the impact of a reform that subsidizes formal out-of-home care. By subsidizing formal care, the government essentially subsidizes care of a particular quality. This introduces an alternative consumption frontier which is included in the right-hand panel of Figure 1. As long as the price of formal care is still positive, parents must forego some consumption to enroll their child in formal care. Therefore, for a family that initially prefers no investment in child quality, located at the intersection of the old frontier and the vertical axis, the new frontier will lie below the former. This implies that some parents will not enroll their children in formal care, in which case their children are not affected by the reform. In Figure 1, families who are initially located to the left of point \( A \) are not affected by the reform.

**Figure 1. Parental consumption and child quality before and after a child care reform subsidizing formal care**

Since the program is subsidized, the price of formal care is lower than the price of similar quality care elsewhere. For a family that initially prefers the same quality of care as that offered in the subsidized program, the new frontier must therefore lie above the former. The new frontier must thus cross the former from below. For families who are initially located north-east in Figure 1, this implies that the relative price of child investments in terms of foregone parental consumption is lower in the formal
program than in the alternative. If they opt into the formal care, then parents in this region experience a strong substitution effect and therefore, invest more in child quality. This is illustrated in the figure by the shift from point A to A'. For parents who are initially located immediately to the right of A, this argument shows that subsidizing formal care should have an unambiguously positive effect on their children’s outcome.

Finally, opting into subsidized formal care implies locking the quality of care at the level offered in this program. Relatively affluent parents that care relatively little for their own consumption would initially choose to invest heavily in child quality. For these parents, formal care may be of lower quality than what they would otherwise provide. Depending on how easily parents can offset a lower quality of care with other child goods or investments, the new frontier may therefore cross the initial frontier a second time. The previous arguments now obviously work in reverse: In Figure 1, families who are initially located to the right of point B do not opt into the subsidized program, and are therefore not affected by the reform. Parents who are initially located immediately to the left of B opt into the program because of the low price, substituting some child quality for consumption. For children from these families, subsidizing formal care should have an unambiguously negative effect on their outcome.

In sum, the lower panel of Figure 1 draws the predicted effects on child outcomes of subsidizing formal care. For very low or very high levels of child quality, parents do not opt into the formal care program, and the effects are zero. In the intermediate region, the effects are positive in the lower parts of the distribution, then decrease in size as the initial level increases, turning negative at the top.

2. Preliminary empirical evidence

There is a strong theoretical case for important heterogeneity in how subsidizing formal care affects children’s outcomes. The empirical literature on how child care affects children’s outcomes has so far understandably been focused on the plausible identification of mean effects. Therefore, little attention has been paid to uncovering heterogeneous effects of child care.
However, differences in the counterfactual mode of care may be one explanation for the conflicting results found in parts of the literature. For instance, the impacts on maternal employment suggest that the estimates found by Baker et al. (2008) derive about two thirds from children who are shifted into formal care from informal out-of-home care, and one third from children who are shifted into formal care from parental care. In comparison, the estimated effects in Havnes and Mogstad (2010, 2011) can be interpreted as being almost solely driven (96 percent) by shifts from informal to formal out-of-home care. Therefore, it should not necessarily come as a surprise that the effects in the latter study paint a brighter picture of how child care affects children’s outcomes than the former (though, as emphasized by Ruhm and Waldfogel, this may not be the only reason for such a difference).

In Havnes and Mogstad (2010), we explicitly deal with how the effects of child care may vary systematically over the distribution of the outcomes. Specifically, we ask how the introduction in the late 1970’s of subsidized child care available to the general population in Norway affected the entire earnings distribution of children in their early thirties. Applying the methodology suggested by Firpo et al. (2009) in a difference-in-differences setup, we are able to recover an estimate of the counterfactual earnings distribution that would have prevailed if subsidized child care had not been introduced. In line with the above discussion, we find that subsidized child care raised the earnings in the lower and middle parts of the earnings distribution, but that these effects taper out and turn negative at the upper end of the earnings distribution. While the effects are positive over the bulk of the earnings distribution, the negative effects at the very top turn out to be sufficient to offset these gains, so that the mean effect on earnings is, in fact, negative.

To understand more about why we observe such large differences in the effects, we also estimate how the introduction of subsidized child care affected the earnings distribution in several different subsamples. Not surprisingly, estimates confirm that the impact is generally positive in subsamples that are overrepresented in the lower parts of the earnings distribution, such as among females and among children from lower socio-economic backgrounds, while they are less positive or negative in the converse subsamples. However, we also demonstrate that the effects exhibit the same heterogeneity within subsamples as in the full sample. For
instance, among children from high-educated families, it is still the case that it is the lower part of the earnings distribution that is lifted, while the upper part is depressed. This emphasizes that the heterogeneity in how child care affects outcomes is complicated, and may be related to characteristics of the child that are unobserved to the researcher.

3. Concluding remarks

The empirical literature on early childhood interventions in general and child care and parental leave in particular has provided important knowledge about how they may affect children’s outcomes. It has shown that such interventions can be an important determinant of children’s development.

A large and highly credible literature based on randomized experiments has shown that intensive, targeted programs can be very beneficial in the long run. However, while the effects on early cognitive test scores are generally substantial, these effects tend to dissipate over time (Knudsen et al., 2006). Lack of cognitive effects in the medium and long run suggests that the cognitive channel may not be the main mechanism behind the effect of these targeted programs on long-run outcomes such as income, education and criminal behavior. This has led to an increasing interest in determining the impact via an alternative non-cognitive channel (Heckman et al., 2005). This research will hopefully contribute importantly to our understanding of how and when early childhood interventions may have positive effects on children’s development.

The recent literature surveyed by Ruhm and Waldfogel shows that also large-scale programs available for the general population can have benefits for children in the long run. While the evidence has been much improved in recent years, our knowledge is still rather limited and superficial. This is particularly evident due to the strong heterogeneity predicted by both theory and the preliminary evidence. For instance, we still know little about the effects for children below the age of three. We also know little about what determines the quality of the program itself. For instance, there is a lack of plausible evidence on how the learning through play emphasized in the social pedagogy tradition in the Nordic countries performs compared to the structured learning programs often found in the
Anglo-Saxon kindergartens. The scope for further research is therefore considerable.

References

Recruiting, retaining, and creating quality teachers

C. Kirabo Jackson*

Summary

This article synthesizes the research literature on how to ensure that the teaching workforce is effective. It offers three approaches to improving effectiveness: attracting talented individuals into the profession, creating incentives for exerting optimal effort, and providing professional development so that teachers have the skills to be effective. The research literature reveals that each approach can yield meaningfully improved student outcomes and that no one strategy is clearly more effective. The policy implications of these findings is that a multifaceted approach would improve teacher effectiveness and student outcomes. However, although there are examples of successful policies, there is no consensus on the most effective practices. The paper concludes that schools need to be data driven, flexible in their employment practices, and open to experimentation.

Keywords: Teacher quality, education reform.
JEL classification numbers: J2 , J45, I20, I28.

* Northwestern University, IPR and NBER. kirabo-jackson@northwestern.edu.
Policy makers in many nations are grappling with how to improve the quality of public schools, and there is consensus among policy makers, parents, educators, and researchers that teachers are one of the most important components of the school environment. As such, many education policy reforms have the expressed aim of improving the quality of teachers to whom students are exposed. For example, the No Child Left Behind Act of 2001 (arguably the most significant federal intervention into education in the United States over the last 40 years) required states to ensure that all teachers of the core academic subjects were “highly qualified” by the end of the 2005-2006 school year. Similarly, one of the four specific areas that states must advance to receive federal grant funding under Race to the Top (a USD 4.35 billion US Department of Education competition designed to spur reform in state and local district education) is “recruiting, developing, rewarding, and retaining effective teachers and principals.” Similar policy pushes are under way in the United Kingdom, Australia, Mexico, Chile, and other nations.

Although the objective is clear, what is unclear is how to ensure that all teachers are highly qualified and how to go about “recruiting, developing, rewarding, and retaining effective teachers.” There are several proposed policies aimed at improving teacher quality and several conflicting research findings based on different samples and different research designs speaking to what strategies may be most effective. As such, figuring out exactly what policies one should pursue to improve the quality of teachers is neither straightforward nor obvious. However, because some studies are more conclusive than others, and theory suggests that certain policies may be relatively more or less effective in certain contexts, the academic research does provide some guidance.

There are several approaches to improving teacher quality that can roughly be put into one of three categories. The three main approaches to ensuring that students are exposed to high-quality teachers are motivated by three views of the world: In one view of the world, the primary drivers of differences in teaching performance are differences in the innate ability of teachers, which is determined before entering the profession. Under this view, the most effective policies will be those that attract high-quality prospective teachers to the profession and better retain those effective teachers already in the profession. In the second view of the world, differences in teaching performance are primarily due to differences in
teacher effort. Under this view, the most effective policies will be those that provide incentives for teachers to exert more effort. In the third view, differences in teacher effectiveness are primarily due to differences in teaching skills or know-how. Under this scenario, the most effective policies will be those that facilitate the most rapid acquisition of teacher skills for the greatest number of teachers.

In this article, I call on the existing theoretical and empirical research to provide some guidance on what are likely to be the most fruitful kinds of policies for improving teacher quality. In Section 1, I survey the evidence on the importance of teachers for student outcomes, and in Sections 2, 3, and 4, I discuss the main economic forces at play for each of the preceding proposal types and survey the relevant empirical research in the area. Finally, Section 5 presents an overview of all evidence, discusses the specific need for further research, and highlights those policies that are most likely to succeed in improving teaching quality.

1. The importance of teachers

The notion that teacher quality is the most important component of school quality comes from evidence from value-added models that measure the improvements in test scores that students make from the beginning to the end of the year. In principle, because a student’s test score at the beginning of the school year should reflect all that a student knows before entering the classroom that year, and because a student’s test score at the end of the year should reflect all that a student knows after being in the classroom with his or her teacher for the year, the growth in the student’s test scores reflects the additional learning that took place that year. By focusing on test score growth rather than simply comparing test score levels, one avoids penalizing those teachers who teach students with low levels of academic preparedness. With this in mind, researchers take the average growth in test scores for a given teacher as a measure of his or her value added, that is, how much learning took place among all students in the teacher’s classroom in a given year. These analyses have shown

---

1 See Todd and Wolpin (2006) for a discussion of the specific scenario in which test scores at the beginning of the year are a sufficient statistic for the entire history of educational inputs.
2 Note that most researchers do not merely take the average test-score growth for a given teacher but also make adjustments for differences in the levels of parental education, family
that some teachers are systematically associated with consistently high test-score growth (high value-added teachers), whereas others are systematically associated with consistently low test-score growth (low value-added teachers).

Using these value-added models, Rivkin et al. (2005), Rockoff (2004), and other researchers find that moving from an average elementary school teacher to an elementary school teacher at the 85th percentile of value added leads to between a 0.1 and 0.2 standard deviation improvement in student test scores. This is roughly equal to an improvement of between 4 and 8 percentile points in a single year. While the use of statistical models to measure teacher quality is not uncontroversial, Kane and Staiger (2008) find that these value-added measures, based on test-score gains, have tremendous power in predicting a teacher’s future success in the classroom. Jacob and Lefgren (2008) find that they are correlated with principals’ subjective evaluations of teachers, and Jackson and Bruegmann (2009) find that teachers with a high estimated value added improve the test scores of their colleagues’ students. To put the size of these estimates in context, Hanushek et al. (2005) and Jackson (2010c) find that the test-score improvements associated with being exposed to a great teacher versus an average teacher are larger than the differences associated with being in a great school versus an average school – so that a student would be better off at a bad school with a great teacher than at a great school with a bad teacher.³

While this evidence is compelling, there are some reasons why it may be premature to conclude that teachers are important in all contexts for all students and in all grades. The first is that most studies of teacher value added have focused on elementary school teachers so that it is not obvious that high school teachers have as large an effect on student outcomes as elementary school teachers. Two exceptions are Aaronson et al. (2007) and Koedel (2009) who find similar effects on math and reading income levels, student ethnic origins, and school characteristics when computing average test-score growth. In practice, the additional adjustments do improve the estimate somewhat but do not make a very large practical difference, as long as one focuses on test-score growth.

³ Readers may wonder whether this example is contrived given that the best schools may also be those that attract the best teachers, as suggested in Jackson (2009). However, there is more variation in teacher quality within schools than there is variation in mean teacher quality across schools (Hanushek et al., 2005). Consistent with this, Sass et al. (2010) find that the average effectiveness of teachers in high-poverty schools is in general less than that of teachers in other schools, but only slightly and not in all comparisons.
achievement for high school teachers as for elementary school teachers. However, Jackson (2012) uses detailed information about student class, taking into account additional sources of bias,\textsuperscript{4} and finds that while high school math teachers have important effects on math test scores, high school English teachers do not have any systematic effect on student English test scores, suggesting that while elementary school teachers and high school math teachers are important for student test score outcomes, further research may be needed for understanding the importance of non-math teachers in high school.

The other important reason to question the broad importance of teachers is that test scores are an imperfect measure of student achievement so that while teachers are important for improving test scores, other outcomes may be more important for the overall education of the child. There is mounting evidence that the effects of high schools on test scores are not the best predictors of the effects of high school on overall educational attainment (Booker et al., 2008; Deming et al., 2011). As such, the emphasis on teacher quality as measured by the ability to improve test scores alone might be unwarranted and too narrow.

In response to this second critique, there is a nascent literature showing that teachers have effects on other outcomes than test scores. Jackson (2011) finds that English teachers in 9th grade have effects on attendance and discipline problems. Similarly, Koedel (2008) finds that a student’s algebra teacher in high school affects the likelihood that a student will drop out of high school by between 4.2 and 14.1 percentage points and that those teachers associated with higher test-score growth are also those associated with reduced dropout. Moreover, Chetty et al. (2011) find that high value-added elementary school teachers increase their students’

\textsuperscript{4} Jackson (2012) points out that because elementary and secondary schools differ in important ways, methodologies designed for elementary school teachers may be inappropriate for measuring teacher quality in high school for two reasons. First, while much of the literature on teacher effects has focused on biases due to the nonrandom matching of students to teachers in elementary school settings (Rothstein, 2010; Koedel and Betts, 2009), in a high school setting, \textit{even with random assignment to teachers}, if different teachers teach in different tracks, and students in different tracks are exposed to different treatments (e.g., students in the gifted and talented track receive extra college preparatory sessions or students in the remedial math class all have a bad social studies teacher), there will be bias due to \textit{track treatment effects}; that is, in high schools, there is a possible selection bias due to nonrandom placement of students to tracks and an omitted variables bias due to track-specific treatments. This second source of bias creates additional challenges to identifying teacher effects in high school. The extant literature has not accounted for this second source of bias and may therefore be misleading about the importance of teachers in high school.
probability of college attendance, raise earnings, reduce teenage pregnancy rates, and improve the quality of the neighborhood in which their students live at age 25. Taken together, these studies indicate that both elementary and high school teachers have important effects on long-run outcomes and that estimated value added based on test scores is a reasonable measure of the importance of teachers for a student’s long-term well-being.

Despite strong evidence that teachers make important contributions to student outcomes, there are few observable characteristics that are strongly associated with teacher effectiveness. Clotfelter et al. (2007, 2010) find that a teacher’s experience, test scores, and regular licensure all have positive effects on student achievement, with larger effects on math than on reading. They conclude that, taken together, the various teacher credentials exhibit quite large effects on math achievement. To get a sense of the magnitude of teacher quality that can be explained by observable teacher characteristics, Jackson and Bruegmann (2009) summarize all observable teacher characteristics (experience, licensure, score of license exams, certification, degree level) into a single index and find that a 1 standard deviation increase in own-teacher value added due to observable characteristics is associated with a 3.6 and 2.6 percent of a standard deviation increase in math and reading test scores, respectively. These estimates are about one-third of the size of the variability in overall estimated value added, indicating that although observable teacher characteristics can explain some of the differences in teacher effectiveness, much of what makes a good teacher is still largely unknown. Recent evidence suggests that relating detailed preservice survey information from teachers and classroom observations with student achievement may allow us to better understand why some teachers are more effective in the classroom than others (Lavy, 2011; Rockoff and Speroni, 2011).

In sum, though researchers may quibble about the methods used to estimate teacher quality, and although the results vary across samples, studies, and empirical designs, the empirical research indicates that teachers are quite determinative of student test score outcomes and also longer-term well-being. Even though most typically collected objective measures

---

5 They predict student test score growth as a function of observable teacher characteristics and then take the value-added associated with each characteristic and compute the “predicted” value-added for each teacher based on their observed characteristics. This approach weights the characteristics by their importance in improving test scores.
of teacher qualifications cannot explain the large differences in outcomes across teachers, it is clear that there would be large improvements in student outcomes if one could improve the quality of the teacher to which students are exposed. In Sections 2, 3, and 4, I discuss the three main types of strategies that may achieve this.

2. Attracting and retaining quality teachers

The first group of strategies aimed at improving teacher quality are predicated on the notion that a large part of the teacher quality differences we observe are due to differences in innate teaching ability. Under this view, because the variables that are typically collected before the time of hire for teachers do not allow one to identify who will be an effective teacher with much precision, being able to attract talented individuals into the teaching profession and retaining the most successful teachers will be effective ways of improving teacher quality. In this section, I discuss strategies to attract talented prospective teachers into the profession and also strategies to help retain the most effective teachers.

2.1 Attract talented prospective teachers into the teaching profession

Much recent interest in attracting and retaining quality teachers is driven by the observation that there has been a decline in the size and quality of the pool of available teachers; that is, both the fraction of college graduates going into teaching and the measured pre-service aptitudes of teachers have declined over time in many countries. Corcoran et al. (2004) use survey data of US high school graduates spanning the years 1957-1992 and find that (1) while in 1964, more than half of working female college graduates were teachers, by 2000, this percentage had dropped to 15 percent, and (2) the likelihood that a woman from the top of her high school class would eventually enter teaching had fallen from almost 20 percent in 1964 to below 4 percent in 1992. Consistent with this, using data on female college graduates between 1961 and 1997, Hoxby and Leigh (2004) find that the share of teachers coming from the most selective colleges has fallen over time. Similarly, in Sweden, Fredriksson and Öckert (2008) find that (1) in the cohort born in 1948, approximately 40 per-
cent of the university graduates earned teaching degrees, while for the cohort born in 1975, the corresponding figure was just 14 percent, and that (2) between the birth cohorts of 1948 and 1977, the average rank of teachers in ability tests conducted at age 13 had fallen by 10 percentile points. This decline in teacher aptitudes relative to other occupations is not confined to the United States and Sweden but has also been documented in the United Kingdom (Nickell and Quintini, 2002) and Australia (Leigh and Ryan, 2006).

This decline in teacher aptitude is only problematic insofar as it has deleterious effects on student outcomes. The empirical evidence does suggest that students have better outcomes with teachers who graduate from more selective colleges and have high scores on teacher license examinations (Brewer and Ehrenberg, 1994; Clotfelter et al., 2007; Jackson and Bruegmann, 2009). As such, teacher aptitude does matter. However, the decline in aptitude is also worrisome because it is indicative of a general decline in the quality (broadly construed) of students entering teaching. As such, while much of the empirical evidence is focused on aptitude as measured by achievement tests, the patterns observed are likely related to talent broadly construed (of which test scores and college selectivity are just two measures).

Teacher pay and teacher talent

In debates about teacher quality, the claim is always made that teacher pay is too low. What people really mean by this is that high-ability (talented) individuals make too much more outside of teaching for teaching to be an attractive occupational choice and, as a result, only those who would not earn high wages outside of teaching (the less talented) decide to enter the profession. It is reasonable to look at the reasons for the decline in the supply and aptitude of teachers for guidance on how policy makers might better attract and retain high-aptitude individuals. While the timing of the declines and the exact policy or social changes vary from country to country, there are two explanations for these trends. Specifically, (1) average teacher pay (relative to other occupations) has declined over time and (2) the relative monetary benefits associated with ability have declined in teaching over time. Even though most discussions of teacher pay are focused on the low level of teacher pay on average, eco-
economic theory indicates that while the level of average teacher pay will be important for the number of individuals who decide to enter teaching, it is the dispersion of teacher pay that is likely to have a much larger effect on the aptitudes of prospective teachers.

It is helpful to lay out a stylized and simple model of occupation choice following Roy (1951) to illustrate the forces at play. Consider a world in which there are two occupations: teaching and nonteaching. Students go to college and then decide on their occupation. Each student has some known ability level (motivation, cognitive skills, social skills, scholastic preparation) that is associated with better performance and therefore higher wages in both occupations. As such, those who would be the most effective teachers are also those who would be the most effective in the nonteaching profession. The empirical evidence suggests this to be a reasonable description of the world. Specifically, both higher achievement test scores and graduating from a prestigious college are associated with higher wages outside of teaching and better classroom performance within the teaching profession. Moreover, Chingos and West (2010) examine earnings records for classroom teachers employed by Florida public schools and find that among teachers in grades 4-8 leaving for other industries, a 1 standard deviation increase in estimated value added to student achievement is associated with about 8 percent higher earnings outside of teaching; that is, teachers who increase student achievement by 4 percentile points more than the average teacher earn 8 percent more than the average teacher outside of teaching when they both leave the profession.

Suppose that wages in both occupations follow a bell curve-shaped distribution, and both teaching and nonteaching have the same median wage, but nonteaching has a more dispersed earnings distribution (i.e., the difference in earnings between high-ability types and low-ability types is larger in nonteaching than in teaching). The empirical literature indicates that this is a reasonable description of the world. Chingos and West (2010) find that while among teachers leaving for other industries, a 1 standard deviation increase in estimated value added to student achievement is associated with 6 to 9 percent higher earnings outside of teaching, it is not associated with any higher earnings within teaching. Similarly, Fredriksson and Öckert (2008) find that teacher earnings are less dispersed than in nonteaching professions and that the earnings increase
associated with higher cognitive test scores (a measure of talent) is much smaller in teaching than in other professions. This is depicted in Figures 1 and 2.

**Figure 1. Wage distributions for university graduates, by teacher education**

![Figure 1](image1)


Note: The figure shows smoothed Kernel-density plots using wages in 2004.

**Figure 2. The returns to ability for university graduates, by teacher education**

![Figure 2](image2)


Note: The lines show local average log wages conditional on cohort and gender.

In this stylized world, all students care about when choosing an occupation is how much they will earn so that they will choose the occupation
that will grant them the highest earnings. A large body of empirical work suggests that there are several additional considerations that go into the decision to teach, however, this model is meant to illustrate the economic forces at play rather than being a perfect description of the real world. A key aspect of this model is that from an individual student’s standpoint, what matters is what the student could earn in teaching versus in nonteaching, rather than what the average teacher salary is relative to the average nonteaching salary; that is, a student who knows that he or she is high-ability will respond to the wages paid in teaching compared to nonteaching for high-ability workers, whereas a student who knows that he or she is low-ability will respond to the wages paid in teaching compared to nonteaching for low-ability workers.

Consider a world in which an individual who lies in the \( x \)th percentile of the ability distribution will earn at the \( x \)th percentile of the earnings distribution of whichever occupation the individual chooses. A student who scores at the 10th percentile on some ability test is choosing between being at the bottom of a compressed teacher distribution (where all values are close to the median) and being at the bottom of a dispersed nonteacher distribution (where the 10th percentile lies far below the median). Even where the median wage for teachers and nonteachers is the same, low-ability students will be attracted to teaching because teaching offers higher wages for low-ability workers due to the compressed wage distribution. See Figure 3 (top) for an illustration of this, where median earnings are assumed to be USD 30 000 for both teaching and nonteaching.

By the same logic, a high-aptitude student who scores at the 90th percentile on the ability test is choosing between being at the top of a compressed teacher distribution (where all values are close to the median) and being at the top of a dispersed nonteacher distribution (where the 90th percentile lies far above the median). Even where the median wage for teachers and nonteachers is the same, high-ability students will be attracted to nonteaching because nonteaching offers higher wages for high-ability workers due to the dispersed wage distribution.

---

\(^{6}\) For example, Jackson (2009) finds that, holding salary constant, teachers change their school choices in response to changes in student demographics. Moreover, Ladd (2011) uses rich administrative data for North Carolina and documents that working conditions are highly predictive of teachers’ intended movement away from their schools, independent of other school characteristics such as the racial mix of students.
In this stylized world, even if median wages (and average wages) are the same in both occupations, because wages are more dispersed in non-teaching, high-ability types will choose to enter nonteaching, whereas low-ability types will choose to enter teaching. It is helpful to use this stylized framework to analyze two reforms to teacher pay: (1) an increase in the pay of all teachers by USD 6 000 (about a 25 percent increase for the median-ability student in teaching) and (2) an increase in the pay of high-ability teachers and a decrease in the pay of low-ability teachers, while keeping the pay for the median student in teaching the same.

What happens when all teachers receive an across-the-board increase?

While giving an increase to all teachers would certainly make teaching more attractive for all prospective teachers, it is not obvious that it will lead to a large increase in the ability level of teachers. To illustrate this point, consider the preceding setup, in which salaries follow a normal distribution. The median teacher and nonteacher salaries are each USD 30 000 per year, that is, the median-ability student will earn USD 30 000 in teaching or in nonteaching. In this hypothetical example, nonteacher salaries are more dispersed than teacher salaries such that the standard deviation of teacher salaries is USD 3 000, while that of nonteaching is...
USD 9,000. As such, in this hypothetical world, 95 percent of teacher salaries lie between USD 24,000 and USD 36,000, while 95 percent of nonteacher salaries lie between USD 12,000 and USD 48,000 per year. This is illustrated in Figure 3 (top). If one were to give all teachers a sizable pay increase of USD 6,000, it would shift the entire distribution of teacher earnings to the left by USD 6,000. This shift is illustrated in Figure 3 (bottom).

After the shift, the median-ability student who goes into teaching earns more than the median-ability person who goes into nonteaching by USD 6,000. In this example, after the increase, the median-ability student who teaches (and would now earn USD 36,000 in teaching) will earn the same as the 75th percentile person in nonteaching so that the 75th percentile-ability student will choose to enter teaching over nonteaching after the USD 6,000 across-the-board increase in teacher pay. In fact, one has to go all the way up to the 82nd percentile of the ability distribution for the teacher salary and the nonteacher salary to be the same. As such, all ability levels below the 82nd percentile would earn higher wages in teaching after the increase, and they would all choose to enter teaching; that is, now students with an ability between the 50th and 82nd percentiles will choose to enter teaching after the across-the-board increase in teacher pay but would not have chosen teaching before. This example also illustrates that owing to the higher dispersion in nonteaching, despite the sizable increase in teacher pay, the highest-ability students will still not decide to enter teaching. Clearly, with a large enough increase in teacher pay (such as increasing teacher pay by USD 200,000 per year), virtually all students will be attracted to teaching (both high- and low-ability students). However, even in this implausible and expensive scenario, because students of all ability levels will choose to enter teaching, unless schools and districts can identify and select the most talented applicants, the average teacher will be of a similar ability level as the average student.

What happens when the dispersion in teacher salaries grows?

Another approach to altering teacher pay would be to make it more responsive to ability so that the dispersion of teacher pay is increased. Consider the simple case where median-ability teacher salaries are kept the
same in both occupations but pay is reduced for low-ability teachers and increased for high-ability teachers to the point where the lowest-ability (lowest paid) teachers earn less than the lowest-ability nonteachers and the highest-ability teachers earn more than the highest-ability nonteachers. This situation is analogous to the situation outlined earlier (depicted in Figure 3, top), but with teaching and nonteaching reversed. Because the median-ability student will earn the same in teaching and nonteaching, he or she would be indifferent between the two occupations. However, students with below-median ability will all choose nonteaching, and students with above-median ability will all be attracted to teaching. Under this pay scheme, only students who are above the median will choose to enter teaching so that even if schools and districts cannot identify and select the most talented applicants, the average ability of teachers will be higher than the ability of the average student.

Comparing the two strategies

Whereas increasing teacher pay across the board attracts more students with abilities above the median into teaching, it does not dissuade low-ability students from teaching, and it does not attract the highest-ability students. In contrast, increasing the dispersion in teacher pay by making it more sensitive to ability attracts more high-ability students into teaching (even those of the highest ability) and dissuades low-ability students from entering teaching. As such, by increasing the teacher pay for everyone (even the less effective teachers), one both pays more to all teachers and has a lower-ability teaching pool than under the system that increases pay dispersion and only pays high-ability students high wages in teaching. While these examples are stylized, they highlight the economic forces at play and make it clear that while the average level of teacher pay will affect the ability levels of those going into teaching, the dispersion in teacher pay will generate much larger selection effects and is likely to be more important for the average ability levels of teachers. In the next section, I outline the main empirical evidence supporting this model of the world.
Empirical evidence on wage structure and teacher ability

Across a variety of nations, the decline in the aptitude of teachers is coincident with (1) declines in the level of average teacher pay, (2) a narrowing of the wage distribution in teaching, and (3) a widening of the wage distribution in nonteaching; that is, across a variety of nations, the wages paid to teachers have on average declined over time relative to nonteacher wages, the difference in wages between the highest-ability teachers and the lowest-ability teachers has declined over time, and the difference in wages between the highest-ability nonteachers and the lowest-ability nonteachers has increased over time.\(^7\) The Roy models predicts that each of these should have led to a decline in the aptitudes of teachers. While Hoxby and Leigh (2004) find that pay compression for teachers in the United States increased the share of the lowest-aptitude female college graduates who became teachers by about 9 percentage points and decreased the share of the highest-aptitude female college graduates who become teachers by about 12 percentage points, Leigh and Ryan (2006) find that the rise in pay differentials associated with ability in nonteaching occupations is responsible for the decline in the academic aptitude of new teachers in Australia. Irrespective of exactly how one apports “blame”, fundamentally, what has happened across a variety of nations is that for an individual with the potential of earning a wage at the top of the earnings distribution in any occupation, teaching has become a much less lucrative option today than it was in the past.

Policy implications

The policy implication of the fact that the relative compression of the teacher pay distribution is the primary driving force behind the decline in teacher aptitude over time for a variety of nations is that teacher pay should be structured to attract high-aptitude students and deter low-

\(^7\) In Australia, Leigh and Ryan (2006) document an increase in the ratio of pay at the 90th percentile relative to the 50th percentile (so that the highest-paid nonteachers saw relative increases in pay compared to the average nonteachers). In contrast, they see a decrease in the ratio of pay at the 90th percentile relative to the 10 percentile (so that the highest-paid teachers saw relative decreases in pay compared to the lowest-paid teachers). Similarly, Hoxby and Leigh (2004) document that in the United States, the wages of female teachers from the most selective universities became more similar to those of teachers from the least selective universities between 1963 and 2000, at the same time as the wages of female nonteachers from the most selective universities were rising.
aptitude students. Simply put, teaching should be financially lucrative for those who are good at it (presumably, high-aptitude students) and not lucrative for those who are bad at it (presumably, low-aptitude students). In most countries, however, teacher pay is almost completely unrelated to aptitude or performance on the job. In the United States, for example, in most districts, teachers are paid based on the number of years of experience, being regularly certified, and having a master’s degree. This is also true in Israel (Lavy, 2002), Sweden (Hensvik, 2010) and other nations. Teachers who perform well on the job earn the same amount, have the same likelihood of being fired, and receive the same raises as those who perform poorly, conditional on these three characteristics.

There is a variety of ways one could make teacher pay more responsive to skills such that high- and low-ability workers will be able to earn higher and lower wages in teaching than they currently do, respectively. The first is to use performance-based pay. While much of the research on performance-based teacher pay has focused on the short-term incentive effects (i.e., a teacher may exert more effort and have better outcomes to receive higher pay under performance-based pay), there could also be potential long-term selection effects associated with performance-based teacher pay; that is, if high-ability workers are more likely to have good performance and therefore earn a higher pay that is comparable to what nonteaching professions pay, they will be attracted to teaching. In contrast, low-ability teachers who know that they will not have a good performance have no additional incentive to pursue a career in teaching. While this is certainly a real possibility, to my knowledge, sadly no empirical evidence documenting this exists to date.

Another approach is to use performance-based promotions (such as in academia). In most systems, all teachers receive the same raise based on how long they have been teaching. A simple tweak to this would reward high performers (who are most likely to be high ability) with larger raises and weak performers with no raises. Under such an approach, high-ability students may select entering into teaching even if the starting salaries are low if they know that if they work hard, their earnings will rise rapidly over time. Such an approach has the added benefit of giving an increased incentive for high-ability teachers to remain in the profession. Related to this, another approach could give high-ability teachers retention bonuses to keep those who enter the profession from leaving. Clotfelter et al.
Recruiting, retaining, and creating quality teachers

(2006) find that bonus payments of USD 1 800 (about 2 percent of the average teacher’s salary) were sufficient to reduce the mean turnover rates of the targeted teachers by 17 percent. This suggests that a targeted retention bonus program based on teacher performance could alter the composition of teacher staffs by retaining effective teachers.

The other side of this coin is performance-based firing. If we think of a teacher’s expected wages as being the teacher’s wage if employed multiplied by the likelihood of being employed, one way of making expected teacher pay more sensitive to ability without altering the structure of teacher pay is to make the likelihood of future employment responsive to teacher ability. If low-aptitude or low-quality teachers know that they will be fired if they do not perform well in the classroom, they will be less likely to enter the profession and will likely seek employment elsewhere. It is worth noting that in addition to the selection effect due to a teacher’s own actions, this policy will also have a direct effect on the composition of teachers by removing low-performing teachers from the teaching pool – I discuss this mechanism separately in Section 2.2.

While there is relatively little evidence of the effect of increasing the dispersion in teacher pay on student outcomes, the empirical evidence supports the idea that increasing teacher pay can be beneficial. A number of studies document that geographic areas that experience larger increases in teacher pay over time experience larger improvements in student outcomes (Heckman et al., 1995; Card and Krueger, 1992), and Loeb and Page (2000) find that once one adjusts for labor market factors, raising teacher wages by 10 percent reduces the high school dropout rates by 3 to 4 percent. Because these increases in teacher pay were not random, it is not obvious that this relationship is causal; however, the robustness of the positive association between teacher pay and student outcomes using different models and different samples suggests that the relationship is positive. As such, policies increasing teacher pay will likely have a positive impact on student outcomes.

Because most evaluations of performance-based teacher pay have been on short-run interventions that are unlikely to generate any selection effects, there is little conclusive direct evidence of the selection effects of performance-based pay and the associated improvement in student outcomes. However, Figlio and Kenny (2007) use survey data from across the United States and find that test scores are higher in schools that offer
individual financial incentives to teachers for good performance. Similarly, using country-level performance-pay measures with international achievement microdata, Woessmann (2011) finds that the use of teacher salary adjustments for outstanding performance is significantly associated with better math, science, and reading achievement across countries. Naturally, these studies capture both the incentive effects of performance pay (due to teachers working harder) and the selection effects on the quality of teachers (due to higher-ability students entering teaching) and may be affected by other differences correlated with introducing teacher performance pay. However, the results are suggestive that making teacher pay more responsive to ability and performance might attract better prospective teachers into the profession and, accordingly, may improve student outcomes.

2.2 Identify and retain good teachers and dismiss bad teachers

The previous section deals with how one might use the teacher compensation system to induce self-selection of talented individuals into teaching and encourage good teachers to remain in teaching. Another strategy to change the composition of teachers is to actively retain good teachers and dismiss or fire bad teachers. This strategy is predicated on the notion that we have reliable information about which teachers are good and which teachers are not good. As previously discussed, teachers exhibit considerable heterogeneity in their ability to raise student test scores. In principle, if teacher quality were perfectly observable, one could simply hire only the best teachers, and the problem would be solved. Unfortunately, the world is not so simple because teacher quality is difficult to determine before a teacher is hired. To investigate the extent to which schools might be able to predict teacher effectiveness before teachers are hired, Jacob et al. (2011) administered an in-depth survey to new elementary and middle school math teachers in New York City that assessed a host of teacher qualities at the time of hire, including cognitive ability, content knowledge, personality traits (e.g., openness, conscientiousness, extraversion, agreeableness, and neuroticism), and feelings of self-efficacy (a person’s belief in his or her own competence). They find modest and marginally statistically significant relationships between student achievement and several nontraditional predictors of teacher effectiveness.
and conclude that while schools wishing to increase the effectiveness of their teacher force may benefit from gathering a broad set of information on new applicants, data on job performance will still be a more powerful tool for improving teacher selection than data available at the recruitment stage.

The most straightforward way of using on-the-job performance to remove bad teachers is to determine some cutoff value-added level and fire or dismiss all teachers with an estimated value added below this cutoff level. Rockoff and Staiger (2010) point out that if there were no hiring costs and teachers did not improve with experience because replacement teachers will, on average, be as effective as the average effectiveness in the population, principals should set the cutoff where the effectiveness of the marginal teacher (the least productive teacher not fired) is equal to the effectiveness of the average teacher in the population; that is, principals should keep only those teachers who are expected to be better than the average replacement teacher. If the principal were to set a cutoff above the level of the effectiveness of the average teacher, then by slightly lowering the cutoff, the principal could have more above-average teachers. As such, the cutoff should not be above the average effectiveness. Similarly, if the principal were to set a cutoff below the level of the effectiveness of the average teacher, then by slightly increasing the cutoff, the principal could hire fewer below-average teachers. As such, the cutoff should not be above the average effectiveness. Taken together, in this simple world with no costs to hiring and no experience effects, the cutoff should be at the average effectiveness in the population.

In the real world, there are costs to hiring, and teachers do learn on the job. Specifically, Barnes et al. (2007) and Milanowski and Odden (2007) study recruitment, hiring, and replacement costs and find that replacing a teacher costs a school anywhere between USD 4 000 and USD 15 000, with an average estimate of roughly UDS 8 200. It is worth noting, however, that some of these monetary costs will be defrayed by the lower salaries earned by new first-year teachers. Moreover, because teachers tend to improve substantially in their first year of teaching, there is an additional cost in that students assigned to the replacement rookie teacher over the first two years of teaching lose about 0.07 standard deviations in student achievement. The students’ lost lifetime earnings associated with these achievement losses are likely to be larger than the direct replace-
ment costs. When all is taken into account, the optimal dismissal rule should set the cutoff effectiveness level to maximize student achievement while taking into account the costs of recruiting, training, and hiring rookie teachers.

The effect of this optimal dismissal policy on student outcomes will primarily depend on two important factors: (1) how long teachers stay in the profession and (2) how stable the teacher effectiveness is from year to year. I briefly describe the implications of each of these factors, and I then present the research findings on what kinds of improvements one could expect from imposing the optimal policy of removing the least effective teachers.

How long teachers stay in the profession is important for the effect of the dismissal policy on student achievement. If a large fraction of teachers leave the profession for other employment and have to be replaced with average teachers, then a large share of the above-average teachers not fired will leave the profession and have to be replaced with average teachers—dampening the effect of firing the ineffective teachers. In the extreme hypothetical case in which all teachers in one year leave the profession the following year, firing ineffective teachers will have no effect on the quality of teachers the following year. By the same logic, if all high-quality teachers not fired remain in the profession, the policy will have a maximal effect, all else being equal. The US Department of Education estimates that among public school teachers who were teaching during the 2007-2008 school year, 84.5 percent remained at the same school, 7.6 percent moved to a different school, and 8.0 percent left the profession during the following year. As such, from the perspective of a school, the loss rate of those retained would be approximately 15 percent, while the loss rate for the teaching profession as a whole would be 8 percent. This suggests that while high levels of teacher turnover would mute the effects of any policy of removing bad teachers, the aggregate teacher attrition rate is relatively low such that this is likely to be a second-order concern.

The stability of teacher effectiveness from year to year is also important for the effect of this policy on student achievement. Consider the extreme case in which a teacher’s effectiveness does not change at all over time so that a teacher who is estimated to be very strong in one year

---

will be very strong the following year. In this example, all teachers who are found to be above average today will perform above average next year so that the performance will improve relative to the policy whereby no teachers are fired. Consider the other extreme example where a teacher’s effectiveness in one year is completely unrelated to his or her effectiveness the following year. In this example, all teachers who are found to be above average today will perform like average teachers next year so that the student performance will be the same under the dismissal policy as under a policy in which no teachers are fired (under the simplifying assumption that teachers do not improve with experience). These examples illustrate that the more stable teacher effectiveness is over time, the more effective the optimal dismissal policy will be. A variety of empirical studies find that more than half of the variation in estimated individual teacher impacts on tests in math and English in a given year is not persistent over time. Specifically, estimates of the correlation between estimates for the same teacher from one year to the next range between 0.2 and 0.4 (Sass, 2008; Goldhaber and Hansen, 2008; Rockoff and Staiger, 2010). Put another way, somewhere between about one third and one half of the estimated effectiveness from one year can be expected in another year. This suggests that while the benefits to the optimal dismissal policy may not be enormous, teacher effectiveness estimates are sufficiently reliable over time that the dismissal policy is likely to yield economically meaningful benefits.

Policy implications

While this policy has never been carried out, Rockoff and Staiger (2010) take the available empirical estimates about teacher turnover, the increase in teacher productivity associated with experience, and the persistence of teacher effectiveness over time to simulate the effects on steady-state (long-run) teacher quality under the optimal teacher quality cutoff rule.9

---

9 They set the standard deviation of the persistent teacher effect (in student-level standard deviation units) equal to 0.15 and the reliability of the value-added measure (the ratio of the persistent variance to total variance) equal to 40 percent. For the return to experience, they assume that a first- and second-year teacher’s value added is −0.07 and −0.02 student standard deviations below the value added of teachers in their third year or higher. They ignore the direct costs of hiring a new teacher. Finally, they assume a maximum teaching career of 30 years and an exogenous turnover rate of 5 percent, which is approximately the proportion of experienced teachers who leave the Los Angeles and New York City districts each year.
Using Monte Carlo simulations, they find that teacher value added in steady state could be increased by 0.08 standard deviations (in student achievement units) using the optimal dismissal strategy after the first year of teaching. Note that this estimate is only based on the changing composition of teachers and does not include any benefits due to incentive effects. They also analyze the effects of the optimal dismissal strategy whereby the required dismissal only occurs after a teacher’s second, third, and fourth year. While postponing the required dismissal year makes it possible to get more reliable information on teachers, this comes at the cost of keeping bad teachers on staff longer. As such, if the policy requires that all dismissals take place at a certain point in a teacher’s career, dismissal after the first year is best. Finally, they analyze the effects of a policy that allows dismissal at any point in time up until the fourth year of teaching. This is a more flexible policy as it makes it possible to remove teachers who are clearly below the threshold after one year but makes it possible to gather more reliable estimates of teacher effectiveness for other teachers. With this more flexible optimal dismissal policy, they estimate that steady state teacher effectiveness could on average be increased by 0.1 standard deviations in student achievement units. To put these estimates into perspective, the optimal flexible cutoff dismissal policy would raise the student achievement by about 4 percentile points on average. Under this strategy, 67 percent of the teachers would be dismissed after their first year of teaching, 8 percent after their second year, 4 percent after their third year, and 5 percent after their fourth year.

Adopting a policy that would dismiss more than 80 percent of all early-career teachers may not be politically palatable. As such, it is helpful to describe the gains that one could expect in steady state with a more modest approach. It turns out that the increase in steady-state value added is roughly linear to the proportion of teachers dismissed (below the optimal dismissal rate).\(^\text{10}\) Therefore, as a rule of thumb, steady-state value added increases by approximately 1.1 percentage points for every 10 percent of low-performing early-career teachers dismissed. This suggests that a less aggressive policy of removing the bottom third or half of teachers would increase the steady-state teacher value added by 3.5 and 5.5 percent of a standard deviation in student achievement units, respectively. This is between 1.5 and 2.5 percentile points on average.

\(^\text{10}\) On the basis of the author’s own calculations and those in Rockoff and Staiger (2010).
This simulation is predicated on a variety of assumptions such as no spillovers across teachers, no incentive effects, and that there is an unlimited supply of teachers. If effective teachers improve the outcomes of their peers as suggested in Jackson and Bruegmann (2009), the achievement gains could be as much as 30 percent greater than those calculated. Moreover, if teachers work harder because they know that they might be dismissed for poor performance, there could be some positive incentive effects that would make the increases in student achievement even greater. However, if there is not a large supply of prospective teachers (as may be the case for certain remote locations and for difficult to staff schools), the improvements in teacher quality may be limited. In any case, the simulated effects under reasonable assumptions are sufficiently large that dismissal policies based on estimated effectiveness should be considered.

3. Getting teachers to work harder

The discussion thus far has described teaching quality as an immutable characteristic of teachers and has been focused on altering the composition of teachers. However, there is mounting evidence that how well teachers perform in the classroom varies over time and across classroom and schooling contexts. Specifically, Aaronson et al. (2007) and Koedel and Betts (2009) estimate the teacher-year effects based on student achievement data; they group teachers into performance quintiles based on fixed-effects estimates of teacher performance and find a considerable movement between quintiles—suggesting that an excellent teacher one year is not necessarily a wonderful teacher the following year. Moreover, Jackson (2010c) finds that teacher performance varies systematically across schools. She finds that the match between teacher and school can “explain away” a quarter of the difference and is as economically important as teacher quality. In the light of these findings, it is clear that thinking of teachers as fixed quantities of value added that can be moved around and organized may be overly simplistic. These studies also suggest that there may be substantial benefits to improving teacher performance through mechanisms other than changing the composition of teachers. The variation in individual teacher performance across schooling contexts and time may be due to differences in teacher effort across
contexts and time. Insofar as this is true, there may be sizable gains to increasing the effort levels of the teachers we already have.

3.1 Teacher incentives

One policy aimed at improving student outcomes by increasing teacher effort is teacher incentive pay or pay for performance. Pay for teacher performance has been adopted in numerous districts in the United States since the 1990’s, by the Pay Performance and Management Reform in the United Kingdom, by the Victorian Government Schools Agreement in Australia, by the Carrera Magisterial Program in Mexico, and by National System of School Performance Assessment in Chile as well as through reforms in other nations.

The use of teacher incentive pay is predicated on the notion that teacher outcomes are a function of both teacher ability and teacher effort and that teachers do not expend the optimal level of teacher effort. Data from the United States, Israel, Sweden, and the United Kingdom show that most public school teachers are employed by districts or schools that use salary schedules to determine pay (Podgursky, 2007; Lavy, 2002; Atkinson et al., 2009; Hensvik, 2010). Thus, most teachers are paid primarily based on years of experience and education level. As such, teacher pay is largely unresponsive to actual teacher performance. Consistent with this, Jackson (2010c) and Chingos and West (2010) find that the correlation between teacher pay and teacher performance in the classroom is weak. If a large component of teacher performance is effort, the unresponsiveness of teacher salary to classroom effectiveness would imply that the monetary benefits to exerting more effort to improve student outcomes are close to zero. If teacher effort is responsive to monetary incentives, student outcomes could possibly be improved by changing the compensation scheme such that teachers benefit monetarily from their own effort. In other contexts, researchers have found that worker effort and output are higher when workers are paid for the output directly versus with an hourly wage (Foster and Rosenzweig, 1994; Lazear, 2000). If teaching is like other occupations, rewarding teachers for their performance may increase teacher effort and improve student outcomes.11

---

11 Readers may wonder whether the need for test-based accountability arises from the fact that school principals do not bother to observe teacher performance in the classroom. However,
Where there is a close correspondence between teacher effort and performance, and where performance is well measured, a pay-for-performance contract should elicit more effort and improve student outcomes more than a standard wage or salary contract (as most teachers have). However, in reality, neither condition is met because we cannot observe teacher effort, and teacher performance is imperfectly measured. Instead, principals are forced to infer teacher effort based on student outcomes, such as student test scores, which themselves are not a perfect measure of student learning (the outcome we really care about). This leads to three problems with pay for performance in teaching.

The first problem is that if the outcome measure (say, test scores) can be improved without exerting effort, teachers will have an incentive to do so. This would involve practices, such as “teaching to the test” and cheating, that do not require the teacher to work harder and do not lead to greater learning (which is not measured) but do improve student test scores (which are measured). Evidence of this kind of behavior has been documented in high-stakes test-taking contexts. For example, Figlio (2006) finds that low-performing students are relatively more likely to be suspended than high-performing students around the time period when accountability tests are administered. Jacob (2005) finds that test score gains associated with school-level accountability were largely driven by increases in test-specific skills and student effort and did not lead to comparable gains on a state-administered low-stakes exam. Moreover, Jacob and Levitt (2003) find evidence of teacher cheating on high-stakes student tests. These examples underscore the fact that an effective performance pay system should be based on measures of teacher performance that cannot be easily gamed (i.e., improved without increasing effort).

The second problem stems from the fact that we often care about outcomes other than the measured outcome (such as test scores). Holmstrom and Milgrom (1991) point out that incentives to perform on one dimension may cause agents to withdraw effort from other dimensions. For instance, if teacher pay is based on student standardized test scores,
teachers may spend more time teaching the topics covered on the test to improve test performance (which is easily measured) and less time nurturing a deep understanding of, and interest in, the subject (which are difficult to measure). If these difficult-to-measure aspects of performance are sufficiently important, one may be better off without any performance rewards at all. This point highlights the fact that pay-for-performance systems will not unambiguously lead to improved student outcomes and that any well-designed pay-for-performance scheme must be based on outcomes that are a good measure of student learning.

The third problem stems from the fact that effort and outcome measures may not be closely related. Performance pay is predicated on the notion that when a teacher exerts more effort, student outcomes – in this case, test scores – will tend to improve. This may not be true for two reasons. The first is that student test scores are influenced by a variety of factors outside the control of the teacher. If these outside influences change sufficiently often, the relationship between outcomes and teacher effort will be weak. Knowing this, a teacher may not be compelled to exert effort if he or she knows that it will likely not result in better outcomes and increased pay. This problem can be statistically reduced by accounting for the influence of student attributes and family influences so that one can net out the influence of the teacher (much like the value-added approach). The second problem, raised by Murnane and Cohen (1986), is that merit pay may not be effective at improving teacher performance if individual teachers do not know what to do to improve their teaching performance. If teachers do not know how to improve student outcomes, even if they exert more effort, one may have small or no improvements in student outcomes. The implication of this is that pay for performance is likely to be most effective in contexts in which the outcome measures are well understood (such as a standardized test that has been administered for several years so that there is a general familiarity with how to improve performance) or in which some guidance or coaching is provided so that teachers know how to turn their increased effort into better student outcomes.
3.2 Empirical evidence

The preceding discussion highlights that the effects of performance pay are likely to be context specific and dependent on how the incentive scheme is designed. As such, one might expect that across different studies from different nations and different outcomes, the results will be mixed. This is the case. Where one would expect the effects to be context specific, the correct question to ask is not “does it work?” but rather “can it work?” and “in what contexts does it work?” I review the empirical literature from this perspective.

While there are examples of teacher performance pay schemes that did not improve student outcomes (e.g., Fryer, 2011; Goodman and Turner, 2009; Springer et al., 2010), the empirical evidence indicates that individual teacher performance pay for individual teacher outcomes can work. Lavy (2009) analyzes an experimental program in Israel that offered individual teachers bonus payments on the basis of the performance of their classes on high school graduation exams in English and mathematics. The bonuses were structured so that teachers were rewarded on the basis of their performance relative to other teachers of the same subjects in the same school. The rewards ranged between 6 and 25 percent of the average teacher’s salary. Lavy finds that the intervention increased overall pass rates by 12 percent and average math scores by 10 percent. The effects were less precisely estimated and about half the size in English. Using survey data, he finds that improvements were mediated through changes in teaching methods, enhanced after-school teaching, and increased responsiveness to students’ needs. Similarly, positive results have been found for a pay-for-performance scheme in England that rewarded individual teachers at least an 8 percent permanent salary increase for improving student performance. Atkinson et al. (2009) find that this payment scheme improved secondary school test scores, and value added on average increased by about 40 percent of a grade per pupil. Similarly, Muralidharan and Sundararaman (2009) analyze an experimental program in India that provided bonus payments to individual teachers based on the average improvement of their students’ test scores in independently administered learning assessments (with a mean bonus of 3 percent of annual pay). After two years, students in incentive schools performed significantly better than those in control schools by 0.28 and 0.16 standard devi-
ations in math and language tests, respectively (improvements of 12 and 6 percentage points in math and reading, respectively).

Taken together, these studies show that individual teacher incentive pay can lead to sizable improvements in student outcomes. They demonstrate that this is true both in developing nation contexts (where one may expect large effects due to improvements in teacher absenteeism) and in developed nations, such as Israel and the United Kingdom, where improved outcomes likely come from other changes in teacher behavior.

Where there are positive spillovers across teachers, and teachers learn from each other, systems that reward individual teachers for group outcomes may be more effective. However, Lazear (2000) points out that under such group incentives, teachers may put in less effort because they can free ride off other teachers. Studies have found positive effects of individual teacher pay for group outcomes. Lavy (2002) analyzed a program that provided both the school and its teachers with monetary performance incentives for school-leaving-exam performance and the school dropout rate. The bonuses paid to teachers were between 1 and 3 percent of the average teacher salary and were associated with test score improvements of between 3 and 6 percent and reductions in the dropout rate of about 2 percentage points.

It is well documented that both individual and group level incentives can improve student outcomes in meaningful ways, but this is not always the case. Glewwe et al. (2008) analyze a randomized evaluation that provided primary school teachers in Kenya with group incentives based on test scores and find that, while program schools saw improved test scores in the short run, students did not retain the test-score gains after the end of the incentive program. They conclude that the results are consistent with teachers expending effort towards short-term increases in test scores but not towards deeper longer-term learning.

To help us understand the situations in which performance pay is likely to improve outcomes, I discuss the results from papers that find differences in treatment heterogeneity across different incentive designs within the same study and look for similarities in the contexts among those studies that find positive effects. Muralidharan and Sundararaman (2009) compare individual and group incentives within the same experiment in India and find that group and individual incentive schools perform equally well in the first year of the program, but that the individual incentive
Recruiting, retaining, and creating quality teachers

Schools significantly outperform group incentive schools in the second year. This suggests that incentive pay schemes for individual outcomes are likely to produce better outcomes than those for group incentives. Consistent with this, Sojourner et al. (2011) compare the effects of different kinds of pay-for-performance schemes in Minnesota and find that districts offering greater rewards for teacher-level goals experienced large gains in reading, whereas those offering rewards based on school-wide goals or subjective evaluations did not. Moreover, an evaluation of District Awards for Teacher Excellence, a state-funded program in Texas, finds that the schools and districts with larger maximum awards to individual teachers had larger test score gains in math. Taken together, these results suggest that (1) individual-level incentives are likely to produce better outcomes and (2) higher-powered incentives (i.e., larger bonuses) are likely to produce better outcomes.12

Among studies looking at group-level incentives, Fryer (2011) and Goodman and Turner (2009) analyze group incentives in New York City and find that while the average effects of performance pay were zero, there were some positive effects in smaller schools. This suggests that group incentives may be effective in small-group contexts in which the free-rider problem is less severe.

As previously mentioned, one would expect better outcomes in contexts where there is support for teachers or where there is widespread familiarity with the test because merit pay may lead to improved teacher performance when teachers have ways of figuring out how to improve student outcomes (such as other teachers or test manuals). Although no studies have directly manipulated the level of teacher support, the body of evidence suggests that this may be important. Specifically, some of the most successful studies are based on well-established national tests (Israel and England) for which teachers have a great deal of institutional knowledge at their disposal about how to improve outcomes. In contrast, many of the unsuccessful performance pay schemes in the United States are based on new tests that are not well known so that knowledge of the test is limited. While not discussing a pure teacher performance pay in-

12 Because teachers may not always be able to translate increased effort into improved outcomes, it is not obvious that higher-powered incentives will lead to improved outcomes. In fact, there is evidence that higher-powered incentives could actually lead to worse outcomes either through increasing the stress level of teachers or by exacerbating any undesirable distortionary behaviors.
tervention, Jackson (2010a, 2010b) provides suggestive evidence of this. Specifically, Jackson analyzes the longer-run effects of a high school intervention that includes cash incentives for both teachers and students for each passing score earned on Advanced Placement exams, teacher training, and curricular oversight. Jackson finds the largest positive effects on college going and college performance at those schools that had a long-standing and well-developed Advanced Placement infrastructure before the intervention was adopted. Taken as a whole, the empirical evidence points to a need for teachers knowing how to improve the outcome being rewarded.

3.3 Policy implications

In summary, the empirical evidence on the effectiveness of teacher performance-based pay suggests three things: (1) pay for performance can lead to large improvements in student outcomes, (2) pay for performance is not effective in all contexts, (3) performance schemes that reward individual-level outcomes are likely to be more effective than those that reward group outcomes, (4) performance schemes that reward group outcomes are more likely to be effective when group sizes are small so that teachers do not free ride, (5) schemes that have higher rewards are likely to be more effective than those with lower rewards, and (6) schemes that are based on well-understood outcomes and provide guidance about how to turn increased effort into improved outcomes are likely to be most effective. The range of estimates indicate that a well-designed pay-for-performance policy could have sizable positive effects on student outcomes. To put the range of positive estimated effects in context, Krueger (1999) analyzes the well-known project STAR (Student/Teacher Achievement Ratio) class size reduction experiment and finds that reducing class size by seven students increased average test scores by 4 percentile points. The most successful interventions yield larger effects than this and are likely to be less costly. Moreover, Lavy (2002) compares the costs and benefits of teacher performance pay to those of spending an equivalent amount on school resources and concludes that expenditures

\[13\text{ Note that the causal estimates from the project STAR are among the larger of the estimated class size effects. However, because they are based on an experimental design, they are likely to be the most internally valid.}\]
on teacher performance pay are “much more cost effective” that those on additional school resources.

Because teacher performance pay is based on performance at least in part based on a test, an important consideration is test design. Neal and Schanzenbach (2011) illustrate that test-based accountability systems based on the number of students who score at or above specified proficiency levels provide incentives for teachers and principals to target children near the current proficiency levels for extra attention and weak incentives to devote extra attention to students who are clearly proficient already or who have little chance of becoming proficient. As such, performance pay should ideally be based on an outcome measure that is sensitive to the improvement of all students (such as average test score growth) so that teachers have an incentive to spend extra effort on all students. In fact, Barlevy and Neal (2011) point out that because the effort required to improve test scores by 1 point may differ for low- and high-achievement students, the ideal incentive pay scheme would link teacher compensation to the ranks of their students relative to observationally similar students at the beginning of the school year. Under such a scheme, teachers will have a clear incentive to improve the outcomes of all students and will be evaluated relative to teachers with very similar student populations.

Another practical consideration is how often students need to be tested. Standardized tests are often not administered every year. In principle, one could base teacher performance pay on the results of a single test. The problem with this is that those teachers who have less well-prepared students will be penalized, whereas those with better prepared students will be rewarded. If students were randomly assigned to teachers, then annual testing would not be necessary. However, because students typically sort into schools and into classrooms, it will be important to have some sort of pre-exposure measure of student ability so that teachers can be compared to other teachers with similarly well prepared students. As such, as a practical matter, the use of test-based teacher pay will likely necessitate the use of two tests: one before exposure to teachers to determine pre-exposure proficiency and one after exposure to teachers to be used to determine teacher value added (based on some measure of student achievement growth).
4. Give teachers the knowledge and skills they need to be effective

The final approach to improving the outcomes of existing teachers is to improve their teaching skills. This approach is predicated on the notion that while innate teacher ability and teacher effort are important for student outcomes, teachers can become more effective in the classroom through improving their skills. Several studies find that while there is a wide variation in effectiveness among teachers, the benefits to additional years of experience are small beyond the first few years of teaching (e.g., Rockoff, 2004; Rivkin et al., 2005). However, recent work by Wiswall (2011) and Papay and Kraft (2010) find that teacher effectiveness improves even after the first few years of teaching. Wiswall (2011) in particular finds that after accounting for the quality of teachers who remain in the profession, the returns to later career teacher experience are sizable. This implies that teaching ability is highly malleable so that interventions that aim at increasing teacher knowledge and skills could potentially be effective at increasing average teacher effectiveness and overall student outcomes.

4.1 Evaluation programs

Some argue that teacher evaluation can improve teacher performance due to both the incentives inherent in the evaluation and to increases in teachers’ professional and skill development. Specifically, because teachers may have a desire to have positive performance evaluations, they may exert more effort and be more likely to invest in their skills. Moreover, because evaluations may provide teachers with helpful information on ways in which to improve, teachers may use this information to improve their teaching skills. Despite these possibilities, most of the existing evaluation programs are viewed as being poorly designed and therefore inef-

---

14 He estimates that a teacher with 30 years of experience has over 1 standard deviation higher measured quality than a new, inexperienced teacher and about 0.75 standard deviations higher measured mathematics effectiveness than a teacher with 5 years of experience. In comparison, estimates on the same data sample using previously restricted models suggest that experienced teachers have between 0.1 and 0.2 standard deviations higher quality than new teachers, with almost all of these gains in the first few years of teaching.
Recruiting, retaining, and creating quality teachers

In an overview of teacher evaluation systems in the United States, Donaldson (2009) describes most evaluation systems as follows:

External constraints decrease evaluators’ inclination to evaluate rigorously—vague district standards, poor evaluation instruments, overly restrictive collective bargaining agreements, and a lack of time all contribute to this problem. Internal constraints, such as the absence of high-quality professional development for evaluators, a school culture that discourages critical feedback and negative evaluation ratings, and a district culture that offers little oversight and few incentives for administrators to evaluate accurately, also contribute to inflated ratings. (p. 2)

He also found that in most cases, “evaluators rarely provide teachers with substantive feedback” (p. 2). This highlights the fact that the relative lack of evidence supporting the use of evaluation systems may not be due to them being ineffective per se but rather to the fact that they are often poorly implemented.

Consistent with this notion, Taylor and Tyler (2011) study the effects of a seemingly well-designed teacher evaluation scheme on teacher performance in Cincinnati Public Schools. They find evidence that quality classroom observation-based evaluation and performance measures can improve mid-career teacher performance both during the period of evaluation and in subsequent years, consistent with human capital investment. Specifically, students assigned to a mid-career teacher after he or she participates in evaluation score about 0.1 of a standard deviation higher in math than similar students taught by the same teacher prior to participation. This is the same size effect for replacing an average teacher with a teacher at the 85th percentile of the distribution and is equivalent to improving test scores by 4 percentile points. This is about the same size effect as reducing class size by about 30 percent and is notable given that the sample being studied comprised mid-career teachers.

A few key aspects of this program likely led to its success: (1) teachers were evaluated based on rubric criteria that have been demonstrated to produce higher achievement;15 (2) evaluation outcomes were linked to

---

15 Teachers are evaluated on dozens of specific skills and practices covering classroom management, instruction, content knowledge, and planning, among other topics. Evaluators use a scoring rubric, based on Charlotte Danielson’s (1996) Enhancing Professional Practice: A
career development such that teachers who did not have strong evaluations had to undergo a year-long process of intensive assistance from a mentor that included another full year of evaluation with more frequent classroom observations, along with other meetings and writing assignments; and (3) teachers received detailed feedback about where they were deficient and how to improve. In essence, this program appeared to overcome many of the shortcomings cited as responsible for the failure of other evaluation systems and led to meaningfully improved student outcomes. This suggests that similarly well designed programs might also be successful.

4.2 Peer mentoring

Another approach to improving teacher skills is to use peer mentoring. In the United States in 2003-2004, nearly 70 percent of the recently hired teachers reported that they received help from a mentor in their first year of teaching, and a majority of states require mentoring programs for new teachers (Rockoff, 2008). Despite the prevalence of mentoring programs, many of the empirical evaluations of mentoring programs are based on research methodologies that do not provide credible causal estimates. In one, the first evaluation based on quasi-experimental variation, Rockoff (2008) studies the impact of mentoring in New York City. He finds strong relationships between measures of mentoring quality and teachers’ claims regarding the impact of mentors on their success in the classroom but weaker evidence of effects on teacher absences, retention, and student achievement. However, he does find that retention within a particular school is higher when a mentor has previous experience of working in that school, suggesting that an important part of mentoring may be the provision of school-specific knowledge. While the relationship may not be causal, he also documents that 10 additional hours of mentoring are associated with 0.05 standard deviations higher student achievement in math and 0.04 standard deviations higher achievement in reading. If this suggestive relationship is causal, these are important effect sizes.

Framework for Teaching, which describes the performance of each skill and practice at four levels: Distinguished, Proficient, Basic, and Unsatisfactory

16 See Strong (2005) for a review of this literature.
Although there is only suggestive evidence of benefits to formal peer mentoring, Daly et al. (2010) and Bakkenes et al. (1999) use network analysis to find that teachers are most likely to interact with teachers in the same grade level. As such, informal peer mentoring may be important for how teachers acquire skills. Using other teachers at the same school in the same grade as a measure of peers, Jackson and Bruegmann (2009) study the effect of working around more effective colleagues. They find that exposure to better teacher peers improves a teacher’s own performance. Specifically, a 1 standard deviation increase in mean teacher-peer quality was associated with a 0.04 and 0.02 standard deviation increase in math and reading achievement, respectively. After two years, a 1 standard deviation increase in mean teacher-peer quality that persists over time was associated with a 0.078 standard deviation increase in student math test scores and a 0.072 standard deviation increase in reading test scores. Importantly, these effects persisted even after the working relationships ended, suggesting that they reflected peer learning.

Taken together, these findings suggest that peer learning and peer mentoring can be fostered to improve teacher effectiveness. In fact, Jackson and Bruegmann (2009) find that about 20 percent of teacher effectiveness can be explained by the value added of a teacher’s peers over the previous three years, suggesting a significant role for peer learning. The evidence also suggests that the most effective peer mentors are likely to be teachers from the same school and the same grade who share similar experiences and can impart both pedagogical skills and specific knowledge about the grade and school.

### 4.3 Formal teacher training

The most obvious way of improving teacher quality would be to provide formal teacher training. Indeed, formal training is prevalent among teachers. In the United States, 72 percent of the teachers report having engaged in training related to their main teaching assignment during the previous year (Parsad et al., 2001), and a similar fraction reports receiving training on new teaching methods. Unfortunately, most of the existing research on in-service teacher training suffers from the fact that the training is determined by teachers and schools so that it is unclear whether one can credibly compare the outcomes of teachers who undergo training to the out-
comes of those who do not. There are two notable exceptions. First, Jacob and Lefgren (2004) who use a credible regression discontinuity design and find no statistically significant effect of receiving additional hours of teacher training on student outcomes in Chicago. Moreover, using a quasi-experimental design, Harris and Sass (2007) study the effects of various types of education and training on teacher productivity. Their results suggest that teacher training generally has little influence on productivity but that content-focused teacher professional development appears to be positively associated with productivity in middle and high school math. However, Angrist and Lavy (2001) find that an in-service teacher training program in Israel designed to improve the teaching of language and math rather than content was associated with between 0.2 and 0.4 standard deviations higher student test scores in non-religious schools. While these effects are not very precisely estimated, these are sizable effects of between 4 and 8 percentile points on average. This seemingly successful training program involved a mixture of counseling and feedback sessions for teachers, changes in the organization of class time, and training in the use of instructional aids.

Taken together, the results suggest that although many of the teacher training programs that have been evaluated have not been wildly successful at improving student achievement, the results from Israel indicate that a well-designed in-service training program can be an effective tool for improving teacher quality. The challenge is to identify these effective programs so that best practices can be replicated in other settings.

5. Conclusions

Under the views that (1) the primary drivers of differences in teaching performance are differences in the innate ability of teachers, which is determined before entering the profession; (2) differences in teaching performance are primarily due to differences in teacher effort; and (3) differences in teacher effectiveness are primarily due to differences in teaching skills or know-how, there are three main approaches to improving teacher quality. They are to (1) change the composition of teachers so that only the most able are in the profession, (2) provide appropriate incentives so that teachers exert the optimal amount of effort to improve
student outcomes and become more effective, and (3) provide teachers with the knowledge and skills they need to be successful in the classroom.

The empirical research indicates that there are large and persistent differences in teaching ability such that having a teacher at the 85th percentile of the value-added distribution versus a teacher at the median would be associated with 0.1 standard deviations higher student achievement. This implies that obtaining the correct mix of teachers could lead to sizable achievement gains. At the same time, the empirical research shows that by exerting more effort, teachers can increase their effectiveness by the same amount as having a teacher at the 85th percentile of the value-added distribution versus a teacher at the median — implying that while having a high-ability teacher is important, having a hardworking teacher may be equally important for student achievement. Moreover, the empirical evidence shows that teachers do become more effective over time as they acquire more skills and that the achievement gains associated with well-designed mentoring, training, and evaluation programs are similar in magnitude to the gains associated with having a teacher at the 85th percentile of the value-added distribution versus a teacher at the median. As such, it is clear that all three approaches, if done correctly, may be successful at improving student achievement.

One common element of the different approaches to improving teacher quality is that most of them require that teaching effectiveness be measured in some way; that is, making pay more responsive to performance, using incentive-based pay, evaluation, and targeted peer mentoring all require the collection of data on teacher performance. As such, one of the most important policies that districts and schools should adopt is the use of data on teacher performance that can be used either as a diagnostic tool, as an outcome measure on which to base pay or personnel decisions, or as an evaluative tool. With this infrastructure put in place, most of the proposed education reforms can be implemented at a relatively low cost.

In discussing the different approaches to improving teacher effectiveness, it is also apparent that many strategies are related to each other in meaningful ways such that some policies might work better when others are in place. For example, teacher training may be more effective when teachers are rewarded for their improved classroom effectiveness and therefore have an increased incentive to learn. Similarly, pay for perfor-
mance may be more effective when coupled with teacher evaluations because the evaluations may provide teachers with much-needed information about where they are deficient and what they need to improve. Moreover, optimal dismissal policies and targeted retention bonuses may be more effective where teachers acquire many teaching-specific skills.

It is also the case that some policies might improve outcomes through all three mechanisms (changing the composition of teachers, changing the effort levels of teachers, and increasing the skills of teachers). In light of evidence that all three mechanisms can lead to meaningfully improved student outcomes, strategies operating through all three mechanisms might be most useful. For example, if districts were to provide teachers with performance-based\textsuperscript{17} raises (as in many occupations), while also providing high-quality teacher training, this could induce the selection of high-quality applicants to teaching, lead to an increase in effort for existing teachers, and also give teachers an incentive to invest in their skills and undergo high quality training. The combination of all three mechanisms seems the most likely to be effective. The empirical literature suggests that schools or districts that use performance-based measures in the dismissal and retention of teachers, make teaching financially lucrative for high-ability students, and provide high-quality professional development and training would likely experience large improvements in student outcomes.

Finally, the fact that there are few rigorous studies on best practices for evaluation programs or training programs or for pay-for-performance programs suggests that policy makers might wish to experiment with different versions of these programs and then adopt those practices that yield the best results. While this may not be a short-term solution, if districts and schools were more experimentally minded and were to adopt programs in ways that could be rigorously evaluated, the successful interventions (that often led to improved student outcomes) could be implemented in many schools in the long run.

\textsuperscript{17} These performance measures need not be test based but can be based on classroom observations and take the form of principal or peer evaluation – allowing for the evaluations to serve not only as an evaluative tool but also as a diagnostic tool and for useful feedback.
References


Deming, D., Hastings, J., Kane, T. and Staiger, D. (2011), School choice, school quality and academic achievement, manuscript, YALE University.
Fredriksson, P. and Öckert, B. (2008), The supply of skills to the teacher profession, manuscript, Uppsala University.
Fryer, R. (2011), Teacher incentives and student achievement: Evidence from New York City public schools, manuscript, Harvard University, Cambridge, MA.
Jackson, C.K. (2011), High-school teachers and non-test score outcomes?, manuscript, Northwestern University.


Comment on Jackson: Recruiting, retaining, and creating quality teachers

Torberg Falch*

How can governments improve the teaching quality in schools? The behavior of teachers is essential for teaching quality; they are responsible for the learning environment in class. Student achievement clearly depends on a variety of other factors, such as family background, but teacher behavior can more easily be affected by policy interventions. Kirabo Jackson’s paper gives an insightful and comprehensive review of policies to increase teacher quality. The paper focuses on three main policy areas and provides both theoretical arguments for the impact of these policies and empirical evidence from the US. In conclusion, better recruitment to the teacher profession, financial incentives in the profession, as well as professional development have the potential of increasing teacher quality.

One may wonder about the relevance of evidence on teacher behavior and teacher labor market policies in the US for education policy in other countries. In general, there are many differences in labor market institutions across countries. Institutions that are relevant for the teacher labor market, however, have many similarities in the Nordic countries and the US. Teachers are certified by formal education. There is decentralized responsibility for compulsory education, in the form of school districts in the US and multi-purpose municipalities in the Nordic countries. These local governments are responsible for the hiring (and eventually firing) of teachers. Wages are traditionally determined by experience and formal education, and not related to performance on the job. Teacher unions are

* Department of Economics, Norwegian University of Science and Technology, torberg.falch@svt.ntnu.no.
visible in public debates, seem influential, and take part in collective bargains.\(^1\)

In addition, the available empirical evidence indicates that teacher behavior is similar across countries. This is the case for turnover rates and turnover behavior (Falch and Strøm, 2005; Karbownik, 2011), with a relatively high share of teachers leaving for other jobs. In general, there has been a decline in the quality of novice teachers as measured by IQ scores and relative university performance.\(^2\)

Teacher education is an institutional feature that differs across countries. While teachers in the US are typically educated at regular universities, the Nordic countries have traditionally had specific teacher colleges that enroll students directly from upper secondary education. The study programs are explicitly designed for a teacher career, including practical training. Other countries have larger differences in educational and teacher institutions compared to the US. In southern Europe, for example, the decision-making is more centralized. The hiring of individual teachers at specific schools is done by the state in several countries, including Italy, France and Germany (see, for example, Barbieri et al., 2011), leaving much less leeway for the local governments.

I will argue that the evidence from the US is of particular importance in areas where evidence is missing from other countries. For example, the Nordic register data do not allow matching of individual teachers and students. Since it is only possible to match teachers and students at the school level, variation in teacher quality cannot be estimated as in the US studies. However, the similarities in institutions and the teacher behavior observed imply that a similar variation in teacher quality in the Nordic countries must be expected. The review of the literature shows that the variation in teacher quality – measured as value added in student test scores – is substantial. And since students’ skills have long-term impacts, teacher quality has a high economic value (Hanushek, 2011).

\(^1\) Sweden is an exception with respect to wages since they are determined by individualized bargains, see for example Hensvik (2012). However, the teacher unions seem to be important players in these bargains.

\(^2\) With regard to attracting high-ability individuals to the teaching profession, Finland seems to be an interesting exception. Teacher colleges in Finland are highly oversubscribed, and enrollment requires very high performance in secondary education. At the same time, Finland is among the countries with the very best results on international comparative student tests such as the PISA-test. Why the teaching profession is so attractive in Finland as compared to other countries is unclear and is an important issue for further research.
In the following, I provide some discussion of the three main policies that Jackson’s paper considers. My viewpoints are from a Nordic institutional setting perspective.

The first policy is recruiting and retaining high-quality teachers. One obvious possibility is to increase teacher wages relative to wages in other professions. The evidence clearly indicates that schools that increase the wage improve their recruitment of teachers. For Norway, Falch (2010, 2011) exploits a system with wage bonuses in the three northernmost counties during the 1990’s, and finds that the supply elasticity faced by individual schools is in the range 1.0-1.9. Clotfelter et al. (2008) investigate a similar bonus system in North Carolina, and find an effect of similar size. The implication is that individual schools can use wages strategically to increase the number of teachers interested in working at that particular school. This is an argument for flexibility in wage setting at the school level to the extent that teacher supply varies across schools. In fact, the interventions investigated by Falch (2010, 2011) and Clotfelter et al. (2008) were motivated by teacher shortages. By allowing wages to vary across schools, one should, in principle, achieve a larger wage variation of the form that Jackson’s paper argues for. Then, teachers willing to work in less-popular schools, and arguably performing more demanding tasks, are rewarded for their extra effort or negative amenities.

Interestingly, Sweden introduced wage flexibility not only at the school level but also at the individual level in the 1990’s. Surprisingly, Hensvik (2012) finds that the decentralization of the wage setting did not change the overall distribution of teacher wages across teachers. However, she finds that the average teacher wage in the local labor market is responsive to competition. The entry of private high schools with public funding increases the average wage in public high schools. The estimated effect of competition is not homogenous. The wage increases first and foremost for novice teachers and teachers in fields with the most severe shortages such as math and science teachers. In addition, private competition is also associated with a stronger link between wages and teachers’ cognitive skills. Although the estimated effects are relatively small, they indicate that a compressed wage structure and low wage responsiveness to teacher quality are at least partly a result of weak external pressure.

The theoretical considerations in Jackson’s paper do not take the recruitment behavior of schools into account. Teachers are attractive work-
ers in other workplaces than schools, which is visible by the fact that there are many more certified teachers in the labor force than the teachers actually teaching. Thus, the recruitment policy of the schools is important for overall teacher quality. Since Ballou (1996), it has been an understanding that schools do not recruit the best teachers. However, Ballou uses imperfect measures of teacher quality in his analysis, mainly the quality of the college from which the teacher is educated. An alternative analysis would be to investigate whether recruitment policy affects school outcomes. Indeed, Naper (2010) finds for Norway that decentralizing the recruitment process from the municipal level to the schools increases school effectiveness. Similarly, Hensvik (2012) for Sweden and Hoxby (2002) for the US find that private schools to a larger extent than public schools hire teachers who lack formal skills but who are of higher ability. If higher teacher wages are to have a major impact on teacher quality, it seems like it has to be combined with decentralized recruitment and accountability at the school level.

The second policy discussed is teacher incentive pay. Financial incentives can be divided into two different kinds; high performance can be rewarded by (short-term) bonuses or (long-term) promotions. The latter are far less common for teachers than for other workers since there is basically no career ladder within the teacher profession. Traditionally, the only promotion possibility has been to leave teaching for administrative tasks such as becoming a school principal. The lack of possibilities for promotions within teaching gives weaker incentives to improve the performance than in other professions.

The evidence of the effects of incentive pay for teachers is mixed. The highest returns in terms of student achievement are found for incentives to individual teachers within well-established systems for objective measurements of student outcomes. The evidence of a weak impact of incentives at the group level is not necessarily in accordance with the evidence for other types of work. In a consideration of whether individual or group level incentives are to be used, the type of work to be performed is probably important. In a survey of the literature on incentives in firms, Prendergast (1999) finds that team incentives yield higher productivity growth than what is suggested by economic theory. He argues that an important mechanism is peer pressure, where workers can monitor one another and respond to those who fail to perform adequately. This indi-
cates that group incentives are only effective when the group members cooperate closely and can relatively easily observe each other’s behavior. This is typically not the case for teaching since the teacher typically stands alone in front of the class.

Interestingly, after abolishing the maximum class size rule, team-based teaching has become popular in Norwegian schools. With team-based teaching, several teachers are jointly responsible for all students in a particular grade. Then, it becomes more straightforward to observe the behavior and contributions of other teachers in the group, which might make group incentives more effective. In addition, direct observations of other teachers can improve teaching skills by fostering peer learning.

Increasing the teaching skills of the teachers is the final policy Jackson’s paper considers. In at least Sweden and Norway, there are expanding state programs for formal training of experienced teachers. The type of training varies, and the choice of training program is mainly the choice of the individual teacher. The self-selection in such systems makes them inherently difficult to evaluate, and few evaluation studies exist. However, the evidence presented by Jackson indicates that formal teacher training programs have little impact on subsequent student achievement. This disappointing result can be due to the programs not being well designed, as indicated by Jackson.

The challenge seems to be to create incentives such that (i) the training providers offer effective training programs and (ii) the teachers choose the most effective programs. How can this be achieved? The critical factor is the mechanism for choosing training programs. If the teachers make the choices, one should question what kind of training programs they will demand. Without any incentives in schools related to student achievement, why should the programs most effectively increasing student achievement be chosen? Without financial incentives within schools, it seems necessary that some other actors accountable for the student performance are involved in the choice of training programs, for example the school principal or the school owner. Then, the need for external evaluation of the training programs would also be evident.

This is not, of course, to say that teachers themselves are not interested in student achievement. It is just that from a society viewpoint, other objectives can get too high weights when the profession itself makes the decisions. The evidence for training at the workplace (mentoring, evalua-
tions and feedback from colleagues) is more promising. At least it points to the potential of integrating external training with feedback at the workplace. Overall, how to increase teaching skills is a field with little scientific knowledge and where more research would be highly welcome.

References

On the margin of success? Effects of expanding higher education for marginal students*

Björn Öckert**

Summary

This paper reviews the empirical evidence on the effects of marginal changes in (i) college tuition/grants, (ii) college proximity and (iii) admission standards, respectively, on college enrollment, graduation and labor market performance. It also provides evidence on the marginal return to potential policy changes in Sweden. The review shows that reducing college costs as well as expanding college regionally may have sizable effects on college attendance. The empirical results for Sweden suggest that students from weaker backgrounds – and those with the richest parents – have a higher return to college. It also shows that those who go to college due to college proximity may have relatively high returns. Further, the return to college is highest among top students, why attracting them to college may be worthwhile.

Keywords: Return to college, college enrollment, academic performance, earnings.

JEL classification numbers: I23, J3.

---

*I thank Per-Anders Edin, Peter Fredriksson and Torbjørn Hægeland for valuable comments.

**Institute for Evaluation of Labour Market and Education Policy, Uppsala University and UCLS, bjorn.ockert@ifau.uu.se.
The share of the population with a college degree (or higher) has increased in the OECD over the last few decades (Peracchi, 2006; OECD, 2010). At the same time, the return to human capital investments has remained remarkably stable, and has even increased, in most countries since the 1980’s, suggesting that the demand for skills has risen faster than the supply (Katz and Autor, 1999; Peracchi, 2006; Acemoglu and Autor, 2011). Still, some researchers are concerned that too many go to college, resulting in a (growing) mismatch between skills and jobs (see Leuven and Oosterbeek, 2011 for a critical review on the literature on “overeducation”). Clearly, the return to further expansions of higher education depends on the future demand and supply of skills on the labor market.

Given the macroeconomic conditions, the return to marginal expansions of higher education depends on which students respond to different policies. In general, the pay-off from attending college differs between individuals, and students self-select (or are selected) partly based on the expected return to their investments. If those with the highest return to human capital investments already go to college, individuals at the margin will have lower returns (see e.g. Carneiro et al., 2011; Murray, 2008). On the other hand, if individuals are credit constrained, or if they discount the future too much, the return for the marginal student may well be as high as for the average student (see e.g. Card, 1999, 2001). Policies targeted toward individuals at different margins may, thus, have very different returns.

This paper reviews the empirical evidence on the effects of marginal changes in (i) college tuition and student grants, (ii) college distance and (iii) admission standards, respectively, on college attendance, graduation and labor market success. It also provides evidence on the marginal return to potential policy changes in Sweden. The main focus of the paper is the economic return to college for students at different margins, but I also discuss how effective different policies may be in attracting new groups to college and if the marginal students are likely to complete their studies.

The literature review shows that lower tuition and generous grants programs may be effective in attracting students to colleges. The opening up of new colleges in areas where the supply of places has been scarce may also have a large effect on college enrollment. There is, however, less evidence on how different policies affect academic performance and
economic returns. The review also shows that most of the credible evidence is based on US data and that less is known about the effect of these policies in Europe.

The empirical analysis reveals that college enrollment and academic performance does not vary by family income in Sweden, once controlling for family background and different measures of ability. Individuals from poorer families have higher returns to college, as do those at the very top of the family income distribution. Individuals living close to a college at the age of 16 are more likely to enroll at college, but it is those who have to move to go to college that are more likely to complete their studies. Taken together, the return to college is highest for those living close to college and for those living far away. A higher upper secondary school GPA increases the probability of going to college. Individuals with a higher GPA are also more successful among those who enroll, and they have higher returns to college. Attracting high-ability students to college, for instance by lowering the admissions standards to popular programs, would therefore be worthwhile.

The rest of the paper is organized as follows. Section 2 outlines a simple human capital framework describing the college enrollment decision. It also discusses some potential market failures and the impact of policies trying to address these failures. Section 3 reviews the existing empirical evidence on the effects of different policies aimed at promoting college attendance, on enrollment, graduation and labor market performance. Section 4 describes the institutional setting, discusses some empirical considerations and describes the data used in the analysis. Section 5 presents the results and Section 6 concludes the paper.

1. A simple model of college choice

The human capital model (Becker, 1964) treats educational choices as investment decisions and individuals choose whether to go to college or not by comparing the (expected) benefits of college with the costs. The benefits may include the economic gains from the investment, the consumption value of college studies and possible transfers to the student (student grants); whereas the costs may be tuition, foregone earnings while studying, and effort. In a very simple formalization of the human
capital model, individual $i$ chooses to attend college ($D_i$) so as to maximize the following utility function:

$$U(Y_i, D_i) = Y_i (D_i) - c_i D_i = Y_{0i} - (Y_{1i} - Y_{0i}) D_i - c_i D_i,$$

(1)

where $Y_{0i}$ are the earnings when not going to college, $(Y_{1i} - Y_{0i})$ is the economic gain – or the earnings premium – of going to college and $c_i$ is the cost of college studies.\(^1\) In general, the gains and costs of going to college differ between individuals, and individuals self-select to college based on comparative advantages net of costs (as in the Roy, 1951 model). In particular, individuals go to college if the benefits exceed the costs:

$$Y_{1i} - Y_{0i} \geq c_i.$$

(2)

Thus, individuals choose not to go to college either because their return from doing so is low, or because their (perceived) cost of college studies is high (or both). Under perfect market conditions, the share of the population choosing to go to college corresponds to the socially optimal level of human capital investments. In reality, however, credit constraints may lead individuals to underinvest in college education, even though their return would have been high.\(^2\) In particular, individuals from poorer backgrounds either cannot borrow to cover their study costs, or they meet higher interest rates on the credit market than do those from better backgrounds. Without policy interventions, this may constrain the college choices for talented students from low-income families. Thus, there are both efficiency and equity arguments for addressing the problem of borrowing constraints.\(^3\)

Many countries have introduced policies to meet the problem of imperfect credit markets, either by setting caps on how much colleges are allowed to charge the students and/or by giving grants and offering loans to students. In some countries, the colleges themselves have set up their

---

\(^1\) Note that the consumption value of college and student grants can enter the model as negative costs.

\(^2\) Credit constraints would imply higher costs of college ($c_i$) for constrained individuals.

\(^3\) The term credit constraints need not only reflect problems to borrow for college studies, it can also reflect other barriers such as lack of information, risk aversion, social costs, etc., that make individuals underinvest in education.
own student grant systems. However, tuition restrictions and student grants may generate an excess demand for college. To restrict the supply, most colleges impose enrollment constraints. This generates the problem — or opportunity — of selecting which students to admit. Most colleges select students based on some measure of ability (test scores, grades, etc). A simple formalization of the college admission process is that individuals are admitted to college if they have merits exceeding a certain admission threshold:

\[ \gamma_1 \text{GPA}_i + \gamma_2 \text{SAT}_i \geq \gamma_0, \]  

where \( \text{GPA}_i \) is the upper secondary school grade point average and \( \text{SAT}_i \) is the scholastic aptitude test score. Individuals are admitted to college if their weighted sum of GPA and SAT exceeds the admission threshold \( (\gamma_0) \), where the weights \( (\gamma_1 \) and \( \gamma_2 \) ) depend on how colleges value different ability indicators in the admission process.

The aim of this paper is to review the evidence on the marginal returns to expanding college and provide some evidence from Sweden on how different policies may affect the economic gains of going to college. However, the marginal effects depend on the policy in mind. I will focus on three different policies that promote college attendance: (i) reduced tuition/increased student grants, (ii) reduced distance to college (expanding college regionally) and (iii) lowered admission standards (expanding college nationally). The first type of policy aims at increasing the demand for college, while the second type of policy increases both the supply of places and stimulates demand. The third policy entirely focuses on shifting the college supply. In the following, I will discuss which students may be affected by marginal changes in these policies.

---

4 Colleges may use student grants to generate excess demand that allows them to select which students to admit (Winston, 1999). They may also use need-based grants to price discriminate students by income (Epplle et al., 2003).

5 Colleges may care about student ability since it is likely to be an important input in the production of skills. Able students probably learn more (and faster) for a given amount of educational resources, and they may also affect the skill formation of their peers.

6 Colleges could, of course, also use other inputs in the college admissions process, e.g. letters of recommendation or interviews.

7 In the simple model sketched above, the return to college is given by the benefits less the costs. As is common in the literature on the returns to schooling, however, I will sometimes use the term “return” for the economic gains of expanding higher education, e.g. \( (Y_i - Y_0) \).
A general price reduction – either by reducing tuition or increasing general student grants – lowers the costs of going to college \((c_i)\) for everyone. This will induce individuals who would otherwise not have found it worthwhile to go to college. The marginal student affected by the policy, thus, either has modest returns from college studies and/or high costs. Note, however, that even general price reductions may be relatively more important for the college decisions of individuals who would otherwise have been credit constrained than for unconstrained individuals. Need-based (i.e. means-tested) financial aid policies are designed to induce individuals from poorer backgrounds – whose parents cannot pay for their studies – to go to college. These are individuals who would otherwise face high costs of college and, thus, who can have high returns. Unless the targeting is perfect, however, even such policies may affect the college decisions also for individuals who are not directly credit constrained, and whose return may be modest.

College education can be expanded both nationally and regionally. To attract groups that typically do not go to college, many countries choose to expand colleges in areas where the supply of places is scarce. A regional expansion of college education reduces the distance to college for those living in these areas, thereby potentially reducing the cost of going to college \((c_i)\). There are at least two reasons why college proximity may affect the costs of going to college. First, individuals living far from college may have less information on the return to college investments and may also face higher social costs of going to college. Second, individuals who live further away from college typically have to move (from their parents) to enroll which substantially increases the costs of living. Both these channels suggest that individuals living far away from a college are less likely to enroll and that expanding colleges regionally may substantially reduce individuals’ costs and, thus, attract individuals with relatively high returns to college.

A simpler (and at first sight cheaper) way of promoting college attendance would instead be to reduce the admission standards \((\gamma_i)\) – i.e. to expand the number of places at existing colleges. Such policies would affect individuals who find it worthwhile to go to college, but who have too weak formal merits to gain admission. The return to college for individuals at the margin of admission depends on how economic gains are related to the admission determinants (typically different measures of the
student’s ability). Since individuals’ abilities are also rewarded on the labor market, the opportunity costs of going to college are probably higher for top students. Thus, those with the highest abilities may not necessarily have the highest returns to college.

2. Short review of earlier evidence

This section provides a short literature review of the effects of the three policies I consider. Since the success of different policies does not only depend on to which extent they can attract students to college, but also if the students complete their studies and succeed on the labor market, I will summarize the evidence on how different policies affect all these margins. Only studies with a reasonably credible strategy for estimating causal effects are considered.

Ideally, I would have liked to focus the review on European studies, both since the empirical evidence provided in this paper is from Sweden and since the effect of different policies in the US has recently been reviewed (see e.g. the surveys by Kane, 2006 and Deming and Dynarski, 2010). As it turns out, however, there is little (credible) evidence from Europe, so I will also give short summaries of the US findings.

2.1 The effects of tuition and grants

The literature on the effects of college tuition and student grants on college attendance is extensive. Most of the studies are based on American data. Studies using credible identification strategies to identify the effects of tuition on attendance typically find that a reduction in costs (lower tuition) of USD 1 000 increases the probability of college enrollment by roughly 3-5 percentage points (see Kane, 2006; Deming and Dynarski, 2010).

The effect of student grants typically depends on the program design. Broadly speaking there are three types of programs; (1) general student grants, (2) need-based (targeted) grants and (3) merit-based student grants (scholarships). Quite surprisingly, highly targeted programs often turn out
to have a lower effect on enrollment than do more general programs.\textsuperscript{8} The more general student aid programs have estimated enrollment effects in the same ballpark as tuition, i.e., a general grant of USD 1 000 increases the probability of college enrollment by roughly 3-5 percentage points. The merit-based programs – which condition grants on school performance – seem to be at least as effective in attracting students as are the general grants; some studies suggest that they may even be more successful (Dynarski, 2000, 2004).

Skyt Nielsen et al. (2010) provide one of the few credible estimates of the effects of student aid on college enrollment for a European country, where tuition typically is lower (or free) and grant systems are more general (and sometimes also more generous). They estimate the effects of a student aid reform in Denmark, where a need-based student grant was replaced with a more generous and general student grant. In contrast to many other student aid reforms, the increase in student grants was thus concentrated among children from wealthier families.

Using a regression-kink design (together with the change induced by the reform), they find that a USD 1 000 increase in the annual grant increases college enrollment by 1.35 percentage points on average. They find some evidence suggesting that the enrollment response (for a given student aid increase) is higher for children from poorer backgrounds, but the difference between the groups is not statistically significant. The weaker enrollment response to the reform in Denmark, compared to some of the reforms in the US, may be due to the fact that the reform increased the student grant mainly for students with better family backgrounds and that the response for them is generally weaker.\textsuperscript{9}

Baumgartner and Steiner (2005) evaluate the effects of a student aid reform in Germany, where half of the student loan was replaced by a student grant. Using a difference-in-differences approach on a small sample, they find no significant effect of the reform on college enrollment. Baumgartner and Steiner (2006) evaluate the effects of a later student aid reform in Germany, where the amount of student aid for eligible students

---

\textsuperscript{8} One possible explanation for weaker effects for targeted programs is that the paperwork required to identify the neediest students is so demanding that this group is deterred from applying for grants (Deming and Dynarski, 2010).

\textsuperscript{9} Arendt (2012) examines the effects of another feature of the Danish student aid reform, where the grant increased relatively more for younger than for older students (independently of family income). However, this change did not significantly affect the grant received (i.e. no first stage) or any measures of study success.
increased by on average 10 percent. Analyzing the change in enrollment before and after the reform, for those who were eligible for the grants and those who were not (difference-in-differences), they find no significant effect of the reform on college attendance.

Steiner and Wrohlich (2012) exploit the non-linear features of the German student aid system, where the grant received tapers off at different threshold values of family income (depending on family size), to estimate the effect on college enrollment (regression-kink design). Estimating a competing risk model and controlling linearly for family income on a small sample, they find that an increase in student aid by USD 1 000 leads to an enrollment effect of about 1.1 percentage points. As a sensitivity check, they account for non-linear effects of both student aid and family income, which yield about the same results. However, since identification in the regression-kink design relies on the assumption that any direct effect of family income on college enrollment can adequately be accounted for, the rather restrictive controls for family income may lead to biased estimates.

A couple of US studies examine whether lower tuition/higher grants also affect educational attainment. Dynarski (2003) finds that the elimination of a social security student benefit program – aimed at helping children with deceased parents – reduced the enrollment by about 20 percentage points for eligible children. The probability of college completion fell by more than 15 percentage points when the program was eliminated. This corresponds to an enrollment effect of about 4 percentage points, and a graduation effect of about 3 percentage points, for an increase in student grants of USD 1 000.

Dynarski (2008) analyzes the effects of introducing two state-wide scholarship programs in the US on college completion. Students performing above a certain level in high school were eligible for the grant, while renewals were based on college performance. Using a difference-in-differences approach, she finds that the programs affected college enrollment by 1.6 percent and college completion by almost 3 percent. The larger effect on completion than on enrollment is probably driven by the fact that renewal of grants was conditioned on college performance, and Dynarski (2008) argues that the increased degree probability among those

10 Even though the grant was, in principle, general, the eligible group was predominantly from weaker backgrounds (not the least with respect to family income).
who would have gone to college anyway is at least 1.4 percentage points.\footnote{Häkkinen and Uusitalo (2003) evaluate the effects of a student grant reform in Finland in 1992 where student loans were replaced by student grants. They find small or no effects of the reform on times-to-degree. Heineck et al. (2006) study the effects of introducing tuition for long-term studies at the University of Kastanze in Germany. They find large increases in the probability of taking a degree for those enrolling after the reform. Avdic and Gartell (2011) estimate the effects of a student aid reform in Sweden, where the grants increased while loan repayment became stricter. They find small or no effects of the reform on the number of credits taken per year. A problem with these studies, which the authors also typically acknowledge, is that the reforms affected all students simultaneously, and that no natural comparison group is available. Thus, the causal interpretation of the results relies on quite strong assumptions.}

For Europe, Garibaldi et al. (2012) estimate the effect of college tuition on the probability of graduating on time at a large private Italian university (Bocconi). At Bocconi, students are assigned to one of 12 tuition levels based on their family income, and students below and above each threshold face different tuitions to enroll even though they have a similar family income. Garibaldi et al. (2012) exploit this source of variation in college tuition in a regression-discontinuity design to estimate the effect of tuition on the probability of completing college for students in the last regular (4th) year in a program. They find that increasing college tuition with USD 1 000 boosts the probability of completing college in time by almost 10 percentage points. A potential problem with the study, however, is that the analysis is based on enrolled fourth-year students only. Since college tuition is expected to affect both college enrollment and course progression – with the relatively more able students being more likely to accept a higher level of tuition upon enrollment – the estimates may be biased upward.

Less is known about the effect of tuition/grants on the return to college for the marginal student. Kane and Rouse (1993) exploit cross-sectional variation in college tuition as an instrument for the number of college credits taken. Conditional on a number of controls for family background and ability, they found that the return to college for those who were induced to go to college due to lower college tuition is about 11 percent. This should be compared to the descriptive relation between years of college and earnings of about 8 percent.
2.2 The effects of college proximity

Starting with the influential studies by Kane and Rouse (1993) and Card (1995), distance to college has been used as an instrument for college attendance in many studies attempting to estimate the return to college. Carneiro and Heckman (2002), Cameron and Taber (2004) and Frenette (2009) show that distance to college is, however, correlated with different measures of ability, seriously questioning the use of college proximity as an instrument for college attendance in cross-sectional data. Therefore, a few recent papers exploit college openings and expansions as a source of exogenous variation, where it is possible to account for both time and regional fixed effects (difference-in-differences).

Currie and Moretti (2003) exploit college openings in the US during the second half of the twentieth century as a source of exogenous variation in mothers’ educational attainment, to estimate the effect of maternal education on infant health. They compare the change in educational attainment for mothers living in counties where a new college was established, to the change in education for mothers growing up in other areas, and find that opening up a two-year or a four-year higher education institution affected the probability of going to college. The effect of opening up a new four-year college (per 1000 18-22 year-olds in the county) increased the probability of attending a four-year college by almost 20 percentage points, while the opening up of a two-year college increased college attendance at such an institution by roughly 3 percentage points.

Frenette (2009) applies a similar approach as Currie and Moretti (2003) to estimate the effect of university openings on college and university attendance, as well as on employment and earnings in Canada. He finds that opening up a new university increases the probability of attending a university among people living in the area by 6.4 percentage points. However, this comes at the expense of a reduction in the probability of attending a college by 5 percentage points, implying that the net increase in post-secondary school attendance is just 1.3 percentage points. The response to opening up a new local university is strongest among individ-

---

12 This falsification test is not as strong as it may seem. Individuals living far from college who, thus, know that the costs of going to college will be high have fewer incentives to perform well at school. This is likely to also affect the performance on different ability tests, since they have been found to be strongly affected by schooling (see e.g. Hansen et al., 2003). Thus, the lower test results for those living far away from college may be a treatment effect of college distance rather than a selection effect.
uals with a lower family income, where the net increase in post-secondary school attainment is more than 5 percentage points. The effects on employment and earnings are somewhat unclear, which is probably explained by the fact that the labor market outcomes are measured as 20-24 year olds.

Holzer (2009) estimates the effects of a regional expansion of college education in Sweden during the 1990’s. The results show that the change in the probability of going to college for individuals living in the expanding areas increased by 5.6 percentage points more as compared to individuals in other areas. Even though the increase in the number of study places was larger at newer colleges, she does not find that the probability of enrolling at an old institution changed more for individuals living in the expanding areas than elsewhere.

Oppedisano (2011) evaluates the effects of university openings in Italy on college attendance and academic performance. During the 1990’s, the number of public campuses increased from 69 to 80, with more college openings in areas where there had previously been no higher education institutions. Studying upper secondary school leavers in 1995 and 1998, who either entered college directly or not at all, she finds that opening up a new college (per 1000 secondary school leavers) increased college attendance by 8 percentage points. The response was strongest among medium-ability students and those from middle-class backgrounds.

In sum, distance to college seems to affect attendance by reducing the costs of going to college. Opening up colleges in areas where there had previously been no nearby institution of higher education may have substantial effects on college attendance. This has been found using historical data (Currie and Moretti, 1993), as well as for more recent expansions (Holzer, 2009; Oppedisano, 2011). The type of institution of higher education that is opened up also seems to matter. The results in Frenette (2009) indicate that expanding the number of universities may shift away students from colleges. Possibly, the opposite effect can be expected from opening up a new college, although this is not found for Sweden (Holzer, 2009). The interest in studying the effects of college proximity originally arose as a way of identifying the return to college (Card, 1995; Kane and Rouse, 1993). Recent studies exploiting more credible identifying strate-
gies are, however, silent about the effects of regional college expansions on academic success and prime age labor market performance.\textsuperscript{13}

\section*{2.3 The effects of college admission standards}

It is well-known that brighter students – measured as their upper secondary school GPA, SAT or ACT – are more likely to go to college.\textsuperscript{14} This is driven both by a higher probability that these students apply for college and by the use of different measures of past school performance in the college admissions. There is also a mass of evidence showing that upper secondary school grades and SAT predict college success (Camara and Echternacht, 2000).

Even though much is known about the selection to higher education and the predictive value of different admission instruments, less is known about the return to college for students of different ability. There is a number of reasons why the returns to college may vary by student ability. Able students learn faster and, thus, face lower costs of going to college. The return to skills may also be higher in college jobs than in upper secondary school jobs. On the other hand, able students may face better outside options, which would increase the cost of going to college. Another reason why high-ability college graduates may earn more than less able graduates is that they are more likely to go to high-quality institutions and read highly demanded college majors. Thus, a higher return to college by student ability may just reflect differences in college quality and type of program, rather than a higher pay-off to attending college for able individuals in general.

Altonji and Dunn (1996) estimate the return to schooling by IQ test scores in the US, both in a cross section and with family fixed effects. The results indicate that the return to schooling increases by ability, but

\begin{footnotesize}
\textsuperscript{13} Oppedisano (2011) presents some evidence on the effect of college openings in Italy on academic performance. Correcting for college selection using a Heckman selection model, the study reports that the drop-out rates decreased while the academic performance fell for those who enrolled at college. However, the quality of the instruments used (the number of available degrees and the unemployment rate in the region) is unclear and the results should probably be interpreted with caution.

\textsuperscript{14} A large literature uses measures of student ability to account for the possible endogeneity of schooling in earnings regressions (see e.g. Griliches, 1977; Blackburn and Newmark, 1995; Murnane et al., 1995), and they show that more able individuals acquire more schooling (which may bias the estimated returns to schooling upwards).
\end{footnotesize}
that this mainly reflects differences in the return to schooling by family background. Using a structural framework, Carneiro et al. (2003) find that the return to schooling increases by ability, and that the return to schooling could even be negative for low-ability students.

Nordin (2007) estimates heterogeneous returns to schooling by cognitive ability in Sweden. Making use of data from the military draft – covering more or less all Swedish men – he estimates the return to schooling for individuals with different ability levels. He finds that the return to education increases with ability, and that this cannot be explained by differences in the return by family background (as in Altonji and Dunn, 1996). The results do not change when controlling for broad categories of type of education. However, it is quite possible that a more detailed control for type of education and type of institution may affect the results. I return to this issue in the next section.

There is some direct evidence on the effect of lowering admission standards as a way of increasing college attendance. Maurin and McNally (2008) analyze the effects of lowering grading standards at selective stages of the university system in France, following the student revolts in 1968. In that year, exams became less stringent than usual, in an attempt not to harm the students who had struggled for a better university. Since passing the exams automatically gives the students the right to advance within the higher educational system, the lowered grading standards are comparable to lowered admissions standards.

Maurin and McNally (2008) study the effect of lower grading standards for the cohort that was most likely to be studying at the key stage of selection to higher education in 1968, compared to earlier and later cohorts that were less likely to have been affected. They find that men in the cohort affected by the lowered standards were almost 3 percentage points more likely to obtain at least a two-year university diploma than men in the surrounding cohorts. The effect was concentrated among men from middle-class families.

Men in the affected cohort also received significantly higher wages than men in the other cohorts. The return to college is estimated at about 14 percent, which should be compared to a descriptive return of about 10 percent. Thus, lowering the admission standards may have substantial returns for individuals at the margin of admission. It should be noted, however, that the share of a cohort of men who studied at a university in
France in the late 1960’s was about 20 percent. The effects of lowering the admission standards to college today may be very different than fifty years ago.

Öckert (2010) estimates the return to college by exploiting discontinuities and randomness in college admissions in Sweden in the early 1980’s. In Sweden, college applicants to oversubscribed programs and colleges are ranked on the basis of their formal merit, and admission is given to applicants ranked higher than some threshold value. Comparing applicants at the margin of admission – the students affected by small changes in the admission standards – Öckert (2010) finds that the return to college for this group is close to zero. This should, however, be compared to a very low descriptive return to college education in Sweden.

In all, there is a large number of studies documenting that student ability is an important determinant of college attendance. Different performance indicators – such as upper secondary school GPA and the SAT – have also been shown to predict college success. The evidence on the economic return to college for students of different ability is somewhat mixed, but the result in Nordin (2007) suggests that the return to schooling increases by ability. Whether this is driven by the fact that more able individuals are more likely to attend programs and colleges with a higher return or that the return to college of the same type increases by ability is unclear.

It is also worth noting that there is less direct evidence on the effects of lowering the admission standards to college. Maurin and McNally (2008) estimate a high return to college for those affected by the lowered grading standards following the student revolts in France in 1968. Since the share of a cohort entering higher education at that time was quite low, it is quite possible that the effect of lowering grading standards today would be different. The results in Öckert (2010) suggest that the return for the students at the margin of admission to college in Sweden is low, although this may very well be due to a low descriptive return to college on average.
3. The effects of potential policies to promote college attendance in Sweden

The previous section illustrated that the evidence on the effects of different policies aiming at stimulating college attendance is not evenly spread. We know a great deal about the effects of different policies on college enrollment, in particular about the effects of tuition and grants on the probability of going to college. A great deal is also known about the importance of student ability as a determinant of college choice. However, it is important to note that most of the evidence on the success of different college promoting policies comes from the US. Less is known about the effects of tuition and grants in the European context. In general, there is also less evidence on the effects of college stimulating policies on academic and labor market performance than on college enrollment.

Another shortcoming of the existing evidence is that it evaluates the effects of different policies in different time periods and different regions. As such, it is hard to compare the estimates from different studies, and decide which policies are the most efficient in attracting students to college, how likely it is that the new students complete their studies and what the returns of going to college are for the marginal student. This section offers an analysis of how expanding college education along different margins in Sweden may affect college enrollment, graduation and economic returns. The analysis is based on the same cohorts during the same time period for the same country and institutional context. This will highlight the advantages and disadvantages of different policies to promote college attendance.

3.1 College education in Sweden

Sweden has nine years of compulsory schooling, and children typically start school at the age of seven. In the 1980’s, upper secondary schools offered short (two-year) and long (three- or four-year) programs. The shorter programs were vocational, while the longer programs were theoretical and prepared students for further studies. In 1991, the vocational upper secondary school programs were extended from two to three years. After the completion of upper secondary school, students can choose to go to college.
In the 1980’s, there were one private and nine public universities as well as 17 public university colleges in Sweden. In the 1990’s, the number of study places expanded dramatically, in particular in regions that earlier had no or little supply of college places. There is no tuition at Swedish colleges and generous student grants and loans are available for all students. College students can choose between programs and (single-subject) courses. Programs vary in length (typically 3-4 years) and students are awarded a degree on successful completion of a full program. The courses are usually shorter, but can be combined to achieve a full degree. All programs and courses are subject to enrollment constraints.

To be eligible for college, students have to complete upper secondary school. College admissions take place twice a year, with roughly three quarters of the places being allocated in the fall semester. Typically, competition is fierce and at least half of the applicants are rejected. College applicants for oversubscribed programs and colleges are ranked with respect to their upper secondary school grade point average (GPA) and Scholastic Aptitude Test (SAT), and admission is given to applicants ranked higher than some threshold value (established by the number of places available). In the period studied, however, only few places were allocated on the basis of the SAT.

3.2 Empirical considerations

Ideally, a number of policies to promote college attendance would have been introduced for the same cohorts (and in a way that makes evaluation possible). It would then be possible to assess the relative effectiveness of different policies when it comes to attracting new college students, the probability that they complete their studies, and their performance on the labor market. In Sweden, there has been a number of policy changes that could, in principle, be evaluated (the regional expansion of college education during the 1990’s being one). Still, many policy instruments have remained unchanged throughout the period. Tuition is still free at all colleges. The student grant system did change in the year 2000. However,

---

15 The main difference between universities and university colleges is that university colleges generally do not provide graduate education. I will use the term college for both types of institutions throughout the paper.
the reform was introduced simultaneously for all students, which makes evaluation complicated (and the follow-up period too short).

In the absence of direct evidence on the effects of different policy changes, the approach taken here is to describe how enrollment, graduation and labor market returns vary along margins that different policies are likely to affect. Even though it is not clear who the marginal students may be, different policies typically aim at attracting different types of students. Thus, by studying the design of different policies, it is possible to decide what the relevant margins may be. This approach has the advantage that it can illustrate the effects of different policies along the full support (all values) of the underlying margins affected. However, it has the obvious shortcoming that it relies on the ability to identify the “right” marginal students.

Policies that change the direct costs of college by reducing tuition or increasing grants attempt to weaken the link between family income and college attendance. Thus, to assess the effects of changes in the grant system, I study how enrollment, completion and earnings relate to family income. Clearly, the effect of giving general subsidies to families (which is what the treatment is most closely related to) may be different than giving targeted subsidies to their children if they go to college. Nevertheless, studying heterogeneous effects by family income will still give some insights into the scope for such policies to attract students to college and what the returns may be.\(^\text{16}\)

Policies that aim at attracting students to college by expanding the number of places in regions where the supply has been restricted effectively reduce the distance to college for the individuals living in the expanding areas. College proximity may affect attendance by reducing the costs of going to college, either since it determines whether it is possible to live at home with one’s parents while studying, or since it proxies for the social costs of going to college. Individuals living far away from a college may be less likely to be exposed to highly educated individuals, and fewer of their peers proceed to college. Thus, to assess the effects of

\(^{16}\) Moreover, the relation between family income and college access has been studied in many other countries to assess the existence of credit constraints. Clearly, few Swedish students should literally be credit constrained, given free tuition and the generous student grants. The extent to which college attendance differs by family background can thus be informative about the lower bound of the potential for tuition and student grant policies to counteract the socio-economic bias at colleges.
regional college expansions, I study heterogeneous effects by college distance.

Policies that lower the admissions standards to accommodate more students will affect applicants with slightly worse merits to oversubscribed programs and colleges. In Sweden, college admissions are mainly based on upper secondary school GPA and SAT. Since few students took the SAT in the period studied, I will focus on the relation between GPA and college enrollment, graduation and labor market success. It is important to note, however, that upper secondary school GPA does not only affect the probability of gaining admission to college in general, but also the chances of gaining admission to popular programs and colleges. Since the return to colleges and programs that are in high demand may be high regardless of the quality of their students – either due to high college quality or due to high labor demand – part of the relation between GPA and economics returns may be driven by differences in college quality and the return to different programs. Thus, to isolate the effect of admitting weaker students, it is necessary to condition on the program and college attended.

Armed with the underlying variables representing the margins of different policies, the objective is to estimate how college enrollment, academic success and labor market performance vary along these dimensions. Ideally, I would like to know how the outcomes vary by the underlying variables for those who go to college (treated on the treated). This parameter would be informative about the consequences of small policy changes to attract more students, under the assumption that the effects of college (along the relevant margins) do not differ between those who are induced to go to college and those who would have gone to college anyway.

There are two major empirical problems when estimating how the return to college varies by the underlying variables. First, the underlying variables – like family income and college proximity – are not randomly allocated in the population. Thus, there may be other differences between individuals with, e.g., richer or poorer parents, which make a causal interpretation of the correlation between, e.g., family income and different outcomes problematic. Second, individuals are not randomly allocated to college, and comparing individuals who go to college with those who
have an upper secondary school degree may be very misleading about the causal effects of going to college.

The approach taken here is to control for a set of observed characteristics – most importantly different measures of the individuals’ cognitive and non-cognitive skills – that are strongly related to college attendance, academic success and labor market performance. The hope is that controlling for variables that simultaneously determine college enrollment and different outcomes, i.e. assuming selection-on-observables, shall free the estimates from bias. In reality, however, a number of important determinants of the underlying variables and college attendance are probably missing from the data at hand. Thus, a somewhat more realistic assumption is that controlling for observed covariates should reduce any bias in the estimates.

The identifying assumption required for giving the conditional difference between college enrollees and upper secondary school leavers with different family background or distance to college a causal interpretation depends on the outcome of interest. The difference in enrollment for individuals from, e.g., richer or poorer homes can be interpreted causally under the assumption that there are no other relevant differences between individuals with different backgrounds, once the observed covariates have been controlled for. To give the relation between, e.g., family income and academic performance a causal interpretation, the same type of assumption has to be made. The identifying assumption for interpreting differences in economic outcomes between college students and upper secondary school leavers with a given, e.g., family background, is stronger, since it also requires that the mean earnings of secondary school leavers is an unbiased estimate of the mean earnings of college enrollees had they not enrolled (the counterfactual outcome) conditional on the observed covariates.

The reason for studying the relation between upper secondary school GPA and the outcomes is different from the analysis of family income and college distance. For the latter underlying variables, the parameter of interest is the causal relation between the underlying variables and college enrollment, graduation and earnings, while the parameter of interest for upper secondary school GPA is the descriptive relation with the outcomes. This is because lowering the admission standards implies enrolling students with worse formal merits. The performance of the marginal
students affected by such a policy does not only depend on their lower grades per se but also on everything else that correlates with lower grades. The identifying assumptions for analyzing the effects of upper secondary school GPA are therefore milder than for the other underlying variables, since they do not require that GPA is randomly allocated in the population.

3.3. Estimation

I will use regression analysis (OLS) to adjust for observed differences between individuals with different values of the underlying variables as well as for differences between college enrollees and secondary school leavers. The effect of different underlying variables on college enrollment can be described by the following model:

\[ D_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 X_{1i} + \alpha_3 X_{2i} + \varepsilon_i, \]

where \( D_i \) is college enrollment, \( Z_i \) is the underlying variable of interest (family income, distance to college or high school GPA, respectively), \( X_{1i} \) is a vector of basic control variables (gender, year of birth, labor market region at the age of 16 and upper secondary school program attended), \( X_{2i} \) is a vector of additional covariates (e.g. cognitive skills and personality traits, and parental education), and \( \varepsilon_i \) is an error term. To allow the effects to vary more freely by different values of the underlying variable, I will estimate separate effects for each decile of the underlying variable. Under the selection-on-observables assumption, \( \alpha_i \) is the average causal effect of the underlying variable on college enrollment. When upper secondary school GPA is the variable of interest, I restrict the conditioning set to the baseline covariates, \( X_{1i} \), only.

The completion probability for individuals enrolled in program × college can be expressed as:

\[ g(D_i) = \frac{1}{1 + e^{-D_i}} \]

17 In other words, when analyzing how different outcomes vary by upper secondary school performance, it is mainly the selection effect that is of interest. This is very different from the causal effect of lower grades, e.g. by randomly assigning GPA to upper secondary school students.
\[ Y_{ig} = \beta_0 + \beta_1 Z_i + \beta_2'X_{1i} + \beta_3'X_{2i} + \lambda_g + u_{ig}, \]

where \( \lambda_g \) is a program by college fixed effect. Under the assumption that individuals with different values on the underlying variables are comparable conditional on the observed covariates, \( \beta_i \) is the causal effect of the underlying variable on college completion for those who go to college. Again, when estimating the descriptive relation between upper secondary school GPA and academic performance, I use the smaller conditioning set, \( X_{1i} \).

The earnings effects for different underlying variables are estimated separately for college enrollees and upper secondary school leavers. This effectively means that the effects of all variables are allowed to differ by college enrollment status. Estimating the effects separately by enrollment status also makes it possible to control for the program and college attended for those who go to college (which is not observed for those who do not enroll). I will estimate the following earnings equation:

\[ Y_{ig} = \pi_0 + \pi_1 Z_i + \pi_2'X_{1i} + \pi_3'X_{2i} + \lambda_g + \eta_{ig}, \]

where conditioning on the program by college fixed effect is removed when estimating the earnings effects of the underlying variables for upper secondary school leavers.

Under the selection-on-observables assumption, the earnings for high school leavers constitute an unbiased estimate of the relationship between earnings and, e.g., family income, for college enrollees had they not enrolled (the counterfactual outcome). Estimating the counterfactual relation between high school GPA and earnings for college enrollees is trickier, since the objective is to estimate the descriptive rather than the causal relationship. Thus, controlling for all observed covariates would probably understate the descriptive relationship. At the same time, college enrollees and secondary school leavers may differ in many important ways and, therefore, the descriptive relationship between earnings and GPA for those who do not enroll is probably a biased estimate of the counterfactual relation between earnings and GPA for college enrollees.

However, it is possible to estimate the descriptive counterfactual relationship between GPA and earnings for college enrollees using data for secondary school leavers, and still account for observed differences be-
tween college enrollees and secondary school leavers. I do this in two steps. First, I use the estimated coefficients from a regression of earnings for secondary school leavers conditional on the full set of covariates, to predict the non-enrollee earnings in the full population (i.e. also for those who enroll at college):

$$\hat{Y}_i = \hat{\pi}_0 + \hat{\pi}_1Z_i + \hat{\pi}_2X_{1t} + \hat{\pi}_3X_{2t}. \quad (7)$$

Second, I use the predicted earnings for college enrollees as the outcome, to estimate the descriptive relation between GPA and (counterfactual) earnings for them (conditional only on the smaller set of covariates;)

$$\hat{Y}_i = \gamma_0 + \gamma_1Z_i + \gamma_2X_{1t} + v_i. \quad (8)$$

Under the selection-on-observables assumption, this will be an unbiased estimate of the descriptive relationship between GPA and earnings for college enrollees, had they not enrolled. I use this two-step procedure to estimate the counterfactual forgone earnings and mid-age earnings for college enrollees.

### 3.4. Data

The data come from population-wide registers maintained by Statistics Sweden and the Military Archives of Sweden. The analysis is based on all male upper secondary school leavers born 1967-1976. It includes information on the type of upper secondary school program attended and the upper secondary school grade point average (GPA) achieved. Information on programs/courses and the college attended has been added to these data. College credits are available for cohorts born 1974-1976. All men have been linked to their biological parents by means of their unique personal identifiers, and a number of background characteristics have been added. Data on annual earnings in 1985-2009 come from the

---

18 The analysis is restricted to men, since I only have information on cognitive skills and personality traits for them.

19 During this period, roughly 80 percent of each cohort of men completed upper secondary school.
Income Tax Register, and are based on income statements made by employers.

To evaluate the success of policies aiming at stimulating college attendance, I study the effect on enrollment, graduation and labor market performance. An individual is regarded as having enrolled at college if he has ever been registered for a course or a program at any Swedish college up to the year 2009. College completion is defined as having obtained at least a Bachelor’s degree (or having credits enough for a degree) within five years after enrollment (and is available for the cohorts born 1974-1976). The economic returns are measured as annual earnings in 2009, i.e. when the individuals are 33-42 years of age. Earnings are therefore measured at an age when they correlate highly with lifetime income (Böhlmark and Lindquist, 2006). To measure forgone earnings while studying, I use the average annual earnings at ages 20-25 for individuals who have not (yet) enrolled at college.

The three key variables of interest are family income, college proximity, and upper secondary school GPA. Family income is the sum of the mother’s and father’s average prime age (35-45 years) real annual earnings. The measure of family income is therefore to be seen more as a permanent income measure. College proximity is based on the neighborhoods where the individuals lived when they were sixteen years old, i.e. long before some of them entered college. Each neighborhood is coupled with information on the exact location of the neighborhood center (i.e. the geographical coordinates). Together with information on the location of all institutions of higher education, I calculate distances from each neighborhood to each college (using Pythagoras’ theorem). Then, I assign the distance to the closest college to individuals living in each neighborhood. The GPA is just the raw mean of the grades in all subjects in upper secondary school. To make the estimate for different underlying variables comparable, all underlying variables have been percentile ranked. This also facilitates comparisons with other studies using this metric.

To account for differences between individuals with different values of the underlying variables, as well as those who go to college and those

---

20 In an attempt to generate a measure of earnings potential that does not reflect labor supply decisions too much, I use the average of the three highest annual earnings for ages 34-45 years.

21 I use the Small Area Market Statistics (SAMS) to define neighborhoods. The SAMS areas are stable homogenous geographical regions with about 1000 cohabitants on average.
who do not, it is crucial to control for important determinants of family income, college proximity and college enrollment. The baseline controls are a number of fixed effects for birth cohort, the type of upper secondary school program attended and the labor market area of residence at the age of 16. The additional controls include immigrant status, dummy variables for the mother’s and father’s highest educational attainment, the mother’s and father’s age at the birth of the first child and an indicator for whether the parents have separated before the child was 16 years old. Thus, the family income estimates will be identified from earnings variation independent of educational attainment.

I have also constructed a measure reflecting the potential demand for places at the closest college, defined as the number of individuals in the cohort living closest to the same college when they were 16. By the same token, I have generated a measure of the supply of places at the closest college by summing the number of individuals in a cohort entering the nearest college (from all over the country) at ages 19-24 years. Additionally, I have information on the unemployment rate in the resident municipality at the age of 16, which reflects both the opportunity costs of going to college and the demand for slots. These variables should hold the potential demand and supply of places at the closest college constant, purging the college proximity estimates from any such correlation.

What is unique with the data set is the population-wide information on cognitive and non-cognitive abilities measured at the military draft. The enlistment consists of a series of physical, psychological and intellectual tests and evaluations. The evaluation of cognitive ability includes tests of logical, verbal, and spatial abilities, as well as a test of the conscript’s technical comprehension. The tests have been found to be a good measure of general intelligence (Carlstedt, 2000). I use the percentile ranked sum of raw scores by birth cohort as a measure of cognitive skills.

---

22 Öckert (2011) shows that the competition for college places is higher in economic downturns than in upturns, which leads to a more favorable student composition at the colleges in bad times than in good times.

23 All Swedish men are obliged by law to attend the draft. During the 1980’s and the early 1990’s, more than 90 percent of the men went through the draft, with only those with mental or physical handicaps being exempted.
### Table 1. Descriptive statistics

<table>
<thead>
<tr>
<th></th>
<th>Did not attend college</th>
<th>Attended college</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome variables:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College enrollment, up to 2009</td>
<td>0.00 (0.00)</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>Bachelor’s degree, within 5 years</td>
<td>0.00 (0.00)</td>
<td>0.60 (0.49)</td>
</tr>
<tr>
<td>Earnings, 2009 (SEK)</td>
<td>320,725 (169,400)</td>
<td>421,898 (297,045)</td>
</tr>
<tr>
<td><strong>Variables of interest:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family income (SEK)</td>
<td>425,860 (157,909)</td>
<td>511,697 (210,745)</td>
</tr>
<tr>
<td>Distance to college, at age 16 (km)</td>
<td>37.03 (35.11)</td>
<td>31.93 (33.57)</td>
</tr>
<tr>
<td>Upper secondary school GPA, at ages 18-19</td>
<td>40.71 (26.69)</td>
<td>62.94 (26.72)</td>
</tr>
<tr>
<td><strong>Control variables:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive ability, at age 18</td>
<td>38.12 (25.66)</td>
<td>66.55 (24.63)</td>
</tr>
<tr>
<td>Non-cognitive ability, at age 18</td>
<td>45.27 (28.17)</td>
<td>56.59 (28.29)</td>
</tr>
<tr>
<td>N</td>
<td>230,501</td>
<td>165,439</td>
</tr>
</tbody>
</table>

Source: Own calculations based on data from Statistics Sweden and the Military Archives of Sweden.

Notes: Standard deviations are in parentheses. Information on college graduation is restricted to the 1974-1976 birth cohorts. Family income is the sum of the mother’s and father’s average prime age (35-45 years) real annual earnings. The upper secondary school grades, cognitive abilities and non-cognitive abilities have been percentile ranked by birth cohort.

The measure of non-cognitive abilities is based on a psychological evaluation aimed at determining the conscripts’ capacity to fulfill the requirements of military duty and armed combat. Central to this are the abilities to cope with stress and contribute to group cohesion. A conscript is given a high score if considered emotionally stable, persistent, socially outgoing, willing to assume responsibility, and able to take initiatives. The evaluation is performed by a certified psychologist who conducts a structured interview with the conscript. The interview follows a specific, and secret, manual that states topics to discuss and also how to grade different answers. The grades are given on four different sub-scales rang-
ing from 1 to 5. I use the percentile ranked subtests by birth cohort as measures of non-cognitive abilities.24

Table 1 reports descriptive statistics for the central variables in the analysis, by college enrollment status. Roughly 40 percent of the men born in 1967-1976 have at some point been registered at a college. About 60 percent of those who enroll complete a full degree within five years. This figure is quite low, but can partly be explained by the fact that almost 40 percent start their college studies in a single-subject course, which does not always lead to a full degree. Those who enroll at college have more than 30 percent higher earnings in their mid-labor market carrier. College enrollees generally also have a better family background (higher family income, more well-educated parents, are less likely to have separated parents) than upper secondary school leavers. Those who go to college lived closer to a college at the age of 16, indicating that college distance may affect enrollment. College entrants are also strongly selected on upper secondary school grades, which is unsurprising since grades are the main selection criteria in college admissions. College enrollees are also positively selected by cognitive and non-cognitive skills, with a stronger selection on cognitive than on non-cognitive skills.

4. Results

The estimated effects of the underlying variables on college enrollment, graduation and labor market success will be presented graphically. All figures have the same basic structure, where the point estimates for each decile of the underlying variables are plotted. To give an idea of the levels of the outcomes variables, I normalize the graphs so that the 50th percentile is set to the average value of the outcome in the population. This means that the average of the treatment effects at all deciles is the average treatment effect for college enrollees.25

24 The measures of cognitive and non-cognitive skills have been shown to be strong predictors of both school and labor market performance (Lindqvist and Vestman, 2011; Grönqvist et al., 2010)

25 To obtain a measure of the average counterfactual earnings for college enrollees had they not enrolled (the counterfactual outcome), I take the observed earnings for college enrollees and subtract the average earnings difference between college enrollees and upper secondary school leavers (obtained from a regression where earnings are regressed on a college attendance indicator and all other observed covariates).
As pointed out in the section on empirical considerations, the credibility of the empirical strategy crucially depends on how well controlling for observed covariates – most importantly individuals’ cognitive skills and personality traits – can account for differences between individuals with different values of the underlying variables and those who go to college and those who do not. The identifying assumption for giving the estimates a causal interpretation cannot be formally tested. However, it is possible to assess the credibility of the strategy by studying if there are any important remaining differences between individuals with different e.g. family incomes who go to college, and those who do not, after controlling for the full set of covariates.

For individuals born 1969-1976, I have information on the compulsory school GPA. Therefore, I can test whether individuals who go to college have a different compulsory school GPA than those who do not, once controlling for the other observed covariates. I can also study if the difference between college enrollees and upper secondary school leavers differs along the underlying variables. As it turns out, however, when controlling for the observed covariates, individuals who go to college only have 0.5 percentile ranks higher compulsory school GPA than secondary school leavers on average. Even though the difference is statistically significant, it still suggests that there are not any large systematic differences between the groups conditional on the observed covariates.

What is more important for this paper, however, is that the difference between college students and upper secondary school leavers at different values of the underlying variables does not vary too much, and this does seem to be the case. The difference in compulsory school GPA between college enrollees and upper secondary school leavers at the median of the family income distribution is less than 0.7 percentile points higher than the corresponding difference at the bottom of the distribution, and less than 0.3 percentile points lower than those at the top.

For distance to college, the difference in compulsory school GPA between those who enroll at college and those who do not does not vary to any larger extent between the 1st and 5th deciles, while the grade difference between college goers and upper secondary school leavers at the median of the distance distribution is about 1 percentile point higher than for those at the top of the distance distribution. The differences in compulsory school GPA between college students and upper secondary
school leavers are somewhat larger at different points of the secondary school GPA distribution, where the difference for those at the median is at most 2.5 percentile ranks higher than the difference for those at the bottom or the top of the distribution.

The large sample size typically makes the compulsory school GPA differences between college students and high school leavers at different points of the distributions of the underlying variables statistically significant, but the magnitudes are still small. This suggests that the observed covariates do a relatively good job in accounting for much of the selection to college as well as differences along the values of underlying variables, and that the remaining bias in the estimates should probably be minor.

4.1 College enrollment

Figures 1a-1c show how college enrollment varies by family income, distance to college and GPA, respectively. All models control for the baseline covariates (dummy variables for birth cohort, upper secondary school program and labor market region of residence at the age of 16). The effects of the underlying treatment variables are presented with and without the full set of covariates. Since it is the descriptive relation that is of interest for upper secondary school GPA, this graph only shows the estimates with the basic controls.

The solid line in Figure 1a shows that there is a strong association between family income and college attendance. Thus, even in Sweden where tuition is free and there are generous grants for all, children from richer backgrounds are significantly more likely to go to college than children from poorer backgrounds. The difference in college enrollment between children at the top and bottom of the family income distribution is about 11 percentage points, and highly significant. However, the enrollment probability rises steeply only above the median, with an enrollment difference of about 10 percentage points.\footnote{The estimates are based on upper secondary school leavers and condition on the type of upper secondary school program attended. If I instead use the full sample (also those who have not completed upper secondary school), the enrollment difference by family background is much higher. The difference in college attendance at the top and the bottom of the family income distribution is then 40 percentage points. If I instead compare the top and bottom income quartile, the enrollment difference is 30 percentage points. Belley et al. (2011) report that the difference in post-secondary school enrollment between the highest and lowest family income quartile}
Figure 1a. Enrollment probability by family income

Source: Own calculations based on data from Statistics Sweden and the Military Archives of Sweden.

Notes: The figure shows college enrollment by family income decile, obtained from a linear probability model where an indicator for having enrolled at college is regressed on a smaller set of covariates (solid line) and a larger set of covariates (dashed line), respectively. The smaller conditioning set includes dummy variables for year of birth, local labor market of residence at the age of 16, and the upper secondary school program attended. The larger set of controls also includes dummy variables for each decile of the other underlying variables and of the overall measures of cognitive and non-cognitive abilities, cubic functions of the sub-tests underlying the measures of cognitive and non-cognitive abilities, dummy variables for the mother’s and father’s highest educational attainment, the mother’s and father’s age at birth of their first child and an indicator for if the parents have separated before the child was 16, immigration status, the number of college places at the closest college, the number of potential applicants to the closest college and the local unemployment rate. The unadjusted difference in college enrollment between the median and the 1st decile is not statistically significant, while the difference between the top and the median of the distribution is significant at the 1-percent level.

Family income is closely related to parental education and the children’s cognitive abilities. Therefore, many papers that study the effect of family income also control for measures of cognitive ability and other family background characteristics. The dashed line in Figure 1a shows the relation between family income and college enrollment in Sweden after controlling for different dimensions of the individuals’ cognitive and non-cognitive skills, upper secondary school GPA, parents’ education and other background characteristics (described above). This almost wipes out the relation between family income and college attendance. The remaining college attendance difference between the richest and the poorest is about 20 percentage points in Canada and 45 percentage points in the US. Thus, even though tuition in Sweden is free and generous student grants are available for all – which effectively leave few (no) individuals truly credit constrained – college attendance differs more by family income in Sweden than in Canada (but not as much as in the US).
families is no more than one percentage point, although still significantly different from zero.

**Figure 1b. Enrollment probability by distance to college**

![Graph showing enrollment probability by distance to college]

Source: Own calculations based on data from Statistics Sweden and the Military Archives of Sweden.

Notes: The figure shows college enrollment by college distance decile, obtained from a linear probability model where an indicator for having enrolled at college is regressed on a smaller set of covariates (solid line) and a larger set of covariates (dashed line), respectively. See Figure 1a for details. The unadjusted difference (the adjusted difference) in college enrollment between the median and the lowest decile is significantly different from zero at the 1-percent (1-percent) level of confidence, while the difference between the top and the median of the distribution is not statistically significant.

That family income is unimportant for explaining college attendance conditional on upper secondary school completion, upper secondary school grades, cognitive and non-cognitive skills, and parental education, does not necessarily mean that borrowing constraints (in the broad meaning of the word) are absent. It is likely that many of these controls are endogenous. Individuals who know that they will not afford higher education, or who find the social costs of going to college too high, would, for example, be less motivated to work hard at school (which results in lower grades and less development of cognitive skills). The evidence from this exercise does, however, indicate that increasing family resources – e.g. by raising student grants – may have small effects on college enrollment for individuals who have already completed upper secondary school. It is less informative about the effects of such a policy on college attendance for coming generations.

The solid line in Figure 1b shows the relation between college proximity at the age of 16 and college attendance, controlling for just the baseline covariates. Individuals living closer to a higher education institu-
tion are more likely to go to college. The relationship is (again) highly non-linear, with higher probabilities of going to college for those with a college distance below the median (which corresponds to living less than 25 kilometers from a college). The difference between individuals in the 1st decile and those in the 5th is about 8 percentage points, and highly significant. The non-existing differences in college attendance for those living further away from college indicate that it is the reduced costs of living rather than the social or informational costs that are driving the relation between college attendance and distance.

The dashed line in Figure 1b shows the relation between college distance and enrollment conditional on the full set of control variables. Controlling for all covariates reduces the probability of college attendance for those living within walking distance to college almost by half. This confirms what has also been found in earlier studies; distance to college is not randomly allocated in the population. Still, even with detailed information on individuals’ skills, those who lived nearby a college as teenagers on average have a 4 percentage point higher probability of going to college than those having more than 25 kilometers to college. The difference is statistically significantly different from zero. Thus, there may be some scope for regional college expansions to attract more students.

Finally, Figure 1c shows the descriptive relation between upper secondary school GPA and college enrollment. Upper secondary school performance is an extremely strong predictor of attendance. Of those with the lowest grades, only 25 percent go to college. The corresponding figure for those at the top of the grade distribution is 65 percent. The differences are also highly statistically significant. This is unsurprising given that GPA is the main basis for admitting students to college. Moreover, high-ability students probably self-select to college to a greater extent anyway. Since many programs and colleges are oversubscribed, lowering the admission standards is an obvious way of increasing enrollment.

Figure 1c. Enrollment probability by GPA

27 Once more, it is worth stressing that some of the control variables are probably endogenous. Individuals who live far from a college at the age of 16, thus knowing that the costs of college attendance will be high, may have less incentives to work hard at school.

28 It may be surprising that these individuals gain admission to college at all. However, they may have enrolled in programs with excess supply. They may also have gained admission through the SAT, which was open for a smaller group of individuals.
On the margin of success? Effects of expanding higher education

Source: Own calculations based on data from Statistics Sweden and the Military Archives of Sweden.

Notes: The figure shows college enrollment by secondary school GPA decile, obtained from a linear probability model where an indicator for having enrolled at college is regressed on dummy variables for year of birth, local labor market of residence at the age of 16, and upper secondary school program attended. College enrollment at the first (last) decile is significantly different from the median at the 1-percent (1-percent) level of confidence.

4.2 College graduation

For enrollment-promoting policies to be efficient, they should not only attract students to college. If the new group of students drop-out or perform badly, the value to society of such a policy may be limited. Therefore, it is crucial to examine to which extent different policies can attract students that have the potential to complete their studies. Figures 2a-2c show the estimated relation between the underlying treatment variables and the probability of taking a Bachelor’s degree within five years upon enrollment. These are based on men born 1974-1976 who have enrolled at college, and all models control for the interaction between the college and the program attended. Thus, it essentially compares students enrolled in the same program (or single-subject course) at the same college.

The solid line in Figure 2a shows the descriptive relation between family income and the probability of graduating, and indicates that those from richer backgrounds perform slightly better than those from poorer backgrounds. However, once controlling for the full set of covariates (the dashed line), there are only small differences in the graduation probabilities by family income. The difference in academic performance for individuals in the upper and lower tails of the family income distribution is then just one percentage point, and not statistically significant. The new
students attracted by policies that reduce the direct costs of going to college will probably perform as well as the average student. Thus, in this respect, there is no obvious equity-efficiency trade-off when designing policies to attract students from poorer backgrounds.

**Figure 2a. Graduation probability by family income**

![Graph showing graduation probability by family income](image)

Source: Own calculations based on data from Statistics Sweden and the Military Archives of Sweden.

Notes: The figure shows the probability of obtaining a Bachelor’s degree within five years upon enrollment by family income decile, obtained from a linear probability model where an indicator for having a Bachelor’s degree is regressed on a smaller set of covariates (solid line) and a larger set of covariates (dashed line), respectively. All models control for college by program fixed effects. See Figure 1a for details. The unadjusted difference in graduation probabilities between the median and the lowest decile (the highest decile and the median) is significantly different from zero at the 1-percent (1-percent) level of confidence. The adjusted graduation difference between the median and the lowest decile, and between the highest decile and the median, is not statistically significant.

Figure 2b shows the relation between college distance and the probability of having obtained a Bachelor’s degree within five years from enrollment. Earlier, we saw that individuals who lived closer to a college at the age of 16 were more likely to attend college. Those who enroll, however, are significantly less likely to complete their studies than college students who live further away. This is true both with and without controlling for the full set of covariates. Actually, conditioning on observed characteristics does not change the estimated graduation probabilities to any considerable extent. If anything, those who lived close to a college perform slightly worse when accounting for family background and abilities.

**Figure 2b. Graduation probability by distance to college**
What is interesting to note is that the graduation probability increases the most for college distances above the median, which is contrary to what was found for the enrollment probabilities. This could possibly be due to selection, but adding the full set of controls actually strengthens this finding. Another explanation for this result would be that individuals who live far away from college must invest more to go to college. They may therefore be more eager to collect the returns from their investment.

When it comes to regional expansions of college education – which reduces the distance to college – there is potentially a trade-off between equity and efficiency. Those who live close to college may be more likely to enroll, but this may come at the expense of lower graduation probabilities. Since individuals living close to colleges face lower costs from enrolling at college, they may find it worthwhile to study even though the return is not that high; I examine whether this is true in the next section.

Figure 2c. Graduation probability by GPA
As discussed in the literature review, upper secondary school grades and SAT are strong predictors of academic success. This is confirmed in Figure 2c which shows the descriptive relation between upper secondary school GPA and college completion. Thus, past school performance is a very strong predictor of future performance. Less than 25 percent of the individuals with the lowest grades in upper secondary school are expected to complete a degree within five years from enrollment, while 65 percent of the top upper secondary school students are predicted to complete college. Clearly, if it were possible to attract more top students to college, the graduation rates would be boosted. The problem is, however, that most high-performance students already go to college, so lowering the admission standards would more likely affect the average upper secondary school student. One way of attracting relatively more top students would be to lower admissions standards only for highly demanded programs and colleges.

4.3 Economic returns to college

The costs of human capital investments come early in life while the pay-off is expected to come later. To estimate the internal rate of return to one year of college for those who enroll, at different deciles of the under-
lying variables, I need information on (1) the forgone earnings while studying for those who enroll, (2) the mid-age earnings for college enrollees and (3) the counterfactual mid-age earnings for enrollees had they not enrolled.

To measure forgone earnings while studying, I use the average annual earnings at ages 20-24 years for those who have not (yet) enrolled at college, while the mid-age earnings are measured in 2009 for all. The effects of each decile of the underlying variables on mid-age earnings are estimated separately for college enrollees and upper secondary school leavers. The models control for the full set of covariates. To normalize the earnings gain to one year of college, I use the number of credits taken as a measure of the time spent in college.

Using the information on forgone earnings and mid-age earnings gains, I calculate the internal rate of return to college, i.e. the discount rate when the two earnings streams are equalized, for each decile of the underlying variables. This is based on the assumptions that (1) the mid-age earnings gain is a reasonable approximation of the life-cycle earnings gain, (2) individuals start college at the age of 20 and (3) leave the labor market at the age of 65 (the legal retirement age). This leads to an average internal rate of return to college for those who enroll of about 3.5 percent. This result is in line with the findings in Öckert (2010).

Figure 3a shows the estimated internal rate of return to one year of college by deciles of the family income distribution. The return to college is estimated to be higher for children from poorer families and at the very top of the family income distribution. The higher returns at the bottom of the family income distribution are consistent with credit constraints (in the broad sense). Thus, individuals from poorer backgrounds who go to college have higher returns than the average student. Thus, there are both equity and efficiency arguments for attracting these individuals to college.

29 The estimated earnings profiles for college students and upper secondary school leavers are presented in Figures A1a-A1c in the Appendix.
30 The effects of the different underlying variables on years of college mirror the results found for college graduation (although the scale is different).
31 The internal rate of return does not account for student grants, transfers and taxes.
The spike in the return to college for children from the very richest families may be consistent with family networks, where the parents (the father) either own or have a leading position in a company or organization. Corak and Piraino (2011) and Bingley et al. (2011) show that sons inherit employers from their fathers, in particular this is true at the very top of the income distribution. Access to the very best paid jobs, however, often requires formal merits as well. Thus, it may only be when the children acquire a college degree that parents can use their position to influence the hiring process.

Figure 3b illustrates the return to college by college proximity at the age of 16. It is in the lower tail and the upper tail of the distance distribution where the return to college is highest. This pattern is consistent with some type of borrowing constraints. If individuals were not constrained by their family background, we would expect the return to college for those who choose to enroll to increase by college distance. It is only those with a higher return to college who would find it worthwhile to enroll if they lived far away from college. The return to college also seems to be higher for those living further away. The higher return for those who live close to college is consistent with credit constraints, where the reduced costs of living may attract individuals who face high social costs or who discount the future too much. Even though this group of students does not
perform as well as the average student, the return to going to college for them is higher than the average. Thus, a regional expansion of college education that reduces the distance to college may be an effective way of attracting individuals with a high return to college.

**Figure 3b. Return to college by distance to college**

![Graph showing return to college by distance to college](image)

Source: Own calculations based on data from Statistics Sweden and the Military Archives of Sweden.

Notes: The figure shows estimated internal rates of return to one year of college by distance to college decile. See details in the text.

Figure 3c shows the return to college by upper secondary school GPA. Students with higher grades have a higher probability of going to college, are more successful at college, and have higher monetary returns to college. For the very top students, the return to college is 10 percent, which is on par with the return to schooling in, e.g., the US. The weakest students are less likely to complete their studies and the return is low. For those at the lower end of the GPA distribution, the return to college is even negative. About 40 percent of the students have internal rates of return exceeding 4 percent. Lowering the admission standards for very popular programs can be one way of attracting these students. However, since applicants who are rejected from popular programs often gain admission to other programs, the policy may not be that efficient in attracting more high-ability students to college.
5. Summary and discussion

By reviewing evidence from earlier studies and providing new evidence for Sweden, this paper analyzes the effects of different policies to promote college attendance on enrollment, graduation and labor market performance. Reducing the direct costs of going to college – by reducing tuition or increasing student grants – may have sizable effects on college enrollment. However, in countries where tuition is already low (or even free) and generous student grants are available for all, there is little leeway for boosting enrollment by reducing direct costs. In Sweden, no students seem to be credit constrained in the literal meaning of the word. Still, there are substantial differences in earlier school performance and the choice of upper secondary school programs between children from richer and poorer backgrounds. Thus, individuals may be constrained by their family background more broadly, and policies aimed at weakening the observed link between family income and college attendance should probably be implemented earlier rather than later.

Studies from the US find that students who are induced to go to college due to lower costs do not perform worse than other students. Thus, it
is possible to attract students from poorer backgrounds without reducing the academic performance of students. In agreement with this, evidence for Sweden shows that the probability of taking a degree does not vary by family income. Earlier studies have not been able to present any credible evidence on the effects of reducing the direct costs of college on labor market performance. The evidence presented in this paper suggests that the return to college in Sweden is highest among children from the poorest backgrounds; then the return typically falls with family income. Thus, policies that attract children from poorer backgrounds may have high returns. At the very top of the family income distribution, however, the return to college peaks, possibly driven by family networks.

Individuals living close to a college are more likely to enroll at college. This has been documented in many studies, and is also found for Sweden. Thus, expanding college education regionally – either by opening up new colleges or by increasing the number of slots at existing institutions – may attract new students to college. The more credible studies on college expansions are, however, silent about the effects on college completion and labor market returns. Evidence for Sweden suggests that it is the students living further away from college that are more likely to graduate, probably because they need to complete their studies to collect their (possibly) higher return to college.

In the absence of credit constraints, we would expect the return to college to increase by college distance, since it is only those with high returns to college who live far away that would find it worthwhile to go. In line with this argument, evidence for Sweden suggests that the return to college is higher for those living far away from a higher education institution. However, the return to college is also high for those living close to college, indicating that reducing the costs of living may attract individuals who would otherwise have been credit constrained. Thus, regional expansions of higher education in areas where the supply has been scarce may have high returns.

It is well-documented that high-ability students are more likely to go to college and are also more successful at college. A couple of studies also report that the return to college increases by student ability. The evidence for Sweden presented in this paper also shows that the return to college increases sharply by upper secondary school GPA. For the very top students, the return to college in Sweden is on par with the returns to
college reported in, e.g., the US. There is, however, not that much direct evidence on the effects of lowering admissions standards; the study by Maurin and McNally suggests that the returns for the marginal students in France in the late 1960’s were substantial, whereas Öckert (2010) finds very low returns for the students on the margin of admission in Sweden in the 1980’s. To some extent, the diverging results can probably be explained by contextual differences between studies.

Taken together, the evidence presented in this paper suggests a lowering of admissions standards only in oversubscribed programs. This could potentially attract more high-ability students to colleges, and the return to them may be high. Thus, when deciding upon the optimal mix of programs at colleges, the supply of places should probably be adjusted more to student demand than what is now the case.\footnote{If tuition is low, the supply of places should probably not be completely adjusted to the demand, since students do not account for the costs of different programs when making their choices.} A limitation of this policy, however, is that applicants who are rejected from popular programs often gain admission to other programs, why the policy may turn out not be that efficient in attracting more high-ability students to college.

References


Dynarski, S. (2008), Building the stock of college-educated labor, Journal of Human Resources 93, 576-610.
Öckert, B. (2010), What’s the value of an acceptance letter? Using admissions data to estimate the return to college, Economics of Education Review 29, 504-516.
Appendix

Figure A1a. Earnings by college attendance and family income

Source: Own calculations based on data from Statistics Sweden and the Military Archives of Sweden.
Notes: The figure shows estimated earnings in 2009 by family income decile for college enrollees and upper secondary school leavers, respectively. The estimates come from a regression where the earnings for college students and upper secondary school leavers, respectively, are regressed on the full set of covariates. The model for college students controls for college by program fixed effects. See Figure 1a for additional details.

Figure A1b. Earnings by college attendance and distance to college

Source: Own calculations based on data from Statistics Sweden and the Military Archives of Sweden.
Notes: The figure shows estimated earnings in 2009 by college distance decile for college enrollees and upper secondary school leavers, respectively. The estimates come from a regression where the earnings for college students and upper secondary school leavers, respectively, are regressed on the full set of covariates. The model for college students controls for college by program fixed effects. See Figure 1a for additional details.
Figure A1c. Earnings by college attendance and GPA

Source: Own calculations based on data from Statistics Sweden and the Military Archives of Sweden.

Notes: The figure shows estimated earnings in 2009 by upper secondary school GPA for college enrollees and upper secondary school leavers, respectively. The estimates for college students come from a regression where the earnings for college enrollees are regressed on the full set of covariates. The model for college students controls for college by program fixed effects. See Figure 1a for additional details. The estimates for upper secondary school leavers are obtained using the two-step procedures described in section 4.3.
Comment on Öckert: On the margin of success? Effects on expanding higher education for marginal students

Torbjørn Hægeland*

In most countries, the share of the population with higher education has increased substantially in recent decades. Despite this increase in the supply of highly educated workers, there has not been any secular decline in the economic returns to education, implying that there has been a similar increase in the demand for educated workers. If this demand continues to increase, there is a need for higher capacity of higher education in the future. However, stable returns to education in the past do not imply that the labour market will accommodate any future increase in the share of highly educated workers. Whether further expansion of higher education is worthwhile from society’s point of view depends on future demands for skills produced by colleges and universities. This demand is to a large extent reflected in the private returns to schooling. Absent externalities and other market imperfections creating wedges between private and social gross returns to schooling, the private return is therefore a substantial component in the social accounting of expanding higher education. Key requirements for a successful policy for expanding higher education is that it manages to attract more students, that the students complete the education and that they earn a high enough return to justify the investments made by themselves and society. To the extent that prospective students differ in any of these dimensions, the social return to a policy

* Statistics Norway, thd@ssb.no.
reform of expanding higher education depends on which individuals that actually do respond to that reform.

The expansion of higher education is often seen as a transformation from elite education to mass education. A central objective for authorities has in many countries been to recruit new groups to higher education, typically young people from low-income families and/or rural areas, believing that there was an unexploited pool of talent in the population. Policies to induce more people to take higher education often focused on making education affordable, by increasing grants and subsidies, and available, by establishing institutions for higher education.

The point of departure for Björn Öckert’s paper is to what extent further expansion of these traditional policy measures may be expected to work in today’s Sweden, where higher education is relatively widespread and the direct economic costs of schooling are quite modest. A key issue is that in a situation where a large fraction of each cohort takes higher education, there may no longer be a large pool of unexploited talent in the sense that all individuals attracted to higher education by a certain policy will necessarily earn a high return on their investments. Individuals may differ in their returns to schooling. There may also be differences between policy measures with respect to which students they attract. The paper discusses possible expansions of higher education along three different margins: (i) reducing the direct costs of schooling by lowering college tuition or increasing general student aid, (ii) increasing college proximity by opening up new institutions for higher education, and (iii) allowing for more students by lowering the admission standards. This is a highly relevant policy issue. The paper is well written and contributes to our understanding of important problems.

There is a relatively large literature on each of these topics, but as Öckert concludes in his literature survey, it is hard to draw any general conclusions on which type of policies that may be effective from this literature as the studies differ with respect to identification strategy, time and institutional context. A main strength of Öckert’s paper is that it attempts to evaluate different policy opportunities for expanding higher education within the same dataset, i.e. for the same cohorts in the same country and institutional context. The absence of an exogenous variation in variables like family income and college proximity makes a direct causal interpretation of the results somewhat problematic. However, the
Comment on Öckert

A rich set of covariates makes the conditional independence assumption required for causal interpretation of the estimates less heroic.

As pointed out by Öckert, there are quite substantial differences with respect to family income in the propensity to take higher education in Sweden. When not restricting the sample to those that have completed secondary school, he finds that the differences are larger than in Canada and somewhat smaller than in the United States. This does not imply that there are large economic obstacles to higher education. The positive relationship between family income and higher education may be the result of many mechanisms leading to parents that do well in the labour market having children that do well at school. These mechanisms are seldom affected by economic incentives for higher education. Controlling for a rich set of covariates (including ability scores and parents’ education), Öckert finds no relation between family income and college enrolment in Sweden, leading to a conclusion that few students are directly credit constrained and that increasing student aid may not attract many new students to higher education, given their performance in secondary education. If more generous grants make students study harder in secondary school, this may, however, have an indirect effect.

Öckert finds that distance to the nearest college matters for enrolment into higher education, and that the returns to college for those who live far away from a college are as high as for those who live nearby. He concludes that an expansion of regional colleges may be worthwhile in terms of attracting more students to higher education. To reduce the distance to college for many students, new colleges must be located quite far away from other colleges. Öckert’s argument implicitly hinges on an assumption that these new colleges will be of the same quality as existing educational institutions. It is not obvious that this will be the case, and there may well be a tradeoff between accessibility and quality of higher education.

Not surprisingly, there is a strong relationship between academic performance in high school and both success in higher education and return to college. By assuming that students (conditional on their secondary school GPA and other covariates) do not self-select into higher education based on their own returns to schooling, comparing Figures 1c and 3c in the paper suggests that there may be a significant number of high-return individuals that, in fact, do not enter into higher education. To the extent
that the high-return non-students are effectively rationed by restricted access, there is a large scope for expanding higher education. However, at least for the upper half of the GPA distribution, it is unlikely that a large fraction is rationed in the sense that their applications to colleges and universities are rejected. The question is then why they choose not to take higher education. High costs of schooling may prevent individuals with high gross returns to enter. As indicated by the empirical evidence presented in the paper, high pecuniary costs of schooling are unlikely to be a major driving factor. However, there may, of course, be differences in non-pecuniary costs and benefits to higher education that affect the choice of higher education. Persistent differences in economic returns between different fields of higher education may reflect sorting by ability and admission restrictions, but the fact that investments in many types of higher education earn a negative economic return (see e.g. Kirkeboen, 2010, for Norwegian evidence) indicates that noneconomic factors may substantially affect choices of investments in higher education. Such factors may be aspects of the education per se or the types of jobs for which the education qualifies. The academic literature on the determinants of educational choice, both in terms of levels and fields of study, is relatively scarce, but see e.g. Zafar (2011) and the reference cited there. Accordingly, we only have limited knowledge of the relative importance of economic versus non-economic factors in educational choice. A deeper understanding of what governs educational choices will, of course, be helpful for providing the right policies for attracting students to higher education.

Another interpretation of the somewhat puzzling result that high-return individuals choose not to take education (and vice versa) is that young people have limited information on the returns to higher education. To what extent do prospective students know how much different educations pay off economically?

The paper is concerned with policies that may attract new students into higher education. How today’s cohorts of secondary school graduates are allocated in terms of further education is, of course, an important policy question that merits such an analysis and Öckert mainly discusses his results with respect to the relevance for policy measures directly affecting higher education. In a more general setting, however, policies for attracting students to higher education should not be discussed in isola-
tion. Young people considering higher education respond to incentives and policy initiatives, but how they respond or are able to respond very much depends on their ability and willingness to pursue further schooling. Many factors contribute to this, but the functioning of the education system at lower levels is probably the most important among these factors that can be affected by government policy. Human capital policy, including policies aimed at stimulating students to take higher education, should therefore have a life cycle perspective. As emphasized by e.g. Heckman (2000), skills beget skills, and early learning fosters later learning. The dynamic complementarities in human capital accumulation imply that the most effective policy measures for stimulating higher education may not be measures aimed directly at higher education per se, but rather at stimulating learning earlier in life. Öckert finds that academic performance in high school is a strong predictor for success in higher education. To the extent that the skill level reflected by secondary school GPA is not purely determined by “innate ability”, and may at least in part be affected by school practice and education policies at young ages suggests that in the longer run, the most efficient way of attracting marginal students to higher education may be to address the issue earlier, thus affecting the processes leading up to the skill level at the end of high school (measured by high school GPA). At least in countries where the direct costs of studying are relatively modest and the economic resources fairly evenly distributed, preparing students for higher education by providing them with the necessary skills may be at least as effective as a tool for attracting more students to higher education as reducing the direct costs of studying even more. It is an open question how large is the scope for educational policies to raise average pupil performance and mitigate performance differences between socioeconomic groups within reasonable costs. However, as documented by e.g. OECD (2008), in many countries there are substantial differences between schools in how they contribute to student learning. Calculations based on value-added estimates for lower secondary schools suggest that if low-performing schools adopted something close to a “best practice”, this would substantially raise the average performance of pupils, thereby increasing their readiness for higher education.
References

Gender differences in education*

Tuomas Pekkarinen**

Summary

The paper surveys the trends in gender gaps in education, their causes and potential policy implications. I show that female educational attainment has surpassed, or is about to surpass, male educational attainment in most industrialized countries. These gaps reflect male overrepresentation among secondary school drop-outs and female overrepresentation among tertiary education students and graduates. Existing evidence suggests that this pattern is a result of a combination of increasing returns to education and lower female effort costs of education. A widening gender gap in education combined with recent wage and employment polarization will likely lead to widening inequalities and is linked to declining male labor force participation. The paper discusses evidence on educational policies that both widen and reduce gender gaps in educational outcomes.

Keywords: Education, gender differences, test scores.
JEL classification numbers: I20, J160, J240.

* The author thanks the editors and Anna Sjögren for helpful comments as well as the Academy of Finland for financial support.
** Aalto University School of Economics, IZA, IFAU, tuomas.pekkarinen@aalto.fi.
One of the most striking trends in education in the past two or three decades has been the rapid increase in female educational attainment. From the cohorts born in the 1950’s and onwards, women quickly caught up with men in educational investment. Furthermore, it is clear that the growth in the educational attainment of women has not stalled at gender parity. At the moment, female educational attainment clearly dominates male educational attainment in a majority of industrialized countries. This is true for several measures of attainment. Women are in a clear majority among secondary school graduates, among students enrolled in tertiary education, and among tertiary graduates. Furthermore, judging from recent trends, it seems likely that the gender gap in educational attainment will keep on widening in favor of women in the future.

These trends imply a dramatic change in the composition of the skill supply in industrialized economies. Whereas in the post World War II industrialized world, men were substantially overrepresented among highly skilled workers, this will not be true in the coming decades. And this change happens at a time when educational investments are becoming more and more important for labor market outcomes. In particular, the recent trend of employment polarization will likely lead to more pronounced differences between the labor market fortunes of high- and low-skilled workers. Gender differences in educational attainment imply that these highly rewarded high-skill workers will be predominantly female and increasingly disadvantaged low-skill workers will be predominantly male. The gender gap in educational attainment can therefore have far reaching labor market implications.

The aim of this article is to provide an overview of the trends in gender differences in educational attainment and examine their potential causes as well as implications for labor market outcomes. When it comes to documenting the trends, the emphasis in this article will be on cross-country comparisons. This is partly because most of the previous surveys on gender gaps in education have focused on the trends in the United States and the timing and the patterns differ across countries in interesting ways.¹ Moreover, I strongly believe that focusing on cross-country data may shed some new light on the factors underlying the observed gender

¹ On American surveys, see in particular Buchmann et al. (2008) as well as Goldin et al. (2006).
differences in attainment. Finally, I will also survey available evidence on policies that have affected the gender balance in educational attainment.

The rest of the article is structured in the following way. In the next section, I document the trends in gender gaps in educational attainment. I will mostly focus on industrialized countries and put particular emphasis on the comparison of Nordic countries and the United States, which is the country that has traditionally led trends in educational attainment. In Section 2, I examine the potential causes behind these trends. As a starting point, I use the standard economic model of educational investments and look for gender differences in the trends of benefits and costs of education. The conclusion that arises from this analysis is that a combination of increasing returns to education, the removal of barriers to women’s careers and long-standing gender differences in the effort cost of education, has made educational investment more attractive for women than for men. Section 3 explores the potential labor market implications. The recent trend of employment and wage polarization implies that increasing female and stalling male educational attainment will change the traditional patterns of gender gaps in labor market outcomes. In Section 4, I survey evidence on the effects of different policies on the gender gap in educational attainment. I argue that the most relevant policies deal with achievement in secondary education or even earlier. Section 5 concludes the paper.

1. Changing gender differences in educational attainment

Both stock and flow measures can be used when measuring educational attainment. Stock measures reflect the pool of human capital available in the economy at a specific point in time whereas flows reflect the contributions of incoming cohorts to the stock of human capital. A number of researchers have constructed international data sets that attempt to measure both stocks and flows of human capital across a large number of countries. Perhaps the most famous of these data sets is the Barro and Lee data set (2010) which is based on census/survey information originally compiled by various international organizations such as Eurostat, OECD
Here I use the latest 2010 version of the Barro and Lee data to examine cross-country trends in the gender gap in educational attainment.

In this section, I use cross-country data to examine gender differences in both stocks and flows of human capital. In both of these measures, I start by looking at the mean outcomes. However, since the tails of the attainment distribution are in many ways more interesting, I put a particular emphasis on gender differences in tertiary graduation and enrollment rates as well as on gender differences in the likelihood of not obtaining a secondary school diploma.

1.1 Gender differences in stock measures of educational attainment

Stock measures obviously take a longer time to reflect more recent changes in educational attainment. However, the process of female dominance in educational attainment has been going on for long enough so that it can already be seen in human capital stocks in many countries. Whereas in 1950, according to Barro-Lee data, women had more years of education than men in only 11 countries (Canada, UK, and the US among them) out of 146, in 2010 this figure was already 43. These 43 countries include all Nordic countries as well as Australia, the United States, Canada, and the United Kingdom. Overall, in 9 out of 24 advanced economies in the Barro and Lee data, the overall female population is on average more educated than the male population.

The pool of university graduates has also become increasingly female in many countries. In 2010, there were more female university graduates than male university graduates in 13 out of 24 advanced economies in the Barro and Lee data set. In 1950, male university graduates still outnumbered women in all of these countries. The increase in the average educational attainment of women thus reflects the fact that increasing numbers of women are choosing to continue their education to the tertiary level.

---

2 According to Barro and Lee (2010), the latest version of their data addresses some of the measurement issues which were raised, among others, by de la Fuente (2011) in the previous issue of this review.
1.2 Gender differences in flow measures of educational attainment

The trends in the stock of human capital mask interesting changes in the flow measures of human capital. Focusing on birth cohorts makes it clear that female educational attainment will surpass that of men in a majority of countries in the Barro and Lee data. Among the individuals born between 1975 and 1979, women obtained more years of education than men already in 74 out of 144 countries. Hence, in many countries, where men still have a higher average educational attainment than women, this situation will change in the near future. Among advanced economies, men born between 1975 and 1979 had a higher educational attainment than women only in Austria, Germany, Switzerland and Turkey.

Figure 1. Female and male average years of education by birth cohort and gender in the Nordic countries and the United States

In Figure 1, I have plotted the average years of education by gender and birth cohort in the Nordic countries and the United States. Barro and Lee data makes it possible to examine the educational attainment from cohorts born in the late nineteenth century onwards. Figure 1 reveals that

---

3 The Danish statistics for cohorts born after 1970 in the Barro and Lee data imply implausibly large drops in the level of educational attainment for both men and women. This data was not used in Figures 1 and 2 and this is why the Danish figures end with the cohort born in 1970.
in the United States, men had a higher educational attainment in the cohorts born between the 1920’s and the 1940’s. However, in all of these countries, with the exception of Denmark, women caught up with men in educational attainment by the cohorts born in the 1950’s. After this, the gender gap in average educational attainment has widened in favor of women. This widening of the gap reflects the fact that in most industrialized countries, the growth in male educational attainment decelerated whereas female educational attainment kept on growing. This deceleration of the growth of male educational attainment is particularly dramatic in the United States where the growth has virtually stopped since the cohorts born in the 1940’s.

Naturally, the average educational attainment simply gives us the mean of the distribution of human capital. The tails of the human capital distribution are often more relevant for labor market outcomes. Therefore, it is interesting to focus on university education. These measures reveal even more dramatic gender differences. In the Barro and Lee data, women born between 1975 and 1979 completed university education at a higher rate than men in 21 out of 24 advanced economies. The three countries where this did not take place were Germany, Switzerland and Turkey. In Figure 2, I have plotted the university education completion rates by birth year and gender in the Barro and Lee data. The gender differences are even clearer in Figure 2 than in Figure 1. In the United States, for example, male university attainment started to drop after the cohorts born in the 1950’s. At the same time, female university attainment continued to increase. The gaps in university attainment have become even larger in the Nordic countries. Yet in these countries, the attainment levels of both men and women have kept on increasing. The increasing gender gap in the Nordic countries reflects a particularly fast growth of female university attainment.
Another way of looking at the top end of the attainment distribution is to focus on enrollment in tertiary education. In the most recent data available from UNESCO for the year 2008, out of 119 countries for which there are data on female enrollment in tertiary education, this was higher than the male enrollment in 81 countries. This is true for all OECD countries with the exception of Japan, South Korea, Switzerland and Turkey. But female enrollment in tertiary education is also higher in some large non OECD countries such as the Philippines, Iran and Thailand. Indeed, according to UNESCO, globally the number of female students in tertiary education is almost 2.5 million higher than the number of male students, making the ratio of female to male students slightly higher than one at 1.03. It seems that female dominance in educational attainment is becoming a global phenomenon.
In Figure 3, I have plotted the ratios of female to male tertiary students in the United States, Switzerland and the Nordic countries. The gender ratios in the US and the Nordic countries seem to increase almost in parallel fashion. The first country where the number of enrolled female students at the tertiary level surpasses the number of male students in this figure is the United States where this happened in 1980. The Nordic countries followed the US approximately five years later and in all of these countries, the ratio of female to male students has remained above one and has kept increasing.

It is insightful to contrast the American and the Nordic enrollment ratios to those in Switzerland which, as was already explained above, is a country where female educational attainment has lagged behind the other OECD countries. The Swiss enrollment ratios are clearly below the US and Nordic ones. However, also in Switzerland the ratio of women to men among university students has kept steadily increasing in the past 40 years. In fact, all OECD countries repeat the same pattern although they lag behind the US and the Nordic countries somewhat. Judging from
enrollment patterns, female dominance in tertiary education will become stronger in the future in virtually all advanced economies.

1.3 Gender differences in the choice of university major

Despite the fact that women are now in majority among students at the tertiary level, they still choose different major subjects than men. Here, it is difficult to draw any conclusions on international long-term trends since no long, internationally comparable time series are available. However, some studies on individual countries suggest that gender differences in choices of major subjects at the tertiary level may be diminishing. Both Turner and Bowen (1999) and Goldin (2006) show that the dissimilarity index of college majors in the United States declined from close to 50 in 1970 to around 20 in 1994. According to the World Bank data, the US dissimilarity index was still at 22 in 2009 which is the lowest figure in the OECD. The average level of the dissimilarity index in the OECD in 2009 is 29 with the highest level in Finland at 42. The data that is available from 1999 onwards does not reveal any striking trends in the gender differences in the choice of university major. However, in the majority of OECD countries, the dissimilarity index declined during this period with Denmark, Iceland, Japan, and Switzerland showing declines of more than 5 points.

1.4 Gender differences in secondary school drop-out rates

At the other end of the educational attainment distribution are individuals who do not finish upper secondary education. Although the compulsory education ends with lower secondary school in many industrialized countries, individuals without upper secondary school education are usually seen as a particularly disadvantaged group when it comes to labor market outcomes. Men have traditionally been overrepresented in this group.

Figure 4. Fraction of male and female cohorts without upper secondary education in the Nordic countries and the United States in 2008

---

4 The dissimilarity index is calculated by summing the absolute values of the differences between the percentages of women and men in each field and dividing this sum by 2. A value of 100 for this index indicates total segregation by gender across fields and a value of 0 means that male and female choices are identically distributed.
In the OECD member countries in 2008, women are left without an upper secondary education more often than men only in Germany and Switzerland (OECD, 2011). In Figure 4, I have plotted the fractions of both men and women who do not obtain an upper secondary school diploma in the Nordic countries and the United States. This figure reveals that there are large differences in the overall drop-out rates across countries with particularly high rates in Sweden and the United States. However, what is really striking in Figure 4 is the gender difference. For example, in Iceland, only 2 percent of the girls are left without an upper secondary education whereas the corresponding rate for boys is 21 percent, making males more than ten times more likely leave school before obtaining an upper secondary school diploma. The difference in Finland, where the female and male drop-out rates are 2 percent and 8 percent, respectively, is also very high. In fact, Finland and Iceland have the highest gender inequalities in the drop-out rates in the OECD. As I will argue below, the gender gap in secondary school drop-outs should be a cause of particular concern for policymakers.
2. Factors behind the emerging female dominance in educational attainment

The standard economics textbook model of human capital investment views rational individuals as weighing the costs of additional education against the benefits from this investment. The costs of human capital investment include direct monetary costs such as tuition fees and the forgone labor market earnings during the time the individual is studying and indirect costs in the form of the effort that the individual has to exert in order to acquire human capital through education. Even though education can yield some direct pleasure in the form of consumption value, the human capital theory assumes that the most important benefits to the individual from education are the direct labor market returns. Education should increase the productivity of individuals and therefore lead to higher earnings.

The traditional human capital model is a useful starting point for considering the potential causes underlying the dramatic shift in the gender gap in educational attainment. In particular, since male and female educational investments have performed differently over the past three decades or so, it seems logical to look for explanations for this divergence in gender differences in the marginal benefits and costs of education. In this section, I will discuss how gender differences in both benefits and costs of education have developed in industrialized countries in recent decades and whether they can be seen as a potential cause for the observed patterns in the gender gap in educational attainment.

2.1 Gender differences in returns to education

It is a well-documented fact that the returns to education, and to university education in particular, have been increasing in most OECD countries since the early 1980’s (Acemoglu and Autor, 2011; Van Reenen, 2011). In the light of the large returns, one would expect to see strong supply responses in the form of increasing educational attainment. However, the evidence discussed in the previous section suggested that this is only true for female educational attainment whereas male educational attainment has only grown sluggishly or not at all. This has led many researchers to
speculate that the returns to investment in education are higher for women than for men.

There is a huge number of studies that report returns to education estimates separately for men and women. A fairly consistent pattern in these studies has been that the average return to an additional year of education is somewhat higher for women than for men. Dougherty (2005) cites 27 US studies of which 18 report unambiguously higher returns for women. Trostel et al. (2002) report estimates for 28 different, mostly European, countries where the returns are higher for women in 24.

It is not clear, however, which level of schooling is responsible for different estimates of returns to schooling by gender. Whereas several earlier US studies seemed to imply that the returns to university education were higher for women, these results have recently been called into question. Hubbard (2011) shows that previous studies fail to account for the top-coding of earnings in the available data and that once this is done, there is no evidence on gender differences in returns to university education. Furthermore, it seems that the returns to university education have increased in the same way for both men and women since the early 1980’s. Whether the failure to account for top-coding in wage data also biases the results from other countries is unclear, but it seems fair to conclude that the evidence on gender differences in returns to university education is mixed.

However, it seems to be unambiguously true that the rates of returns comparisons mask large differences in favor of men in total earnings benefits of university education. Women still work less hours over the life cycle than men at all levels of education. Hence, even if there are no large differences in university premiums between men and women, men are still in a better position to reap the full financial benefits of university education than women.

Yet, in this respect, the changes over the past decades have been substantial. Female labor force participation has increased considerably in all OECD countries since the 1970’s. The capacity of women to fully benefit from the financial rewards of educational investments has therefore improved in recent decades. Moreover, as reported by Mulligan and Rubinstein (2008), the composition of the female labor force has changed. Whereas the selection into the female labor force was negative in the past in the sense that lower educated women had a higher probability of par-
ticipating than high-educated women, the selection has become positive since the 1980’s so that high-educated women are more likely to participate in the labor force.

In the same way as higher labor force participation has increased the benefits of education for women, the changing role of family has also improved the chances for women of reaping the benefits from educational investments. The median age at first marriage has increased in virtually all industrialized countries. At the same time, the divorce rates have increased, thereby making marriage more fragile. As a consequence, the fraction of their lifetime that individuals spend married has decreased substantially. This has meant that the value of economic independence has increased for women and the time during which they can benefit from educational investments in the form of labor market returns has become longer.

Goldin (2006) as well as Goldin et al. (2006) argue that these kinds of secular long-run changes in the labor market and in the role of the family in the United States increased the overall benefits of education for women and led young women to revise their expectations about their future role in the labor market somewhere in the 1970’s. Citing survey evidence in the National Longitudinal Survey of Youth (NLSY), Goldin et al. (2006) show that the fraction of young women who expected to be employed at the age of 35 increased substantially from the end of the 1960’s to the beginning of the 1980’s. They argue that the women who saw their mothers’ generation participating in the labor market and postponing their marriage to an increasing extent decided to invest in education since they expected to be better able to benefit from investment in education. Indeed, Goldin et al. (2006) show that those teenage girls in the NLSY that expected to be employed at the age of 35 were more likely to go to university.

While there is no international longitudinal data on the teenage expectations of women and their subsequent educational investments, cross-country evidence can be used to see whether the current female/male enrollment ratios at the tertiary level correlate with the labor market participation of working age women and the median age at first marriage in a way that is consistent with the argument that female educational investment increased at least partly as a response to an increase in the total benefits of education.
Currently (in 2008) among the OECD countries, female participation rates range from 26 percent in Turkey to 81 percent in Iceland. The Turkish figure corresponds to the US situation in the 1930’s whereas the Icelandic figure is the highest that any industrialized country has reached in recorded history. In Figure 5, I have plotted the female participation rates and the female/male ratios in tertiary enrollment in the OECD countries in 2007 which is the last year for which there is data on enrollment. There is indeed a clear, positive correlation between female labor force participation and the ratio of women to men in tertiary enrollment. Furthermore, as both participation rates and enrollment ratios are available from 1980 for most countries, one can examine whether this positive relationship also holds within countries. For most OECD countries, there is a clear positive correlation between female labor market participation and the female/male ratio in tertiary enrollment.
A similar pattern is found between age at first marriage and enrollment ratios across OECD countries. OECD’s family indicators provide a snapshot of the situation in the OECD in 2002 and reveal that there is a higher variation in the age at which individuals first marry across the OECD than within time in the United States. The median age at first marriage varies from 22 and 23 in Turkey and Mexico to 31 and 32 in Norway and Sweden, respectively. In figure 6, I have plotted the median ages at first marriage in the OECD against the tertiary enrollment ratios in 2002. The age at which individuals marry for the first time is clearly positively correlated with female/male enrollment ratios at the tertiary level.

Neither of the patterns in Figures 5 or 6 indicates any causality. It seems likely that both the rise in female labor market participation and educational attainment may well be caused by some third factor. There is little agreement, let alone solid evidence, on what caused the changes in the labor market that enabled women to reap the benefits of educational investment. Some authors, such as Goldin et al. (2006), claim that decreased discrimination removed barriers to women’s careers. Others, such as Chiappori et al. (2009), build on the assumption that the technological
change in household production decreased the time spent on these activities. Perhaps the only piece of solid evidence on the underlying factors is the effect of contraceptive innovations. Goldin and Katz (2002) show that the diffusion of oral contraceptives across US states led to a postponement of marriage.

However, for the purposes of this survey, the key lesson from the changes in the returns to education for men and women is that the total returns to education have increased more for women than for men. Yet the level of total returns is still likely to be lower for women. Hence, changes in the total returns are probably not enough to explain why women have overtaken men in educational attainment in most industrialized countries.

2.2 Gender differences in the costs of education

Since there are apparently no gender differences in the benefits of education that would be consistent with the observed gender gaps in attainment, it is logical to look for gender differences in the costs of education. As explained above, the main monetary costs of education, especially at the tertiary level, consist of forgone labor market earnings as well as, in some countries, tuition and other fees. Non-monetary costs of education, on the other hand, stem from the effort required to reach the desired attainment levels. Here, I go over the existing evidence on gender differences in these costs.

The literature quite clearly shows that there are no gender differences in the forgone earnings that would explain observed attainment differences. Indeed, as will be shown below, the recent changes in the labor markets in industrialized countries would actually imply that there has been a decrease in forgone earnings for men since the wages and employment in medium skilled male jobs have decreased. Becker et al. (2010) show that gender differences in the hourly earnings of college graduates relative to high school graduates are minor in the United States. Nor is there any evidence that direct monetary costs in the form of tuition or fees would be different for men and women.

However, there is a large literature showing that there may be considerable gender differences in the effort costs of education. The evidence on these differences is often indirect. For example, the pattern that girls
achieve higher grades in secondary school is often cited to support this argument.\footnote{There is no international evidence on gender differences in grades. American studies show a consistent pattern that girls achieve higher grades from kindergarten to high school. Buchmann et al. (2008) cite evidence starting from the 1950’s.}

More direct evidence for gender differences in the effort costs of education is related to differences in underlying cognitive and non-cognitive abilities. Differences in cognitive abilities have been a subject of long and contentious debate starting from the work by Maccoby and Jacklin (1974). A more recent survey on this literature by Spelke (2005) concludes that while there are some gender differences in intrinsic cognitive abilities, they are small and do not consistently favor either gender in a way that would make men or women cognitively more adept for schoolwork.\footnote{This debate was contentious mainly because many commentators were interested in whether the gender differences in cognitive ability are biological or not. This debate has not been resolved and is beyond the scope of this paper.}

Even if the gender differences in tests measuring cognitive abilities are mixed, there are clear and consistent gender differences in the incidence of behavioral problems that imply differences in non-cognitive abilities. According to the medical and psychological literature surveyed by Buchmann et al. (2008), males are overrepresented in populations with reading disabilities, antisocial behavior, mental retardation, attention disorders, dyslexia, stuttering and delayed speech. There is also evidence that adolescent girls score higher in tests measuring non-cognitive skills such as attentiveness, organizational skills, and self-discipline.

One would expect these differences in non-cognitive abilities to show up in educational attainment. Recent studies in economics (Heckman, 2007) suggest that non-cognitive abilities have direct and sizable effects on schooling. Indeed, Jacob (2002) shows that controlling for non-cognitive behavioral factors in a data set on US high school graduates from 1992 can explain the entire female advantage in the probability of continuing to college.

Standardized achievement tests on skills that are learned in schools are a useful composite measure of school performance and preparedness for further education. These tests do show consistent gender differences. Typically, male scores are, on average, somewhat higher in tests measuring mathematical skills and women score clearly higher in tests measur-
ing verbal skills. However, irrespective of the subject, the variance in tests scores seems to be higher for men (Hedges and Nowell, 1995; Machin and Pekkarinen, 2008).

Figure 7. (Left) The relationship between female-male gaps in reading and mathematics in 2000 PISA test scores. (Right) The relationship between the average female-male gap in PISA 2000 test scores and the female/male enrollment ratio in tertiary education.

Source: OECD and UNESCO.

Interestingly, there is a considerable variation in the size of gender gaps in achievement tests across countries. Here, I use OECD’s PISA tests to examine cross-country differences in gender gaps more closely. PISA measures skills learned in lower secondary school across OECD countries in a comparable way. On the left-hand side figure of Figure 7, I show that the gender gaps in mathematics and reading are highly correlated across countries in PISA 2000. When the female-male gender gaps in reading are high, the gaps in mathematics tend to be close to zero or even positive as in Iceland and New Zealand. This pattern suggests that there are cross-country differences in the relative performance of women irrespective of the subject matter being tested.

On the right-hand side figure of Figure 7, I have plotted the average female-male gender gap against the female/male ratio in university enrollment. The pattern in this figure is very clear. The higher the gap in
favor of women, the higher is the female/male enrollment ratio in universities. This positive correlation suggests that in countries where women perform relatively better in secondary school achievement, women also participate in higher education in higher numbers. Since the skills that are learned in secondary schools should be directly related to the effort costs of education at the tertiary level, the evidence in Figure 7 can be interpreted as support for the claim that the gender differences in the effort costs of education are related to gender gaps in educational investment.

2.3 Gender differences in the net returns

The combination of increasing returns to education for both men and women and the lower effort cost of education for women imply that the net returns to education have become higher for women than for men. Whereas the effort costs of education have probably been lower for women already before the female educational attainment surpassed male attainment, the fact that the total benefits of education have increased more for women through increasing labor force participation and delayed marriage means that female educational investment started to reflect lower effort costs only after these barriers to women’s careers were diminished.

Becker et al. (2010) argue that this combination of demand-driven increases in returns and a long-standing female advantage in costs explains the emerging female dominance in educational attainment. They also argue that the evidence on the gender differences in the distributions of both cognitive and non-cognitive test scores implies that the supply of female human capital is more elastic than the male supply at the higher education levels. The lower female variance in school achievement implies that as the returns to university education increase, there will more women at the threshold where investment in university education becomes profitable.

3. Labor market consequences of the female dominance in educational attainment

When considering the labor market consequences of the changes in the gender balance of educational attainment discussed above, it is important
to remember that according to the patterns in Figures 1 and 2, the gender gaps in educational attainment are increasing because the growth of male attainment has either stalled or started to decline and because female attainment has kept increasing. This pattern implies a dramatic change in the composition of skill supply in industrialized economies. Moreover, this change happens at a time when demand-driven changes in the labor market have made the skills acquired through education more important than in the past.

Although increasing returns to education and the growth in overall earnings inequality in industrialized countries over the past three decades is a well-documented trend, the recent literature in labor economics has highlighted some important patterns in the labor market that further underline the importance of skills for labor market outcomes. In a recent survey of the literature, Acemoglu and Autor (2011) show that in addition to the increasing returns to college education, the level of real earnings of low skilled workers has actually declined since the early 1980’s in the United States. At the same time, both wage growth and employment have become “polarized” so that wages at the lower and upper tail of the wage distribution have grown more relative to the middle and the employment shares of low-skill and high-skill occupations have grown whereas the employment share of occupations in the middle of the skill distribution have decreased. Many of these trends, although not identical, are also visible in the European Union as shown by Goos et al. (2009).

What these changes imply is that the wages and employment in traditional middle skill jobs are declining both in the United States and the European Union. These jobs include titles such as sales and office workers, production workers and operatives. According to Acemoglu and Autor (2011), the employment share of these job titles decreased from 57 to 46 percent between 1979 and 2009 in the United States. At the same time, there was a considerable increase in the employment in low-skill service occupations and high-skill managerial jobs.

The most likely explanation for these changes is demand-driven. Rapid technological change, and computerization in particular, has made it possible to substitute technology for labor and this has been most feasible in tasks for which modern computers are suitable. They are typically routine tasks that were traditionally performed by workers in middle-skill occupations. At the same time, technological change has increased the
demand for workers who perform tasks that computers cannot. These involve abstract tasks performed by workers in high-skill tasks as well as manual tasks performed by workers in low-skill tasks.

Figure 8. Male and female labor force participation rates in 1980 and 2007 in the United States and the Nordic countries

![Graph showing labor force participation rates](image)


Naturally, albeit important at the aggregate level, these changes need not imply anything for gender differences. And indeed it is true that, at least in the United States, the employment in middle-skill occupations has decreased for both men and women. However, Acemoglu and Autor (2011) show that the disappearance of middle-skill jobs has had different consequences for male and female employment. Whereas female employment has predominantly shifted to high-skill and highly paid jobs, male employment has shifted to both tails of the occupational distribution. Women have moved upward in the occupation distribution with barely any growth in low-skill employment. On the other hand, the shift for men has been more evenly split with actually a slightly higher increase in low rather than in high-skill employment.

Changes in earnings mirror changes in the employment shares in the United States. For male secondary school graduates, real earnings declined by more than 10 percent between 1979 and 2009 whereas the real earnings of university graduates increased by approximately the same
share. Once more, the change in the female distribution is more positive. There was actually a small increase of 6 percent for female secondary school graduates whereas the real earnings of female university graduates increased by almost 30 percent. Overall, these changes imply a significant worsening of earnings and employment prospects for male workers with mediocre skills.

Although the aggregate changes in the employment shares are likely caused by demand side factors, it seems likely that the differential consequences of these changes by gender are linked to gender differences in educational attainment. Indeed, long-run trends in male and female participation rates seem to support this connection, as is shown in Figure 8. Between 1980 and 2007, male labor force participation declined in all OECD countries except Iceland. At the same time, female participation rates increased in all of these countries. Furthermore, Autor (2010) shows that in the United States, the decline in labor force participation was strongest for men with only secondary school education whose participation rate declined by more than 10 percentage points. It seems likely that these changes in participation rates reflect responses to declining earnings and employment opportunities. The fact that male educational attainment is lagging behind female educational attainment at a time when the prospects for low educated workers are particularly bleak is causing low-educated men to drop out of the labor force.

4. Policy implications

Naturally, it is not entirely obvious that the gender gap in educational attainment should be a policy concern. First of all, the gender gap may simply reflect different rates of growth of educational attainment. Yet, the evidence cited above suggests that in many countries, the widening gender gap in educational attainment reflects a stalling or even declining male educational attainment. In particular, the fraction of male cohorts continuing to post-secondary education has stalled or started to decline in a number of countries. This is a cause for concern since there is a large body of evidence showing that the returns to university education are large and increasing (Acemoglu and Autor, 2011; Van Reenen, 2011).
According to this evidence, a significant fraction of male cohorts are missing out on education that would benefit them.

A second reason for why policies should perhaps not primarily focus on the gender gap is that male gender may simply be a proxy for poor performing students. Any policies that would help students at the lower end of the achievement distribution would necessarily decrease the gender gap in achievement, since boys are overrepresented among the low achievers. While this is partly true, there are reasons to specifically focus on poor performing males. In particular, there is some basis for the argument that the negative externalities from low male educational attainment are large. Lochner and Moretti (2004) show that education has a negative causal effect on male criminality and calculate that a 1 percent increase in male high-school graduation rates in the US would save as much as USD 1.4 billion. Since women are much less likely to commit crimes, low female educational attainment does not generate these kinds of negative externalities. Furthermore, Autor (2010) argues that declining male educational attainment is at least partly responsible for the fragile state of lower-income families in the US.

As showed by the evidence presented above, the gender gap in performance opens up earlier than at university entry. Hence, it is not surprising that existing studies show that women respond more strongly to lowering the costs of university education (Dynarski, 2008) or that women respond more to incentives to perform well at university (Angrist et al., 2009). This is why I will focus on policies that affect performance at the earlier stages of education. There are very few studies evaluating the effects of policies that specifically target the gender gap in achievement. Notable exceptions are studies that focus on the effects of the gender of the teacher and the peer groups on student achievement. Otherwise, there are numerous studies on the effects of policies that attempt to improve the outcomes of poorly performing students. These policies can be classified as policies that change school resources or anticipate educational interventions. Jacob and Ludwig (2008) provide an elegant survey of the overall effects of these kinds of policies. Finally, the structure of the educational system, in particular the timing of school tracking into academic and vocational tracks, has been shown to affect gender differences in educational attainment. Here, we focus on the effects of each of these policies on poorly performing males and the gender gap in achievement.
4.1 Gender of teachers, peer groups and teaching practices

The fact that females are hugely overrepresented in the teacher profession in most industrialized countries has led many commentators to speculate that one of the sexes (which one it is depends on the commentator) is discriminated against in assessments of students. The best available evidence on this issue is mixed. There are some studies that hint at a potential bias against boys. In particular, Lavy (2008) finds that in Israeli high schools, boys score higher grades in blind-tests where the gender of the student is not revealed than in non-blind tests where the grader observes the gender. This result would suggest that boys are discriminated against. However, it is not clear whether the blind and non-blind tests actually test the same thing. In a recent Swedish study, where exactly the same tests are graded blindly and non-blindly, Tyrefors Hinnerich et al. (2011) find no evidence of discrimination against either sex.

There are also signs that single-sex education is attracting more support in some countries. For example, Buchmann et al. (2008) report that in the United States, there were 233 public schools offering gender-separate education in 2006 whereas in 1998 there were just 4. The effect of single-sex education on student achievement is a difficult topic because selection to these schools is far from random. However, the best available studies on the effect of the gender composition of the peer groups (Lavy and Schlosser, 2011; Hoxby, 2000) suggest that single-sex schooling would exacerbate gender differences in attainment rather than reducing them. The achievement of both boys and girls is increasing in the fraction of females in their peer groups.

Changes in teaching practices have also been suggested as a potential strategy for reducing gender gaps in achievement. It is often argued that boys are more sensitive to poor teaching. This concern led to a policy experiment in the UK, where primary schools were given highly structured instructions on teaching objectives and class management for daily “literacy and numeracy hours”. Machin and McNally (2008) is an evaluation of this experiment and they show that this strategy was successful in improving the test scores of both boys and girls. However, the effects were always larger in magnitude for the sex that is generally weaker in the particular subject. Hence, these literacy and numeracy hours were successful in reducing gender gaps in achievement.
4.2 School resources

Whereas the characteristics of the teachers and teaching practices are often only indirectly affected by policies, the school resources are a fairly simple instrument for the politicians to manipulate. Especially reducing class size by increasing the number of teachers is a very attractive policy tool. The effectiveness of these policies, however, has been subject to a rather heated debate in the research literature for a long time. The raw correlation between class size and student achievement is usually very weak and major reductions in class size over time have not led to any major increases in test scores. Still, careful studies of randomized and natural experiments on class size reductions do show that reducing class size may improve student achievement and is also often effective in reducing gender gaps.

The most carefully studied experiment in class size reductions is the STAR experiment which randomly assigned students and teachers to classes of different sizes within schools in Tennessee. The short-term effects of this experiment on test scores were studied by Krueger (1999) whereas Chetty et al. (2011) provide an analysis of long-term outcomes such as completed schooling and earnings. Interestingly, these studies show that reducing class size by one third improved both the short-term (test scores) and long-term (completed schooling) outcomes more for men than for women. However, it is not clear whether the class size reduction was effective in targeting the long-term outcomes of men at the lower end of the achievement distribution. Whereas the results in Krueger (1999) suggest that the experiment also improved the test scores of students from poor families (students receiving free school meals), Chetty et al. (2011) report that the long-term effects were stronger for high income students.

Fredriksson et al. (2011) show similar results from a natural experiment in Sweden where administrative rules create exogenous changes in class sizes. Using this variation, the authors find that class size reductions, also in this case, have stronger effects on men but at the same time they increase the achievement gaps between students from rich and poor families. Hence, even though it seems that reducing class size may be an effective tool for reducing gender gaps in achievement, we cannot be certain whether they are successful in targeting the achievement of poorly performing males which should be the main target group, if we are concerned about the gender gap.
4.3 Early intervention

Early intervention programs have become very popular, especially in countries, such as the United States, where publicly provided pre-school is not widely available. The popularity of the early intervention programs is due to their impressive short-term and long-term outcomes on average. Indeed, the importance of early education is one topic on which most of the researchers in economics of education agree.

However, there has been some controversy regarding the effects of these policies on gender gaps in achievement. Whereas the early studies seemed to suggest that early interventions were more effective for boys, more recent studies that pay more attention to the identification of the effects find that the early intervention programs actually tend to widen the gender gap in schooling outcomes. Ludwig and Miller (2007) focus on the Head Start which is the largest of the American early intervention programs and provides preschool, health and other social services to over 900 000 children. According to their results, there is no consistent support for the claim that Head Start would reduce gender gaps in long-term scholastic achievement. Similarly, Anderson (2008) who looks at much smaller scale Abecedarian, Perry Preschool, and Early Training projects finds that, if anything, these programs increased the gender gap in achievement by significantly improving later-life academic outcomes of women while failing to have any effect on male outcomes.7

The Nordic countries do, of course, have a much longer experience with universal, publicly provided preschool than the US. Therefore, it is not clear how the results of the American studies generalize to Nordic countries. Havnes and Mogstad (2011) analyze the long-term effects of the publicly provided preschool in Norway. When it comes to educational outcomes, the Norwegian preschool benefitted both men and women. However, the effects were slightly larger for men and clearly concentrated on children from low-educated families. Hence, this study strongly suggests that the Nordic style, national scale preschool programs can be effective in targeting the poorly performing males and in increasing their long-term educational attainment.

7 Heckman et al. (2010) is a reanalysis of the Perry Preschool programs which confirms the findings in Anderson (2008) when it comes to academic outcomes. However, Heckman et al. (2010) do find that the program had significant crime reducing effects on men.
4.4 Tracking

As can be seen from the papers cited in this chapter, most of the policy discussion in this literature is from the United States and naturally deals with the type of institutions that are of interest to American policymakers. Therefore, many institutional features that are typical in the European context and may affect gender gaps in achievement are not dealt with in this literature. One important institutional dimension in which European educational systems differ from each other is the timing of tracking in secondary school to academic and vocational tracks. In some countries, such as Germany and Austria, this is done early at ages 10 to 11 whereas in other countries, such as the Nordic countries, the tracking takes place when the students are 15 to 16 years old.

Some authors have speculated that the timing of tracking may explain why the gender gaps in attainment are low in early tracking countries and high in late tracking countries. As we have already seen, Germany and Austria have female/male enrollment ratios of around one in tertiary education. Perhaps even more interestingly, though, according to the OECD (2011), there are large differences across countries in the female/male enrollment ratios in the academic track at secondary school. In Germany and Austria, these ratios are actually below one whereas in the Nordic countries, they vary from 1.06 in Denmark, Norway and Sweden to 1.12 in Iceland. One explanation for these differences is that the gender differences in maturity are larger at ages 15 to 16 when many boys are still going through adolescence than at ages 10 to 11. Pekkarinen (2008) studies the effect of the Finnish comprehensive school reform on the gender gap in educational attainment and finds that moving from an early to a late tracking system exacerbated the gender gap in the likelihood of choosing the academic track in upper secondary school.

5. Conclusion

The rapid growth in female educational attainment is one of the most striking trends in education statistics in the post World War II world. In an increasing number of industrialized countries, female educational attainment is now higher than male educational attainment. Women are in majority among secondary school graduates, among tertiary level stu-
dents, and among tertiary level graduates. Judging from recent trends in international data, it seems likely that the female dominance in educational attainment will become stronger in the coming decades.

Evidence on the returns and costs of education suggests that the emerging female dominance in education is caused by a combination of increasing returns to both men and women and lower female effort costs of education which has meant that the net returns to education, particularly at the higher levels, have increased more for women than for men. Whereas the level of total benefits of education are probably still higher for men, they have increased more for women over the past three decades through the removal of barriers to women’s careers. The effort costs of education, on the other hand, have been lower for women for a long time due to gender differences in non-cognitive abilities.

The fact that the widening gender gap in education partly reflects stalling or even falling male educational attainment is a cause for concern for policymakers since recent trends in the labor market make education ever more important for labor market outcomes. There is plenty of evidence showing that the medium-skill well-paid jobs are rapidly disappearing and that low-educated men are increasingly employed in low-paid low-skill jobs. Furthermore, there are signs that male participation rates are declining across the whole OECD and that this decline is particularly strong for low-educated men.

Although there is scarce evidence on the effects of policies that would directly target gender inequalities in education, one can indirectly draw some conclusions on the effectiveness of different policies. First of all, it seems to be the case that single-sex education is more likely to widen gender gaps in education than to decrease them. However, there are indications that male students are more responsive to school resource investments such as reductions in class size. Early intervention policies, on the other hand, seem to be more effective in improving both female and male long-term outcomes without reducing gender gaps. There is also suggestive evidence that the structure of the educational system, and the timing of tracking in secondary school in particular, affects male and female students differently.
References

Autor, D. (2010), U.S. labor market challenges over the longer term, manuscript, Massachusetts Institute of Technology, Boston.
Dougherty, C. (2005), Why are returns to schooling higher for women than for men?, Journal of Human Resources 40, 969-988.


Jacob, B.A. (2002), Where the boys aren’t: Non-cognitive skills, returns to school and the gender gap in higher education, Economics of Education Review 21, 589-598.


Spelke, E.S. (2005), Sex differences in intrinsic aptitude for mathematics and science? A critical review, American Psychologist 60, 950-958.


Comment on Pekkarinen: Gender differences in education

Anna Sjögren*

Women’s educational attainment has caught up and overtaken that of men. Women are increasingly overrepresented in tertiary education and men are increasingly overrepresented among high school drop outs. Pekkarinen (2012) argues that a reason for this development is a combination of increased returns to education for both men and women, reduced barriers to women’s careers and lower effort costs of education for women. He further argues that the widening gender gap in education is a cause of concern because it is likely to lead to economic inequalities and he also discusses how a number of educational policy measures may reduce or widen the gender education gap.

I agree with much of Pekkarinen’s general discussion of the returns and costs of education, largely based on arguments put forth in the literature by authors such as Becker et al. (2010), Chiappori et al. (2009), Goldin (2006) and Goldin and Katz (2002) regarding the driving forces behind women’s continued investments in higher education. It is harder to square why men increasingly opt out or drop out of education with these broad explanations; if the returns increase, why are not more men opting for higher education? What constrains them?

I much sympathize with the strengths of cross-country comparison in analyzing these universal trends in education of men and women. Yet, some important patterns may be hidden when the analysis is restricted to broad education categories due to lack of detailed data. After discussing some aspects of the gender education gap, I therefore think it is useful to

* IFAU and SOFI, Anna.Sjogren@ifau.uu.se.
complement Pekkarinen’s analysis by taking a closer look at Swedish data. This allows me to consider other margins than the senior secondary and tertiary margins, and it will hopefully shed some further light on male behavior. With some simple graphs, I illustrate how the short-run and long-run consequences of educational choices differ for men and women. I draw the conclusion that policy should not necessarily focus on the gender gap in education per se, but rather on raising achievements and addressing barriers to learning at the bottom of the performance distribution, regardless of gender. Policy should also be concerned with the efficiency of educational investments.

1. Why worry about a growing gender education gap?

At the bottom of the educational distribution, I believe the gender gap to be a symptom of a more serious underlying problem, namely that poor achievers, both boys and girls, have a hard time getting through high school. I agree that poor labor market prospects and increasing social problems for individuals who fail to complete high school and thus enter adulthood with too little human capital is a cause of concern. Although boys are overrepresented among drop outs, I am not convinced, however, that it is wise for policy to focus on reducing the education gender gap per se. Instead, education policy should focus on improving the human capital development of children at risk of falling behind. An important reason for this is that the consequences of poor school performance are dire for both boys and girls. I will present evidence from Sweden which suggests that labor market and social outcomes are particularly weak for women at the lower end of the education distribution. Hence, although, as pointed out by Pekkarinen, the societal costs in terms of criminality and delinquent behavior may be more important for boys, the costs of welfare, teen motherhood and child poverty associated with poor school outcomes for girls are also likely to be important.

At the other end of the educational distribution, the margin for tertiary education, the gender gap is a cause of concern if there are reasons to believe that society is losing out on male talent because men can no longer compete for scarce education resources and if the educational investments in women are for some reason less productive. As long as some of
the costs of education are borne by taxpayers and education is rationed, it is a valid concern that the education system should not promote an inefficient allocation of human capital investments between boys and girls. However, it is not quite clear in which ways education reforms have systematically made schools and universities more favorable to girls over time. Education research is full of results pointing to heterogeneous effects, not only along gender lines, but also over the ability distribution, and across social backgrounds (see e.g. Machin and McNally, 2008; Grönqvist and Vlachos, 2008; Sjögren, 2010).

However, when reforms, and the design of the education system, have a differential impact on boys and girls and thereby have an impact on gender gaps in education, it is important to understand why. Sweden’s very large gender gap in high school graduation opened up as a likely result of some major high school reforms in the 1990’s. However, I will present evidence suggesting that boys and girls were hit in the same way, if we adjust for compulsory school grades. Pekkarinen brings up the impact of tracking age and suggests that reforms delaying tracking may hurt boys relative to girls. Once more, it is important to understand the mechanisms. Do these gender differences stem from a shift in who is actually making the tracking choices or from different maturity gaps between boys and girls at different ages? For instance, the reasons why early tracking in German speaking countries seems to favor academic studies for boys relative to girls may well be related to the fact that parental preferences for their sons’ and daughters’ future education reflect the traditional gender roles in these countries rather than academic maturity or aptitude. Hence, it is far from clear that earlier tracking would narrow the gender education gap if introduced in a society where children are granted large influence over their decisions.

Although there is a growing gender gap in tertiary education in many countries, highly prestigious universities and university programs still attract many men and the gender ratios are relatively balanced. According to the WIA Report (2011a, 2011b), men made up between 47 and 51 percent of the students enrolled at IVY league schools in 2009 and in the top 29 US research universities, 49 percent of all first year students were men in 2011. In the work force, men still outnumber women among top-professionals even if the gender ratios have become more balanced.
Among Swedish 45-year olds, women only constitute about a third of the workforce holding professional degrees.\(^1\)

Analyzing the consequences of a more gender balanced supply of professionals for the organizational choices of firms is an interesting topic for future research. When firms compete for top talent and many of these individuals are women, it may prove rational for firms to adjust the way they organize production. In particular, it may become rational to reduce the productivity loss associated with child bearing. It is already evident that policies regarding parental leave and flexible work hours are important aspects of employment packages considered by young men and women on the job market.

However, empirical evidence suggests that the gender earnings gap is still particularly high for top professionals (Bertrand et al., 2010, for the US and Boschini et al., 2011, for Sweden). An important reason is that over their life cycle, women supply less labor and hence lose out on experience and wage growth. Large penalties for taking time off in professional occupations are thus likely to be one reason why although women are taking over at universities, they are still not dominating the very top.

So far, the evidence does not suggest that society is losing out on top male talent. However, it is interesting to understand why women increasingly dominate less prestigious universities and tertiary education programs and hence make up the majority of university non-professionals. A possibility is that the returns in terms of employment and earnings at this level are more important for women. A reason may be that this part of the education system caters for the education and welfare sectors of the economy, which are female dominated in many countries. The organization of the labor market is also likely to be of importance. In the Nordic countries, large monopsonist public employers can require formal training in a way that the private sector which employs most men has less interest in doing. In what follows, I present some evidence based on Swedish register data on the long-run consequences of the educational choices of men and women. The evidence suggests that while professional degrees pay off very much for men, other non-professional university degrees are much less lucrative options.

\(^1\) Based on data from Boschini et al. (2011). Professional degrees include medicine, law, business and economics, and engineering.
2. Educational choices in Sweden – long-run and short-run consequences

One way of trying to further understand the driving forces behind the widening education gender gap, in particular the reasons why men increasingly opt out of education, is to take a closer look at the long- and short-run consequences of dropping out and at the long-run consequences of various educational choices. In what follows, I do this in a Swedish context. I present some simple descriptive graphs illustrating the long-run and short-run consequences of educational choices. I first consider labor market outcomes at the age of 45 for different educational categories in 2006-2007, i.e. for the cohorts born 1961-1962. I present evidence on the probability of making a decent living and on annual earnings for individuals without high school education, individuals with high school education, more than high school, i.e. shorter vocational or other tertiary education programs, university graduates (at least 15 years of education) with non-professional degrees and university graduates with professional degrees, i.e. medicine, law, economics and business administration, and engineering. The idea is that long-run outcomes associated with different educational choices are likely to influence the educational decisions of younger generations and may thus provide clues to what drives the education gender gap. Second, I also present some evidence on the emerging gender gap in high school graduation and compulsory school grades for the cohorts born 1974-1984. These young cohorts illustrate that the gender education gap is wide already in compulsory school but also that the way in which the education system is organized matters for how this initial gap is manifested in different high school completion rates. I also illustrate some gender differences in the short-run consequences of opting out of high school for young men and women 23 years of age in 2006-2007, i.e. for the cohorts 1983-1984. It needs to be stressed that the graphs are descriptive and should not be interpreted to reflect the causal effect of education on outcomes since selection issues are ignored. The data used is register data from Statistics Sweden.

---

2 I take the average over two years to reduce the influence of annual fluctuations.
3 For a more detailed description of the 45-year old data set, see Boschini et al. (2011) and for the 23-year-old data set, Sjögren and Svaleryd (2012).
2.1 Long-run consequences of educational choices

First, consider the education of men and women in Figure 1.

**Figure 1. The education distribution of 45-year olds in 2006-2007**

![Education distribution chart](image)

Source: Statistics Sweden.

While some 7 percent of men are university educated with professional degrees, only some 4 percent of women belong to this education category. Women are, however, overrepresented among university non-professionals and among those with some education beyond high school. Men, on the other hand, are overrepresented among high school graduates. There is a relative balance among those with only compulsory education.

Next, we take a look at employment and earnings of men and women in Figures 2a and 2b. In particular, in Figure 2.a, I have chosen to show a measure of employment for 45-year olds in 2006-2007, based on the fraction of individual with annual earnings above SEK 100 000, varying by education category. The cut off level SEK 100 000 in 2007 prices is chosen because it roughly corresponds to the national norm for welfare eligibility. Individuals with only compulsory education have a substantially higher probability of having low earnings. For men with compulsory schooling, only around two thirds have annual earnings above the welfare eligibility norm. For women, the fraction having poor earnings is as high as 40 percent in recent cohorts. Note that the divergence between the
compulsory and the secondary school category is more pronounced for women. Furthermore, for men, university educated men with professional degrees stick out and have a far lower probability of having poor earnings than other groups of men. Moreover, for men, there is very little difference in employment between men with some education beyond high school and those having a non-professional university degree. For women, however, university graduates with a non-professional degree do better than women with some post secondary education in terms of employment.

Figure 2. Education, employment and earnings at 45

a) Employment – the fraction of 45-year olds earning at least SEK 100 000 in 2006-2007

b) Earnings of working 45-year olds, SEK ’000 in 2006-2007

Source: Statistics Sweden.
Next consider earnings at the age of 45 among those with positive earnings, displayed in Figure 2b. It is clear that holders of professional degrees earn significantly more than other groups and men have much higher earnings than women. Furthermore, it is interesting to note that also when it comes to earnings, men with some post secondary education make almost as much as men with non-professional university degrees, while both these groups have higher earnings than those who only have secondary education (high school graduates). 4

Some tentative conclusions to draw from these figures are that having low education is associated with relatively worse outcomes for women than for men. Women with only compulsory schooling have a very high probability of having very low earnings at the age of 45 and even among those that work, earnings are very low. Second, while getting some education beyond secondary school seems to pay off for men, completing a non-professional university degree does not make any significant difference. There is evidence of a somewhat greater return at this margin for women both in terms of employment and in terms of annual earnings. Some of these differences in labor market outcomes are likely to stem from the fact that men and women, within these education categories, are not in the same educational fields and hence, face different labor markets.

2.2 Short-run consequences of opting out of high school

Next, I compare labor market and social outcomes of 23 year-olds with and without a high school diploma. First, I show the recent dramatic decline in high school graduation rates. This decline has been analyzed in Hall (2010) and Björklund et al. (2010). The decline is largely explained by some major school reforms in the 1990’s that extended vocational high school programs by one year and by the introduction of absolute grading standards and course requirements.

Figure 3 shows how the Swedish gender gap in high school graduation, which Pekkarinen notes is one of the widest, opened up as high school graduation rates began to drop following the 1990’s reforms. Men were over represented in the two-year vocational programs prior to the

4 The pattern is very similar if average earnings, including those with zero earnings, are considered.
reforms. The extension of these programs therefore hit male graduation rates particularly hard.

**Figure 3. The fraction of 23-year-olds holding a high school diploma**

![Graph showing the fraction of 23-year-olds holding a high school diploma over time.](image)

*Source: Statistics Sweden.*

However, if we take a closer look in Figure 4 at how high school graduation rates of men and women relate to compulsory school grades, it is clear that conditional on the position in the compulsory school GPA distribution, the drop in high school graduation rates is rather similar for boys and girls, and in the most recent cohorts they have very similar graduation rates. In fact, girls in the bottom GPA quintile group fail to complete high school slightly more often than boys. Hence, the absence of a graduation gap before implied that some boys got through high school in spite of very poor compulsory school grades. The figure also shows graduation rates for pupils without complete compulsory school grades.

The performance gap between boys and girls in compulsory school has been stable over time. This is evident from relative grades of boys and girls which were rather stable for the cohorts studied. The average boy ranked about 43 in the GPA percentile distribution while the average girl ranked about 54 or 55. From Figure 5, which shows the fraction of boys in the five quintile groups as well as among those that do not complete compulsory school, it is very clear that boys are over represented in the lower quintiles, while girls are strongly over represented in the top quintiles. There is a weak downward trend in the fraction of boys in the top
quintile group. From the cohorts born in the 1980’s there is, however, a slight tendency that boys are slightly less overrepresented at the bottom, while they are more overrepresented in the second quintile group. The share of boys has also grown among those that have not completed compulsory school. In this category, there is a balance between boys and girls in the more recent cohorts.

**Figure 4. High school graduation rates by quintile of the compulsory school GPA distribution**

![Graph showing high school graduation rates by quintile of the compulsory school GPA distribution](image)

*Source: Statistics Sweden.*

**Figure 5. Fraction of boys at different quintiles in the compulsory school GPA distribution**

![Graph showing fraction of boys at different quintiles](image)

*Source: Statistics Sweden.*
Next, when comparing the outcomes of young men and women with and without a high school diploma, the gender gap in high school graduation rates implies that we are in some respect comparing different selections of men and women. Furthermore, earnings or employment comparisons at this young age should be interpreted with caution since many young high school graduates are still enrolled in post secondary education.

Figure 6. Employment - the fraction of 23-year-olds earning more than SEK 100 000 2006-2007

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>No high school</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school graduate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Statistics Sweden.

Figure 6 shows employment rates of 23 year old male and female high school graduates and drop outs in 2006 and 2007, i.e. the cohorts born in 1983 and 1984. It is interesting to note that, just as for the 45 year-olds, the employment rates are substantially higher for men. While almost half of the male drop outs are employed, only about a third of the female high school drop outs earn more than the welfare norm. Even if many high school graduates are likely to be enrolled in higher education, they are nevertheless more likely to be employed at the age of 23 than high school drop outs. The gap between high school graduates and drop outs is larger for women and hence, the relative advantage of high school graduates compared to drop outs is slightly more important for women.

A similar pattern emerges when I consider the fraction on welfare. Figure 7 shows that the fraction on welfare is low and similar for young men and women with a high school diploma. However, in the group of
high school drop outs, it is more common for young women to be on welfare. This partly reflects that female drop outs are worse selected than male: a larger fraction of each cohort of men drops out of high school. At the same time, Figure 4 showed that selection on school grades is, in fact, similar for boys and girls since the drop out rates were very similar holding compulsory school grades constant.

Figure 7. Fraction of 23-year-olds receiving welfare benefits in 2006-2007

3. Conclusion

I have argued that although the education gender gap is an important phenomenon, our main policy concern should be focused on raising achievements and addressing barriers to learning and at the bottom of the performance distribution, regardless of gender. The main reason is that social and labor market outcomes of both men and women at the bottom of the education distribution are poor. There is evidence that women with low education do particularly poorly.

Because educational investments are fairly gender balanced at the very top, there is still not much evidence that society is losing out of top male talent. There may be more reasons to worry about the efficiency of educational investments in the middle of the distribution, where women dominate greatly.
Balance at the top, male dominance at the bottom, and female dominance in the middle of the education distribution have interesting implications also beyond the labor market. Of particular interest is to understand the long-run effects on fertility and population growth. Changes in the male and female education distributions are likely to influence not only the level and timing of fertility and thus future labor supply, but also the distribution of children, with important implications for both inter- and intragender inequality.

References

Sjögren, A. and Svaleryd, H. (2012), Mental health and school reforms, manuscript, IFAU, Uppsala.
Educating children of immigrants: Closing the gap in Norwegian schools

Bernt Bratsberg, Oddbjørn Raaum and Knut Røed

Summary

Children of immigrant parents constitute a growing share of school cohorts in many OECD countries, and their educational performance is vital for long-term integration. We examine educational outcomes of first and second generation non-OECD immigrants in Norway. Children of immigrants, and particularly those born outside Norway, are much more likely to leave school early than native children. Importantly, this gap shrunk sharply over the past two decades and second generation immigrants are now approaching the educational performance of natives. For childhood immigrants, upper secondary completion rates decline with age at arrival, with a steep gradient after age seven. Finally, we find that immigrant-native attainment gaps disappear when we condition on grade points from compulsory school.

Keywords: Immigrant children, education, school performance.

JEL classification numbers: J15, I21, I24.

* We are grateful to Lena Nekby and the editors for helpful comments. We acknowledge funding from the Norwegian Research Council (grant #202513) and NORFACE (grant #415). The paper is part of the activities of the Centre for the Study of Equality, Social Organization, and Performance (ESOP), University of Oslo. Data made available by Statistics Norway have been essential for this research.

** Ragnar Frisch Centre for Economic Research, bernt.bratsberg@frisch.uio.no, oddbjorn.raaum@frisch.uio.no, knut.roed@frisch.uio.no.
The labor market performance of immigrants plays a central role in the overall economic impacts and the distributional effects of international mobility. The labor market outcomes of immigrants themselves and their children are likely to feed back into the barriers that receiving countries invoke to restrict immigration inflows. Over the last few decades, the immigrant population shares have grown substantially in most high-income countries. Even more striking, the composition of the immigrant population has changed radically with increasing inflows from low-income source countries (Bauer et al., 2000; Blau et al., 2008). In most Northern European host countries, employment and wages of immigrants from low-income countries fall short of those of natives with comparable levels of education and experience.¹

From a long-term perspective, the consequences of immigration will fundamentally depend on the performance of the next generations. Offspring of immigrant parents constitute a significant and steadily increasing fraction of the present and future populations of many of the high-income receiving countries. Human capital accumulation is crucial for adult economic performance and social integration, even outside the labor market. Indeed, the convergence of educational attainment across generations to that of natives is commonly seen as a key indicator of successful integration and several analysts emphasize education as the key pathway for economic integration of immigrants and their descendants (Card and Schmidt, 2003; Card, 2004).

Among immigrant groups from most low-income source countries, the arrival in Europe or North America involves a dramatic change in access to and quality of educational institutions across generations. If the educational environment matters, one would expect children of immigrants to outperform their parents and the association of attainment across generations should be weaker for immigrants than for natives (Dustmann and Glitz, 2011).² Studies that compare educational attainments among first and second generation immigrants confirm these predictions. Algan et al. (2010), for example, show that across the first two generations, most

¹ Recent studies that document large labor market disparities between natives and immigrants from low-income source countries include Algan et al. (2010) for France, Germany and the UK, Nielsen et al. (2003) for Denmark, Barth et al. (2004) and Bratsberg, Raun and Reed (2010) for Norway, Åslund and Rooth (2007) for Sweden, and Sarvimäki (2011) for Finland.
² Tests score gaps between majority and ethnic minority children in the UK decline with age, suggesting a role for teacher and school behavior in explaining ethnic differences in educational attainment (Dustmann et al., 2010).
immigrant groups in France and Germany catch up with natives, and that
the superior attainment of immigrants in the UK is maintained in the se-
cond generation. In the United States, education levels are typically much
lower among immigrants from developing countries than in other groups,
with smaller differences between natives and US born children of immi-
grant parents from these same countries (Card et al., 2000; Trejo, 2003;
Chiswick and DebBurman, 2004). Similar patterns are documented for
the immigrant populations of Canada (Sweetman and Dicks, 1999), Swe-
den (Urban, 2011), and Switzerland (Bauer and Riphahn, 2007).

Across host countries, the differences in educational outcomes by
country of origin are large. Using international achievement data, Schnepf
(2007) and Dustmann et al. (2012) find that test scores measured at ages 9
to 15 of children born to immigrant parents are typically much lower than
those of children of native parents. Comparisons across host countries
show that the gap is wider the larger is the difference in attainment be-
tween immigrant and native parents, reflecting intergenerational correla-
tions in educational outcomes. Schneeweis (2011) concludes that test
score differences between immigrant and native children are smaller in
host countries with an early school starting age. In studies from Northern
and Central Europe, a substantial part of the immigrant-native differential
at age 15 disappears when the authors condition on parental education
and occupational status; see, e.g., Tables 4.2 and 4.3 of Dustmann et al.
(2012) and Table 2 of Liebig and Widmaier (2009).

Because childhood immigrants arrive at different ages, the heteroge-
neity in educational outcomes among child immigrants extends beyond
differences associated with country of origin and parental socioeconomic
position. Late arrival means fewer years of exposure to the higher quality
learning environment and that the prime age for language adaption may
have passed. Using data from Sweden, Böhlmark (2008, 2009) identifies
a substantial negative effect of age at immigration on educational perfor-

---

3 A broader literature studies the intergenerational economic progress of children of US im-
migrants; see, e.g., Borjas (1994; 2006), Card et al. (2000) and Zhou (1997).
4 Several of the papers published in the Journal of Population Economics symposium on
“Second-generation immigrants and the transition to ethnic minorities” address educational attainedments across generations in the immigrant population; see, in particular, Nielsen et al.
(2003), Riphahn (2003), and van Ours and Veenman (2003).
Böhlmark also identifies a critical arrival age of nine years since childhood and youth immigrants who arrive at an older age have substantially lower grade points than their younger peers. Those arriving after the age of nine appear to catch up with the younger arrivals, however, as educational attainment in young adulthood exhibits less co-variation with age at arrival in the Swedish data than grade points at the age of 16 (Böhlmark, 2009).

This paper examines educational outcomes in Norway for children of immigrants from low-income countries, focusing on the importance of being born in Norway relative to arriving as a childhood immigrant. Because children of immigrants from low-income source countries still form a young population group in Norway, our choice of outcome measure is completion of upper secondary education within five years of compulsory education (typically observed at the age of 21). A key question is whether any native-immigrant differential has changed over time, which we study by looking at upper secondary completion rates of the 15 cohorts that graduated from compulsory education between 1990 and 2004. Access to individual level data for two generations enables us to study the transmission of educational attainment across generations. A fundamental challenge is to disentangle the effects of immigrant background and influences from other family-related factors. Many prior studies ignore essential problems associated with controlling for other observable socio-economic background characteristics. For example, we show that imposing a similar coefficient structure on family characteristics (identified by variation across majority group families) exaggerates estimates of the impact of differential family resource environments.

Our most central finding is that there is indeed evidence of the educational performance of children of immigrants gradually catching up with that of native children during our data period. This is particularly evident for Norwegian-born children of immigrant parents, but we also uncover some indication of progress for children born abroad. Not surprisingly, children’s school outcomes are closely related to their parents’ resources, as captured by their earnings or educational attainment. But changes in the composition of these resources cannot explain the relative improve-

---

5 See also the US evidence on age at immigration and adult wages in Bratsberg and Ragan (2002), Gonzalez (2003), and Bleakley and Chin (2004); Cortes (2006) examines age at immigration and test scores of US immigrants.
ment in school performance of children of immigrants; on the contrary, we find evidence that immigrant parents have fallen behind native growth in family earnings, if anything contributing to a negative development in the educational performance of their children. For children born abroad, the age at immigration is of critical importance. In particular, educational performance declines steeply with each year the child does not attend Norwegian schools. Finally, we provide evidence that the probability of completing upper secondary education relates closely to the performance during the final year of compulsory education, as measured by grade points. Conditional on grade points from compulsory school, there is, in fact, virtually no difference between immigrants and natives, and family background becomes much less important (and irrelevant for the children of immigrants). By implication, if policy makers were to succeed in improving the compulsory school performance of children of immigrants, they would also come far in terms of getting immigrant youth successfully through upper secondary education.

1. Overview

In Norway, like in most other Northern European countries, large-scale immigration from low-income countries commenced three to four decades ago. Today’s cohorts with immigrant parents therefore remain relatively young, but the numbers that graduate from compulsory education are now growing rapidly; see Figure 1. In 1980, the share with immigrant parents among youth coming out of Norwegian compulsory schools was less than one percent and dominated by those with parents from the OECD area. Twenty-nine years later, eight percent of the compulsory school graduating cohort consisted of children of immigrant parents born in a non-OECD country. Within the non-OECD group, the composition is rapidly changing as children born in Norway to immigrant parents (second generation), who accounted for less than one percent of the overall graduating cohort as recently as 1995, made up 3.5 percent in 2009 and, based on the resident population as of 2008, is projected to constitute ten percent of the graduating cohort in 2024. In comparison, offspring of immigrants also made up ten percent of all native-born children in the United States in the mid-1990’s (Card et al., 2000). Hence, successful
integration of this group of children in the school system is important, not only because it will facilitate the economic assimilation of the immigrant population, but because it will also affect the overall economy.

Figure 1. Children of immigrants in compulsory school graduating cohorts 1975-2009 and projection 2010-2024

An important “early indicator” of final educational attainment is completion of upper secondary education. Completion of the upper secondary level is viewed as a key to success in the Norwegian labor market, and raising the completion rate is a major policy priority. Upper secondary education is a requirement for enrollment in higher education, and studies find high labor market returns to the upper secondary diploma (e.g., Dagsvik et al., 2011). Since 1994, all children in Norway have been legally entitled to free upper secondary education upon graduation from compulsory school. As a result of the 1994 reform, the fraction of pupils dropping out directly after compulsory education fell from five to three percent (Raaum et al., 2009, Table 1). The statutory duration of upper secondary education is three or four years, depending on academic or vocational track (see also the descriptions of the Norwegian educational system in Fekjær, 2007 and Støren and Helland, 2010). Any discrepancy
Educating children of immigrants: Closing the gap in Norwegian schools

in attainment across groups will, of course, reflect a combination of differences in both entry and completion patterns.

**Figure 2. Share of the compulsory school cohort with completed upper secondary education within five years, by immigrant status**

Many students delay their completion of upper secondary education beyond the statutory duration (see, e.g., Bratsberg, Raaum, Roed and Gjefsen, 2010). For this reason, we will use *completed upper secondary education within five years* (of graduating from compulsory school) as the key educational attainment indicator in our analyses. Figure 2 displays how this indicator has evolved for the compulsory school graduating cohorts from 1990 through 2004 for three different groups: natives (those born in Norway with two native-born parents) and children of immigrant parents born in Norway and abroad, respectively. (The sample construction is explained in detail in Section 2.) Because of high out-migration rates that complicate the study of long-term integration (Bratsberg, Raaum and Sørlie, 2007), and because of their relatively low numbers

---

6 None of our key findings are affected if we instead choose a longer interval such as seven or eight years, but sample sizes are severely reduced as we are forced to drop one graduating cohort from the data for each additional year of the interval.
(refer back to Figure 1), we exclude children of immigrants from pre-1994 OECD member countries (except Turkey).

The figure illustrates a number of important empirical patterns. First, there are significant differences in completion rates between the three groups. The completion rate is much higher for natives than for childhood immigrants, with children born in Norway to immigrant parents somewhere between the two. Second, the completion rate for Norwegian-born children of immigrant parents seems to be catching up with that of native children, particularly during the last 5-10 years of the observation period. And third, the completion rate for those born abroad improved during the latter half of the 1990’s, but has deteriorated after that. A fourth and particularly interesting pattern is that, for natives, the completion rate has been remarkably stable around 70 percent over the 15-year period. Despite growing up in more educated and wealthier families, several education reforms aimed at raising the completion rate, and significant increases in school expenditures, the completion rates among native children hardly budged over the period and remain low when compared to other similar countries (OECD, 2011).

Educational attainment convergence across generations is illustrated in Figure 3, where the center of each circle shows the fraction by source country of immigrant parents and their children who have completed upper secondary education. While natives are located near the 45-degree line with completion rates around 70 percent in both generations, the attainments tend to be significantly higher among immigrant children born in Norway as compared to their parents (Panel A). The improvement is particularly large for groups within low parental attainment, suggesting that the move to a new environment actually enhanced the opportunities for their offspring even if they do not completely match the attainment of children of native parents. For childhood immigrants, the overall educational progress across generations is less pronounced although the figure illustrates significant intergenerational mobility in that educational differences among children are much smaller than among their parents. High intergenerational mobility implies a low association between children’s and parents’ outcomes. In Figure 3, the regression line relating the completion rates of children to those of their parents has a slope of 0.3 in both
panels, showing substantial intergenerational education mobility in the Norwegian immigrant population. 7

Figure 3. Intergenerational mobility by source country. Fraction with completed upper secondary education, immigrant parents and their children

The purpose of the present paper is to shed some light on the mechanisms that have determined the developments shown in Figure 2 and the intergenerational mobility patterns implicit in Figure 3. Policy makers have made large efforts to speed up the integration processes in Norwegian schools; inter alia, by legislating the right to upper secondary education for all adolescents, by providing compensating resources to schools with large shares of ethnic minority children, by offering instruction in the mother language, and by expanding the supply of subsidized institutional child care (there is now 100 percent coverage for children aged above one). In addition, to encourage minority participation, some school districts with particularly large fractions of immigrant children provide

7 Because parental education is measured by the maximum attainment of mother and father, the data in Figure 3 understate the overall educational progress across generations. In their cross-country study of intergenerational mobility, Raaum et al. (2008) find that the correlation coefficient between the educational attainment of mothers and daughters, and that between fathers and sons, is approximately 0.4 in Norway, the Unites Kingdom, and the United States.
pre-school child care free of charge during core hours (Drange and Telle, 2010).

An important question is whether the apparent improvement over time in the schooling performance of immigrant-background children born in Norway shows that such policy efforts eventually pay off. Or can the development illustrated in Figure 2 be explained by other factors, such as better integrated and more resourceful parents, with longer experiences in Norway, or perhaps a secular change in the mix of source countries? Why do immigrant outcomes fall below those of natives in the first place? Is it related to their parents’ performance in the Norwegian labor market or to their country of origin? How important are such factors as age at immigration and the duration of their parents’ stay in Norway?

We address these questions by means of merged administrative register data for the cohorts leaving compulsory education between 1990 and 2004. Prior studies that examine early school leaving in Norway identify immigrant background as a predictor of non-completion of upper secondary education; see Støren and Helland (2010) and Markussen et al. (2011). Both these studies are based on cohorts of upper secondary school entrants or pupils leaving compulsory education around the year 2000. Two distinguishing features of the present study is that our data cover a total of 15 cohorts and that we follow pupils who leave compulsory schooling and will therefore capture any differential dropout in the transition from compulsory to upper secondary education.

2. Data

The data we use in the present paper are based on merged administrative registers, primarily the population register and the national database for education statistics (Vangen, 2007). From the education database, we first extracted records for all children who graduated from compulsory education (i.e., 9th grade) between 1975 and 2009. Then, we merged these records with the central population register, which contains data on country of birth and date of entry supplied by the immigration authorities. The

---

population register also lists immigrant status of parents as well as “country background,” for our purposes defined as own country of birth if born abroad or mother’s country of birth if born in Norway. (The variable gives country of birth of mother’s mother if the mother is Norwegian born and, in some cases, country of birth of father, but we do not use this information in the present study.) From these records, we keep individuals with both parents born abroad, which we label “immigrant background,” and those born in Norway with both parents also Norwegian born, labeled “natives.” Children with one foreign-born and one native parent are not included in the analysis sample. We further split the children of immigrant parents into two groups, those born in Norway (i.e., second generation) and those born abroad. Finally, we drop children with an immigrant background from the OECD area (but keep those from Turkey).

Because of small immigrant sample sizes in the graduating cohorts from the 1970’s and 1980’s (see Figure 1), we restrict our analysis samples to those who left compulsory school after 1989. And, because we track individuals for five years post lower secondary school, the data availability through 2009 limits the samples to the graduating cohorts from before 2005. As a result, the analysis data consist of the 15 cohorts that left compulsory education between 1990 and 2004. We further limit the analysis samples to those who graduated between the ages of 15 and 17 (in fact, 98 percent of the sample turned 16 during the graduation year), and to those who did not spend a full calendar year abroad nor died within five years of compulsory school. For those born abroad, we also require that they must have been present in Norway for at least three years at the time of compulsory graduation, which means that the maximum age at immigration in our sample is thirteen.

The data thus include information about upper secondary school completion (within five years of compulsory education) collected from the education database, gender, and family background. From the income register of the tax authority, we next extracted annual earnings for each of the parents covering the ten-year period ending in the year of graduation from compulsory school, inflated earnings to 2009 values using the base amount of the national pension system, took the average for each parent (accounting for any parental mortality), and computed a measure of parental earnings as the average earnings of mother and father (or as the
earnings of the single parent when only one was present during the ten-year period). For immigrant parents, we followed exactly the same procedure and computed average earnings covering the full ten-year compulsory school period. The exception is parents who arrived in Norway after their child turned seven (who make up 34 percent of the immigrant sample). For these parents, we only use the years that they were actually present in Norway when computing parental earnings.9

We further collected data on educational attainment for parents from the national education database. For native parents, education data typically stem from the records of Norwegian educational institutions or the state educational loan fund. For immigrant parents without Norwegian schooling or student loans, the education data come from various additional sources, including self-reported attainment in immigrant admission and census records as well as two surveys administered by Statistics Norway to all residents with missing education records in 1990 and 1999, and from the agency that certifies foreign education of health workers. We collapse parental education into a single measure based on the highest observed attainment of mother or father. For the last four graduation cohorts (i.e., 2001-2004), we also have access to the grade points obtained in compulsory school. Grade points are given as the total of grades in 11 subjects, each on a scale from 1 to 6.

---

9 The reasons for considering a ten-year period when computing parental earnings are two-fold. First, we seek to measure parental resources for the full duration of compulsory schooling. And, second, using a ten-year period – as opposed to parental earnings the year the child left compulsory schooling, as is common practice in the literature – has been shown to reduce the attenuation bias from measurement error in estimates of intergenerational associations (see, e.g., Bratsberg, et al., 2007). The fact that 34 percent of the immigrant parents arrived after their child reached school age, and therefore are observed for less than ten years, hints that measurement error might be a greater concern for our estimates of effects of parental resources in the immigrant sample than for natives. We find no indication, however, that the estimates are impeded by attenuation bias as coefficient estimates of immigrant parental earnings largely remain unaffected when we restrict the immigrant sample to those observed for at least ten years in Norway. Note also that no parents in our sample are observed for less than three years and that 90 percent of the immigrant parents are observed for at least five years.
Table 1. Descriptive statistics

<table>
<thead>
<tr>
<th></th>
<th>Native children</th>
<th>Children of immigrants born in Norway</th>
<th>Born abroad</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td></td>
</tr>
<tr>
<td>Completed upper secondary education within 5 years of compulsory education</td>
<td>0.706</td>
<td>0.626</td>
<td>0.539</td>
</tr>
<tr>
<td>Female</td>
<td>0.489</td>
<td>0.489</td>
<td>0.482</td>
</tr>
<tr>
<td><strong>Parental earnings</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>336,782</td>
<td>197,601</td>
<td>136,800</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>166,353</td>
<td>141,460</td>
<td>122,821</td>
</tr>
<tr>
<td>Share in 1st decile of native earnings distribution</td>
<td>0.100</td>
<td>0.460</td>
<td>0.654</td>
</tr>
<tr>
<td><strong>Parents’ highest educ. attainment:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary or less</td>
<td>0.095</td>
<td>0.420</td>
<td>0.374</td>
</tr>
<tr>
<td>Lower secondary</td>
<td>0.213</td>
<td>0.076</td>
<td>0.050</td>
</tr>
<tr>
<td>Upper secondary</td>
<td>0.319</td>
<td>0.200</td>
<td>0.226</td>
</tr>
<tr>
<td>Tertiary, first stage</td>
<td>0.274</td>
<td>0.201</td>
<td>0.197</td>
</tr>
<tr>
<td>Tertiary, second stage</td>
<td>0.099</td>
<td>0.057</td>
<td>0.082</td>
</tr>
<tr>
<td>Missing</td>
<td>0.000</td>
<td>0.047</td>
<td>0.070</td>
</tr>
<tr>
<td><strong>Grade points compulsory school</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>43.5</td>
<td>41.6</td>
<td>38.6</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>9.3</td>
<td>9.3</td>
<td>9.6</td>
</tr>
<tr>
<td>Age at immigration</td>
<td></td>
<td></td>
<td>7.4</td>
</tr>
<tr>
<td><strong>Major source countries:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pakistan</td>
<td>0.411</td>
<td>0.095</td>
<td></td>
</tr>
<tr>
<td>Vietnam</td>
<td>0.114</td>
<td>0.119</td>
<td></td>
</tr>
<tr>
<td>Turkey</td>
<td>0.091</td>
<td>0.060</td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>0.081</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td>Morocco</td>
<td>0.052</td>
<td>0.019</td>
<td></td>
</tr>
<tr>
<td>Chile</td>
<td>0.024</td>
<td>0.059</td>
<td></td>
</tr>
<tr>
<td>Iran</td>
<td>0.008</td>
<td>0.092</td>
<td></td>
</tr>
<tr>
<td>Bosnia and Herzegovina</td>
<td>0.006</td>
<td>0.097</td>
<td></td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>709 768</td>
<td>8 975</td>
<td>18 102</td>
</tr>
</tbody>
</table>

Source: Authors’ own tabulations from administrative register data (see Section 2).

Note: Samples consist of youths leaving compulsory education (9th grade) at the age of 15-17 between 1990 and 2004. The native sample is restricted to those with two Norwegian-born parents and immigrant samples to those with two foreign-born parents from non-OECD countries. Samples exclude those who died or spent a full calendar year abroad within five years of compulsory education, and immigrant samples are further limited to those below 14 at entry and with at least three years in Norway at the time of compulsory school graduation. Parental earnings are inflated to 2009 values using the base amount of the Norwegian pension system and are measured during the ten-year period prior to graduation from compulsory school, accounting for any parental mortality. The earnings are averaged across parents. For immigrant parents who arrived before their child reached school age (66 percent of the immigrant sample), earnings are computed using the full ten-year period. For those who arrived after their child turned seven, earnings are averaged across the years they are observed in Norway. Grade points from compulsory school are available for the 2001-2004 cohorts only; observation counts are 186 231, 3 905, and 6 966 for the native and two immigrant subsamples with grade points data, respectively, covering 98 percent of the native and 97 of the immigrant samples in the relevant years.
Descriptive statistics for our analysis samples are provided in Table 1. In total, we have 736,845 observations, of which 8,975 are for those with a non-OECD immigrant background and born in Norway, and 18,102 are for non-OECD immigrant children born abroad. As shown by the table, the rates of upper secondary education completion within five years of compulsory education range from 54 percent for children born abroad to 71 percent for native children. The three groups differ markedly in terms of parental background characteristics, with the immigrant samples tending to have parents with much lower education and earnings compared to the native sample. In fact, fully 65 percent of the immigrant children born abroad and 46 percent of those born in Norway have parental earnings that fall in the bottom ten percent of the native parental earnings distribution. Similarly, immigrant children are overrepresented in the lowest parental education bracket. For the four cohorts with grade point data, the average for natives (43.5) exceeds that of immigrant-background children born in Norway by about two points, whose average again exceeds that of those born abroad by three grade points. Finally, Pakistan, Vietnam, Turkey, Bosnia and Herzegovina, and Iran constitute the major source countries in the immigrant samples.

In sum, the descriptive statistics document important differences in educational outcomes by immigrant status, but also large differences in parental resources. In the next section, we examine the relationships between parental resources and educational outcomes in detail.

3. Regression analysis

In this section, we present results from regression analyses aimed at identifying the determinants of lower and upper secondary school performance among immigrant and native children. We first examine upper secondary school completion for all cohorts in our data set, focusing on the role of family resources and the identification of catching-up trends for immigrant children. Then, we turn to the determinants of school performance at the age of 16 measured by grade points for the last four cohorts of the data period, and also examine the relationship between grade points at the age of 16 and the subsequent completion of upper secondary
3.1 Completion of upper secondary education

In this subsection, we first present the results from linear probability regression models with upper secondary education completion (within five years of compulsory school) as the dependent variable. We focus on four parameters:

- the average differential (over all cohorts) between children born in Norway to immigrant and native parents,
- the additional average differential for children born abroad,
- the average annual change in the impact of immigrant background for those born in Norway (the catching-up rate), and
- the average additional annual change in the impact for immigrant children born abroad.

We estimate these parameters and examine their estimation sensitivity in a series of regressions distinguished by the selection of control variables and the way they are allowed to affect the outcome of interest. Our main results are presented in Table 2.

For the baseline model reported in column (1), we have only included gender and 15 cohort dummy variables in addition to the four variables of direct interest. The cohort dummies are included to control for general time developments in completion rates; i.e., changes in the school system and labor markets that are common to those with native and immigrant backgrounds. As shown by the column, the estimated average difference in completion rates between immigrant-background children born in Norway and natives is 7.9 percentage points, while the difference between children born abroad and natives is 16.6 percentage points (7.9+8.7). Turning to the estimated trend effects, we note that the completion rate for immigrant children born in Norway catches up with that of native children by 0.7 percentage point per year, whereas for those born abroad the catching-up rate is 0.25 (0.69-0.44) percentage points per year. These numbers are significant, both from a substantive and from a purely statistical point of view. When we add 131 dummy variables for country
background to the regression in column (2), it is evident that Norwegian-born children of immigrants perform somewhat better, while immigrants born abroad perform somewhat worse than indicated by the estimates in column (1). In other words, childhood immigrants are, on balance, born in countries with slightly higher completion rates than the immigrant parents of children born in Norway. Importantly, the positive catching-up estimates in column (1) and the positive trends depicted in Figure 2 are not the consequence of a secular change in the source-country composition of the immigrant population.

Columns (3)-(6) list results from regression models where we control for parental resources, in terms of earnings or educational attainment; see the table note for details. In columns (3) and (5), the specification follows the common practice in the literature and imposes the restriction that parental resources play exactly the same role in forming educational outcomes for immigrant and native children (see, e.g., Dustmann et al., 2012; van Ours and Veenman, 2003; Schnepf, 2007). As shown by the columns, controlling for parental resources with the restrictive parameter specification has a huge effect on the estimate of the average impact of having an immigrant background. To illustrate this, the estimated disadvantage of 6.7 percentage points reported in column (2) is turned around to an advantage of as much as 5.1 percentage points in column (3), where we have controlled for parental earnings. Taken at face value, differences in parental earnings do not only explain the disparity of immigrant attainment – when we compare immigrant and native children with similar parental earnings, immigrant children do significantly better than native children. Can we conclude from this that children born in Norway to immigrant parents would actually outperform native children if families were provided with similar economic resources? The answer to this question is no, and the reason why we cannot rely on the results reported in columns (3) and (5) is that the models turn out to be grossly misspecified. The problem is that parental resources do not affect immigrants and natives in the same way. Imposing a common parameter structure will therefore “over-control” for differences in family resources.
Table 2. Upper secondary completion regression results

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immigrant</td>
<td>-0.079</td>
<td>-0.067</td>
<td>0.051</td>
<td>0.003</td>
<td>0.002</td>
<td>-0.038</td>
<td>-0.106</td>
</tr>
<tr>
<td>background</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Immigrant*</td>
<td>-0.087</td>
<td>-0.106</td>
<td>-0.060</td>
<td>-0.076</td>
<td>-0.075</td>
<td>-0.089</td>
<td>-0.031</td>
</tr>
<tr>
<td>born abroad</td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Immigrant*</td>
<td>0.069</td>
<td>0.063</td>
<td>0.083</td>
<td>0.075</td>
<td>0.058</td>
<td>0.068</td>
<td>0.101</td>
</tr>
<tr>
<td>Trend/10</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Born abroad*</td>
<td>-0.044</td>
<td>-0.034</td>
<td>-0.038</td>
<td>-0.037</td>
<td>-0.001</td>
<td>-0.014</td>
<td>-0.023</td>
</tr>
<tr>
<td>Trend/10</td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Controls</td>
<td>Baseline + country</td>
<td>Baseline + country + parental earnings (common coeffs)</td>
<td>Baseline + country + parental earnings (separate coeffs)</td>
<td>Baseline + country + parental education (common coeffs)</td>
<td>Baseline + country + parental education (separate coeffs)</td>
<td>Baseline + family fixed effects</td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ own tabulations from administrative register data (see Section 2).

Note: Standard errors are listed in parentheses. Baseline controls include gender and 15 cohort effects; “country” 131 country-fixed effects; “parental earnings” 10 earnings intervals given by deciles of the native distribution; and “parental education” 5 levels plus missing education. In columns (2)-(7), the coefficient of “immigrant background” is computed as the difference between the sample mean fixed effects in the immigrant and native subsamples. In addition, in columns (4) and (6) the coefficient of “immigrant background” is averaged across ten income and five education levels, weighted by the native frequency distribution. The regressions have 736,845 observations, except column (7) which is based on 512,874 children from 224,077 families with multiple siblings in the data.
In Figures 4 and 5, we show how parental resources are predicted to affect the completion rates of immigrant and native children when the effects are estimated separately for each group. It is clear that whether we look at earnings or education as indicators of parental resources, their effects on the offsprings’ completion rates are much larger for natives than for immigrants. For example, comparing children in the ninth and second deciles of the (native) earnings distribution, native children in the upper tail have a 22 percentage point higher completion rate than those in the lower tail, as compared to a ten percentage point difference among immigrant children. Numerous studies of family background and offspring educational outcomes remind us that correlations between the two only partly reflect causal impacts of parental earnings or education on offspring outcomes (see, e.g., the discussions in Björklund and Salvanes, 2011 and Holmlund et al., 2011). These observed family characteristics are highly correlated with other, perhaps more fundamental, resources, such as parents’ ability — which is again highly correlated with the offspring’s own ability, both through genetic and social transmission mechanisms. Why should this imply a weaker resource gradient in Figures 4 and 5 for immigrant children than for native children? A probable explanation is that the correlation between ability and earnings/education in the parent generation is much weaker for immigrants than for natives, as immigrant parents may not have had the same opportunity to accomplish their potential in the Norwegian labor market as natives, and because the variation in immigrant parental attainment also reflects differences in educational systems of source countries.10 If such explanations are correct, we would expect the difference in gradients to become smaller if we also control for the offspring’s own ability. We return to this issue below.

Figures 4 and 5 show that children of immigrants have higher completion rates than native children at very low levels of parental earnings and education, but lower completion rates at high levels of parental earnings and education. Therefore, it is not clear how one would evaluate the difference between immigrant and native children when controlling for parental resources. One approach is to account for differences in the impacts of parental resources, and to evaluate the differential across the native parental distribution (in other words, create a counterfactual where immi-

---

10 Differences in measurement error in the parental earnings and education variables might also yield weaker correlations for immigrants as compared to natives.
grant children have exactly the same parental resource distribution as native children). This is the approach underlying the results reported in columns (4) and (6) of Table 2. We then find that immigrant-background children born in Norway on average perform approximately as natives when the differences in the distribution of parental resources are controlled for. Moreover, the estimate of the rate of catching-up becomes slightly larger than in the models that do not control for parental resources. The explanation for this finding is that immigrant parents have fallen behind native parents over time in the distribution of resources (particularly their earnings) which, ceteris paribus, has pulled the completion rates of immigrant youths downwards.

**Figure 4. Parental earnings and upper secondary school completion**

Source: Authors’ own tabulations from administrative register data (see Section 2).

Note: Scatter points give the predicted completion rate and the sample mean of parental earnings within ten earnings brackets defined by year-specific deciles of the native earnings distribution. Shaded areas indicate 95 percent confidence intervals around the point estimates. Estimates are based on a regression that controls for gender, 15 cohort effects, 131 country of origin effects, as well as born abroad and trends in immigrant completion (see Table 2, column 4). Immigrant intercept is evaluated at the weighted average of born abroad, time trends, and country effects.
Although the controls for parental earnings and education go a long way in accounting for underlying differences in family resources, unobserved differences between immigrant and native families are bound to remain. One way of controlling even for unobserved family characteristics is to estimate models with family fixed effects (i.e., a dummy variable for each family included in the data set). For this purpose, we obviously identify coefficients using families with more than one offspring in the data (otherwise all differences will be absorbed by the family fixed effect). Such a model can nevertheless be useful for two purposes; first, to obtain an unbiased estimate of the difference between immigrant-background children born in Norway and those born abroad (as some families contain both types of children) and, second, to obtain unbiased estimates of catching-up trends (as we can compare siblings belonging to different graduation cohorts). The results from the model estimated with family fixed effects are reported in column (7). It should be noted that the reported average difference between immigrant-background children born in Norway and natives of 10.6 percentage points now incorporates differ-
ences due to, e.g., country of origin and family resources (i.e., nothing is “controlled away” in the estimated intercepts). The difference between immigrant-background children born in Norway and abroad is much smaller within than across families. To some extent, this reflects the fact that childhood immigrants with siblings born in Norway tend to be relatively young at the time of arrival, and, as we will show below, this greatly improves their educational prospects. The most important result to emerge from column (7), however, is that the catching-up trend is even larger in this model than in the models where we only control for observed family characteristics. Hence, the finding of a rapid convergence in the completion rates for immigrant and native children appears to be a very robust result.

3.2 The role and determination of grade points in compulsory school

As indicated above, an important methodological lesson to be learned from our analysis is that the common practice of “controlling for parental resources” without allowing these resources to affect immigrant and native children differently may lead to a highly misleading inference. A possible reason for the discrepancy in family resource impacts between children of immigrant and native families is that these variables to some extent operate as proxies for unobserved abilities. If this is the case, we would expect the impact of family resources to become weaker and more equal, the more we are able to control for student ability. One way of (imperfectly) controlling for student ability is to condition the analysis of upper secondary school completion on the academic results obtained at the compulsory level. As explained in Section 2, in our data, we have access to grade points obtained in the last year of compulsory school only for the last four cohorts of the sample.

Table 3 presents regression results based on these four cohorts (note that we do not estimate catching-up trends in these models, as the four-year observation period is too short to make such an exercise meaningful).
Table 3. Upper secondary completion, 2001-2004 cohorts

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immigrant background</td>
<td>-.052</td>
<td>.001</td>
<td>-.018</td>
</tr>
<tr>
<td></td>
<td>(.009)</td>
<td>(.008)</td>
<td>(.030)</td>
</tr>
<tr>
<td>Immigrant* born abroad</td>
<td>-.098</td>
<td>-.016</td>
<td>-.021</td>
</tr>
<tr>
<td></td>
<td>(.012)</td>
<td>(.010)</td>
<td>(.038)</td>
</tr>
<tr>
<td>Female</td>
<td>.100</td>
<td>-.021</td>
<td>.008</td>
</tr>
<tr>
<td></td>
<td>(.002)</td>
<td>(.002)</td>
<td>(.005)</td>
</tr>
<tr>
<td>Constant</td>
<td>.663</td>
<td>.193</td>
<td>.309</td>
</tr>
<tr>
<td></td>
<td>(.002)</td>
<td>(.003)</td>
<td>(.009)</td>
</tr>
<tr>
<td><strong>Grade points</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd decile</td>
<td></td>
<td>.201</td>
<td>.151</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.004)</td>
<td>(.010)</td>
</tr>
<tr>
<td>3rd decile</td>
<td>.351</td>
<td>.266</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.011)</td>
<td></td>
</tr>
<tr>
<td>4th decile</td>
<td>.476</td>
<td>.358</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.011)</td>
<td></td>
</tr>
<tr>
<td>5th decile</td>
<td>.585</td>
<td>.459</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.011)</td>
<td></td>
</tr>
<tr>
<td>6th decile</td>
<td>.662</td>
<td>.516</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.012)</td>
<td></td>
</tr>
<tr>
<td>7th decile</td>
<td>.722</td>
<td>.567</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.012)</td>
<td></td>
</tr>
<tr>
<td>8th decile</td>
<td>.764</td>
<td>.578</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.012)</td>
<td></td>
</tr>
<tr>
<td>9th decile</td>
<td>.796</td>
<td>.612</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.012)</td>
<td></td>
</tr>
<tr>
<td>10th decile</td>
<td>.811</td>
<td>.622</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.013)</td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>Cohort +</td>
<td>Cohort +</td>
<td>Cohort +</td>
</tr>
<tr>
<td></td>
<td>country</td>
<td>country</td>
<td>family fixed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>effects</td>
</tr>
</tbody>
</table>

Source: Authors’ own tabulations from administrative register data (see Section 2).

Note: Standard errors are listed in parentheses. The coefficient of “immigrant background” is computed as the difference between the sample mean fixed effects in the immigrant and native subsamples. There are 197,076 observations, except in column (3) which is based on 49,615 children from 24,399 families with multiple siblings in the data. The samples are restricted to students with grade point data and cover 98 percent of the native and 97 percent of the immigrant-background children in the relevant cohorts.

A first point to note from the table is that controlling for grade points completely eliminates the difference in upper secondary school completion between immigrant children (of both types) and native children, even without controlling for family resources. When we also include family resources in these models, they turn out to have very moderate effects for natives, and no effects at all for children of immigrants – and, as ex-
pected, the difference between the two gradients becomes smaller.\textsuperscript{11} The grade point achievements in compulsory school, on the other hand, turn out to have a huge effect on the completion propensity. Comparing the estimated effects of grade points in column (2) (with controls for cohort and origin country only) and column (3) (which also includes family fixed effects), we note that the latter are significantly smaller. We interpret this as evidence that the grade point estimates reported in column (2) to some extent reflect the correlation between grade points and unobserved family resources and therefore contain an upward bias. Yet, the causal effect estimates reported in column (3) remain large; moving from the second to the ninth decile of the grade point distribution raises the completion probability by 46 percentage points.

Note that the coefficient of “female” switches signs between columns (1) and (2) – a pattern that calls for an interpretation. While girls in general have significantly higher completion rates than boys, accounting for their superior performance in terms of grade points at the age of 16 alters the gender differential and, conditional on grade points, boys appear to do better than girls in terms of upper secondary completion; see column (2). The latter result most likely reflects the large and probably inflated estimates of the effect of grade points on upper secondary completion in column (2). When estimated within families, as in column (3), there is no difference in the completion rates between boys and girls once we control for grade points in compulsory school.

The regressions reported in Table 3 build on the assumption that grade points have the same influence on upper secondary completion for immigrant and native children. In light of our finding that family resources have very different effects for the two groups, a legitimate concern is that this restriction might be false. The assumption turns out to be valid, however. This is illustrated in Figure 6, where we show the estimated impacts of grade points when we allow these to differ for children of immigrants and natives (otherwise based on the underlying models in columns (2)

\textsuperscript{11} Due to space concerns, we do not report the coefficients of parental earnings intervals in tables. The estimated gradients between the lower and upper parts of the earnings distribution, represented by the difference between completion rates of children at the ninth and second deciles of the (native) earnings distribution and shown in Figure 4 to be 22 percentage points for native children and ten percentage points for immigrant children, become seven percentage points for native children and zero for immigrant children when we control for grade points.
and (3) of Table 3). The impacts are strikingly similar for immigrants and natives, both with and without family fixed effects included in the model.

Figure 6. Compulsory school grade points and completion of upper secondary education

![Graph showing the relationship between grade points and upper secondary completion]

Source: Authors’ own tabulations from administrative register data (see Section 2).
Note: Scatter points give the predicted upper secondary education completion rate and the mean grade points for each of 10 intervals defined by deciles of the grade point distribution, and shaded areas indicate 95 percent confidence intervals around the point estimates. The estimates in panel A are based on a regression that controls for gender, cohort, born abroad, and country-fixed effects. The regression underlying panel B controls for gender, cohort, born abroad, and family fixed effects. See also Table 3, columns 2 and 3, which impose the additional restriction that the relationship between grade points and completion is the same for immigrants and natives.

What about the relationship between family resources and grade points? Given the argument that observed parental resources form a poorer proxy for ability among immigrants than among natives, we would expect the association between parental resources and the offspring’s grade points to be stronger among natives. This is indeed the case; see Figure 7.

More complete results from grade point regressions are provided in Table 4. In this table, we simply replicate the regression specifications underlying the columns in Table 2 above, only this time with grade points serving as the dependent variable instead of upper secondary school completion. The starting point is once more a significant disadvantage for immigrant-background children (see column 2). Controlling for parental resources with common coefficients (clearly the wrong model; refer back to Figure 7) again shifts the unconditional disadvantage towards a signifi-
cant conditional advantage for Norwegian-born children of immigrant parents, particularly when the model includes parental earnings (column 3). But, using the more appropriate strategy of letting family resource coefficients vary between immigrant and native parents reduces the apparent advantage held by native-born children of immigrants over other groups; see column (4). A robust result to emerge from Table 4 is that childhood immigrants achieve significantly lower grade points at the age of 16 than their peers born in Norway. In the next subsection, we take a closer look at the importance of arriving in Norway as a child, with a particular focus on the role of age at arrival.

**Figure 7. Parental education and compulsory school grade points**

![Graph showing parental education and grade points](image_url)

Source: Authors’ own tabulations from administrative register data (see Section 2).

Note: Scatter points give predicted grade points for each of five levels of parental educational attainment and shaded areas 95 percent confidence intervals around the point estimates. See also notes to Figure 4 and Table 4.
Table 4. Grade points regressions, 2001-2004 cohorts

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immigrant background</td>
<td>-2.005</td>
<td>-2.007</td>
<td>1.031</td>
<td>.450</td>
<td>-567</td>
<td>-1.703</td>
<td>-2.790</td>
</tr>
<tr>
<td></td>
<td>(.145)</td>
<td>(.185)</td>
<td>(.177)</td>
<td>(.212)</td>
<td>(.176)</td>
<td>(.144)</td>
<td>(.518)</td>
</tr>
<tr>
<td>Immigrant* born abroad</td>
<td>-2.900</td>
<td>-2.898</td>
<td>-1.734</td>
<td>-1.948</td>
<td>-1.804</td>
<td>-2.244</td>
<td>-1.467</td>
</tr>
<tr>
<td></td>
<td>(.180)</td>
<td>(.229)</td>
<td>(.217)</td>
<td>(.221)</td>
<td>(.211)</td>
<td>(.212)</td>
<td>(.651)</td>
</tr>
<tr>
<td>Controls</td>
<td>Gender +</td>
<td>Gender +</td>
<td>Gender +</td>
<td>Gender +</td>
<td>Gender +</td>
<td>Gender +</td>
<td>Gender +</td>
</tr>
<tr>
<td></td>
<td>cohort</td>
<td>cohort +</td>
<td>cohort +</td>
<td>cohort +</td>
<td>cohort +</td>
<td>cohort +</td>
<td>cohort +</td>
</tr>
<tr>
<td></td>
<td>cohort +</td>
<td>country +</td>
<td>country +</td>
<td>country +</td>
<td>country +</td>
<td>country +</td>
<td>country +</td>
</tr>
<tr>
<td></td>
<td>country +</td>
<td>parental</td>
<td>parental</td>
<td>parental</td>
<td>parental</td>
<td>parental</td>
<td>parental</td>
</tr>
<tr>
<td></td>
<td>earnings</td>
<td>earnings</td>
<td>earnings</td>
<td>earnings</td>
<td>education</td>
<td>education</td>
<td>education</td>
</tr>
<tr>
<td></td>
<td>(common</td>
<td>(common</td>
<td>(common</td>
<td>(common</td>
<td>(separate</td>
<td>(separate</td>
<td>(separate</td>
</tr>
<tr>
<td></td>
<td>coeffs)</td>
<td>coeffs)</td>
<td>coeffs)</td>
<td>coeffs)</td>
<td>coeffs)</td>
<td>coeffs)</td>
<td>coeffs)</td>
</tr>
</tbody>
</table>

Source: Authors’ own tabulations from administrative register data (see Section 2).

Note: Standard errors are listed in parentheses. Regressions have 197 076 observations, except in column (7) which is based on 49 615 children from 24 399 families with multiple siblings in the data. In columns (4) and (6) the coefficient of “immigrant background” is averaged across ten income and five education levels, weighted by the native frequency distribution. See also notes to Tables 2 and 3.
3.3 Age at immigration

For immigrant children born abroad, prior research based on Swedish and US data shows that the age at which they actually enter the host country is of critical importance. To examine this issue, we have re-estimated the key regressions of the prior sections replacing the “immigrant born abroad” indicator with a complete set of dummy variables for each age at arrival; see Table 5. The listed coefficients give the estimated differential completion rate compared to an immigrant-background child born in Norway. The models are estimated with and without accounting for parental years of residence in Norway, separately for boys and girls and, finally, in the limited sample with and without controlling for grade points from compulsory school.

As shown by the first five columns, the completion rates decline sharply with age at arrival. For example, a childhood immigrant arriving at age 11 faces a 13 to 15 percentage point lower likelihood of completing upper secondary education as compared to a child with an immigrant background but born in Norway. These estimates are robust whether or not we control for parental years since arrival and whether we look at boys or girls. But, as illustrated by column (6), the disadvantage of late arrival is fully accounted for by differences in compulsory school grade points.

One possible explanation for the catching-up in educational performance among immigrant children documented in prior sections and replicated in column (1) of Table 5, is that over time, their parents have become more socially integrated in Norwegian society or have acquired better language skills due to longer residency in Norway. The results in columns (2)-(6) examine this possibility by including parental years since migration in the regression models. As shown by the columns, parental time in the country has a minimal effect on the offspring’s educational performance, and accounting for parental years since arrival does not alter the estimated catching-up trend.
Table 5. The role of age at arrival

<table>
<thead>
<tr>
<th>Age at immigration</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-0.034</td>
<td>-0.035</td>
<td>-0.025</td>
<td>-0.045</td>
<td>-0.028</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.033)</td>
<td>(0.031)</td>
<td>(0.029)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>1</td>
<td>-0.051</td>
<td>-0.053</td>
<td>-0.041</td>
<td>-0.067</td>
<td>-0.070</td>
<td>-0.037</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.016)</td>
<td>(0.025)</td>
<td>(0.022)</td>
<td>(0.024)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>2</td>
<td>-0.059</td>
<td>-0.061</td>
<td>-0.048</td>
<td>-0.069</td>
<td>-0.034</td>
<td>-0.018</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.022)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>3</td>
<td>-0.065</td>
<td>-0.068</td>
<td>-0.062</td>
<td>-0.076</td>
<td>-0.087</td>
<td>-0.037</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.022)</td>
<td>(0.021)</td>
<td>(0.023)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>4</td>
<td>-0.064</td>
<td>-0.067</td>
<td>-0.075</td>
<td>-0.059</td>
<td>-0.075</td>
<td>-0.045</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.015)</td>
<td>(0.021)</td>
<td>(0.020)</td>
<td>(0.023)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>5</td>
<td>-0.080</td>
<td>-0.084</td>
<td>-0.092</td>
<td>-0.074</td>
<td>-0.077</td>
<td>-0.027</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.023)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>6</td>
<td>-0.081</td>
<td>-0.085</td>
<td>-0.096</td>
<td>-0.071</td>
<td>-0.084</td>
<td>-0.034</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.021)</td>
<td>(0.020)</td>
<td>(0.024)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>7</td>
<td>-0.085</td>
<td>-0.090</td>
<td>-0.092</td>
<td>-0.085</td>
<td>-0.088</td>
<td>-0.033</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.015)</td>
<td>(0.021)</td>
<td>(0.020)</td>
<td>(0.024)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>8</td>
<td>-0.100</td>
<td>-0.106</td>
<td>-0.097</td>
<td>-0.114</td>
<td>-0.080</td>
<td>-0.017</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.015)</td>
<td>(0.022)</td>
<td>(0.021)</td>
<td>(0.026)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>9</td>
<td>-0.093</td>
<td>-0.099</td>
<td>-0.120</td>
<td>-0.072</td>
<td>-0.167</td>
<td>-0.045</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.016)</td>
<td>(0.022)</td>
<td>(0.022)</td>
<td>(0.028)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>10</td>
<td>-0.126</td>
<td>-0.132</td>
<td>-0.159</td>
<td>-0.100</td>
<td>-0.146</td>
<td>-0.010</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.016)</td>
<td>(0.023)</td>
<td>(0.022)</td>
<td>(0.029)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>11</td>
<td>-0.148</td>
<td>-0.155</td>
<td>-0.158</td>
<td>-0.153</td>
<td>-0.129</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.016)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.028)</td>
<td>(0.024)</td>
</tr>
</tbody>
</table>
Table 5. Continued....

<table>
<thead>
<tr>
<th>Age at immigration</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>-0.147</td>
<td>-0.154</td>
<td>-0.181</td>
<td>-0.125</td>
<td>-0.140</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.016)</td>
<td>(0.024)</td>
<td>(0.023)</td>
<td>(0.029)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>13</td>
<td>-0.164</td>
<td>-0.171</td>
<td>-0.202</td>
<td>-0.138</td>
<td>-0.133</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.017)</td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.029)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Immigrant backgr</td>
<td>-0.032</td>
<td>-0.021</td>
<td>-0.029</td>
<td>-0.013</td>
<td>-0.002</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.010)</td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.015)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Immigrant backgr*</td>
<td>0.068</td>
<td>0.071</td>
<td>0.075</td>
<td>0.066</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trend/10</td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.018)</td>
<td>(0.017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Born abroad*</td>
<td>-0.041</td>
<td>-0.044</td>
<td>-0.083</td>
<td>-0.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trend/10</td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.023)</td>
<td>(0.022)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years since migr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>parents/10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Observations</td>
<td>736 845</td>
<td>736 845</td>
<td>376 858</td>
<td>359 987</td>
<td>197 076</td>
<td>197 076</td>
</tr>
<tr>
<td>Control for grade points?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Sample</td>
<td>Full</td>
<td>Full</td>
<td>Boys</td>
<td>Girls</td>
<td>2001-2004 cohorts</td>
<td>2001-2004 cohorts</td>
</tr>
</tbody>
</table>

Source: Authors’ own tabulations from administrative register data (see Section 2).

Note: Standard errors are listed in parentheses. The dependent variable is an indicator variable for whether or not the student completed upper secondary education within five years of compulsory education. All regressions control for cohort, country, gender, parental education, and the interaction of parental education and immigrant background (ref Table 2, column 6).
So far, we have not considered whether the empirical patterns of upper secondary education completion might differ by gender (other than noting that there is an overall advantage for girls of about ten percentage points that disappears when we account for grade points from compulsory school; refer back to Table 3). In Table 5, columns (3) and (4), the extended specification of the completion regression is estimated separately for boys and girls. The listed coefficients show that, in general, boys and girls face very similar coefficient structures. For example, the disadvantage of late arrival is quite similar by gender and for neither boys nor girls is there any effect of parental years of residence in Norway.

The one exception is that the, albeit moderate, catching-up trend for childhood immigrants uncovered in prior tables, appears to be limited to girls. In fact, the results in Table 5 indicate that immigrant girls born abroad have benefitted from the same rising trend in upper secondary completion as their sisters born in Norway, whereas boys born abroad have not seen any rising trend in completion rates. If the positive trend in upper secondary completion is linked to early school interventions, these results appear to fit with findings elsewhere in the literature showing that early interventions may have larger effects for girls than for boys (Drange and Telle, 2010; Havnes and Mogstad, 2011). On the other hand, for those with an immigrant background but born in Norway, Table 5 documents positive and statistically significant trend coefficients for both genders and there is no difference between the estimates for boys and girls. This finding shows that the processes behind the more favorable outcomes for immigrant-background children over time are unlikely to be gender related.

Figures 8 and 9 give visual illustrations of the estimated age at immigration effects in the two models explaining upper secondary school completion (columns 6 and 7 of Table 2) and grade point achievement (columns 6 and 7 of Table 4). Once more, all impact estimates are evaluated relative to immigrant-background children born in Norway (represented by the zero-line).

The figures reiterate the finding that age at immigration is of critical importance, with a particularly steep gradient after age seven, which was the school starting age for the cohorts covered in this analysis (the school starting age was reduced to six years in 1997). The right-hand panels of the figures show estimates of age at arrival effects based on within-family
comparisons of siblings (see Figures 8 and 9, panel B, and also Böhlmark, 2008). As is evident from the figures, within-family estimates are very imprecise, and with wide confidence intervals, we are typically unable to reject both the null hypothesis that age at arrival does not matter and the null hypothesis that the true effect of age at arrival is given by the estimate in panel A, computed without family fixed effects. It should be acknowledged that within-family identification of the age-at-immigration effects underlying Figure 8 is thin as long as the model specification contains time trends, as it is difficult to attribute the higher completion rates of younger siblings to their younger age at arrival or to the underlying positive trend in completion rates for immigrant-background children. In fact, when we omit the trend variables from the within-family model, siblings-based coefficient estimates are very similar to those in panel A, paralleling the findings of Böhlmark (2008) based on Swedish data.

Figure 8. Differential upper secondary completion of immigrant children born in Norway and abroad by age at immigration

Source: Authors’ own tabulations from administrative register data (see Section 2).

Note: Scatter points give the estimated difference in upper secondary education completion rates of immigrant-background children born in Norway and those born abroad conditional on age at arrival, and shaded areas indicate 95 percent confidence intervals around point estimates. The estimates in panel A are based on a regression that controls for gender, parental education (five levels plus missing), parental education interacted with immigrant background, 131 country-fixed effects, 15 cohort effects, and time trends for immigrants born in Norway and abroad. The regression underlying panel B controls for gender, cohort, time trends, and family fixed effects. See also Table 2, columns 6 and 7 (which impose the restriction that completion rates do not vary with age at immigration).
3.4 Country of origin

In the analyses so far, country of origin has only entered into our models as dummy control variables. We now briefly examine what the resultant estimated “country-specific effects” look like. Figure 10 shows the scatter plot of the country-fixed effects estimated in the upper secondary school completion regression (Table 2, column 6) and those estimated in the compulsory school grade points regression (Table 4, column 6). Scatter point values are scaled relative to the weighted average in the immigrant data. For example, the scatter point for Pakistan (labeled “PK”) shows that children of Pakistani immigrants score 0.7 grade points below the sample average in the immigrant data, and that their upper secondary education completion rate is 0.033 (3.3 percentage points) below the average for children of immigrants (see also the fixed effect estimates listed in the appendix).

There are three points to note from the figure. The first is that there are huge differences in school outcomes across source countries. For example, the poorest performing groups (Somalia, Kosovo, Morocco, Chile, Iraq, and Turkey) have upper secondary school completion rates
that are nine to twelve percentage points below the overall average for immigrant children, ceteris paribus, while the best performing groups (China, Bosnia and Herzegovina, Vietnam, and Sri Lanka) have completion rates that are 13 to 20 percentage points higher than the immigrant average. A second point to note is that the country-specific fixed effects estimated in the grade point regression exhibit a similar dispersion across source countries and there is a strong positive correlation between the fixed effects estimated in the grade point regression and those estimated in the upper secondary education completion regression. Finally, this relationship is virtually the same across source countries as it is for regression residuals among natives.

**Figure 10. Country of origin heterogeneity**

Source: Authors’ own tabulations from administrative register data (see Section 2).

Note: Scatter points show the country-of-origin fixed effects from regressions presented in Table 2, column (6), and Table 4, column (6). The size of the scatter point reflects cell size. Only cells with at least 30 observations are included.

4. Concluding remarks

Children of immigrants constitute a rapidly growing share of school cohorts in Norway, with the Norwegian-born children of immigrant parents from the non-OECD area projected to account for ten percent of the cohort graduating from compulsory school by the year 2024. The educa-
tional outcomes of immigrant children are important for future growth and inequality. Our study of the 15 cohorts that graduated from compulsory education between 1990 and 2004 shows that children of immigrants five years after compulsory school already have substantially higher educational attainments than their parents, suggesting that moving from a low to a high-income country indeed presents improved opportunities for education. Even so, children of immigrants are more likely to leave school early than children of natives, as the estimated average difference in completion rates from upper secondary education relative to native children is 16.7 percentage points for childhood immigrants and 8.0 percentage points for Norwegian-born children of immigrant parents. However, we find robust evidence of a positive trend for immigrant-background children with a catching-up rate of 0.7 percentage points per year for those born in Norway. Such developments have contributed importantly to a marked reduction in the attainment gap between immigrant and native children over time. In fact, the difference in upper secondary education completion rates among Norwegian-born children of native and immigrant parents declined from 13 percentage points in the 1992-1994 compulsory school graduating cohorts to four percentage points in the 2002-2004 cohorts. For children born abroad, we find that the completion rates decline sharply with age at immigration. This pattern holds for boys and girls, within families (between siblings), and controlling for parental years of residence in Norway.

For the four most recent cohorts of our sample, the data include school performance marks at the end of compulsory education. Children of immigrants obtain lower scores and controlling for grade points completely eliminates the observed differences in upper secondary school completion between immigrant-background and native children, even without accounting for differences in parental earnings and education. The disadvantage of late arrival is also fully accounted for by differences in compulsory school grade points.

Policy conclusions are bound to be speculative as our empirical analyses are not designed to identify the causal effects of educational reforms or policies. Nonetheless, the patterns we uncover do suggest that policy initiatives aimed at improving immigrant children’s educational performance – such as the allocation of extra resources to target schools and to child care institutions in communities with high minority densities – may
actually have contributed to a relative improvement in the educational outcomes of children of immigrant parents. Prior studies provide strong indications that differential resource allocation across schools plays an important compensating role in the Norwegian educational system. Schools with high shares of disadvantaged pupils in terms of family background tend to have higher teacher-pupil ratios than other schools (Hægeland et al., 2004), and schools with many children of immigrant background have a higher incidence of teacher’s aides for pupils with special needs (Hægeland et al., 2009). Further, Hægeland et al. (2008) identify a positive influence of additional school resources on pupil performance at the age of 16, and Havnes and Mogstad (2011) show that the large-scale expansion of subsidized child care in Norway during the late 1970’s and 1980’s had strong positive effects on children’s educational attainment.

Of particular relevance is the finding of Drange and Telle (2010) that the introduction of free core-hour child care in school districts with high immigrant densities led to improved school outcomes of children from immigrant families when compared to developments in districts without free child care. The cohorts under study experienced a marked expansion of subsidized child care institutions nationwide, with child care coverage for three to six year olds increasing from below 30 percent for the oldest cohorts to above 60 percent for the youngest cohorts of our analysis data (Havnes and Mogstad, 2011). It is possible that the child care expansion had a greater impact on subsequent educational outcomes of children of non-OECD immigrant parents than native children, even if immigrant parents use pre-school child care less than native parents. In general, policies and interventions targeted at youth at risk of early school leaving are likely to have disproportionally positive effects for children of immigrants. Consistent with this view, Brinch et al. (2012) conclude that the 1994 reform granting entitlement to upper secondary education for all compulsory school graduates brought about a greater boost in the transition rate from compulsory to upper secondary education for immigrant youth than for native youth. Our empirical evidence therefore broadly harmonizes with an emerging European literature pointing to a role of the educational system in reducing disparities within the immigrant population as well as between children of immigrant and native parents (Dustmann et al., 2010; Schneeweis, 2011).
Issues of integration of immigrants from low-income countries extend beyond schooling, and it remains to be seen whether the positive trends in educational attainment documented in this study are transmitted into more favorable labor market outcomes as growing numbers of immigrant children enter their thirties. Evidence from other countries indicates that, in spite of comparable and even superior education, wages and employment outcomes among children of immigrants often fall substantially below those of their majority-background classmates.

References


Urban, S. (2011), University education as a compensating strategy among second generation immigrants, forthcoming in International Migration Review.
Appendix

Table A1. Labels and scatter point values for major countries listed in Figures 2 and 10

<table>
<thead>
<tr>
<th>Label</th>
<th>Country</th>
<th>Figure 2, panel A (born in Norway)</th>
<th>Figure 2, panel B (born abroad)</th>
<th>Figure 10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cmpl rate child (1)</td>
<td>Cmpl rate prnts (2)</td>
<td>Cell size (3)</td>
</tr>
<tr>
<td>HR</td>
<td>Croatia</td>
<td>0.726</td>
<td>0.427</td>
<td>117</td>
</tr>
<tr>
<td>PL</td>
<td>Poland</td>
<td>0.749</td>
<td>0.890</td>
<td>227</td>
</tr>
<tr>
<td>RU</td>
<td>Russia</td>
<td>0.637</td>
<td>0.750</td>
<td>320</td>
</tr>
<tr>
<td>TR</td>
<td>Turkey</td>
<td>0.494</td>
<td>0.195</td>
<td>815</td>
</tr>
<tr>
<td>BA</td>
<td>Bosnia Herz</td>
<td>0.760</td>
<td>0.540</td>
<td>50</td>
</tr>
<tr>
<td>MK</td>
<td>Macedonia</td>
<td>0.537</td>
<td>0.306</td>
<td>147</td>
</tr>
<tr>
<td>RS</td>
<td>Serbia</td>
<td>0.575</td>
<td>0.552</td>
<td>87</td>
</tr>
<tr>
<td>XK</td>
<td>Kosovo</td>
<td>0.561</td>
<td>0.390</td>
<td>41</td>
</tr>
<tr>
<td>ER</td>
<td>Eritrea</td>
<td>0.695</td>
<td>0.729</td>
<td>59</td>
</tr>
<tr>
<td>MA</td>
<td>Morocco</td>
<td>0.495</td>
<td>0.180</td>
<td>471</td>
</tr>
<tr>
<td>SO</td>
<td>Somalia</td>
<td>0.365</td>
<td>0.364</td>
<td>931</td>
</tr>
<tr>
<td>Label</td>
<td>Country</td>
<td>Cmpl rate child (1)</td>
<td>Cmpl rate prnts (2)</td>
<td>Cell size (3)</td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>LK</td>
<td>Sri Lanka</td>
<td>0.790</td>
<td>0.824</td>
<td>176</td>
</tr>
<tr>
<td>PH</td>
<td>Philippines</td>
<td>0.727</td>
<td>0.936</td>
<td>220</td>
</tr>
<tr>
<td>IN</td>
<td>India</td>
<td>0.769</td>
<td>0.763</td>
<td>727</td>
</tr>
<tr>
<td>IQ</td>
<td>Iraq</td>
<td>0.693</td>
<td>0.893</td>
<td>75</td>
</tr>
<tr>
<td>IR</td>
<td>Iran</td>
<td>0.815</td>
<td>0.531</td>
<td>81</td>
</tr>
<tr>
<td>CN</td>
<td>China</td>
<td>0.586</td>
<td>0.356</td>
<td>3691</td>
</tr>
<tr>
<td>PK</td>
<td>Pakistan</td>
<td>0.516</td>
<td>0.799</td>
<td>219</td>
</tr>
<tr>
<td>TH</td>
<td>Thailand</td>
<td>0.741</td>
<td>0.410</td>
<td>1023</td>
</tr>
</tbody>
</table>

Source: Authors’ own tabulations from administrative register data (see Section 2).

Note: Country-specific fixed effects listed in columns (7) and (8) are based on regressions presented in Table 2, column (6), and Table 4, column (6), and show the deviation from the weighted mean immigrant constant term of the regression.
Comment on Bratsberg, Raaum and Røed: Educating children of immigrants: Closing the gap in Norwegian schools

Lena Nekby*

A better understanding of why there are differences in educational outcomes between native and immigrant children is extremely important not only because a growing proportion of school-age children have a foreign background but also because education is a crucial individual factor for later labor market success. Given the demographic shift towards rapidly ageing populations in the Nordic countries, the successful integration of immigrants and their children in the labor market has become an important policy question. For foreign-born children as well as native-born children of immigrants, successful integration starts within the educational system. As such, this paper is an important contribution to the literature providing new and policy relevant information on the role of family background and how it relates to diverging educational outcomes between children with native and immigrant backgrounds.

The focus of the study is on the role of parental characteristics for children’s educational outcomes. The results show that native-born children with immigrant parents have significantly higher upper-secondary school completion rates than native children with native parents once differences in parental income are accounted for while foreign-born children do as well as native children (with native parents). This result is

* Department of Economics, Stockholm University and SU Linneaus Center for Integration Studies (SULCIS), lena.nekby@ne.su.se.
based on the assumption that parental income has a similar effect on children’s outcomes regardless of parents’ origin (immigrant or native). When the parents’ levels of education are instead controlled for, native-born children with immigrant backgrounds do as well as native children with Norwegian backgrounds while foreign-born children are shown to have significantly lower completion rates.\(^1\)

The assumption that parental characteristics have a similar impact on children’s outcomes regardless of background is shown to be false in the paper. With the exception of very low levels of parental earnings and education, the impact of parental resources on children’s upper-secondary school completion is clearly higher for natives across the distribution of these parental resources. When the influence of parental characteristics is allowed to vary in estimations of upper-secondary school completion, immigrant children born in Norway no longer do better than native children. Instead, children with immigrant backgrounds have similar or slightly lower completion rates than native children while foreign-born children lag behind considerably. The policy relevant question to answer, therefore, is why parental resources have a differential impact on children with native and immigrant backgrounds?\(^2\) In this comment, the role of residential/school segregation on educational outcomes is discussed as well as how the structure of segregation may matter for the transmission of skills between immigrant parents and their children. Potential differences in enrollment in upper-secondary school between native and immigrant children are also highlighted and discussed.

---

\(^1\) Note that these results disregard potential biases that may arise from measurement error due to the fact that education for native parents is determined from register data on highest completed levels of education while for many immigrant parents, education is determined via self-assessed responses to survey questions.

\(^2\) Note that similar levels of education may imply different levels of skills/ability for immigrant and native parents as the majority of immigrant parents have educations acquired abroad, in different countries with different educational institutions, and these educations may not be comparable to similar Norwegian educations. If similar levels of education correspond to different skill levels, then the ceteris paribus comparison in the paper is false to start with due to another type of measurement error. That said, policies aimed at reducing educational gaps between immigrant and native children nonetheless need to address how to compensate children with lower parental resources within the educational system. Examples of such initiatives include extended school hours, after school homework centers, homework tutorials for parents, greater provision of home-language instruction, increased attendance in pre-school, etc.
1. Residential and/or school segregation

The fact that families with immigrant backgrounds tend to live in immigrant-dense neighborhoods and that children of immigrants attend immigrant-dense schools can be a partial explanation for both the remaining unexplained education gaps between native and foreign-born students (after controlling for the differential effects of parental characteristics) and for why parental characteristics have a differential impact on children’s educational outcomes. In Sweden (2007), 70 percent of the students with immigrant backgrounds attend compulsory schools with an over-representation of immigrant students, defined as a larger proportion of immigrants in school than the national average (Böhlmark and Holmlund, 2011). A high immigrant concentration has in several studies been found to have a negative effect on educational outcomes, especially for first generation immigrant students (Åslund et al., 2011; Jensen and Rasmussen, 2011; Szulkin and Jonsson, 2007). Free school choice which may potentially break the correlation between residential and school segregation can, however, result in increased segregation in other dimensions such as parental income, ability and gender. Evidence from the school voucher system in Sweden (established in 1991), which considerably increased the school choice set for students, is mixed. Most of the evidence points to small positive effects on student outcomes while, at the same time, there is increased school segregation implying a larger variation in outcomes between schools which may have an especially detrimental impact on educationally disadvantaged immigrant students in segregated suburbs (Fredriksson and Vlachos, 2011).

Although segregation in the immigrant/native dimension has been shown to be negatively correlated with student outcomes, it is important to note that ethnic concentration, i.e., a higher proportion of co-nationals in the neighborhood, is not problematic for student outcomes. On the contrary, both ethnic concentration and ethnic capital (the average education level of co-nationals in the neighborhood) have been found in a recent study to have a positive causal effect on the educational outcomes of immigrant children (Åslund et al., 2011). A positive effect of ethnic segregation on educational outcomes is found especially for educationally disadvantaged students.
2. Ethnic vs. immigrant segregation

The direct effect of segregation on children’s outcomes points at a trade-off between having many different immigrant groups in a residential area and having many co-nationals in the area. A larger fraction of many different immigrant groups appears to be detrimental for school outcomes while a lower fraction of other immigrant groups and a higher fraction of co-nationals have positive implications. Understanding the nature of residential segregation and how this pattern changes over time can therefore help us understand why foreign-born children have a lower propensity to complete secondary school. Is it, for example, the case that foreign-born children to a larger degree than native-born children with foreign backgrounds live in areas, and attend schools, with a high concentration of many other immigrant groups? Has this pattern changed over time so that sorting is based more on ethnic rather than immigrant lines? Does ethnic concentration rather than immigrant concentration increase with duration of residence for the foreign born? Note that a higher degree of ethnic rather than immigrant concentration in the neighborhood over time is consistent with the positive catch-up rate reported in the paper, as immigrant children reap the benefits of living in increasingly ethnic enclaves.

The effect of segregation on the transmission of skills between parents and children is less understood and less studied. One hypothesis is that there is a trade-off between immigrant and ethnic segregation similar to that found concerning the direct impact of segregation on children’s outcomes. There are a number of potential mechanisms behind a relatively lower transmission of parental resources between parents and children with immigrant backgrounds which may be correlated with the structure of residential segregation; these include social networks/ethnic capital and ethnic identity. Living in an ethnic enclave has been shown to have a positive causal effect on the labor market outcomes of less skilled immigrants (Edin et al., 2006). This positive effect is, in turn, positively associated with the ethnic capital of the enclave, implying that members belonging to, on average, high-income ethnic groups gain more from living in an enclave than members of low-income ethnic groups (Edin et al., 2006). Parents that reap economic and social gains from living in ethnic enclaves may have a higher transmission of skills to children than parents that live in more ethnically divided communities. Parents in ethnic en-
claves may also have access to better information on school relevant issues via ethnic networks and a larger opportunity to develop and participate in community-level programs aimed at helping children through the host country educational system, both of which may be enhanced by lower language barriers within ethnic enclaves.

Living in ethnic (rather than immigrant) enclaves may also foster the ethnic identity of families which, in the cross-cultural psychology literature, has been found to be important not only for individual well-being and self-esteem but also for adaptation to the host country (Berry, 1997; Berry and Sam, 1997; Phinney, 1990). Integrated parents appear to motivate children more than assimilated parents (Portes and Rumbaut, 1990) and immigrant children who maintain a strong ethnic identity have been associated with stronger school performance (Olneck, 1995). Students with immigrant backgrounds that feel integrated, i.e., feel an affinity with both the ethnic group and the majority culture, have also been found to have higher levels of education than other students (Nekby et al., 2009). These mechanisms suggest that living in segregated areas and especially the structure of segregation (ethnic vs. immigrant) may play a role in explaining differential skill transmission from parents to children, especially if the nature of segregation varies between native and foreign-born children with immigrant backgrounds.

3. Lower enrollment rates?

Another policy relevant question is to what degree lower upper-secondary school completion rates for students with immigrant backgrounds are attributable to lower enrollment rates to start with. Students with immigrant backgrounds in Sweden (2007-2008) were considerably less likely to qualify for national secondary school programs than children with Swedish backgrounds (77 percent vs 91 percent) (Taguma et al., 2010). Qualification to national programs does not only concern grades but also the successful completion of core subjects (Math, English and Swedish). Students who do not qualify for national programs are sorted into individual (preparatory) programs with much higher drop-out rates, as high as 75 percent in one follow-up survey of students (Statistics Sweden, 2007). In addition, students in individual programs have a higher proportion of
interruptions and postponements of upper-secondary school educations than students in national programs (28 percent in comparison to 3 percent) and these delays and disruptions have negative consequences for subsequent labor market outcomes (Skolverket, 2006). In order to formulate relevant policy responses to diverging school outcomes between immigrants and natives, it is essential to have a better understanding of where in their educational careers that students with immigrant backgrounds begin to fall behind their native counterparts.

Research tracking the development of skills across the life cycle demonstrates the importance of both cognitive and non-cognitive skills formed early in life in accounting for ethnic (as well as racial and family background) gaps in schooling (Heckman 2006a, 2006b; Nusche, 2009). Investment in early childhood education is therefore a crucial building block for later school achievement, especially for children with disadvantaged/immigrant backgrounds. Research in Sweden confirms that participation in pre-school decreases differences in language ability between children with immigrant and native backgrounds (Fredriksson et al., 2010). These results suggest that encouraging participation in early childhood education and/or daycare among children with immigrant backgrounds and foreign-born children arriving in host countries at young ages is an important policy measure for decreasing later differences in educational outcomes. For this purpose, one may have to factor in how parental leave policies influence participation in early childhood education. In Sweden, for example, parental leave benefits do not depend on whether the children were born in Sweden or not. This implies that for parents immigrating to Sweden with young children, there is the possibility of receiving the (lower) guaranteed level of parental support for any and all children under the age of eight. An unintended consequence of this legislation is that young foreign-born children may be kept at home to a larger degree than native-born children and may enroll in daycare at later ages (Olli Segendorf and Teljosuo, 2011).

Upper-secondary school enrollment involves two choices. The first-order choice is whether or not to enroll. Controlling for differences in grades as well as differences in incomplete grades within core subjects, students with non-European backgrounds in Sweden are less likely to (immediately) enroll in upper-secondary schools than native students (Jonsson and Rudolphi, 2011). The second-order choice concerns what
type of program to enroll in, vocational or academic. At given levels of performance, students with non-European backgrounds have a higher propensity to choose more challenging academic tracks over vocational tracks (Jonsson and Rudolphi, 2011). This is found to be the case across the grade distribution. So although students with non-European backgrounds have lower grades on average and a higher proportion of incomplete grades in core subjects, their educational ambitions are high. Policies aimed at supporting immigrant students with high aspirations but weak results may therefore lead to diminishing educational gaps. Indeed, adult education programs can serve as a second chance for immigrant students and should allow students not only to attain complete grades in core subjects but also to improve earlier acquired grades. High aspirations could then be realized in the labor market via improved access to higher education for immigrant students.

Numerous earlier studies have pointed at the importance of family background for children’s educational outcomes. The importance of this study is in highlighting that similar levels of parental income and education have differential impacts on children’s educational outcomes between immigrants and natives. The challenge is to understand the mechanisms behind these differential transmission rates and formulate relevant policy responses that will allow students with immigrant backgrounds to overcome the disadvantage in parental resources associated with immigration.

References


Heckman, J. (2006b), Skill formation and the economics of investing in disadvantaged children, Science 312, 1900-1902.


The effects of education on health and mortality

Bhashkar Mazumder

Summary

This article reviews recent studies on the effects of education on health and mortality, focusing on studies that might plausibly yield causal estimates. The best evidence from studies that use compulsory schooling laws as a source of identification provides little support for a causal link. Other credible research designs have in some cases yielded promising results of a causal effect, but more research is needed to supplement these studies. Recent comprehensive analyses using either statistical decompositions or a developmental model of how early life endowments give rise to health disparities complement the studies that have relied on narrowly focused experimental evaluations. Another innovation in the literature that shows promise for future analysis includes the use of school quality measures.

Keywords: Education, health, mortality, compulsory schooling, causal effects, school quality.

JEL classification numbers: JEL classification codes: I1, I2.

* I thank Deanna Becker for excellent research assistance and Anders Björklund, Kjell Salvanes and Emilia Simeonova for helpful comments.

** Federal Reserve Bank of Chicago, bmazumder@frbchi.org.
The striking difference in health status and rates of mortality by socioeconomic status is one of the most compelling and well established facts in social science research. Of the various measures of socioeconomic status, the so-called “gradient” in health by completed years of schooling has been found to be particularly robust. This is apparent in Figure 1 which plots the results of a study of eight European countries by Huisman et al. (2005) published in the *Lancet* which found that the age-adjusted mortality risk of individuals with low education (pre-primary, primary or lower secondary education) was between 20 to 40 percent higher than the comparable rate for those with at least an upper-secondary school education. Cutler et al. (2011) have also shown a similar pattern using the same methodology on US data.

Figure 1. Educational disparities in mortality in Europe

Source: From Huisman et al. (2005).

Note: Samples use men and women 45 years or older. Bars represent the ratio of the age adjusted mortality risk based on Poisson regressions, for those with less than an upper secondary school education (ISCE, 0 to 3) to those with an upper secondary education or higher (ISCE, 4 to 6).

The sharp disparities in health by education levels are especially noteworthy given the rising importance of health care costs in national budgets and the fact that the health of a population is generally one of the top priorities for policymakers. According to the OECD, the per-capita expenditures on health have risen by about 50 percent over the last ten years and the health care sector is a growing share of the economy in
many countries (OECD, 2010). Given the evidence shown in Figure 1, it may be tempting for policymakers to conclude that efforts to promote educational attainment may be a primary vehicle to improve the long-term health of societies – that “education policy is health policy”. Naturally, such a conclusion is only warranted if the relationship between schooling and health is causal. In fact, it could be the case that for a variety of reasons (e.g. parental influences or personal attitudes), people who tend to stay in school longer also tend to be in better health. Alternatively, it could be that causality works in the opposite direction, that children in better health stay in school longer. In the absence of evidence of a causal link, there is no reason to expect that policies aimed at increasing educational attainment will result in improvements in health.

Although there is a long history of studies like that of Huisman et al. (2005) that have documented the correlation between schooling and health with observational data, they have typically not utilized research designs that could be used to identify causal effects without imposing very strong assumptions. Therefore, while they may be valuable for descriptive purposes, they are of more limited usefulness for policy discussion. In recent years, researchers have introduced ever more compelling research designs in order to discern whether the relationship is truly causal. This article begins by reviewing some of the recent evidence that uses more credible identification strategies. In particular, I focus on a flurry of recent studies that use changes in compulsory schooling laws in Denmark, France, Germany, the Netherlands, Sweden, United Kingdom and the United States. For example, laws that increase the age at which students must continue to attend school will potentially create sharp discontinuities in educational attainment by birth cohort. Under the seemingly reasonable assumption that other determinants of long-term health and mortality do not exhibit any sharp discontinuities, one can leverage such laws to identify causal effects.

In addition to reviewing the recent evidence generated by leveraging compulsory schooling laws across countries, I also review other recent studies that have utilized other creative research strategies that may also

---

1 See Grossman and Kaestner (1997) or Grossman (2006) for reviews of the older literature. A few studies prior to 2000 did attempt to make stronger causal claims by exploiting sources of variation in education that were considered to be exogenous (e.g. state-level unemployment rates, parent schooling levels), but it is clear that these could very plausibly affect long-term health independently of their effects through schooling.
plausibly generate causal estimates. These include models that only exploit differences in educational attainment between siblings or twins in order to control for many potentially confounding factors that vary across families. A few US studies have also exploited differences in how the military draft affected the college-going rate of some birth cohorts versus others. Other studies have exploited the timing of the availability of new health information to examine how the diffusion of such information may lead to educational disparities in health.

The empirical literature has also broadened and sought to more comprehensively examine the education-health relationship. Several studies have sought to understand exactly how education affects health by considering how the effects vary across specific health outcomes or by directly examining specific pathways linking education and health. Recent work has also sought to explicitly model how early life endowments affect educational choices and how this contributes to educational disparities in health. The literature has also begun to more carefully consider heterogeneity in the effects of education. Perhaps education matters more at certain levels of schooling or for particular population subgroups. I review two recent articles that have used more comprehensive approaches to shed some light on these questions. One notable development is that a number of studies have begun to examine the effects of school quality on health. As was the case with the literature on the returns to schooling, a few studies have tried to better understand what dimensions of schooling (e.g. teachers, resources) are associated with long-term health outcomes.

In the remainder of this article, I briefly summarize some of the leading theories regarding how educational attainment might causally influence health in Section 1. Then, I review the recent empirical studies that utilize changes in compulsory schooling laws to identify causal effects on health in Section 2. In Section 3, I review evidence from other recent studies that use other research approaches for identification. Section 4 discusses two recent studies that have used more comprehensive approaches to generate new insights concerning the pathways underlying the education health gradient. Section 5 discusses new studies that have assessed the effects of school quality improvements on health and Section 6 concludes the paper.

---

2 This review is certainly not meant to be exhaustive and does not cover all recent developments in the literature but seeks to highlight some of the more innovative new studies.
1. The pathways by which education could affect health

“Education ... is a particularly powerful factor in both life expectancy and health expectancy, though truthfully, we’re not quite sure why.”
Richard Suzman, National Institute of Aging, US

The uncertainty regarding exactly how education influences health is not due to a lack of theories as many links have certainly been proposed in the economics literature. These theories often emphasize the role of education in affecting various proximate determinants of health, including financial resources, knowledge and decision-making ability, and other behavioral characteristics that could lead to better health outcomes. Financial resources come into play because better educated individuals may obtain higher paying and more stable jobs and may thereby be able to afford better quality health care and health insurance. With greater economic resources, they may also choose safer and more secure living and work environments and associate with peers who could, in turn, influence health promoting norms and behaviors.

A second explanation is that higher levels of schooling may lead to greater knowledge and an improved ability to process information and make better choices or take better advantage of technological improvements. Some of these ideas are formalized in Grossman’s (1972) model of health production. As I discuss in Section 4, Cutler and Lleras-Muney (2010) argue that actual knowledge about the effects of drinking and smoking can account for some of the education gradients in these behaviors. They also suggest that education may lead to improved cognitive ability which, in turn, can improve the decision-making ability regarding health. Goldman and Smith (2002) suggest that better educated patients may manage chronic conditions better. For example, they may more closely adhere to treatment regimens for human immunodeficiency virus (HIV) infection and diabetes, which can be fairly complex. For such conditions, the ability to form independent judgments and comprehend treatments is important, and is apparently fostered by schooling. Accordingly, Goldman and Smith (2002, pp. 10, 934) state that “self-maintenance is an important reason for the very steep SES [socioeconomic status] gradient in health outcomes.” Glied and Lleras-Muney (2008)

---

3 See Lyman (2006).
argue that “the most educated make the best initial use of new information about different aspects of health”, permitting them to respond more adeptly to evolving medical technologies.

Finally, it could be that education induces other kinds of behavioral changes. For example, the better educated may value the future more than the present compared to those with less education and may therefore take better care of their health (Becker and Mulligan, 1997). Others have argued that education improves people’s perception of their relative status in society and that improved social standing is associated with better health (Marmot, 1994). However, it is important to note that for some of these proposed pathways, the direction of causality could be reversed. In particular, it could be that those with strong cognitive skills or those who are more future-oriented choose to invest more both in their schooling and in their health (e.g. Fuchs, 1982).

2. What can we learn from the use of compulsory schooling laws

2.1 US evidence

Perhaps the most important innovation in the recent literature exploring the effects of education on health has been the use of compulsory schooling laws as a means of potentially identifying causal effects. A brief summary of the various studies described in this section is shown in Table 1. In the US context, Lleras-Muney (2005) was the first to make use of such laws and her study has been particularly influential.4 Specifically, Lleras-Muney uses changes in compulsory schooling and child labor laws across different US states early in the twentieth century that governed the ages at which children were required to attend school. The identification strategy is premised on the idea that variation in these laws induced people born in different states and different years to obtain different levels of schooling for reasons that are plausibly unrelated to other determinants of health. If this assumption holds, one can estimate a causal effect and

---

4 For example, the New York Times featured this study in a provocatively titled article “A surprising secret to a long life: Stay in school” (Kolata, 2007).
The effects of education on health and mortality

overcome the endogeneity concerns that have traditionally plagued the literature.

Table 1. A brief summary of the various studies

<table>
<thead>
<tr>
<th>Country</th>
<th>Study</th>
<th>Description</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>Lleras-Muney (2005, 2006)</td>
<td>Uses compulsory schooling laws affecting cohorts born between 1900 and 1925 to study effects on mortality based on cell level Census data. Uses an IV strategy</td>
<td>IV estimates imply large effects of education on mortality. RD estimates are too imprecise to detect an effect</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>Clark and Royer (2010)</td>
<td>Uses 1947 and 1973 compulsory schooling reforms. RD estimates based on cohorts born a month or a quarter apart. Examine a range of health outcomes and mortality</td>
<td>Finds a consistent pattern of no significant effects across a range of health outcomes including mortality. Estimates are precise enough to reject quantitatively meaningful effects. They do find effects on earnings consistent with other studies.</td>
</tr>
<tr>
<td></td>
<td>Braakmann (2011)</td>
<td>Uses the interaction of the minimum school leaving age along with timing of educational qualification exams among cohorts born between 1957 and 1970.</td>
<td>Finds no effect of education on various health related measures or on health behaviors such as smoking, drinking or food consumption</td>
</tr>
<tr>
<td>Sweden</td>
<td>Meghir, Palme and Simeonova (2012)</td>
<td>Uses the rollout of compulsory schooling increases over 1000 municipalities between 1948 and 1962 using an IV strategy. Use a range of health outcomes including mortality on very large samples based on population registers.</td>
<td>Generally finds no beneficial effects of the reform on long-term health or mortality. Evidence of delayed mortality from age 40 to 49 to age 50 to 59 for men. No corresponding effect for women. Estimates are precise enough to reject quantitatively meaningful effects.</td>
</tr>
</tbody>
</table>

Table 1. Continued....
<table>
<thead>
<tr>
<th>Country</th>
<th>Authors</th>
<th>Methodology</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>Kemptner, Jurges and Reinhold (2011)</td>
<td>Uses compulsory schooling laws implemented at different times by German states between 1949 and 1970 using an IV strategy. Limited health outcomes available in the German census.</td>
<td>Finds that education lowers illness for men but not women. Find little effect on weight and no effect on smoking behavior.</td>
</tr>
<tr>
<td>Denmark</td>
<td>Arendt (2005)</td>
<td>Utilizes national level variation from schooling reforms in 1958 and 1975 to estimate effects on self reported health and BMI from survey data. Uses an IV strategy.</td>
<td>Finds no statistically significant effects, but the standard errors are large and the point estimates suggest beneficial effects of education on health</td>
</tr>
<tr>
<td>Netherlands</td>
<td>van Kippersluis et al. (2011)</td>
<td>Uses a 1928 compulsory schooling reform affecting cohorts born beginning in 1917. Uses tax register data linked to death register data in addition to survey data. Uses RD design.</td>
<td>Finds that a year of education reduces the risk of death by 2 to 3 percent for those surviving to the age of 80.</td>
</tr>
</tbody>
</table>

In order to isolate the variation in years of schooling arising only from the law changes, Lleras-Muney’s main analysis uses a statistical technique known as instrumental variables or “IV”. The maintained assumption is that the compulsory schooling laws only affected long-term health through their effects on education. Census data is used to construct mortality rates for groups of individuals defined by their year and state of birth linked to the prevailing laws at the time of adolescence. The study finds that increased schooling due to these laws led to dramatic reductions in the mortality rates during the 1960’s and 1970’s. In fact, the main IV results (see Table 2) imply that one more year of schooling would lower  

---

5 The study uses the 1960, 1970 and 1980 1 percent Census samples and calculates ten-year mortality rates for cells of individuals defined by state of birth, birth year, gender, and Census year. Mortality rates are defined by the change in the number of individuals in a cell between Censuses divided by the population in the base year. The regressions are weighted by the number of individuals in a cell.
the mortality rate over a ten-year period by nearly 60 percent – a result that is perhaps implausibly large.6

Table 2. US estimates of the effects of education on health using compulsory schooling laws

<table>
<thead>
<tr>
<th></th>
<th>(1) Weighted least squares</th>
<th>(2) Instrumental variables</th>
<th>(3) Number of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimates from Lleras-Muney (2005, 2006) 1960-1980 1%</td>
<td>No age controls, region X cohort</td>
<td>-0.036 (0.004)</td>
<td>-0.063 (0.024)</td>
</tr>
<tr>
<td></td>
<td>Estimates from Mazumder (2008, 2010) 1960-1980 1%:</td>
<td>No age controls, region X cohort</td>
<td>-0.036 (0.004)</td>
</tr>
<tr>
<td></td>
<td>1960 1%, 1970 2%, and 1980 5%:</td>
<td>No age controls, region × cohort</td>
<td>-0.045 (0.004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>With age cubic, region × cohort</td>
<td>-0.039 (0.004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>With age cubic × Census year, region × cohort</td>
<td>-0.039 (0.004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>With age cubic × Census year, state × cohort trend</td>
<td>-0.040 (0.004)</td>
</tr>
<tr>
<td></td>
<td>1960 1%, 1970 2%, and 1980-2000 5%:</td>
<td>With age cubic × Census year</td>
<td>-0.034 (0.003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>With age cubic × Census year, state × cohort trend</td>
<td>-0.035 (0.003)</td>
</tr>
</tbody>
</table>

Although at a first glance, the approach appears to be superior to a standard analysis using observational data, there is reason to be concerned about relying on these laws for identification. This is because the validity of the estimates rests on the assumption that there were no other changes at the state level that were contemporaneous with changes in these laws. During this progressive era in American history, however, it is not difficult to imagine that in the same time period that states were encouraging

---

6 This is based on the revised IV estimate of −0.063 reported in Lleras-Muney (2006).
greater schooling, they may also have been undertaking other policy changes that might have led to long-term health benefits. One example is that many cities introduced food programs in schools, recognizing that compulsory schooling was pointless if children were malnourished. More generally, other public health policies were enacted during this period. For example, Cutler and Miller (2005) show that clean water technologies introduced during this period sharply reduced concurrent mortality.

One approach to assessing the robustness of the Lleras-Muney results to other state-level policies is to include state-specific time trends in the statistical model. In this case, identification arises purely from deviations from broader state trends, thereby capturing effects due to the precise changes in the timing of state laws. Mazumder (2008, 2010) first replicates the Lleras-Muney analysis and then shows that the effect of schooling on reducing mortality disappears once additional controls for state trends are added. This is shown in Table 2. The first row shows the estimates from Lleras-Muney (2006) and the second row shows the results from an attempt to replicate the results. The entries in column (1) show that when using weighted least squares, the coefficient on schooling is -0.036. Using instrumental variables (column 2), the effect rises sharply to around -0.06 to -0.07. In row 3, larger samples are used and the estimates converge to -0.045. Adding age controls and region of birth trends (rows 4 and 5) does not alter the estimates to any considerable extent. However, in row 6, once state-specific birth cohort trends are added, the IV coefficient is 0.003 and is statistically insignificant. The fact that the standard error is 0.038 suggests that the IV approach lacks sufficient power to identify reasonably sized effects once one controls for state-specific trends. In rows 6 and 7 of Table 2, the sample is expanded to include 1990 and 2000 but the same pattern emerges: the inclusion of state-

---

7 Robert Hunter (1904) wrote in the book Poverty: “There must be thousands – very likely sixty or seventy thousand children – in New York City alone who often arrive at school hungry and unfitted to do well the work required. It is utter folly, from the point of view of learning, to have a compulsory school law which compels children, in that weak physical and mental state which results from poverty, to drag themselves to school and to sit at their desks, day in and day out, for several years, learning little or nothing.” In addition to New York, other cities such as Philadelphia, Boston, Milwaukee, Cleveland, Cincinnati, and St. Louis also began large-scale programs to provide food in public schools during the 1900’s and 1910’s (Gunderson, 1971).

8 These results also shift to using data on compulsory schooling laws from Goldin and Katz (2003).
specific trends yields a small, albeit imprecisely estimated, positive IV coefficient.\(^9\)

The lack of a robust finding is not surprising, given that the law changes affected fewer than five percent of the cohorts examined. Further, the study did not detect a statistically significant effect when using an alternative approach known as a regression-discontinuity (“RD”) design that is arguably better suited to identifying the effects of law changes that create sharp discontinuities by birth cohort. The RD method is a conceptually more simple and arguably a more transparent method that simply compares cohorts in a narrow window around the law change. This approach can identify a causal effect under the assumption that other potentially confounding factors do not move discontinuously around the time of the law changes. Although the RD point estimates in Lleras-Muney (2005) suggest slightly lower mortality rates for cohorts born just after law changes that increased schooling, the \(t\)-statistics are only around 0.2. This suggests that there is insufficient power for proper identification using IV or RD methods with this data.

In a different exercise using individual level data from a household survey, Mazumder (2008) does detect positive effects of education on general health status using compulsory schooling laws as instruments. It is troubling, however, that no effects are found for the vast majority of outcomes (e.g. heart disease, kidney disease, hypertension). Further, the fact that beneficial effects are found for certain outcomes such as vision, hearing and back pain suggests that the laws may have affected health through mechanisms unrelated to education. For example, increased access to schooling at earlier ages may have enabled children to obtain vision and hearing treatment and reductions in the use of child labor could explain long-term reductions in back pain. Therefore, even if significant effects on health are identified, the use of early century US compulsory schooling laws may not be valid instruments. Therefore, further evidence is needed before we can conclude much about the causal effects of schooling on health.

---

\(^9\) Adding 1990 and 2000 and tracking the cohorts into later ages when mortality is much higher yields much smaller effect sizes. Further analysis shows that the IV effects (without state trends) are only found for cohorts born before 1915.
2.2 European evidence

In the last ten years, a number of studies that exploit European compulsory schooling laws for identification of the causal effect of education on health have emerged. In addition, to provide evidence from more countries, the effects of law changes in Europe may provide better evidence. Unlike the US where the laws did not affect most students, many students in Europe have historically stopped attending school at the upper age cutoffs. Furthermore, with European data, it is often feasible to link health outcomes at the individual level to compulsory changes in schooling laws for large samples. Finally, the estimated effects might be more relevant for current policy since many of the European law changes occurred after World War II and in a period that coincides less with other public health improvements.

United Kingdom

The increase in the minimum school leaving age from 14 to 15 in 1947 in the UK has provided one of the best opportunities for identifying causal effects. Oreopoulos (2006, 2008) finds that close to half of the students who would otherwise have left school at the age of 14 continued to attend school for an additional year because of the law and that these individuals experienced large gains in log annual earnings. While earnings constituted the main focus of the analysis, Oreopoulos finds no effect on self-reported health in a supplemental analysis using an IV strategy. Clark and Royer (2010) conduct a much more comprehensive analysis of the health effects of UK compulsory schooling laws using a very tightly defined RD comparison. In addition to analyzing the 1947 law change, they also include a change in 1972 that further increased the school-leaving age to 16. The 1972 law had an impact on the educational attainment of about 15 percent of the relevant cohort. Importantly, both laws affected individuals depending on their exact date of birth. The 1947 law change only affected those born after April 1st 1933 and the 1972 law

---

10 Although the results in the original published paper (Oreopoulos, 2006) suggest that the reforms had a positive effect on health, in subsequent results that correct for coding errors (Oreopoulos, 2008), the corrected point estimates are of the wrong sign and are statistically insignificant. The revised results are available in the data archive at the American Economic Review website (http://www.aeaweb.org/aer/data/mar06_data_20030407.zip). Oreopoulos also examines effects of the laws on earnings and health outcomes in Canada and the United States.
affected those born after September 1st 1957. In contrast to some other studies (e.g. Oreopoulos, 2006) that use year of birth to identify the affected cohorts, Clark and Royer use month of birth to more accurately assign treatment status. Clark and Royer’s RD approach is much more conservative in that identification is solely based on the discontinuity among those born just a month apart and essentially removes the effects of continuous trends. For example, the effects of the 1947 law are identified based on a comparison of those born in March 1933 who were untreated to those born in April 1933 who were forced to stay in school until the age of 15. The beauty of this approach is that it is difficult to imagine any other confounding factor that would behave discontinuously between these groups.

Using the RD approach, Clark and Royer first confirm that there are large and very precisely estimated effects on completed education. For example, they show that the fraction of individuals completing nine or fewer years of education fell by 50 percent. They also demonstrate earnings effects that are in line with previous studies. What is striking, however, is the clear absence of a decline in mortality for the affected cohorts. Using a hazard model and using the 1947 law change, they find that for both men and women, the likelihood of death in the 1970 to 2007 period actually increased slightly, rather than declined. However, these effects are very small and statistically insignificant at the five-percent level. They also show that using a cross-sectional model on grouped data by cohort also yields a positive rather than a negative effect of the law change on mortality. Using simulations, they show that the absence of findings is not due to imprecision and that were the true effects to be of the magnitude reported by Lleras-Muney (2005), they would be easily detectable using their methodology. They further show that this pattern is consistent across three categories of death (respiratory, circulatory, and all other causes).

Similarly, they also find no consistent evidence suggesting that the 1972 law changes significantly reduced mortality. This is useful additional evidence since the 1972 law change affected later cohorts that are

---

11 For males, using the hazard model, they find small but statistically insignificant beneficial effects of the laws on mortality. For females, their estimates show beneficial effects that are borderline statistically significant but still small when compared to Lleras-Muney (2005). Using the cross-sectional model and breaking out the estimates by age, they show that for most age groups, the effects on males are positive rather than negative. For females, the effects are highest at the ages of 20-24 and they find no effects on the mortality of 35 to 44 year olds.
more representative of the current population. Further, it affected individuals with a higher level of education than the earlier law. Turning to other outcomes, Clark and Royer find no effects of the laws on: self-reported health; objective health measures such as obesity, being overweight or high blood pressure; and health related behaviors such as smoking, vitamin intake or consumption of fruits and vegetables. Finally, they point out that using year of birth rather than month of birth in an IV framework yields erroneously large beneficial estimates.\footnote{Similar to the original findings in Oreopoulus (2006), Silles (2009) and Powdthavee (2010) also find beneficial health effects when using year of birth in an IV framework. Silles (2009) has also been criticized by Clark and Royer (2010) and Powdthavee (2010) for not adequately controlling for cohort/age trends.}

Another study of the UK by Braakmann (2011) reaches a similar conclusion but uses a different source of variation that is related to compulsory schooling laws. Braakmann follows the empirical approach of Andenberg and Zhu (2010) who exploit differences in the likelihood of obtaining an educational qualification between people born in January versus those born in February to study the effects of education on the likelihood of being married. Among those born between 1957 to about 1970, rules regarding the exact date individuals could actually leave school along with the timing of exams for educational qualifications (e.g. “O-levels” or “CSE”) influenced the likelihood of attaining these qualifications. Braakmann (2011) uses this same identification strategy to estimate the effects of education on health. He first shows that for these cohorts, individuals born in February have a 2-3 percent higher probability of obtaining these credentials and that this is entirely driven by the likelihood of obtaining the lowest possible qualification, while no differences exist at higher qualifications. Braakmann then uses survey data to examine the difference in health effects due to this quirk between individuals born in February versus January. He finds no effect of education on various health related measures (e.g. health limitations, difficulty in seeing or hearing) or on health behaviors such as smoking, drinking or food consumption.

Sweden

While the UK compulsory schooling reforms were sizable and allow for very narrow cohort comparisons across the entire nation, the timing of the
implementation of compulsory schooling increases in Sweden occurred gradually from 1948 to 1962 and varied across more than 1,000 municipalities. This allows for the identification of effects arising from both geographic and temporal variation. The availability of rich administrative data sources also presents the opportunity for estimating very precise effects as compared to estimates relying on survey data. There is also the potential for analyzing how causal effects of education on health might vary by socioeconomic characteristics. A key part of the educational reform was the expansion of the minimum number of years of schooling from seven to nine years.\textsuperscript{13} Although the timing of when a municipality implemented a reform was not random, Holmlund (2007) presents evidence suggesting that the municipality-level characteristics of individuals cannot predict the timing of reforms. This is consistent with the idea that the program was intended to be rolled out in a way that was nationally representative. Meghir and Palme (2005) utilize this variation in order to estimate the causal effects of the reforms on earnings. They find that the reforms increased the educational attainment of males by about a quarter of a year and that of females by about a third of a year. With respect to earnings, they find that while the overall effects were small, there was a substantial heterogeneity in the effects. The earnings gains were largest among those from lower socioeconomic backgrounds.

Meghir et al. (2012) use population registers to assemble data on 1.4 million individuals born in Sweden between 1945 and 1957. These individuals are then linked to the timing of the reform by their municipality at birth. Using personal identifiers, individuals are also matched to register data on education, hospital admissions and cause of death.\textsuperscript{14} The creation of such an extraordinary database is impressive and may represent the most useful data thus far that has been used to address this question. They use standard hazard models to identify the effects of the reform on mortality and control for municipality and year of birth to address any concerns regarding the non-randomness of the reform. They produce separate estimates by gender and also break out effects by age group and the SES status of parents. Like Clark and Royer (2010), they begin by showing

\textsuperscript{13} Other reforms included abolishing the academic tracking system before grade 9 and the implementation of a universal curriculum across all of Sweden.

\textsuperscript{14} They also link individuals to their parents using a multi-generational register which enables them to classify individuals by their parents’ socioeconomic status.
that standard estimates that simply relate mortality to education yield very large estimates.

Overall, their estimates show little evidence of a statistically significant effect of the reform on mortality with many coefficients suggesting an increase rather than a reduction in mortality risk. Further, their design along with the large samples provide sufficient power to clearly rule out any large beneficial effects of the kind found by Lleras-Muney (2005). For example, they can rule out that the reform reduced the risk of male mortality by more than five percent. While they find some evidence that the reform led to a decline in male mortality for 40-49 year olds, this is completely offset by an increase in mortality among men aged 50-59, thus suggesting the possibility of delayed mortality among men. No corresponding patterns are found for women. With respect to hospitalizations, they find a very similar pattern of results with many precisely estimated effects centered around zero that allow them to rule out any quantitatively meaningful beneficial health effects of the educational reform. Overall, Meghir et al. conclude that their results paint a very similar picture to what is found by Clark and Royer (2010).15

France

France increased its minimum school leaving age from 13 to 14 for cohorts born after 1923 and, subsequently, raised the age again to 16 for cohorts born beginning in 1953. Albouy and Lequien (2009) exploit these reforms to study effects on health. Specifically, they use a dataset that began with a one percent sample of the population in 1968 and link individuals to death certificate data in order to measure mortality. This was among the first papers to argue strongly for using an RD approach in the context of compulsory schooling laws and points out the potential pitfalls of using a wide, rather than a narrow range of cohorts for identification. For the earlier law change, they examine mortality at around the age of 80 and for the later law change, they examine mortality around the age of 50. Despite having relatively large samples and a seemingly “clean” identifi-

15 In another study of Sweden, Spasojevic (2010) finds positive effects of the reform on self-reported health and having a BMI in a healthy range, but these results are based on very small samples ranging from 400 to 900 and also rely on one-tailed tests of significance. Further, the first-stage effects of the reform on education are also vastly larger than what was found by Meghir and Palme (2005), thus suggesting that the data might not be totally reliable.
cation strategy, they find no statistically significant effects of either law change on mortality. Even when they broaden their sample to include more cohorts, the estimate remains small and statistically insignificant. They speculate that perhaps the age at which one leaves school might be a poor proxy for the changes in human capital that are relevant for health production. One concern with their analysis is that the “first-stage” effects of the reforms on educational attainment are small and thus may lack sufficient power to estimate effects.

Germany

In Germany, increases in compulsory schooling from eight to nine years along with other educational reforms were gradually enacted throughout the country over the period from 1949 to 1970. Pischke and von Wachter (2008) use this variation to estimate the effects of these reforms on earnings using an IV strategy. Unlike most other countries, they find essentially no effect of these laws on labor market outcomes. They suggest that this may be due to the timing of when job relevant skills are taught in the German system compared to other countries. As a short aside, they show that the reforms were associated with a statistically significant decline in BMI but do not discuss these health effects in any depth. While Pischke and Von Wachter include various controls including state-specific trends in their analysis, there may be some lingering concern about the validity of the research design since the authors do not present direct evidence showing that the timing of the reform across states was unrelated to other factors.

Kemptner et al. (2011) exploit the same variation as Pischke and von Wachter but focus their analysis exclusively on a set of health outcomes available in the German Census. They find that the laws had significant effects in lowering long-term illness for men but not for women. They also find weaker evidence of beneficial effects on weight and little evidence of effects on smoking behavior. Overall, while the study provides some evidence of beneficial health effects, the results are not universally consistent and thus, further evidence for Germany using alternative research designs would be welcome. It would be especially useful for future research to analyze mortality as an outcome.
Braakmann (2010) exploits a different German schooling reform to infer effects on health. The approach is based on Pischke (2007) who exploits changes in 1966 and 1967 that changed the length of the school year by as much as eight months depending on an individual’s state, birth year and the type of school track. Pischke finds no effects of these reductions on employment and earnings. Braakman utilizes the same design on data from the German Socioeconomic Panel and finds no significant effects of this reduction in education on physical health, mental health or BMI. A concern is that the sample sizes are relatively small.

Denmark

In a pair of studies, Jacob Arendt examines the effects of compulsory schooling reforms on health in Denmark. Arendt (2005) was amongst the first studies to exploit compulsory schooling laws for this purpose. He uses school reforms that occurred in 1958 and 1975. The 1958 reform had two features that could have increased educational attainment: it eliminated a test that governed whether students would be tracked to middle school which, in turn, was necessary for more advanced schooling, and it reduced the disparities between schooling in urban and rural areas. Arendt (2005) only utilizes variation at the national level and does not exploit the relative improvement in rural schooling. The 1975 reform increased the compulsory years of education from seven to nine years but, more importantly, also removed a tracking system during the eighth to tenth form. Arendt finds that the 1958 reform increased educational attainment by about 0.4 years but finds no significant effect for the 1975 reform. Rather than using an RD design, Arendt (2005) uses IV and while the point estimates suggest a beneficial effect of schooling on self reported health, they are not statistically significant.

Arendt (2007) only uses the 1958 reform and exploits the relative improvement in educational attainment in rural areas for identification. Further, the study uses a ten percent sample of the Danish population linked to register data on education and hospitalization. As expected, the reform reduced the rural-urban gap in the share of those with an education beyond primary school, with more pronounced effects for women than men. Arendt (2007) finds a reduction in overall hospitalizations for women but not for men. Given the mixed evidence, it would be useful for future re-
search to revisit the Danish reform with an RD design and with a broader set of health outcomes using similarly large samples.

**Netherlands**

Van Kippersluis et al. (2011) exploit a 1928 law in the Netherlands that increased the years of compulsory schooling from six to seven years in order to estimate the effects of education on mortality. They use two samples that are each linked to mortality register data containing cause of death over 1998 to 2005. The first sample (“POLS”) pools household survey data and contains educational attainment. The second sample (“RIO”) is an administrative tax register containing one-third of the Dutch population. Using the POLS, the authors demonstrate that the 1928 reform increased the years of schooling of males by between 0.6 to 1 years. They find no effect for females so they restrict their analysis to males. However, they show that only using the POLS provides insufficient power to identify the effects of education on mortality using an RD design within an IV framework. This motivates them to use the RIO data instead of the POLS data to estimate the reduced form effect of the 1928 reform on mortality. They use a two sample IV procedure which combines the first-stage estimate from the POLS with the reduced form estimate from the RIO. Using this approach yields estimates that are statistically significant and shows that a year of education lowers the mortality of those who survived until the age of 80 by somewhere between two and three percentage points relative to a baseline rate of 50 percent.

The study’s use of very large samples and a convincing RD design arguably provides the best supporting evidence of beneficial effects of compulsory schooling laws on mortality. Nevertheless, there are several reasons why the results may not be fully convincing. To begin with, the magnitude and statistical significance of the mortality results are sensitive to whether they use quarter of birth, where the discontinuity is more precise, instead of year of birth. The lack of a first stage among women is also somewhat surprising and is at odds with the findings in most of the other countries, raising some concerns about the design. Further due to data limitations, the estimates are only for the mortality of those who survived to the age of 80. The authors argue that if anything, this would understate the effects that would be observed for a younger sample. But it
would be useful for future research to confirm this empirically. Finally, this is one of the few studies to use reforms that occurred prior to World War II. The first cohort affected by this reform was born in 1917 and one might be concerned about other historical factors that might be coincident with this reform. Therefore, it would be useful to see results using more recent reforms. Despite these concerns, van Kippersluis et al. (2011) bring new and provocative evidence to the debate.

2.3 Discussion

A close reading of these studies leads to several emerging themes. It is useful to begin by briefly discussing some of those that relate to methodological issues. First, in most cases, it is clear that a regression discontinuity (RD) approach is generally to be preferred to a more traditional IV framework since it imposes the fewest assumptions for identification and reduces the scope for confounding factors such as contemporaneous trends. In that sense, the literature has clearly made some advances in how it utilizes compulsory schooling laws for identification. Second, when possible and when appropriate, researchers should use the month or quarter of birth rather than year of birth to more accurately assign treatment status. Third, to be successful, these strategies clearly require very large samples combined with laws that unambiguously raised the schooling levels.

In my view, the most convincing studies so far are Clark and Royer’s (2010) analysis of the UK and Meghir et al.’s (2012) study of Sweden. The educational reforms in each case have been shown to have robust effects on educational attainment as well as earnings. When the same methodology is applied to the realm of health and mortality, however, there is a surprisingly consistent finding of zero effect. Both studies also use health outcomes other than mortality and investigate differences by cause of death. Albouy and Lequien’s (2009) analysis is also fairly convincing but there is some concern that the reform may not have exerted a strong enough effect on educational attainment, and mortality is the only outcome examined. The results for the Netherlands by van Kippersluis et al. (2011) also appear to be fairly convincing owing to the large samples and the use of an RD design, but require further supporting evidence for
the reasons described earlier (e.g. the use of much older birth cohorts and a finding only for men).

**Figure 2. Lack of access to health care by country**

Of this admittedly selective group of “relatively convincing” studies, three find no effect of compulsory schooling laws on health, and one does. Therefore, my reading of the current state of this literature is that there is not yet any clear evidence to infer that education has a causal effect on health. This conclusion comes with the caveat (that many of the authors are also quick to point out) that these findings may be specific to the contexts studied. The compulsory schooling reforms typically affected the lower end of the educational distribution among cohorts born in Europe after World War II. It is for that reason that we should not exclusively rely on evidence from compulsory schooling laws.

An alternative hypothesis worth considering is that the mixed findings could be due to cross-country differences. Suppose, for the sake of argument, that we also took the findings from Lleras-Muney (2005, 2006) for the US and Kemptner et al. (2011) for Germany at face value, and also included them with van Kippersluis et al. (2011) to the list of studies.
finding beneficial effects of education on health effects. Then, a possible argument could be that education causally affects health in the US, Germany and the Netherlands but not in the UK, Sweden and France. One possible hypothesis to explain the disparate findings is that the health care systems in the latter three countries are more generous or equitable in their treatment of lower socioeconomic groups. In Figure 2, I compare survey data on the lack of access to health care by country, highlighting the data points for the six countries. There is some evidence that two of the countries that show a causal effect of education, US and Germany, also fare worse with respect to access to care. On the other hand, perhaps the most convincing study of a causal effect is from the Netherlands which has the second best access to care behind the United Kingdom. In Figure 3, I compare a measure of health inequality (OECD, 2004) across countries. While this comparison shows that the US has the most inequality, it also shows Sweden as having more inequality than Germany or the Netherlands which would also be inconsistent with this hypothesis. In any event, these comparisons are not meant to be taken especially seriously given the small number of countries and the multitude of other factors which are not controlled for.

---

16 I do not consider Denmark for the analysis that follows since I do not have any data on the measures of health inequality.

17 Specifically, lack of access to care is defined as whether individuals agreed with the following: “Did not fill prescription, did not visit doctor with medical problem, and/or did not get recommended care.”
Figure 3. Inequality in probability of a doctor visit by country


Note: Conceptually, the horizontal index compares the actual observed distribution of medical care by income with the distribution of need. The study cannot address differences in overall provision between countries: it assumes that the average treatment rates for each country, and the average treatment differences between individuals in unequal need, reflect the accepted overall “norm” for that country. In order to statistically equalize needs for the groups or individuals to be compared, we use the average relationship between need and treatment for the population as a whole as the vertical equity “norm” and we investigate to what extent there are any systematic deviations from this norm by income level. Black squares represent countries where there is evidence that compulsory schooling laws have improved health. White squares represent countries where current evidence suggests that compulsory schooling does not improve health. The smaller grey boxes represent countries for which we do not yet have evidence.

3. Other strategies to identify causal effects of education on health

While an analysis of the effects of compulsory schooling laws lends itself to relatively clean identification strategies that do not require strong assumptions, they may only identify effects that are local to the population being treated. Other empirical approaches may have the advantage of identifying effects that apply more broadly to the population or to other margins of the education distribution (e.g. the decision to attend college). In this section, I survey recent papers that use a variety of other approaches.
3.1 Twin comparisons

Social scientists have often attempted to exploit differences within families as a way of controlling for effects that vary across families. For example, labor economists have used variation between twins in order to purge estimates of the effects of schooling on earnings of “ability” bias (e.g. Ashenfelter and Krueger, 1994). In this section, I discuss three recent studies that have introduced this approach to estimate the effects of education on health.

Fujiwara and Kawachi (2009) use a nationally representative sample of US twins containing data on both educational attainment and several health measures including self reported global, physical and mental health, BMI, smoking and physical activity. The study uses only same sex twin pairs and most of the results use samples stratified by sex and zygosity. The baseline models of twin differences contain samples of only 300 to 400 individuals. Further, the number of twin pairs that have discordant levels of education is not reported so it is unclear how much variation the data really has to offer. The authors argue that twin differences “cancel” common factors such as inherited ability, personality traits such as patience, and other early family circumstances and therefore give the estimates a causal interpretation.

The estimates from a model without twin fixed effects yields estimates pointing to statistically significant beneficial effects of education on all health outcomes. Once the authors turn to the twins fixed effects estimator, however, they find that they have insufficient power to identify statistically significant effects although the parameters are typically in the hypothesized direction and are larger than in the models without twin fixed effects. In an effort to improve power, they pool the twin pairs across both sexes but still find that this is still largely insufficient to identify effects. With the pooled sample, they just barely find a statistically significant effect on self reported global health among monozygotic twins but find an effect of the incorrect sign for dizygotic twins.

In addition to lack of power, the study does not address the fundamental criticism raised in the economics literature by Bound and Solon (1999) that twin differences could exacerbate an omitted variable bias rather than
solving it. Bound and Solon raise an important conceptual point which is that one needs to understand the underlying reason for why one twin has a higher educational attainment than the other and whether the cause of this difference is truly exogenous. They suggest that twin differences, in fact, arise as early as in the womb and that these differences can persist. Therefore, even with larger samples, there may be considerable doubt about the validity of a twin fixed effects model.

Lundborg (2008) uses the same data as Fujiwara and Kawachi (2009) and also estimates twins fixed effects models. Not surprisingly, Lundborg’s empirical results are similar to those of Fujiwara and Kawachi. The point estimates suggest that more education leads to better health but the confidence intervals are very wide. However, Lundborg attempts to directly address the Bound and Solon critique by using additional information about the twins such as early life health, birth weight, classroom placement, peer choices and parent child relations to gauge the degree of bias from these sources. He shows that the correlation between these factors and twin differences in education is very small and that the bias arising from twin differences in measures of early life characteristics cannot account for the size of his point estimates. He concludes by pointing to the importance of re-analyzing his model with larger samples from register data.

A third study by Madsen et al. (2010) uses Danish twin registry data. The authors assemble a sample of over 16,000 monozygotic or dizygotic twin pairs born between 1921 and 1950. Individuals were linked to education levels from a demographic database and to mortality records from death registry data covering the 1980 to 2008 period. In contrast to the two other studies reviewed here, they report the number of twin pairs (~2,200) that actually contribute to the variation in education. They estimate Cox hazard models and begin by showing the mortality risk of those with seven or fewer years of education compared to those with more than seven years. Similar to Fujiwari and Kawachi, they show breakdowns by same sex pairs and by zygosity. They also break down the analysis by two groups of cohorts, those born between 1921 and 1935 and those born between 1936 and 1950.

\[^{18}\text{Specifically, Bound and Solon show that if there are unobserved differences between twins that cause them to opt for different years of schooling then a differenced estimator, could actually reduce the share of the total variation that is exogenous.}\]
They begin by showing estimates without using the twin fixed effects and as with the previous studies, they find uniformly beneficial health effects of education. When they turn to the models using twin differences using a dichotomous education breakdown, they do not show any estimates that are statistically significant at the five-percent level and their point estimates generally imply less beneficial effects than the models without twin fixed effects. When they pool across all groups, they obtain estimates suggesting beneficial effects of education that are very close to statistical significance. In their discussion, they note the possibility that their results could be affected by “selection” on twins with a discordant educational attainment and appeal to the findings in Lundborg to argue that early life health factors do not appear to predict twin differences in education. This suggests that more evidence on the life course pattern of twins and a better understanding of the sources of variation within twin pairs is needed before we can be confident in inferences obtained from the twin design.

3.2 Draft avoidance behavior in the US

Card and Lemieux (2001) show that males in the US exposed to the Vietnam era draft were more likely to attend college than those who were unexposed. In this section, I discuss two papers that have exploited this “quasi-experimental” design to infer the effects of attending college on smoking behavior. De Walque (2007) uses data from the National Health Interview Survey (NHIS) to estimate the effects for cohorts of males born between 1937 and 1956. Draft induction risk is measured as the fraction of the male population, by cohort, that was inducted into the military between the ages of 19 to 22 which peaks for those born around 1946. De Walque shows that the draft induction risk is strongly related to the number of years of education attained beyond high school and uses it as an instrument to estimate the effect of education on current smoking and whether individuals stopped smoking.

An important issue with this identification strategy is that a higher induction risk also makes it more likely that one will serve in the military, which may make individuals more likely to smoke. Bedard and Deschénes (2006) show that this was the case for US servicemen during World War II and the Korean War. De Walque addresses this by using
data on veteran status reported in the NHIS as a control. The baseline model using OLS shows that the college educated are less likely to smoke and to stop smoking if they start. For example, the coefficients imply that an additional year of schooling reduces the incidence of smoking by about four percentage points.

Turning to the IV results, de Walque begins by showing that without controlling for veteran status, the effect of education on current smoking is positive and marginally statistically significant. When he controls for veteran status, the coefficient becomes negative and marginally significant, though larger in absolute value (−0.06) than in the baseline model.19

The study also considers a more speculative exercise that interacts the draft induction risk with the risk of death in Vietnam under the assumption that individuals were more concerned about being killed in the war than simply avoiding serving. Using this instrument the results are statistically significant at the five-percent level. De Walque also conducts an exercise where he attempts to instrument for veteran status as well as education, since veteran status can also be endogenous and finds that the results become even stronger. The study notes that the results should be interpreted as a local effect for those who were induced to enroll in college simply in order to avoid the draft and may not be reflective of the effect for the average individual.

In a separate study that is highly complementary to that of de Walque, Grimard and Parent (2007) use the same basic idea with a different data source, the Current Population Survey. They use cohorts of Whites born between 1935 and 1974 who were at least 25 years old when surveyed. In contrast to de Walque, Grimard and Parent use a much blunter instrument: whether individuals were born between 1945 and 1950 and male. Like de Walque, they report baseline OLS estimates suggesting that an additional year of schooling reduces the likelihood of smoking by about four percent. Their IV estimates suggest an even larger effect of about eight percent. They also conduct an exercise where they try to instrument for veteran status and find that their conclusions are largely unaffected.

While these studies provide some new insights, the methodological approach has some limitations that suggest the need for additional studies that leverage other sources of plausibly exogenous changes in college

19 The IV effect falls to about four percent and is comparable with the OLS results when income is controlled for.
attendance. First, the use of instruments based purely on cohort and gender raises the concern that there could be other cohort and gender-specific factors that had independent effects other than through college-going. These cohorts were the leading edge of the baby boom and there may have been common gender-specific experiences that affected those born just after the end of World War II. In addition, there are concerns about whether these studies are adequately able to find exogenous sources of identification for both college attendance and veteran status.

3.3 Diffusion of health information

Aizer and Stroud (2010) examine the effects of the 1964 US Surgeon General’s report on the subsequent smoking patterns of women by education level. This study explores one of the hypothesized channels by which health disparities by education are created, namely the ability of the better educated to “obtain, process and act upon medical knowledge”. The authors hypothesize that the better educated would have responded to the Surgeon General’s report more quickly, leading to an initial increase in health inequality. This initial gap would be expected to decline over time as knowledge of the ills of smoking gradually diffuses to the less educated.

Using public opinion data, they first show that an education gradient in the knowledge of the dangers of smoking did not exist in 1957 but is large in 1969 and continues to grow through at least 1981 before declining by 1990. Then, they use data from the 1959-1966 period to estimate the effect of the Surgeon General report on smoking behavior and birth weight. Specifically, they use data from the National Collaborative Perinatal Project (NCPP), a survey of pregnant women in 12 urban areas who sought medical care from an academic facility. Their total sample consists of about 50,000 women, of whom about 17,500 have had multiple births which makes a fixed effects analysis possible. They find that nearly half of the women smoked during their third trimester.

Simple plots of the data suggest that the rates of smoking and the number of cigarettes smoked per day appear to have diverged somewhat between those with at least a high school degree and those with less than high school beginning in 1964. Using a regression discontinuity analysis (RD), they attempt to quantify the effects more precisely. They find that
prior to 1964, a year of education is associated with a 1.4 percentage point reduction in the likelihood of smoking. Beginning in 1964, this effect rises by 0.6 percentage points to 2 percentage points, which is just statistically significant at the five-percent level. Effects on birth weight, incidence of low birth weight and fetal death are of the correct sign but are small and statistically insignificant.\textsuperscript{20} A nice feature of their data is that for a small subsample, they have third trimester serum cotinine levels so that they can address concerns about self-reported smoking. They show an increase in the education gradient in cotinine levels beginning in 1964. They also show that their results are robust to the inclusion of measures of cognitive ability and income. Finally, they argue that reductions in the gradient in smoking are largest in cities which are the most segregated by education levels. They interpret this finding as a result of peer influences.

Having established that the 1964 report abruptly increased the education gradient in smoking, they turn to estimating long-term trends using data from vital statistics and natality surveys. They find that the gradient was small prior to 1964 and gradually widens until sometime between 1980 and 1990, at which point the gradient weakens to below 1969 levels by 2006, mirroring trends in the gradient in knowledge of the ill effects of smoking.

Price and Simon (2009) examine the effects of an influential study in the \textit{New England Journal of Medicine} published in July of 2001 showing that women undergoing vaginal birth after previously having a C-Section (VBAC) were more likely to experience a uterine rupture than if they had another C-section. While overall, rates of VBAC dropped after the study was published, the authors show that the rates among the college educated dropped by more than those with less than college. For example, among those with a graduate degree, the VBAC rate fell by 22 percent as compared to a fall of just 11 percent among those with less than a high school degree. They do not analyze how the gradient in VBAC evolved over time. They conclude that it is important for future research to consider the mechanisms by which these differences occur.

Overall, these studies of knowledge diffusion point to a useful strategy by which at least one mechanism linking education to health can be em-

\textsuperscript{20} These effects rise to statistical significance if they include women younger than 19 and older than 34.
pirically validated. The accumulation of these kinds of studies may therefore help our understanding of the gradient.

4. Broader explorations of the pathways between education and health

4.1 Decomposing the effects of education on health

A few recent studies in the literature have attempted to take a more comprehensive approach to understanding the connections between education and health, rather than utilizing narrow experimental designs of the kind described thus far. Cutler and Lleras-Muney (2010) explore many of the hypothesized theories linking education to health described in Section 2, using a variety of datasets from the United States along with the National Child Development Survey (NCDS) for the United Kingdom. Their approach is to use a straightforward regression framework where they first regress a variety of health outcomes on education while controlling for basic demographic characteristics along with other variables that “determine education but cannot be affected by it”. These include factors such as parent characteristics and health endowments that vary depending on the dataset used. The coefficient on education represents the education gradient which they claim could be thought of as causal if their controls were exhaustive. Then, they add a set of candidate explanatory variables and quantify the reduction of the education coefficient expressed in percent. To summarize the effects of each explanatory variable across the range of health outcomes, they use a variety of methods including a simple average of the reduction of the gradient across health outcomes.21

They begin their analysis by relating education to a set of eight groups of health-related behaviors: smoking, diet/exercise, alcohol use, illegal drugs, automobile safety, household safety, preventive care and care for people with chronic diseases (hypertension or diabetes) using data from the National Health Interview Survey (NHIS). They show, for example, that in the US, a year of education reduces the risk of smoking by 3 per-

---

21 They use two other methods. They construct weights based on the contribution of each health outcome to mortality reduction and they use equal weights after each health outcome has first been standardized to have a mean of zero and a standard deviation of one.
The effects of education on health and mortality

...percentage points and the probability of being obese by 1.4 percentage points. Then, they consider the extent to which command over resources reduces these education gradients by adding measures of family income, labor market activity, health insurance coverage, geographic measures, family size and marital status. They find that these variables can account for roughly 20 percent of the education gradient across health behaviors with much of the effect accounted for by family income.\(^\text{22}\)

They also consider how knowledge might affect health behaviors. They use a set of questions in the NHIS that asks about respondents’ knowledge of the health risks of smoking and alcohol. They find that knowledge only accounts for between 5 percent and 18 percent of the education gradient.

In order to estimate the effects of cognition, they turn to data from the National Longitudinal Survey of Youth (NLSY) which contains military test scores. They show that the inclusion of these test scores can account for 15 percent of the education gradient in smoking, 9 percent of the gradient in obesity and 10 percent of the gradient in heavy drinking. Overall, they find that knowledge and cognition account for between 5 to 30 percent of the gradient in behaviors.

They also consider how the effects of time preference, risk aversion, how individuals value the future, and the ability to translate intentions into actions impact the education gradients. They find that they explain very little of the education gradients using questions from several surveys that try to elicit these characteristics. In contrast, they find that measures of social integration reduce the education gradients by about 7 percent.

Finally, using the NCDS data from the United Kingdom, they conduct a similar exercise. They find that economic controls explain a somewhat larger share of the education gradient than in the US. For cognitive skills, they find that the inclusion of tests given at age 7, 11, and 16 can explain between 15 and 44 percent of the education gradients. Consistent with the US results, they find that measures of future orientation and personality characteristics explain little of the education gradients but that measures of social integration make a meaningful contribution.

They conclude that using their framework, they can account for a surprisingly large share of the education gradient of between 60 and 80 percent.

\(^\text{22}\) These results are also based on a similar set of calculations using the Health and Retirement Survey (HRS).
cent. They particularly point to the importance of cognitive skills which they suggest explain as much as economic resources.

Cutler and Lleras-Muney (2010) marshal a large amount of data to provide an extremely informative descriptive analysis of the relative importance of various pathways underlying the education gradient. While some conclusions such as the relative importance of resources and knowledge seem to fit the conventional wisdom, other findings such as the importance of cognitive skill and the lack of importance of patience and forward looking behavior are more striking. These findings may help advance the literature by pointing to which potential pathways may be of first-order importance. Naturally, an important caveat is that the analysis is not set in a causal framework and so efforts to conduct similar decompositions using convincing sources of identification is an important goal for future research.

4.2 Developmental framework for understanding the gradient

One drawback of a regression based accounting framework is that it does not provide a systematic model for understanding the potential dynamic effects of changes to health inputs at particular stages of the life cycle and, as such, may understate or overstate some factors. For example, if certain personality traits revealed early in life shape investments in cognitive skill and formal schooling, then it is unclear how successfully the regression framework will parse the contributions of the various inputs. Further, while regressions are useful for measuring effects at the mean, they may obscure interesting patterns at different parts of the distribution.

In a series of papers Gabriella Conti, James Heckman and Sergio Urzua have developed sophisticated models that allow for such kinds of analysis. Their modeling framework and key findings are described in Conti and Heckman (2010) which I briefly review here. They motivate their analysis by citing the growing research across many disciplines supporting a developmental model of health. In particular, they note that differences in physical, genetic, cognitive and psychosocial dimensions of child development play important roles in determining health. These factors also affect the educational choices of children and parents. This process can be obscured in simple statistical models of the educational gradient which will typically ignore this form of selection. To address this, the
authors develop a general framework of latent factors that explicitly models the selection process. Specifically, they use a system of equations in which various early life measures of cognitive skills, non-cognitive skills, health, and genetic factors are used to model the structure of the unobserved endowments. As a result, under certain assumptions, this framework can be used to separately identify selection effects from causal effects and thereby provide guidance on the effects of educational interventions.

Conti and Heckman (2010) describe results from applying their model to data from the 1970 British Cohort Study. Their outcomes are measured at the age of 30 and include: log hourly wages, employment status, use of cannabis, smoking, exercise, self reported poor health, obesity and depression. Their measures of early life endowments include seven test scores administered at the age of ten, six measures of non-cognitive ability, height and head circumference at the age of ten as well as various characteristics of mothers and fathers at birth. The schooling choice is measured by whether students obtain schooling beyond the compulsory number of years.

In contrast to much of the cognitive-epidemiology literature, they find that early cognitive factors do not play a very large role in directly affecting later life health even though these factors clearly affect educational attainment. In contrast, early endowments in non-cognitive skills and health appear to affect both educational attainment and later life health. There appear to be strong gender differences in how early life health factors affect health at the age of 30. For men, early life health directly impacts later life health but not through education. In contrast, for women, early life health affects both educational choice and has direct effects on adult health. Turning to the effects of education on health, they decompose the overall effects into causal effects and selection effects. Across most health outcomes they find that, in fact, education does have a causal effect on health for both men and women. For example, they find that the causal effects are much more important than the selection effects in explaining education gradients in smoking behavior for both genders. In a series of charts, they highlight that there is sometimes a significant degree of heterogeneity in treatment effects even when the average effects are not apparent.
Overall, the framework provides a useful method by which new insights may be drawn on how factors early in life affect the formation of the education gradient in health. Given the complexity of the model, assumptions are required in order to impose a sufficient structure for identification and future work should gauge the sensitivity of the results to these assumptions. Ideally, one might like to use measures of endowments obtained even earlier in life than the age of ten and one could envision models of the development process that examine multiple stages of childhood. On the larger question that is the central topic of this review, namely how much of the effect of education on health that is causal, the study clearly makes important advances in explicitly quantifying selection effects. Still, future work may consider how access to more data on different dimensions of endowments and at different stages of life might impact the conclusions.

5. The effects of school quality on health

There has been a clear broadening of the literature on the economic effects of schooling over the last 20 years or so to consider the quality of schooling in addition to the quantity of schooling. For example, some researchers have focused on the importance of class size and teacher quality as key determinants of future labor market earnings capacity. A few studies have begun to follow this same type of progression to consider school quality dimensions with respect to the effects of education on health. One well known approach has exploited the sharp differences in school quality experienced by Blacks who attended segregated schools in the American south as compared to Whites. Card and Krueger (1992) collected times series data of the teacher pupil ratio, teachers’ wages and term length by state and by race covering the 1915 to 1966 period and used these measures to estimate the effects of school quality on the returns to education in later Censuses.

Following this approach, Frisvold and Golberstein (2011) pool samples of individuals from the National Health Interview Survey over the 1984 to 2007 period who were born in Southern states between 1910 and 1950. The data is then aggregated to cells by race, sex, state of birth, survey year and ten-year birth cohort. These cells are then linked to school
quality measures based on state of birth and cohort group. The key identifying assumption is that these school quality improvements were unrelated to other factors that may have independently improved the long-term health of Blacks relative to Whites. One strategy they employ to control for other coincident factors that could have improved long-term health is to include Black-White differences in infant mortality as an added covariate. The health outcomes they examine include mortality, self-reported health status, disability, BMI and smoking behavior. They find beneficial effects of some school quality improvements for Blacks relative to Whites for all outcomes except mortality. The effects are in some cases, but not all, statistically significant and are generally strongest for term length and rarely significant for teachers' wages.

While these results are highly suggestive, they may not be fully convincing for a few reasons. First, the school quality measures used are relatively blunt and the variation is at a very broad level (state by ten-year birth cohort). Ideally, one would like to use a much more detailed variation in school quality, perhaps at a more local level and for sharper differences in birth years and perhaps with family fixed effects. Second, it might be preferable to evaluate a specific school quality program experiment where one might be more confident about the design and timing of the program and where one could create more control groups to test for placebo effects. Third, given the relative bluntness of the treatment, there may still be other coincident factors that could conceivably account for the results which the authors cannot control for, a point which the authors acknowledge but do their best to address with available data.

Johnson (2010) presents evidence on the effects of school quality differences for Blacks and Whites for more recent cohorts that arose due to court ordered school desegregation decisions. Johnson uses data on cohorts born between 1950 and 1975 who were followed over time in the Panel Study of Income Dynamics (PSID). Using a restricted version of the PSID with very detailed geographic location during childhood, individuals are linked to information drawn from an extensive database on every US school district that implemented a court ordered desegregation plan. Importantly, the data contains information on the timing of the initial order. Using an event study approach, Johnson first demonstrates sharp improvements in Black-White gaps in per-pupil spending and class sizes immediately after these court orders. Then, he turns to estimating
health effects by using self-reported health status later in life as the key outcome. Johnson measures the proportion of years between the ages of 5 and 17 that a child was in a school district subject to a court-ordered desegregation plan. This measure is further interacted with a dummy for Blacks so that the effects difference out any effects on Whites. When the model also includes county-fixed effects, the estimated effects are also relative to other Blacks living in the same county. In some specifications, Johnson also includes sibling-fixed effects so that the variation is further restricted to comparisons within families.

Johnson finds a consistent pattern of results across a variety of specifications and empirical approaches suggesting that the Blacks who experienced improvements in school quality due to desegregation plans also experienced notable gains in health. No such comparable effects are found for Whites. He finds, for example, that school desegregation can account for about half of the Black-White gap in health that existed for cohorts born between 1955 and 1963.

Johnson (2010) adds fairly compelling new evidence to the literature suggesting that improvements in school quality may have had long-lasting effects. The use of multiple empirical approaches including sibling-fixed effects is especially notable. Unfortunately, the sample sizes in the PSID are not especially large so in some cases, there is only limited power especially when using the specifications with most controls. It would also be useful to see effects using other health outcomes including objective measures. In addition, while the research design is reasonably compelling, there may still be concerns about the factors that precipitated the desegregation plans and whether selective migration could account for some of the effects.

6. Conclusions

Over the last decade, researchers have made further progress in understanding the nature of the striking disparities in health by education level that are evident in industrialized countries. A focal point of recent studies has been to use compulsory schooling laws as a source of exogenous variation in schooling levels to examine whether there are any causal effects of schooling on health and mortality. A review of studies of the
effects of such laws across seven countries points to little evidence of causal effects. The most convincing studies have used law changes that have clearly demonstrated strong effects on both education and earnings but have shown virtually no effects on a range of health outcomes. Moreover, these studies can rule out quantitatively meaningful beneficial effects of education on health. Still, a few studies have shown beneficial effects of the laws in some countries.

A limitation of using compulsory schooling laws is that they may only identify localized effects on those who are “treated” by the laws and therefore, may not be representative of the effects of schooling at higher or lower levels of schooling. Approaches that use other plausible sources of exogenous variation such as twin differences, military draft avoidance behavior, or responses to new health information provide some evidence of potentially beneficial effects of education. Future research should continue to improve upon these approaches.

Two recent studies have used broader research approaches to develop new insights concerning the pathways underlying the education-health gradient. Cutler and Lleras-Muney (2010) decompose the sources of the gradient and have emphasized the surprising importance of cognitive ability in explaining health disparities by education level. Conti and Heckman (2010) use a framework that explicitly models how early life endowments affect educational choices and the importance of this selection effect in giving rise to the education gradient. They show that early life non-cognitive and health endowments play an important role. They also provide evidence of causal effects of education on health. While these approaches have their own limitations, they serve as a useful complement to the studies focusing on more tightly defined experimental approaches in search of more conclusive evidence concerning causality. Finally, several recent papers have begun to evaluate whether school quality differences may affect long-term health with some suggestive evidence that it is not just the amount of time spent in school that matters.

Overall, as researchers continue to probe into the relationship between education and health, there have been clear incremental gains in our understanding. If nothing else, these gains in knowledge have provided a more rounded view of this highly compelling and policy relevant topic.
References


Conti, G. and Heckman, J. (2010), Understanding the early origins of the education–health gradient: A framework that can also be applied to analyze Gene-environment interactions, Perspectives on Psychological Science 5, 585-605.


Johnson, R. (2010), Long-run impacts of school desegregation and school quality on adult health, manuscript, University of California, Berkeley.


Lundborg, P. (2008), The health returns to education – What can we learn from twins?, Tinbergen Institute Discussion Papers 08-027/3.
Spasojevic J. (2010), Effects of education on adult health in Sweden: Results from a natural experiment, in D. Slottje and R. Tchernis (eds.), Contributions to Economic...
Comment on Mazumder: The effects of education on health and mortality

Kjell G. Salvanes*

The question raised in this overview is whether schooling per se makes you different in terms of health behavior and thus longevity, or if there is only a correlation between health outcomes and education. Mazumder gives a very detailed survey of the literature on the effect of schooling on mortality (for the most part) and some health outcomes (hospitalization mostly). The main message of the survey is that even if cross-sectional evidence suggests a strong effect of education on mortality, the effect disappears once pre-reform trends are taken into account and selection into education is controlled for, using – for the most part – educational reforms as instruments. It is especially true when RD techniques are used to assess the effect close to the change on both sides of the discontinuity. So the questions are then: what are the factors that explain the strong correlation between health outcomes and education? And why does it not appear to hold up when more reasonable identification strategies for the causal effect of education on health behavior and health are used?

Mazumder’s paper is a very careful review of evidence and raises a large number of interesting issues. In this comment, I will try to focus on a couple of issues that are not so well covered in the review. An important issue is, of course, why no support is found. Is it due to weak data, something particular with the reforms used, or other factors not taken into account? And next, why is there a strong correlation between health and education that is not picked up by the identification strategies?

* Department of Economics, Norwegian School of Economics, Kjell.Salvanes@nhh.no.
However, first, why would we expect education to affect health behavior and health outcomes? Schooling may in a wide sense improve people’s critical thinking, helping individuals process new problems and situations and therefore make better decisions. Schooling may also facilitate better interaction and communication with others, thus making them more fit to make better decisions as well. More specifically, in the human capital literature, the focus has been on two aspects of improved decision making due to education: 1) Productive efficiency and 2) Allocative efficiency. Both of these channels are relevant when it comes to health behavior and thus health outcomes. The first, improved critical thinking and social skills, allows individuals to get more done in the same amount of time for the same amount of money (e.g. improved multi-tasking skills). And the second channel pertains to situations where the more skilled choose a different mix of inputs in trying to maximize their welfare. In other words, individuals with better skills make better decisions when faced with similar circumstances. If decision making is improved by schooling in this way, it also means that improved decision making should be found in many dimensions of life such as social participation, in the marriage market, criminal behavior etc. (see Oreopoulos and Salvanes, 2011 for a further discussion on evidence of this).

What could be the reason for finding no, or only a very weak, effect of increased education on health? A first issue is, of course, that there may be data limitations of some sort. Data problems may pertain to the health outcomes, the education reforms used in the studies, and how education is measured. Most of the studies so far have used longevity as a health outcome. Clearly, this is an important outcome but when measured too early, e.g. before the age of sixty, mortality is a fairly rare event in the OECD countries, and it is perhaps not so surprising that no or only a weak effect is found. Another outcome used is hospitalization which, of course, is a measure closer to what one wants to measure, but still hospitalization is also a relatively rare outcome. I believe other measures which are not so rare – such as more general measures of health outcomes and health behavior – have the potential of picking up more variation in the population and may provide a better chance of capturing the effect of education.

Another issue is, of course, whether there are any particular issues with the reforms used. The UK reforms used in 1947 have recently been
scritinized by Hart and Devereux (2010), and it is clear from their analysis that the first-order effect of these reforms on income is much weaker than what was found in previous studies. Their results indicate that the real returns are zero for women and three to four percent for men, only when also new earnings data are used. One issue that may be raised then is whether second-order effects such as health outcomes are expected to be strong. For the Swedish mandatory school reform, contrary to the UK reforms, a strong effect on income has already been established and it may be expected also on other outcomes such as health (Meghir et al., 2012). For the outcomes that are being used, mortality before the age of sixty and hospitalization, no effects are found for the population as a whole. However, once the authors split the effect of the reform on health by SES groups, there are effects on mortality although not very strong. I believe that this is an important finding since it is expected that a reform aiming at increasing the mandatory education from seven to nine years will affect the whole population equally and independently of parental educational background. Hence, in addition to the issues that Mazumder focuses on in his survey, weak data, weak reforms and heterogeneous effects are clearly possible explanations for the relatively weak causal effect of education on health outcomes.

In addition to exploring these factors more carefully in future work, other avenues may be to explore which factors are important for health (and other outcomes) and which may or may not be affected by policies. The “third” factor(s) driving the strong cross-sectional correlation in education and health, which may be affecting both education and health, are important areas for research. How family background directly affects both educational choice and health behavior choice are important issues. These other factors may be related to important traits such as risk behavior, the Big Five personality traits, or other preferences. These factors may, be transmitted across generations, which is another interesting avenue for research. There may, of course, also be other factors, such as, for instance, parents’ investment in their children, which affects both children’s education and their health behavior, and which would be interesting for research. These factors may obviously also be affected by polices but not necessarily education policies.
References

Hart, R.A. and Devereux, P.J. (2010), Forced to be rich? Returns to compulsory schooling in Britain, Economic Journal 120, 1345-1364.


ECONOMICS OF EDUCATION

Economics of education: Policies and effects
Anders Björklund and Peter Fredriksson

Long-term effects of early childhood care and education
Christopher Ruhm and Jane Waldfogel

Recruiting, retaining, and creating quality teachers
C. Kirabo Jackson

On the margin of success? Effects of expanding higher education for marginal students
Björn Öckert

Gender differences in education
Tuomas Pekkarinen

Educating children of immigrants: Closing the gap in Norwegian schools
Bernt Bratsberg, Oddbjørn Raaum and Knut Reed

The effects of education on health and mortality
Bhashkar Mazumder