Hot for Teacher

Hot for Teacher:

Studies on Teacher Career and Skill Development in the Netherlands

Christiaan S.C. Vermeulen

Printing: ProefschriftMaken || www.proefschriftmaken.nl
Cover design: ProefschriftMaken || www.proefschriftmaken.nl

ISBN: 978 94 6380 352 6

Hot for Teacher:

Studies on Teacher Career and Skill Development in the Netherlands

Proefschrift

Ter verkrijging van de graad van doctor
aan de Universiteit Maastricht
op gezag van de Rector Magnificus,
Prof. dr. Rianne M. Letschert
volgens het besluit van het College van Decanen
in het openbaar te verdedigen
op donderdag 23 mei 2019 om 12:00

door

Christiaan Stanislaus Caesar Vermeulen

Supervisor

Prof. dr. Inge de Wolf

Co-Supervisor

Dr. Bart Golsteyn

Assessment Committee

Prof. dr. Wim Groot (chairman)

Dr. Ilja Cornelisz (Vrije Universiteit Amsterdam)

Dr. Trudie Schils

Prof. dr. Jan van Tartwijk (Universiteit Utrecht)

Prof. dr. Rolf van der Velden

Acknowledgements

Over the last four years, many people have helped to make the road to the completion of this dissertation an enjoyable and relatively carefree experience. Those who provided comments on specific projects are mentioned in the respective chapters. In this section, I would like to thank some people more in general.

To start, thanks go out to my supervisors Inge de Wolf and Bart Golsteyn for providing me with the chance to embark on this doctoral journey, which has been full of opportunities and low on obligations. Inge, your enthusiasm, optimism, and guidance especially during the early stages as well as the home stretch of the project has been invaluable, and your expertise in navigating policy circles has benefited me greatly.

Bart, thank you for teaching me your approach to academic research and writing, and for your detailed feedback on all projects. I also enjoyed our talks on research, guitars, mouth harps, politics, and tank stations. Further, it is interesting to note that between the times I started and finished my PhD, you rose from an assistant professorship all the way up to full professor. While this relationship is not necessarily causal, it is a fascinating observation that warrants further investigation.

I would also like to thank the reading committee, Prof. dr. Wim Groot, Dr. Ilja Cornelisz, Dr. Trudie Schils, Prof. dr. Jan van Tartwijk, and Prof. dr. Rolf van der Velden for their comments and evaluation of this dissertation.

Thanks also to Tijana Breuer, Nienke Ruijs, and many others at the Inspectorate of Education. Working together on the creation and utilization of the administrative datasets on teacher assignments and other labour market outcomes has been a very fruitful experience, and has added great value to this thesis. I also much appreciate the opportunity to be involved in the yearly iteration of the “Staat van het Onderwijs” as a researcher. I look forward to our future collaborations.

Further thanks go to Lex Borghans and all other members of the EofE group for providing their valuable insights and perspectives on education, either informally or during our not-quite-weekly lunch meetings. A special mention goes out to Corrie of #TeamBKO. Thanks also to my new colleagues at ROA for helpful discussions regarding my work.

All of my office mates at the mandarin office, fridge office, international office, bachata office, party office, and metal office are thanked for their welcome distractions, tolerance of my procrastination induced office shenanigans, and polite smiles at my many attempts at comedy. Special mentions go out to Sergio, a man who gives the phrase "thinking like an economist" a whole new meaning, as well as to Johan, whose disk-jockeying talents are matched only by his encyclopedic knowledge of obscure heavy metal musicians.

Olga and Annelore, I feel honoured that you have agreed to act as paranymphs at my defense. Without your senses of humour, friendship, and existential support, the whole PhD experience would have been much less enjoyable. I hope our careers will continue to cross paths in the future.

Finally, I am grateful to my parents, sister, Lian, and the rest of my family for their support, love, and their reminders that there is a whole world outside of the confines of the University that may warrant some attention every once in a while.

Contents

List of figures

List of tables

1. Introduction

1.1 Motivation

An educational system is only as good as its teachers. Research has shown that teachers are the most important school-based determinant of students' educational outcomes (Hanushek, 2011). The impact of good teachers is sizable: Hanushek (1992) finds that a high quality primary school teacher can achieve 1.5 years' worth of progress in terms of test scores for their students in one academic year, while a low quality teacher achieves only 0.5 years' worth of progress.[1] For the Netherlands, van der Steeg & Gerritsen (2016) show that being assigned to a good teacher rather than a bad one in the final grade of primary education positively affects student achievement in math (reading) by 0.4 (0.25) standard deviations in one year. Other recent studies have shown that teachers can influence their students' lives far beyond the classroom. In a causal study, Chetty, Friedman, & Rockoff (2014) find that students exposed to a good teacher in primary school earn higher wages, end up living in better neighbourhoods, face a reduced risk of teenage pregnancy, and start saving for retirement at an earlier age. Hence, in order to maintain and improve the quality of education, policy makers have to ensure an adequate supply of high quality teachers.

Projected teacher shortages in the primary education sector of 7000 FTE by 2025 (Adriaens et al., 2017), and concerns about the quality (Education Council, 2013) and equitable distribution of teaching resources (Inspectorate of Education, 2018) have given the supply of suitable teachers a prominent place on the Dutch public agenda.[2] Designing the optimal policy intervention in order to increase teacher supply is complicated, as the preferred outcome is to increase their quality and quantity at the same time. Some measures might effectively increase the quantity of teachers, but negatively affect teacher quality or

[1] Teacher quality is defined in terms of a teacher's ability to increase their students' performance on standardized tests. While one could think of other important outcomes partly under a teacher's influence, such as students' non-cognitive skills, or ability to display good citizenship, teachers that are able to help their students reach high test scores have been found to positively influence their students' lives on a variety of outcomes. These include better labour market and residential outcomes, as well as a reduced chance of teenage pregnancy (Chetty et al., 2014). Secondly, standardized testing has become increasingly prevalent, and increasingly important, in determining the educational career of students, especially in primary education. Considering the rising premium of college education (Autor, 2014; Goldin & Katz, 2007), teachers that are able to consistently raise their students' test scores to a level qualifying them to enter the pre-university high school track, are able to markedly improve their students' later life outcomes through increased educational opportunities.

[2] In this thesis, the focus is on the supply side of the teacher labour market in addressing the current mismatch between teacher supply and demand. Alternatively, one could focus on policies that address the demand for teachers, for example increased class sizes, or the creative deployment of ICT-based learning systems, to be able to increase the teacher-student ratio. However, prior research has shown that increasing class size negatively affects student performance (e.g. Angrist & Lavy, 1999; Krueger, 2003; Frederiksson, Öckert, & Oosterbeek, 2012, or see Schanzenbach (2014) for an overview). Second, while ICT-based learning might reduce the demand for teachers in the long run, it is unlikely that it can provide a solution to the mismatch between teacher supply and demand in the short run. Hence, in order to combat the teacher shortage currently faced in the Netherlands without compromising on the quality of education, a focus on teacher supply is warranted.

vice versa. For example, solving quantitative teacher shortages could be as easy as dropping the certification requirement needed to work as a teacher, while the qualitative teacher shortage might be handily solved by only allowing professors of pedagogy to stand at the front of the classroom. These extreme examples illustrate the inherent trade-off between the quantity and quality of teachers. The question necessarily becomes whether a policy's benefits on the targeted aspect of the teacher shortage justifies the consequences on the other side of the coin.

Interventions aimed at increasing the supply of suitable teachers can be targeted at increasing the quantity or quality of the people sorting into teaching, preventing current teachers from dropping out of teaching, or incentivizing incumbent teachers to increase their teaching skills through further training. Designing such policies requires not only the best information possible regarding the current situation, but also insights in the underlying mechanisms. While for some of the aforementioned policy levers plenty of information is available on which decisions can be based, relatively little is known for others.

In terms of the quantity of teachers in the Dutch primary education sector, many facts have already been established. Administrative data tracking students from the start of primary school to the end of higher education is increasingly being made available by DUO, and utilized by other institutions and researchers investigating student sorting into teacher education (e.g. CPB, 2017; Inspectorate of Education, 2017; de Wolf, Vermeulen, & Breuer, 2018). It has become well established that enrolment into primary teacher education has been declining for a number of years, although the trend seems to be reversing as of 2017 (CPB, 2017; de Wolf et al., 2018). In this year, enrolment numbers increased for the first time in over 10 years (Primary Education Council, 2017). In terms of student characteristics, prospective primary school teachers are mainly non-migrant women with a humanities background (CPB, 2017; de Wolf et al., 2018). At the same time, teacher program dropout rates have been steadily increasing, leading to a sharp reduction in the number of available teaching graduates (Inspectorate of Education, 2017), leaving the primary education sector with a prospective teacher shortage of 7000 FTE by 2025 (Adriaens et al., 2017). Furthermore, the introduction of the math and language tests at the end of the first year of primary teacher education in 2006 has been shown to be partly responsible for the drop in enrolment and the increase in dropout rates (van Ruijven, 2016). With the mandatory entry test introduced in 2015, primary teacher education saw a decrease in the number of students with a lower vocational educational background as well as a decrease in the number of non-western migrant student teachers (van Ruijven, 2016; Inspectorate of Education, 2017).

In terms of the quality of teachers, the main challenge facing policy makers aiming to increase teacher quality is that there are few observable characteristics that are related to teacher quality (Hanushek & Rivkin, 2006; Coenen et al., 2018). If teacher quality would be easy to predict based on personal attributes, schools and teacher training programs could select students based on these characteristics in order to increase the quality of the teacher corps. Unfortunately, many international studies have shown that teacher observables such as gender, ethnicity, selectivity of their attended college, advanced degree obtainment, and personality traits are not reliably related to teacher quality (e.g. Kane, Rockoff, & Staiger, 2008; Harris & Sass, 2011; Kim, Dar-Nimrod, & MacCann, 2018; or see Coenen et al. (2018) for a recent overview).

While predicting who will become a good teacher prior to entering the profession is complicated, one of the most consistent findings in the literature is that experienced teachers are better than inexperienced ones (Papay & Kraft, 2015; Wiswall, 2013; Harris & Sass, 2011; Sass et al., 2012). Secondly, the distribution of teachers across schools matters: there is evidence that a match between student and teacher characteristics is beneficial to student learning. Dee (2004, 2005) and Yarnell & Bohrnstedt (2018) find that a match along ethnic lines has a positive impact on student achievement, with benefits extending beyond the classroom on to later life outcomes (Gershenson et al., 2018). Although some other studies find insignificant results (e.g. Ehrenberg, Brewer, & Goldhaber, 1995; Munoz & Chang, 2007). Furthermore, there is evidence of increased student achievement resulting from being taught by a teacher of the same gender (Dee, 2007; Lim & Meer, 2017; Lim & Meer, 2019; Gong, Lu, & Song, 2018). Although here again some studies do find insignificant results (Clotfelter, Ladd, & Vigdor, 2010; Coenen & van Klaveren, 2016; Sokal et al., 2007). Finally, teacher cognitive skills, math skills in particular, have a modest positive relation to student achievement (Hanushek, Piopiunik, & Wiederhold, 2018; Metzler & Woessmann, 2012; Clotfelter Ladd, & Vigdor, 2006; Clotfelter Ladd, & Vigdor, 2007; Boyd et al., 2008).

Knowing that experienced teachers outperform inexperienced teachers, and that there are potential efficiency gains to be made by allocating certain teachers to certain student populations, provides an indication where policies aimed at increasing teacher quality should be targeted. Preventing early career teacher attrition from the profession would mean that more students benefit from being taught by experienced teachers, while ensuring an optimal distribution of teachers across schools could exploit potentially beneficial student-teacher matching effects. Finally, the importance of cognitive skills suggests that a certain baseline level of cognitive skills has to be met in order to be allowed into the teaching profession.

Current knowledge about Dutch teachers is relatively scarce on exactly the points at which policies are most likely to be effective. There is little information on how teachers are distributed across schools, hindering policy makers' ability to design policies aimed at optimizing the distribution of teachers. Second, not much is known about how the careers and skills of teachers develop after their graduation from teacher training programs, hampering the design of interventions aimed to reducing teacher attrition. Until recently, most studies investigating teacher careers in the Netherlands have focused on the start of teacher careers and the transition from teacher training to the labour market (e.g. Cörvers et al., 2017; CPB, 2017; Fontein et al., 2016; see also den Brok, Wubbels, & van Tartwijk (2017) for an overview). Likewise, while there is some information about the cognitive skills of Dutch teachers on average (Hanushek et al., 2018), we do not know how the cognitive skills of the Dutch teacher corps are distributed relative to the rest of the population. This impedes policy makers' ability to set the optimal threshold of cognitive skills required to enter the teaching profession while still ensuring an adequate number of eligible students.

Finally, in terms of incentivizing incumbent teachers to increase their teaching skills through further training, not much is known about how teachers decide to improve their skills over the course of their career. While there are plenty of studies evaluating the effectiveness of, and the organizational factors fostering, professional development programs aimed at developing the skills of incumbent teachers (Evers et al., 2011; Yoon et al., 2008; Carrillo, Maassen van den Brink, & Groot, 2016), less is known about the decision to sign up for particular professional development programs from the teachers' perspective. Without this understanding, it is difficult for policy makers to design an optimal incentive scheme to ensure professional development courses are taken by those teachers that would improve their teaching ability the most by participating.

1.2 Aim

This thesis aims to improve the understanding of how the careers and skills of Dutch teachers progress. These insights are relevant for policy makers aiming to ensure an adequate supply of high quality teachers. When addressing teacher careers this thesis focusses on the primary education sector, while the chapters on skills concern both primary and secondary school teachers. The contribution of this thesis is twofold. First, this thesis adds to the understanding of primary school teacher careers in the Netherlands by employing large administrative datasets containing the entire population of early career primary school teachers. Teachers are followed from as early as secondary school to as late as nine years in the labour market, and linked to school and student characteristics of the school at which they are teaching. With this wealth of data, career patterns of Dutch primary school teachers can be studied at a level of detail and representativeness that was not possible before. Second, the thesis adds to the understanding of how primary and secondary school teachers choose to develop their skills through professional development programs, and on how the literacy and numeracy skills of Dutch primary and secondary school teachers compare to the general population as well as to the relative literacy and numeracy performance of teachers in other countries.

The overarching research question this thesis addresses is therefore:

How do the careers and skills of Dutch teachers develop?

This central research question is addressed in four self-contained chapters answering the following sub-questions:

How do the careers of primary school teachers develop during the first nine years after graduation, and is there a relationship with teacher preparation program dropout rates?

How are primary school teachers distributed across schools in terms of their level of education and migration background, and what are the consequences for student performance?

How do teachers decide to sign up for professional development programs, and what are the consequences for optimal training provision?

How do the literacy and numeracy skills of Dutch primary and secondary school teachers compare to the rest of the population, and to the relative performance of teachers from other countries?

1.3 Outline

The outline of this thesis is as follows. The second chapter relates to the early career development of primary school teachers, and investigates whether dropout rates at the teacher preparation program level are related to early career primary school teachers' attrition rates and other labour market outcomes. The third chapter looks at how primary school teachers sort across schools with differing student populations, and how matching between teacher and student characteristics relates to student outcomes. Chapter four models the decision of teachers to develop their skills through entering professional development programs aimed at increasing their teaching quality. Lastly, chapter five zooms out to take stock of the skills of the Dutch primary and secondary school teacher corps relative to the rest of the population and compared to teachers in 14 other OECD countries.

Chapter 2

Policy makers could focus on increasing primary education teacher supply through interventions aimed at reducing dropout rates during primary teacher training. By signing up for primary teacher training, these students have already revealed a preference for teaching, making it more likely that policies aimed at them would increase teacher supply. However, low teacher training dropout rates can be reached both by raising program quality, and by reducing standards. Therefore, it is a priori unclear whether lower levels of student attrition are related to a higher supply of suitable teachers.

In Chapter 2, I use Dutch registry data on all students entering primary education teacher training programs between 2002 and 2012 to analyze the association between dropout rates at the teacher preparation program level and the labour market outcomes of these programs' graduates during the first nine years after graduation. I find that low teacher preparation program dropout rates are unrelated to the probability of staying employed in the education sector for these programs' graduates. However, controlling for a large set of student observables, entry cohort and graduation year fixed effects, graduates from programs with low dropout rates supply 2.5% fewer monthly hours of teaching after several years in the labour market. Importantly, graduates from both low and high dropout rate programs supply a similar amount of hours when working outside of the education sector. This association is therefore unlikely to be the result of differential preferences for working hours between the types of student that graduate from low (vs. high) dropout rate programs.

The preferred interpretation of the difference in supplied hours of teaching is that the marginal graduate from a low dropout rate program is slightly less suited for teaching and derives less utility from working in the education field because of a lower job-match quality, leading to a willingness to supply fewer hours. Nevertheless, lower dropout rates are associated with a higher supply of teachers even when accounting for the reduction in number of hours worked. While the findings are not causal, they suggest that lowering dropout rates could be a promising channel for policy makers to reduce teacher shortages. However, it is important that provided incentives to reduce student attrition are simultaneously accompanied by a powerful means of program quality assurance.

Chapter 3

Chapter 3 investigates the distribution of primary school teachers across schools in the Netherlands, and looks at how the extent of matching between teachers and students on migration and educational background relates to student achievement. Using administrative data on all primary school teacher assignments between 2008 and 2016, it finds that the share of teachers holding a master's degree is higher at schools where a larger percentage of the student population has highly educated parents. Schools serving a higher percentage of students with a non-western migrant background employ a larger share of non-western migrant teachers. These patterns are especially pronounced in urban areas, where both parents and teachers have more options to act on their preferences for school, student, and teacher characteristics. Positive assortative matching on teacher and parental educational levels is increasing over the period studied, while sorting along migration background is decreasing. Sorting of early career teachers exacerbates the differences in average teacher characteristics between schools.

In terms of student achievement, the results show that assortative matching along migration background is not detrimental to student performance. Students with a migration background perform slightly better in schools with a larger share of non-western migrant teachers, while there is no significant relationship between teacher migrant background and the performance of native students. The results are more pronounced for migrant students from a relatively low socio-economic background, and are driven by increased performance in mathematics, with no matching effects found for language. The share of teachers holding a master's degree is unrelated to the performance of students with university-educated parents on any subject, nor do students with low educated parents perform worse in schools with a larger percentage of master's degree holding teachers. Together, these results suggest no negative consequences of teacher sorting on student characteristics in terms of test scores. Furthermore, considering the small positive association between the share of teachers with a non-western migration background and the performance of migrant students, a potential policy implication could be to stimulate students from a non-western migration background to sort into primary teacher education.

Chapter 4

A promising way to increase the average quality of teachers without reducing the quantity of available teachers is to raise the skills of those that are already teaching through on-the-job training programs. However, an important unanswered question is how teachers select into available professional development programs. Without an understanding of teacher incentives, it is unclear whether training will be taken by those who would benefit the most from participating. This leads to difficulties in evaluating the true added value of the training program, and in designing the most efficient implementation strategy.

Chapter 4 develops a model on the decision to enter on-the-job training for teachers. It argues that, since teacher performance is partly unobservable, teachers incorporate the signalling value of training participation into their decision. Under voluntary participation, sorting into training is inefficient: programs aimed at low ability teachers will be underutilized because of the negative signal participation conveys about their initial ability. Programs aimed at improving the skills of high quality teachers will be overused, attenuating estimations of their impact. Results from the model show that offering training has spillover effects: introducing advanced courses increases participation in basic courses. The intuition behind this result is that improvements in an advanced course stimulate more teachers to sign up for advanced training. This, in turn, decreases the average ability of the group that does not take any training, reducing the negative signal associated with signing up for basic training. By the same logic, an increase in the attractiveness of the basic course reduces participation in advanced courses. Thus, the availability of different types of courses weakens the signal of each course individually.

In terms of optimal training provision, these results have several implications. First, policy makers can affect the participation decision of the entire pool of teachers, and improve average ability overall, by investing in programs targeting the top of the ability distribution. Providing the option of an advanced training course can be effective at inducing low ability teachers to participate into training. Secondly, evaluations of the effectiveness of single programs should consider spillover effects. Even if an advanced course is ineffective at raising the ability of participants on average, its mere existence increases participation of low ability teachers in basic courses. Assuming the basic course is effective, this would lead to an increase in average teacher quality. Finally, policy makers should take the signal the course will give into consideration when designing training programs. The results from the model predict that as a course becomes more basic, fewer teachers will sign up for this course unless it is unrealistically effective.

Chapter 5

Chapter 5 takes stock of how Dutch primary and secondary school teachers perform in terms of their literacy and numeracy skills relative to the rest of the adult population, and compares these results to the relative performance of teachers in other developed countries using the OECD-studies PIAAC and ALL. The results show that both primary and secondary school teachers have better literacy and numeracy skills than the rest of the population in almost all of the 15 countries in the sample on average. Secondly, the comparisons of the skill distributions between teachers and others show that teachers outperform other respondents mainly in the bottom percentiles: the lowest scoring teachers significantly outperform the lowest scoring other respondents. At the high end of the distribution the highest scoring secondary teachers are not strongly outperformed by the highest scoring other respondents; while the highest scoring primary teachers score 0.1 standard deviations lower than the highest scoring other respondents. The results persist when restricting the comparison to the college-educated subsample, are not driven by age or gender, and are insensitive to the inclusion of measures for the frequency of skill use.

Focusing on the Netherlands, the Dutch primary and secondary school teachers are equally skilled as other college-educated respondents on average. Looking at the shape of the distribution, Dutch primary school teachers strongly outperform the rest of the college-educated population at the low end of the ability distribution in terms of literacy, while their numeracy skills are about average. In contrast, at the top of the distribution the highest skilled college-educated respondents outperform the highest skilled primary school teachers. Secondary school teachers are relatively higher skilled in numeracy, where they outperform the college-educated respondents across the ability distribution. Their literacy skills are comparatively weak, scoring about average at the bottom of the distribution, while being outperformed by the highest skilled college-educated respondents at the top of the distribution.

Interestingly, there are significant differences between countries: in some countries teachers outperform others throughout the skills distribution, while in other countries teachers are not even much better at the lower part of the skills distribution. This is important, because policies that work in one institutional setting and for a given teacher skills distribution may not be as effective in a setting in which teacher skills are differently distributed. For example, in sharp contrast to the Netherlands, Danish primary school teachers at the low end of the ability distribution do not outperform other respondents that strongly. In Denmark, it might therefore be an effective policy to focus on the bottom of the distribution (e.g., by raising barriers to enter into teaching, or focusing training on the worst teachers), while in the Netherlands relatively little can be gained in becoming more restrictive at the lower end. These results imply that the scope to improve teachers' skills varies between countries and that policy makers should take the shape of the skills distribution into account when designing interventions in order to most efficiently raise teachers' skills.

2. Plugging the pipeline: Teacher preparation program dropout rates and teacher labour market outcomes

2.1 Introduction

Looming teacher shortages are an increasing cause of concern in many developed countries (Sutcher, Darling-Hammond, & Carver-Thomas, 2016; National Audit Office, 2016).[3] In the Netherlands, the government predicts that by 2025 around 7000 FTEs worth of additional primary school teachers will be needed (Adriaens et al., 2017). Aside from decreasing enrolment numbers in teacher preparation programs (de Wolf, Vermeulen, & Breuer, 2018), a large percentage of students that sort into teacher education never finish their education (Inspectorate of Education, 2017). To combat this "leaking teacher pipeline" policy makers are increasingly focussing on preventing student dropout from teacher preparation programs (van der Aa, Cörvers, & Schoon, 2017). The idea is that higher graduation rates increase the supply of teachers, which in turn decreases shortages.

However, this increase in teacher supply depends on how teacher preparation programs reduce their dropout rates. When programs respond by lowering their graduation requirements, insufficiently prepared students will enter the labour market. If schools are able to predict expected teaching quality at time of hiring, these unfit graduates will be unable to find a teaching job.[4] If schools cannot accurately evaluate prospective teachers' quality at time of hiring, these insufficiently prepared teachers are more likely to be fired once their employer learns about their relatively low ability.[5] Alternatively, programs could reduce dropout by increasing the added value of their training. In this case, they can reach the desired outcome of an increased teacher supply without a decrease in average teacher quality. Because of these different means of decreasing student dropout, it is a priori unclear whether lower observed dropout rates are related to a higher supply of suitable teachers.

In this chapter, I investigate the relationship between dropout rates at the teacher preparation program level and these programs' graduates' labour market outcomes. I use administrative data on the educational career of all students enrolled in primary school teacher preparation programs in the Netherlands between 2002 and 2012, and their labour market outcomes between 2007 and 2016. I show that graduates from teacher preparation programs whose cohort had a low dropout rate are equally likely to work in the education

[3] This chapter has benefited from valuable comments by Tijana Breuer, Bart Golsteyn, Olga Meshcheriakova, Susanne Rijken, Inge de Wolf, and seminar participants at Maastricht University as well as participants at LESE 2019. Thanks to the Inspectorate of Education, DUO, and Statistics Netherlands for sharing their datasets.

[4] While teaching quality is hard to predict, especially for teachers with no prior experience, Boyd et al., (2011) find that schools are able to identify high quality teachers at time of hiring based on pre-service qualifications.

[5] Rockoff, Staiger, Kane, & Taylor (2012) show that school principals form increasingly accurate beliefs about the quality of their teachers over time. This evidence is in line with earlier findings on employer learning in general (e.g. Altonji & Pierret, 2001; Lange, 2007).

sector as graduates from programs with high dropout rates over the first nine years of their career.[6] These results suggest that low dropout rates are not associated with a lower teaching suitability of the average graduate, and that low dropout rates could be interpreted as a sign of high program quality. However, analyses on hours worked show that a one standard deviation decrease in dropout rates is associated with 2.5% fewer hours of teaching supplied per month for the graduates of these programs several years after entering the labour market.

While these results are non-causal and should be interpreted with caution, they are in line with a scenario in which the marginal graduate from a low dropout rate program is slightly less suited for teaching and derives less utility from working in the education field because of a lower job-match quality. This would lead them to be willing to supply fewer hours. However, this interpretation is implausible if these graduates also supply fewer hours when they choose to work outside of the education sector. If this is the case, then their choice of supplied hours is unlikely to be related to the job match quality in the education sector. It would also be implausible if graduates' labour market opportunities outside of education would be better than inside. In this case it would be hard to explain why those graduates do not simply switch careers and drop out of the education sector altogether.

Analyses on hours worked outside of the education sector show that there is no association between monthly hours worked and program dropout rates for teacher training graduates working outside of education. Additionally, I find that those that dropped out of the teaching profession have a lower average hourly wage and a lower probability of obtaining tenure than those that stayed for at least the first nine years of their careers. These results are in line with earlier studies such as Stinebrickner (2002) and Scafidi, Sjoquist, & Stinebrickner (2006), which also find that exiting teachers usually face worse labour market outcomes outside of teaching.

These results suggest that low program dropout rates may lead to fewer hours supplied by the marginal graduate from a low dropout rate program, likely due to a slightly lower job-match quality. However, from a policy perspective it is worth asking whether the increase in the number of teachers due to a low dropout rate offsets the reduction in supplied working hours for these graduates. For example, if 80% of student teachers graduate and go on to supply 30 hours per week on average, it may fulfil a larger share of the total demand for teachers than if only 60% of students graduate and supply 35 hours of teaching per week. In

[6] No results on outcomes later than nine years after graduation can be shown because of the relative recency of the administrative data.

order to inform this debate, I calculate the difference in the amount of FTE supplied between an entry cohort with an average, and with a one standard deviation below average, dropout rate using the results from the previous analyses.

The results of these calculations show that the increase in graduates more than compensates for the reduction in average hours worked. For an average sized program of 400 students, a one standard deviation lower dropout rate is related to a higher teacher supply of 15% for the first couple of years after graduation, and a higher supply of 10% after five years on the labour market, compared to a program with an average dropout rate. These results should be interpreted with caution, as the regressions on which the calculations are based are not causal.

While the relationship between low program dropout rates and a higher supply of teachers holds observationally, the question whether interventions aimed to reduce dropout rates would lead to a sustainable increase in teacher supply cannot be answered based on these analyses alone. The crucial point is how institutions set out to accomplish these lower dropout rates. With a strong accreditation system ensuring a baseline level of program quality, institutions cannot lower their graduation requirements without risking the loss of their accreditation, and would be forced to reduce dropout rates without significantly reducing the quality of their graduates. On the other hand, in a setting where there is little external quality control, programs may be quick to lower their standards. This is easier and less costly than attempting to increase their educational quality. Incentives for programs to reduce their dropout rates should therefore be accompanied by a means through which a certain baseline of program requirements is guaranteed.

This is the first study on how dropout rates at the teacher preparation program level are related to these programs' graduates' labour market outcomes. There has been one prior study that relates teachers' probability to exit the profession to the institution at which they received their teaching certification. Goldhaber & Cowan (2014) use administrative data from the state of Washington to show that teacher retention rates vary substantially between teacher preparation programs. This study adds to their line of research by exploring dropout rates during teacher training as a potential mechanism for their observation that graduates from different institutions may differ in their probability of leaving the teaching profession. Analysing the program dropout rate might be instructive in seeing whether these high teacher retention rates are reached through failing a higher percentage of student teachers before graduation, or through high quality training that is able to raise the teaching ability even of student teachers with relatively low prior ability.

My results on the lack of association between dropout rates and graduates' probability of leaving the teaching profession suggest that differences in dropout rates are unlikely to be responsible for the differences in retention in the labour market between programs. However, the relationship between program dropout rates and hours worked suggests that it is worthwhile to consider both the extensive and the intensive margin of teacher labour supply when analysing teacher shortages and differences between teacher preparation programs. With a similar retention rate but a different propensity of their graduates to work full-time, there could be large differences between programs in terms of their contribution to fulfilling the rising demand for teachers.

The remainder of this chapter is structured as follows. Section 2.2 discusses how dropout rates at the teacher preparation program level could influence graduates' labour market outcomes. In section 2.3, I explain the Dutch institutional setting. Section 2.4 discusses the data and shows some descriptive statistics. Section 2.5 presents the results, and section 2.6 concludes.

2.2 Dropout rates and labour market outcomes

In this section, I create a simple model of student dropout to see how dropout at the teacher preparation program level can have an impact on teacher labour market outcomes. I describe three ways in which programs can influence their dropout rates: selecting students with higher prior aptitude, increasing their added value, and reducing their graduation requirements. Then I show the impact of each of these strategies on their graduates' average productivity. Finally, I discuss their outcomes in a labour market scenario where schools form increasingly accurate beliefs about the productivity of their teachers over time.

The model starts from the situation in which the student has already enrolled in teacher education. I assume that she has made a rational decision, and would strictly prefer graduating to dropping out.[7] In the model, student dropout is a function of student characteristics, and the added value and graduation requirements of the study program:

$$Dropout_{ij} = f(Characteristics_i;\ Added\ value_j;\ Requirements_j),$$

[7] This assumption might be unrealistic, since for example students` outside options may change over time, changing their preferences for graduating to switching study programs. However, teacher preparation programs have no power over their students` outside options. In the model, I abstract away from attrition rate influencing factors that programs have no plausible way of influencing.

where $Dropout_{ij}$ is a binary outcome that takes the value 1 if student i drops out of program j, and 0 otherwise. Student characteristics include students' prior ability, motivation, and personality, all of which combine into a single measure of teaching aptitude prior to teacher education. Added value refers to the quality of the program, which for simplicity is assumed to be independent of student characteristics. Program requirements can be thought of as the minimal level of teaching proficiency and subject knowledge needed to graduate the program. Program requirements are related to the teaching labour market requirements, but since there is no standardized exam at the end of higher education, programs enjoy a certain level of discretion in setting their graduation requirements. In the model, this will be referred to as preferences, hence:

$$Requirements_{jk} = f(Labour\ market\ requirements_k;\ Preferences_j).$$

Students enter teacher training with a certain level of prior aptitude, are exposed to the added value of the program, and evaluated against its requirements. Students graduate when their ability is larger than or equal to the requirements, and drop out otherwise:

$$Dropout_{ij} = \begin{cases} 1, if\ Characteristics_i + Added\ value_j < Requirements_j \\ 0, if\ Characteristics_i + Added\ value_j \geq Requirements_j \end{cases}.$$

Dropout decreases when students' prior aptitude or program added value increases, and becomes higher when graduation requirements increase.

Programs could then use different strategies to decrease their dropout rates depending on which factor they want to target. They could for example impose a barrier to entry in the form of an entrance test, thereby hoping to increase their students' prior aptitude. Alternatively, they could focus on improving their own education program, increasing added value. Finally, they could simply lower the difficulty of their exams, or offer more lenient grading policies, in order to decrease the graduation requirements. While all strategies can be equally effective in reducing dropout rates, they differ in their effect on the teaching ability of the average graduate.

Once the graduates enter the labour market and start working at schools, schools start forming increasingly accurate beliefs about their teachers' expected productivity. Productivity of teacher i from program j in each period t is defined as:

$$Productivity_{ijt} = Characteristics_i + \ Added\ value_j + \varepsilon_t \ ,$$

where in every period *t* some random event $\varepsilon(0, \sigma)$ outside of the teacher's control influences productivity. Employers predict the expected productivity of their teachers by calculating their average productivity over the previous periods:

$$E\left[productivity_{it+1}\right] = \frac{\sum_{k=1}^{t} productivity_{ik}}{t}.$$

Since the random events ε in each time period are unrelated while teacher ability is strongly autocorrelated, schools become more certain about their teachers' abilities over time. Once their expected productivity falls under the labour market requirements they will be fired.[8]

$$Exit\ from\ the\ school_{it+1} = \begin{cases} 1, if \ \ \mathrm{E}[productivity_{it+1}|] \geq Labour\ market\ requirements \\ 0, if \ \ \mathrm{E}[productivity_{it+1}|] < Labour\ market\ requirements \end{cases}.$$

Starting from a scenario in which graduation requirements are equal to the labour market requirements, lowering the graduation requirements implies that at the margin of graduation less suited teachers will enter the labour market. After some time in the labour market, a higher percentage of these graduates will be observed to be insufficiently productive in the eyes of their employer. This will increase teacher turnover rates and exit from the profession, and decrease the probability that these teachers will be tenured or promoted. In contrast, increasing the quality of the education program or selecting students with higher prior aptitude will decrease the dropout rate without an increase in unfit teachers.

Overall, dropout rates can be reduced in different ways with equal success. However, in the labour market, strategies that decreased the dropout rate without increasing the quality of education will result in more unfit teachers. In terms of the supply of suitable teachers, this will ultimately result in more teacher turnover and a higher percentage of teachers leaving the profession. Without considering labour market outcomes, it is therefore unclear whether a low dropout rate at the teacher preparation program level is desirable in and of itself.

[8] In practice, it is unlikely that teachers will be fired so quickly. It might be more appropriate to say that underperformance increases the probability of attrition, and decreases the likelihood of being tenured or promoted. This still implies that on average, labour market outcomes will be worse for graduates of programs that chose to lower their standards.

2.3 The Dutch teacher education system

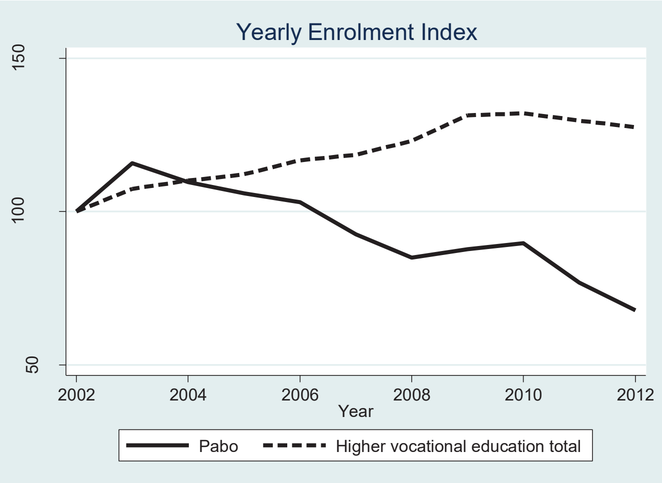
In this section, I describe the Dutch institutional setting in which teachers are trained, and provide some descriptive data on teacher training enrolment over the period studied (2002-2012). In the Netherlands, teacher training is divided into separate tracks for primary and high school teachers. While primary school teacher education is general, high school teacher education is course-specific (e.g. geography teacher education) with some common pedagogical courses. The high school teacher track is further separated into a first and secondary degree license, where the first degree can be thought of as a sort of master's program allowing graduates to teach at the highest grade levels in the middle and high tracks of high school. Additionally, there are possibilities to follow personalized tracks depending on prior degrees and relevant work experience. The latter routes are usually taken by older workers looking to change careers.

In this chapter, I focus on the programs preparing primary school teachers.[9] Over the period studied, there were 23 primary school teacher preparation programs represented in the data, geographically spread across the whole country.[10] The programs are part of the Dutch higher vocational education system, with a nominal length of four years in which students are required to obtain 240 ECTS. Over the period studied they were accessible for all students that passed the highest level of lower vocational education (MBO4), or the middle (Havo) or highest (Vwo) level of high school without any additional entry requirements.[11]

Figure 2.1 shows the index of enrolment in primary school teacher training and the higher vocational education system in total between 2002 and 2012, indexed at 100 in 2002. There is a strong decline in the number of students sorting into primary school teaching programs, with enrolment falling from around 11.000 students in 2002 to around 7000 in 2012. Meanwhile the amount of students in the entire higher vocational education system is steadily increasing until 2009, and remains stable afterwards.

[9] High school teacher training programs are less homogeneous, and therefore less comparable to each other. There are fewer institutes offering high school teacher training programs, and more paths to obtaining a high school teaching license. This makes it almost impossible to correctly define the sample of interest.
[10] One primary school teacher preparation program has been excluded from the analyses because its outlier status made it identifiable, violating the anonymity policy of Statistics Netherlands.
[11] There was one policy change starting in the academic year 2006/2007, where students of primary teacher education had to pass a basic language and math test at the end of their first year of study in order to be allowed to continue their studies.

Figure 2.1: Yearly enrolment index for primary school teacher education (Pabo) and higher vocational education in total between 2002 and 2012



Source: DUO 1cijferHO database

2.4 Data

In this section, I introduce the data sources used, show how the sample and variables of interest are constructed, and provide some descriptive statistics. First, I describe the data on student teachers' educational careers. Second, the labour market data is introduced. Finally, I explain how the final estimation sample and the main dependent variables are constructed.

2.4.1 Student teacher education data

The main educational dataset used is based on the DUO 1cijferHO database. This administrative dataset includes information on all registrations in the Dutch (subsidized) higher education system, both for higher vocational and university level programs. The version of the 1cijferHO data I use has been modified and enriched by the Dutch Inspectorate of Education to a cohort-entry file in which cohorts are defined by taking the enrolment information of October of the first year at which students entered a specific study program at a specific institution. Students are followed throughout their higher education career by updating their status based on their enrolment information in every year.[12] Based on this

[12] For example, if a student is registered at the same institution for the same program in year t and t+1, she is assumed to have continued studying the same program. If a student is not registered in higher education at t+1 but shows up in the graduation registry, she is assumed to have graduated. If she shows up at a different program at t+1 without having graduated from her previous program, she is recorded as having switched programs.

information cohort size, student dropout, and graduation can be determined for each study program.

Aside from enrolment, the 1cijferHO data also contain information about student background characteristics. Among others, age at entry, gender, ethnicity, highest obtained level of prior education, institute of prior education, and place of residence are recorded. For students that graduated high school in the years after the academic year of 2005/2006, I add information on students' end of high school exam grades from DUO's exam subject results data, which can serve as a proxy for student ability when combined with the information on students' highest obtained level of prior education.

From this dataset, I select students that first enrolled in primary teacher training programs between 2002 and 2012. 2002 is the earliest year for which there is data, and 2012 is the final starting year for which it is plausible that students will have finished their studies by 2016, since the nominal length of study is four years. Summary statistics for the total sample are shown in table 2.1. While the number of students sorting into teacher education is declining, their characteristics do not change much over the period under investigation. Assuming stability of observed characteristics is somewhat indicative of stable unobserved characteristics, this suggests that changes in dropout rates and labour market outcomes over time are not strongly related to changes in the type of student that sorts into teacher education.

Table 2.1: Primary school teacher preparation student characteristics

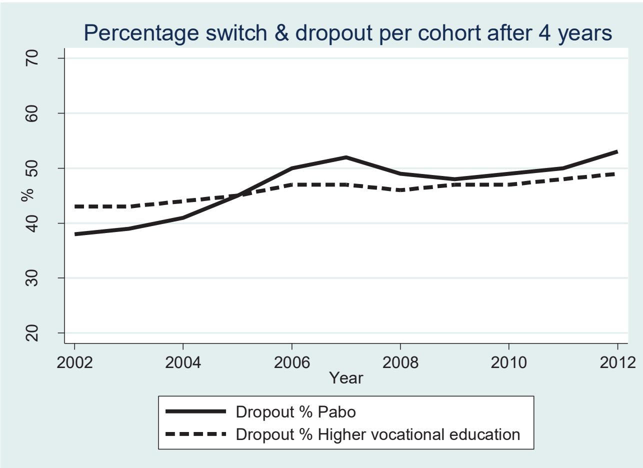
Student Characteristics									
Entry cohort	N	Female %	NW-migrant %	Age mean	MBO %	Havo %	Vwo %	Exam grade Havo mean	Exam grade Vwo mean
2002	10,796	84.8	7.4	23.9	31.7	53.5	14.9	.	.
2003	12,496	82.8	7.8	23.8	34.6	51.2	14.1	.	.
2004	11,839	82	8.2	22.6	36.1	51.5	12.4	.	.
2005	11,382	82	8.7	22.8	37.9	50	12.2	.	.
2006	11,112	81	7.5	21.9	39.4	48.7	11.9	6.13	6.17
2007	9,971	81.5	7.4	21.8	39.5	49.3	11.2	6.19	6.1
2008	9,138	81.9	6.8	21.7	37.1	51.5	11.5	6.21	6.05
2009	9,428	80.7	7.8	22.1	37	49.3	13.6	6.15	6.07
2010	9,656	80.2	8.4	22.3	37	48.4	14.5	6.13	6.05
2011	8,237	79.8	8.9	21.5	35.6	49.8	14.7	6.16	6.03
2012	7,258	78.1	9.2	21.2	34.1	50.7	15.2	6.23	6.17

Source: DUO 1cijferHO database and DUO vakkenanalysebestand. Average exam grades for the cohorts prior to 2006 are not available because of data constraints.

I use this sample to construct *cohort*program* specific dropout rates by taking the share of students who leave their initial teacher preparation program (either switching to a different program or dropping out of higher education altogether) without obtaining a diploma within four years after enrolling.[13] Figure 2.2 shows the development of the average dropout rate for all teacher education students overall, together with the dropout rate in higher vocational education in general. Table 2.2 shows how the dropout rate varies within and between programs over the period studied. Dropout rates are steadily increasing on average, but there is sizable variation between programs over time.

Finally, year of graduation is recorded for those that finish teacher education. Students do not have to graduate within a certain time limit in the Netherlands. Therefore, not all students that enter in the same cohort graduate in the same year. This is important to take into account as we only want to analyse labour market outcomes post-graduation (i.e. no part-time student jobs), and since labour market conditions may differ depending on the year of graduation.

Figure 2.2: 4-year dropout percentage of primary teacher education, and higher vocational education in total



Source: DUO 1cijferHO database

[13] This measure of dropout may underestimate the total dropout rate, since students can still drop out after more than four years. I choose the 4-year dropout rate to be able to make a fair comparison between the earlier and later cohorts. The 2012 cohort has only been observed for four years, so using the total dropout rate for a longer time period would underestimate the dropout rate for the more recent cohorts. In additional analyses, which are available upon request, I use the dropout rate after the first year, and after eight years of enrolment to check the sensitivity of my results.

Table 2.2: 4-year student dropout rate by institution and year-of-entry

	Year										
Institution	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1	0.39	0.42	0.31	0.46	0.42	0.45	0.45	0.36	0.43	0.42	0.42
2	0.45	0.47	0.5	0.5	0.54	0.56	0.52	0.55	0.61	0.7	0.67
3	0.34	0.38	0.39	0.47	0.49	0.68	0.53	0.6	0.46	0.52	0.52
4	0.47	0.35	0.36	0.49	0.58	0.45	0.56	0.56	0.54	0.52	0.58
5	0.39	0.42	0.42	0.52	0.54	0.63	0.43	0.47	0.43	0.41	0.42
6	0.37	0.39	0.44	0.47	0.49	0.52	0.49	0.41	0.49	0.48	0.55
7	0.33	0.36	0.43	0.43	0.55	0.56	0.53	0.58	0.53	0.52	0.58
8	0.4	0.43	0.48	0.49	0.47	0.52	0.56	0.52	0.55	0.52	0.57
9	0.35	0.35	0.38	0.46	0.52	0.54	0.51	0.49	0.51	0.5	0.57
10	0.39	0.38	0.41	0.43	0.49	0.46	0.43	0.42	0.47	0.42	0.64
11	0.43	0.43	0.37	0.39	0.43	0.42	0.48	0.5	0.55	0.5	0.59
12	0.32	0.32	0.32	0.3	0.43	0.33	0.39	0.44	0.41	0.44	0.47
13	0.4	0.33	0.36	0.39	0.46	0.49	0.41	0.45	0.47	0.43	0.55
14	0.39	0.41	0.42	0.45	0.48	0.44	0.43	0.47	0.49	0.52	0.56
15	0.51	0.58	0.59	0.57	0.57	0.62	0.56	0.6	0.55	0.57	0.56
16	0.5	0.45	0.48	0.55	0.58	0.7	0.68	0.64	0.47	0.53	0.55
17	0.41	0.45	0.42	0.56	0.62	0.6	0.52	0.48	0.48	0.49	0.54
18	0.34	0.36	0.35	0.31	0.34	0.41	0.41	0.3	0.41	0.4	0.46
19	0.39	0.4	0.44	0.42	0.56	0.65	0.62	0.62	0.49	0.62	0.67
20	0.33	0.4	0.41	0.45	0.53	0.43	0.45	0.5	0.54	0.57	0.43
21	0.36	0.45	0.41	0.44	0.43	0.52	0.39	0.47	0.46	0.48	0.42
22	0.33	0.33	0.43	0.4	0.52	0.41	0.47	0.47	0.5	0.51	0.43
23	0.32	0.31	0.34	0.33	0.53	0.48	0.44	0.39	0.43	0.47	0.48

Source: DUO 1cijferHO database

2.4.2 Labour market data

I take the information on graduates' labour market data from the Polis administration files of Statistics Netherlands. It contains monthly labour market registry data on all jobs held by people working in the Netherlands. For this study, I use data from 2007 to 2016. The unit of observation is at the *job*month* level, which means that a person can have more than one entry in each month. Among other things, salary, hours worked, tenure status, and employer identification are recorded for each job. Because employers are anonymized, I add information on the economic sector in which the employer operates (SBI2008) using Statistics Netherlands Company activity data. This allows for identification of people working in the education sector.

The sample matched to labour market data consists of those students that graduated primary teacher education between 2002 and 2015 and did not enrol in any other higher education program after finishing teacher education. I take October of each calendar year as the measuring month for all labour market outcomes for computational reasons, and define the first observation in the labour market post-graduation as October of the calendar year after they graduated (e.g. if a student graduates during the academic year 2007/2008, their first labour market observation will be in October of 2008).[14] If graduates cannot be matched to labour market data in a specific year they are either unemployed, self-employed or working outside of the Netherlands.

Table 2.3: Primary school teacher preparation graduate characteristics: matched labour market sample

Graduate Characteristics									
Entry cohort	N	Female %	NW-migrant %	Age mean	MBO %	Havo %	Vwo %	Exam grade Havo mean	Exam grade Vwo mean
2002	5,453	89.5	4.5	23.8	31.1	53	15.8	.	.
2003	5,944	89.4	5.2	24	36.4	49.8	13.8	.	.
2004	5,158	88.4	5.1	23	38.0	49.5	12.5	.	.
2005	4,490	88.2	5.3	22.8	41.6	46.3	12.1	.	.
2006	3,789	88.3	4.0	22.2	41.3	46.4	12.2	6.15	6.15
2007	3,140	89.7	3.6	22	41.3	48.1	10.6	6.23	6.03
2008	2,961	89.6	3.7	21.6	41.1	49.4	9.6	6.22	5.95
2009	2,749	87.2	4.3	22.3	42.4	45.8	11.9	6.15	5.94
2010	2,765	88.1	5	22.7	42.2	46.5	11.4	6.11	5.84
2011	2,101	87.9	3.9	21.5	37.7	50.6	11.7	6.15	5.95
2012	1,442	88.2	3.3	21.1	32.4	53.0	14.5	6.32	6.19

Note: average exam grades for the cohorts prior to 2006 are not available because of data constraints. Source: DUO 1cijferHO database, DUO vakkenanalysebestand, and Statistics Netherlands POLIS salary administration file.

Table 2.3 shows some descriptive statistics of the matched sample. Compared to the teacher preparation program entry sample in table 2.1, there is a higher percentage of women, and a lower percentage of non-western migrants in the labour market sample. This is because men and non-western migrants are more likely to drop out of their study program.[15] The number of observations is also relatively lower in the later cohorts. This is not only due to the increased dropout rate, but also because students that entered in these years are more likely to have still been studying in 2016.

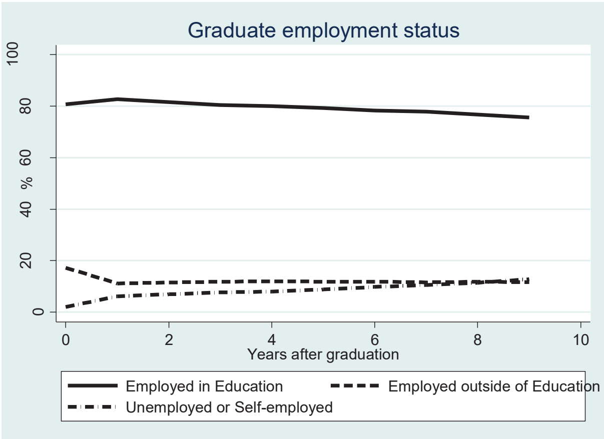
The labour market outcomes of interest are whether graduates are employed in the education sector at certain points after their graduation, monthly hours worked, hourly wages, and tenure status. For all graduates in the sample I create dummies indicating whether they had a job in the education sector for up to nine years after graduation. Since prior studies have shown that teacher dropout is concentrated within the first years of teaching (Hong,

[14] The reason for this is that I do not have information on students' exact date of graduation. However, most students graduate over the summer, at the end of the academic year. Hence, October of the calendar year after graduation corresponds to the third month after graduation for the large majority of students.

[15] To account for the fact that differences in program-level dropout rates are partially related to differences in student body composition between different teacher training programs, I run a sensitivity analysis where I first regress individual students' 4-year dropout probability on all student observables using *cohort*program* fixed effects. Then I use the predicted values of the *cohort*program* fixed effects instead of the raw 4-year dropout rate to explain graduates' labour market outcomes. This exercise leads to qualitatively similar results and does not change the conclusions from the main specifications.

2010; Borman & Maritza-Dowling, 2008; Ingersoll, 2001) retention after nine years is a good predictor for career longevity in education. Figure 2.3 shows the percentage of graduates working in the education sector, outside of the education sector, and the percentage that is not employed by a company in the Netherlands by the number of years after graduation. Around 80% of primary teacher education graduates work in the education sector immediately after graduating. The percentage of graduates working in education stays relatively stable as more years in the labour market are added.[16]

Figure 2.3: Primary teacher education students' employment status *x* years after graduation



Source: DUO 1cijferHO database and Statistics Netherlands POLIS salary administration file.

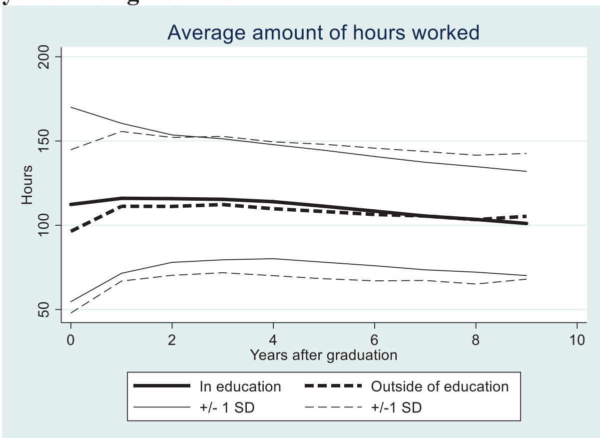
In order to interpret employment outside of the education sector as a sign of a poor suitability for teaching, it would help if teachers face worse labour market opportunities outside of the education sector. If ex-teachers have better labour market outcomes than those that stay in the profession, it might be a sign that the most capable teachers are the ones that quit teaching. If a low dropout rate at the teacher preparation program level would be related to a lower probability to be employed in the education sector, it could just as well be interpreted as a sign of high program quality: graduates of these teacher preparation programs are so well trained that their alternative labour market opportunities rise in accordance. For this reason, I provide some descriptive analyses comparing teacher preparation program

[16] Note that these percentages represent different groups of people at different points on the *x*-axis, as graduates from the more recent cohorts have not been observed for the full nine years after graduation, and the labour market outcomes of the graduates from the oldest cohorts were not observed before 2007. The same caveat applies to figures 2.4, 2.5, and 2.6. Alternative figures holding the cohort fixed at those who are in the labour market for at least five years are available upon request.

graduates working outside of the education sector with those who work inside of the education sector on hours worked, hourly wage, and tenure probability.

Figure 2.4 shows the average number of hours worked per month for those inside and outside of the education sector. Graduates working in the education sector work slightly more hours per month straight after graduating, but the difference between those inside and outside of the education sector decreases over time. What is interesting is that at an average of around 115 hours worked per month, a majority of graduates from primary school teacher education work significantly less than full time (140-160 hours). Working part-time seems to be the norm even early in teachers' careers, and those that find employment outside of the education sector do not appear to do so because of a desire to work more hours.

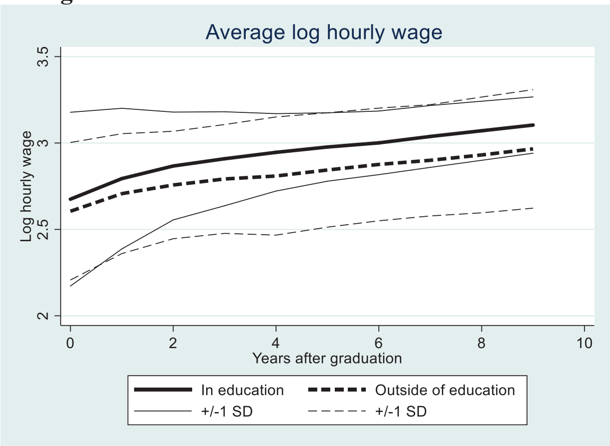
Figure 2.4: Average monthly hours worked for primary teacher education students *x* years after graduation



Source: DUO 1cijferHO database and Statistics Netherlands POLIS salary administration file.

Figure 2.5 plots the average log hourly wage of primary school teacher education graduates working inside and outside of education. On average, those working in education earn a higher wage and the variance of their earnings decreases over time. This is to be expected since in the Netherlands, as in most other countries, primary school teachers earn a fixed salary based on seniority, set by a collective bargaining agreement. Schools have some discretion in allocating teachers to different function profiles, but still the variation in wages is small. Graduates working outside of the education sector earn less on average, but there is more variation in their earnings. Some primary school teacher education graduates working outside of education do earn more than their teaching peers, but this is certainly not the norm.

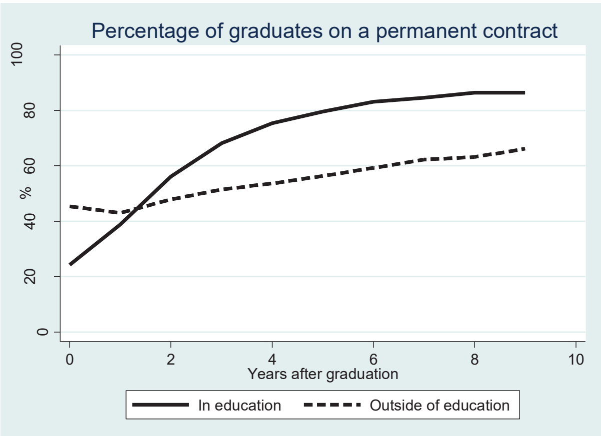
Figure 2.5: Average log hourly wage for primary teacher education students *x* years after graduation



Source: DUO 1cijferHO database and Statistics Netherlands POLIS salary administration file.

Finally, figure 2.6 shows the percentage of primary teacher education graduates working on a permanent contract both inside and outside of the education sector. Straight after graduation, those that work outside of education are more likely to be on a permanent contract. This could be because these graduates are still working the job they had during their studies while looking for a teaching job. However, after several years in the labour market a much larger proportion of those that work in the education field are tenured. In terms of job security, working in the education sector seems to be a better option for graduates of primary school teacher education programs.

Figure 2.6: Percentage of primary teacher education students on a permanent contract *x* years after graduation



Source: DUO 1cijferHO database and Statistics Netherlands POLIS salary administration file.

The overall picture that arises from the descriptives on labour market outcomes is that working in the education sector is associated with better outcomes than working outside of the education sector for those that graduate from primary school teacher education. At least for the first nine years of their career, both wages and job security are higher in the education sector, and the amount of hours worked is comparable. These patterns are in line with earlier studies such as Stinebrickner (2002) and Scafidi, Sjoquist, & Stinebrickner (2006), which also find that exiting teachers usually earn less outside of teaching, and suggest that most teaching graduates do not leave education because of superior outside labour market opportunities.

2.5 Results

In this section, I present the results of the main regressions relating primary school teacher education graduates' probability of employment in the education sector, hours worked, hourly wage, and tenure status by their *cohort*program* 4-year dropout rate. Secondly, using the regression results I conduct some basic calculations to explore the impact of different levels of student dropout on the total amount of hours of teaching supplied by primary school teaching education graduates.

2.5.1 Association between program dropout rates and graduates' labour market outcomes

As previously stated, the independent variable of interest for all outcomes is the *cohort*program* 4-year dropout rate. Additionally, I add controls for student age-at-entry, gender, migration background, highest level of prior completed education, full-time or part-time student status, number of previous studies in higher education, and whether the teacher education program was situated in the Netherlands' most strongly urbanized area (Randstad). Finally, controls are added for year of entry, and year of graduation. For all regressions, standard errors are clustered at the *year of entry*program* level.

The main reported regressions are conducted on the full population of primary teacher education graduates. In additional analyses, I restrict the sample to full-time students, and those graduates that were employed in the education sector immediately after graduation.[17] The former is done because part-time students differ strongly from full-time students in terms of age and prior labour market experience, and their inclusion could potentially make the results more difficult to interpret. Restricting the analyses to full-time students leads to

[17] The results of these additional analyses are available upon request.

qualitatively similar results, and does not change the conclusions from the main specifications.

The analyses restricted to graduates that were employed in the education sector immediately after graduation are conducted to check if the main results are not simply due to differences in labour market opportunities immediately after graduation. Since 85% of Dutch teaching graduates stay in the region in which they studied (Venhorst, van Dijk, & van Wissen, 2010), any positive relationship between *cohort*program* dropout rates and labour market outcomes could be driven by the lower regional supply of teacher graduates as a result of a relatively high dropout rate. There could simply be less competition for teaching jobs, and graduates' labour market outcomes may look better as a result, but this would be unrelated to the quality of the primary school teacher education program. By restricting the analyses to those graduates that find a teaching job immediately after graduation, I can take these differences in local labour market conditions into account. Again, restricting the analyses to this sub-sample does not affect the conclusions drawn from the main specifications.

2.5.1.1 Probability of being employed in the education sector

Table 2.4 shows the marginal effects of a probit regression of the probability of being employed in the education sector immediately after graduation on the *cohort*program* 4-year dropout rate. Column 1 shows that without controlling for individual characteristics, year of entry and year of graduation fixed effects, a one standard deviation increase in the *cohort*program* 4-year dropout rate is associated with a 1 percentage point lower probability of being employed in the education sector immediately after graduation. Considering the baseline probability of 80.7% of working in the education sector, this association is statistically significant but relatively weak.

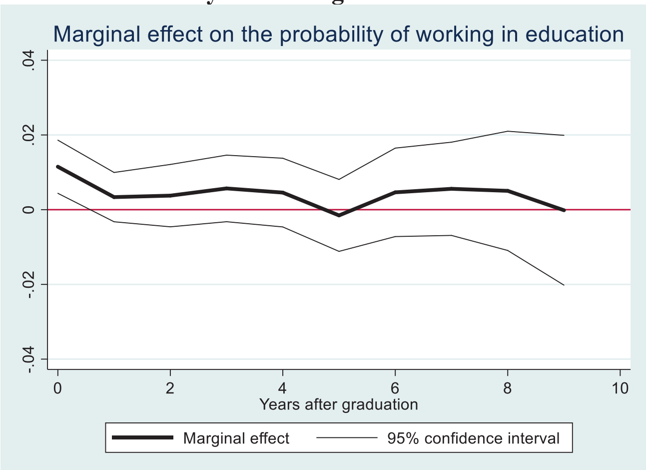
The addition of control variables changes the sign of the coefficient, and reveals a significant positive relationship between dropout rates and the probability to be employed in the education sector straight after graduation. The size of the association is still relatively low at 1.1 percentage points, and it is unclear whether it is due to a higher quality or a lower supply of graduates to fill the teaching jobs that are available at the regional level. Nevertheless, these results are instructive in that they show that policy makers' interpretation of high dropout rates as a sign of poor program quality are not necessarily reflected by the probability of retention in the teaching profession of these programs' graduates.

Table 2.4: Probability of employment in the education sector straight after graduation

VARIABLES	(1) In education after graduation	(2) In education after graduation	(3) In education after graduation
4-year dropout rate	-0.0105	-0.0092	0. 0115
	(0.0043)*	(0.0043)*	(0.0036)**
Gender (1=F)		-0.0127	-0.0146
		(0.0067)	(0.0067)*
Age		-0.0203	-0.0225
		(0.0037)**	(0.0037)**
Age^2		0.0004	0.0004
		(0.0001)**	(0.0001)**
Migration status:			
Western migrant		-0.0425	-0.0456
		(0.0150)**	(0.0147)**
2nd gen NW-migrant		-0.0717	-0.0741
		(0.0143)**	(0.0145)**
1st gen NW-migrant		-0.1563	-0.1679
		(0.0308)**	(0.0316)**
Prior education:			
Havo		-0.0342	-0.0407
		(0.0065)**	(0.0064)**
Vwo		-0.0344	-0.0394
		(0.0101)**	(0.0099)**
Unknown		-0.0856	-0.0884
		(0.0141)**	(0.0142)**
Studied in strongly urbanized area (1=Yes)		0.0294	0.0300
		(0.0082)**	(0.0066)**
Student status (1=Part-time)		0.0221	0.0212
		(0.0108)*	(0.0107)*
Studies before teacher training (1=Yes)		-0.0309	-0.0202
		(0.0085)**	(0.0085)*
Year-of-entry FE			X
Year-of-graduation FE			X
N	34,415	30,233	30,233

*Note: standard errors are clustered at the institution*year-of-entry level. The 4-year dropout rate is standardized to have mean 0 and standard deviation 1. The baseline category for migration status is "non-migrant". The baseline category for prior education is "MBO4". Year of entry*program clustered standard errors in parentheses.*
* $p<0.05$; ** $p<0.01$

Figure 2.7: Marginal effect of the 4-year dropout rate on being employed in the education sector *x* years after graduation:



Source: DUO 1cijferHO database and Statistics Netherlands POLIS salary administration file.

In order to analyse the probability to work in education after more time in the labour market, I run the same probit regressions as before, changing the dependent variable to whether a graduate is working in the education sector *x* years after entering the labour market. Figure 2.7 shows the marginal effect of the 4-year dropout rate for the probability of being employed in the education sector *x* years after graduation using the specification including student characteristics, year of entry and year of graduation dummies.[18]

The graph shows that the association between dropout rates at the program level and the probability of being employed in the education sector disappears after the first year in the labour market. From that point onwards, the dropout rate does not have an influence anymore. If low dropout rates are not associated with higher attrition from the teaching profession, we might draw the conclusion that policies aiming to reduce dropout rates at the teacher training level are a promising way to reduce teacher shortages. However, these analyses do not take into account how many hours of teaching these graduates actually supply. In addressing teacher shortages, it might be preferable to, for example, have 50% of graduates working full-time to having 60% work part-time. In the next section, I therefore look at the average amount of hours worked per month.

2.5.1.2 Monthly hours worked

To analyse hours worked, I run a simple OLS regression of the natural logarithm of monthly hours worked for those graduates working in the education sector at different years after graduation on the *cohort*program* 4-year dropout rate. As with the regressions on the probability of being employed in the education sector, table 2.5 shows the raw correlation and the results when adding control variables immediately after graduations, while figure 2.8 plots the coefficients of the 4-year program dropout rate *x* years after graduation using the specification including student characteristics, year of entry and year of graduation dummies.

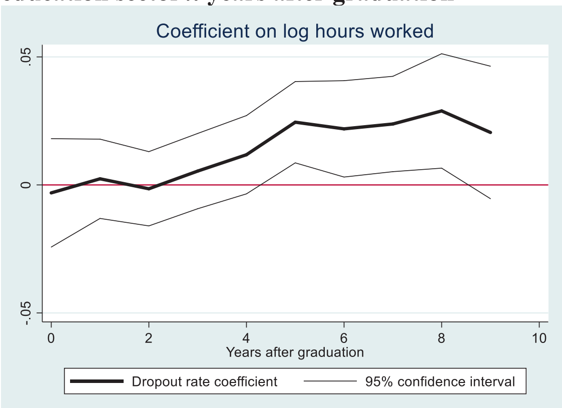
[18] The underlying regression output tables are available upon request.

Table 2.5: Monthly hours worked in the education sector straight after graduation

VARIABLES	(1) Log monthly hours worked	(2) Log monthly hours worked	(3) Log monthly hours worked
4-year dropout rate	-0.122**	-0.132**	-0. 003
	(0.017)	(0.019)	(0.011)
Gender (1=F)		-0.087**	-0.096**
		(0.015)	(0.0013)
Age		-0.003	-0.027**
		(0.009)	(0.008)
Age^2		0.000	0.000*
		(0.000)	(0.000)
Migration status:			
Western migrant		0.023	-0.003
		(0.023)	(0.021)
2nd gen NW-migrant		0.078**	0.051
		(0.025)	(0.024)
1st gen NW-migrant		0.123	0.064
		(0.043)	(0.040)
Prior education:			
Havo		-0.028*	0.072**
		(0.013)	(0.011)
Vwo		0.040*	0.003
		(0.018)	(0.016)
Unknown		-0.058*	-0.069**
		(0.023)	(0.022)
Studied in strongly urbanized area (1=Yes)		0.160**	0.158**
		(0.036)	(0.016)
Student status (1=Part-time)		-0.040	-0.022
		(0.024)	(0.022)
Studies before teacher training (1=Yes)		-0.029	0.036
		(0.0213)	(0.0223)
Year-of-entry FE			X
Year-of-graduation FE			X
Constant	4.5377**	4.641**	5.3073**
	(0.018)	(0.126)	(0.122)
Observations	27,079	23,994	23,994
R-squared	0.023	0.041	0.138

*Note: standard errors are clustered at the institution*year-of-entry level. The 4-year dropout rate is standardized to have mean 0 and standard deviation 1. The baseline category for migration status is "non-migrant". The baseline category for prior education is "MBO4". Year of entry*program clustered standard errors in parentheses.*
*p<0.05, **p<0.01

Figure 2.8: Coefficient of the 4-year dropout rate on log monthly hours worked in the education sector *x* years after graduation



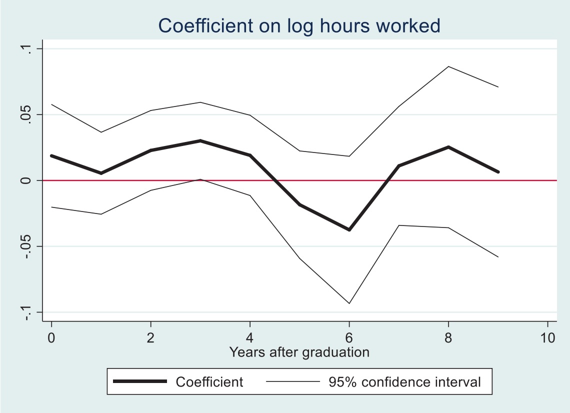
Source: DUO 1cijferHO database and Statistics Netherlands POLIS salary administration file.

The regressions on hours worked show that after about five years in the labour market, those graduates from primary teacher education programs with high dropout rates work significantly more hours per month. The relationship persists until eight years after graduation, and becomes insignificant after nine years. However, this last result might be because of the smaller amount of graduates that can already be observed for nine years in the labour market. The size of the association is relatively small, with a one standard deviation increase in dropout rates being associated with 2.5% more hours worked. At an average of around 111 hours per month, this implies about 2.8 more hours worked per month for those graduates from programs with a one standard deviation higher dropout rate.

While these results are non-causal and should be interpreted with caution, they are in line with a scenario in which the marginal graduate from a program with a low dropout rate is less suited for teaching, and derives less utility from working in the education field due to a lower job-match quality. This would lead them to be willing to supply fewer hours, while not willing to switch careers because of the associated costs and relatively poor alternative labour market options. To investigate whether this interpretation is plausible, we can see if graduates from programs with a low dropout rate also work fewer hours when they find a job outside of education. If this is the case, then it is difficult to argue that the willingness to supply fewer hours is related to the job match quality in the education sector and hence the suitability of these graduates for the teaching profession.

Figure 2.9 shows the coefficients of the 4-year dropout rate *x* years after graduation on log hours worked for those graduates working outside of the education sector. This graph shows that the association between dropout rates and monthly hours worked is specific for the education sector. This result, combined with the results of the analyses on the labour market outcomes of teacher preparation program graduates working outside of the education sector from section 2.4, lends plausibility to the interpretation of poorer teaching suitability of the marginal graduate of primary teaching education programs with low dropout rates.

Figure 2.9: Coefficient of the 4-year dropout rate on log monthly hours worked outside of the education sector *x* years after graduation



Source: DUO 1cijferHO database and Statistics Netherlands POLIS salary administration file.

2.5.1.3 Wage

Table 2.6 shows the results of a standard OLS regression of log hourly wage on the 4-year *cohort*program* dropout rate for graduates working in the education sector immediately after graduation. Without controlling for year of entry and year of graduation fixed effects, there appears to be a positive association between the dropout rate and graduates' hourly wage. However, since both wages and dropout rates are increasing over time, controlling is necessary and renders the correlation zero. This is also to be expected, since primary school teachers earn a salary based on seniority that is relatively inflexible.

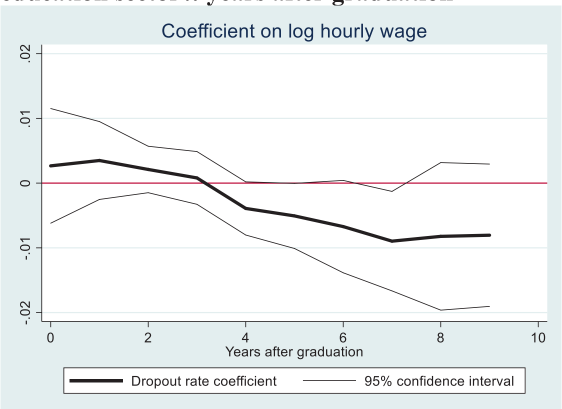
Figure 2.10 shows the coefficients of the 4-year dropout rate *x* years after graduation from the wage regressions. A negative relationship between dropout rates and hourly wage appears to be developing over time. However, the correlation is only significant at seven years after graduation, and the magnitude of the association is rather small. A one standard deviation increase in the dropout rate is associated with a 1% lower hourly wage seven years after graduation. Considering that the average hourly wage at this point is €21,17, this corresponds to a €0,21 decrease.

Table 2.6: Log hourly wage in the education sector straight after graduation

VARIABLES	(1) Log hourly wage	(2) Log hourly wage	(3) Log hourly wage
4-year dropout rate	0.026**	0.24**	0. 003
	(0.006)	(0.005)	(0.011)
Gender (1=F)		-0.054**	-0.035**
		(0.007)	(0.007)
Age		-0.009*	-0.004
		(0.004)	(0.004)
Age^2		0.000*	0.000*
		(0.000)	(0.000)
Migration status:			
Western migrant		-0.003	-0.006
		(0.010)	(0.010)
2nd gen NW-migrant		0.012	0.002
		(0.011)	(0.011)
1st gen NW-migrant		-0.004	0.004
		(0.020)	(0.019)
Prior education:			
Havo		-0.024**	0.038**
		(0.007)	(0.006)
Vwo		0.004	0.000
		(0.010)	(0.009)
Unknown		0.0013	-0.000
		(0.014)	(0.012)
Studied in strongly urbanized area (1=Yes)		0.058**	0.072**
		(0.011)	(0.007)
Student status (1=Part-time)		0.059**	0.060**
		(0.013)	(0.009)
Studies before teacher training (1=Yes)		0.0393**	0.0533**
		(0.0093)	(0.0085)
Year-of-entry FE			X
Year-of-graduation FE			X
Constant	2.746**	2.865**	2.634**
	(0.006)	(0.058)	(0.059)
Observations	27,075	23,992	23,992
R-squared	0.004	0.023	0.150

*Note: standard errors are clustered at the institution*year-of-entry level. The 4-year dropout rate is standardized to have mean 0 and standard deviation 1. The baseline category for migration status is "non-migrant". The baseline category for prior education is "MBO4". Year of entry*program clustered standard errors in parentheses.*
*p<0.05, **p<0.01

Figure 2.10: Coefficient of the 4-year dropout rate on log hourly wage in of the education sector *x* years after graduation



Source: DUO 1cijferHO database and Statistics Netherlands POLIS salary administration file.

Overall, program dropout rates do not seem to have an impact on the wages of the program's graduates. Even if dropout rates would have an impact on the productivity of the average program's graduate, the finding that this is not reflected in their wages is not that surprising because of the seniority based wage structure for primary school teachers. Some differentiation could occur because teachers who take on additional responsibilities within their schools tend to be compensated for this. Therefore, it could have been the case that graduates from high dropout programs would be more likely to take on these additional roles, which would be reflected in their wage. However, this does not appear to be the case either.

2.5.1.4 Contract status

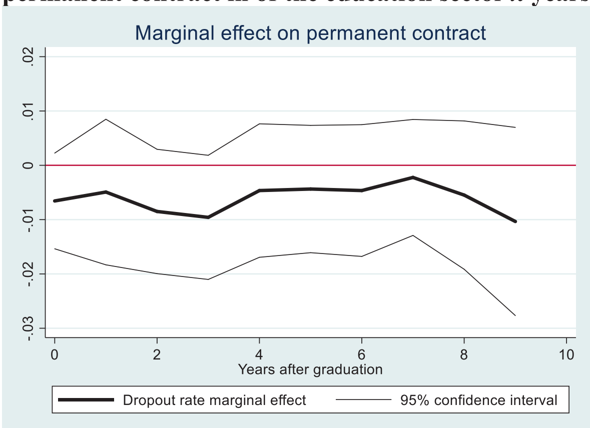
Finally, table 2.7 shows the results of a probit regression of permanent contract, or tenure, status on the 4-year *cohort*program* dropout rate for graduates working in the education sector immediately after graduation. Again, when controlling for year of entry and year of graduation effects, the initial negative association between dropout rates and the probability of working on a permanent contract disappears. Figure 2.11 plots the marginal effects of the dropout rate on permanent contract status *x* years after graduation. After several years in the labour market there appears to be no relationship between dropout rates and the probability of obtaining tenure. Since tenure is mostly granted by default after a certain number of years working for the same employer, this would suggest that graduates from programs with a low dropout rate do not disproportionally switch employers within the education sector.

Table 2.7: Probability of working on a permanent contract in the education sector straight after graduation

VARIABLES	(1) Permanent contract status	(2) Permanent contract status	(3) Permanent contract status
4-year dropout rate	-0.0885	-0.0828	-0. 0066
	(0.0097)**	(0.0094)**	(0.0045)
Gender (1=F)		0.0148	0.0247
		(0.0079)	(0.0067)**
Age		-0.0089	-0.0128
		(0.0043)**	(0.0036)**
Age^2		0.0002	0.0002
		(0.0001)**	(0.0001)**
Migration status:			
Western migrant		0.0051	-0.0123
		(0.0110)	(0.0098)
2nd gen NW-migrant		0.0036	-0.0231
		(0.0147)	(0.0106)*
1st gen NW-migrant		0.0105	-0.213
		(0.0243)	(0.0187)
Prior education:			
Havo		-0.0155	-0.0573
		(0.0065)*	(0.0059)**
Vwo		0.0089	-0.0252
		(0.0105)**	(0.0086)**
Unknown		0.0079	-0.0270
		(0.0155)**	(0.0112)**
Studied in strongly urbanized area (1=Yes)		0.0070	0.0272
		(0.0174)	(0.0057)**
Student status (1=Part-time)		0.1567	0.1651
		(0.0180)**	(0.0107)**
Studies before teacher training (1=Yes)		0.0235	0.0715
		(0.0111)*	(0.0117)*
Year-of-entry FE			X
Year-of-graduation FE			X
N	27,079	23,993	23,993

*Note: standard errors are clustered at the institution*year-of-entry level. The 4-year dropout rate is standardized to have mean 0 and standard deviation 1. The baseline category for migration status is "non-migrant". The baseline category for prior education is "MBO4". Year of entry*program clustered standard errors in parentheses.*
*p<0.05, **p<0.01

Figure 2.11: Marginal effect of the 4-year dropout rate on the probability of having a permanent contract in of the education sector *x* years after graduation



Source: DUO 1cijferHO database and Statistics Netherlands POLIS salary administration file.

2.5.2 Implications for net teacher supply

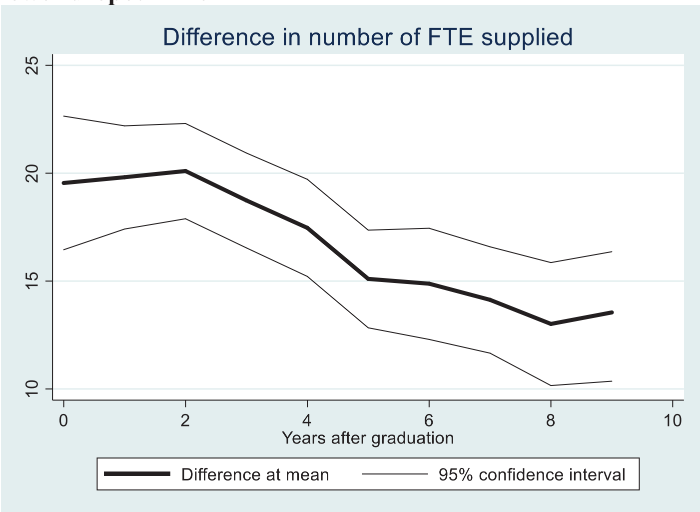
From the analyses above it appears that a lower dropout rate is associated with fewer hours worked in the education sector for these programs' graduates. In terms of overall teacher supply, it is important to know whether the increase in the number of graduates offsets the reduction in average hours worked. This section shows some basic calculations using the results from the regressions on monthly hours worked and teacher retention rates to see whether low dropout rates are associated with a higher supply of FTE working in the education sector.

To illustrate the calculations, I create an entry cohort of 400 students, which is about the average size of the cohort per program over the period studied. In the baseline scenario 47% of students drop out, which is the average 4-year dropout rate over all cohorts in all programs. In the low dropout scenario, the dropout rate is one standard deviation lower, at 39%. For the baseline scenario, I calculate the total number of hours supplied each year after graduation using the average amount of hours worked in the education sector from figure 2.4 multiplied by the amount of graduates and the percentage of graduates working in the education sector from figure 2.3.

In the low dropout scenario, the average amount of hours worked in the education sector is shifted by the percentage resulting from the coefficients of the 4-year dropout rate from the regressions on monthly hours worked in table 2.5 and figure 2.8, and again multiplied by the number of graduates and the percentage of graduates working in the education sector from figure 2.3. Supplied FTE are calculated by taking the sum of monthly hours worked and dividing them by the monthly hours worked by a full time employee.

Figure 2.12 shows the difference in the amount of FTE supplied between the average and one standard deviation lower program dropout rate after *x* years in the labour market for an average sized cohort. The upper and lower lines represent the 95% confidence interval. The figure shows that relative to the baseline scenario with an average dropout rate, a program with a one standard deviation lower dropout rate supplies significantly more FTE. The reduction in average hours worked is more than compensated for by the increase in the amount of graduates. The difference is around 20 FTE for the first couple of years after graduation, decreasing to around 15 FTE after five years. In terms of percentages, this corresponds to a 10-15% difference in the amount of FTEs supplied.

Figure 2.12: Difference in the number of FTE supplied between an average sized program with an average dropout rate and an average sized program with a one SD lower dropout rate



Source: own calculations.

Considering that this calculation is for a single program, reducing dropout rates across all programs could result in a sizable increase in the supply of teachers. However, these results should be interpreted with caution, as the regressions on which the calculations are based are not causal. Therefore, it cannot be said that reducing dropout rates through interventions would definitely result in an increase in the supply of teachers, only that it is plausible based on these observational analyses.

2.6 Conclusions

In order to combat looming teacher shortages, a promising focal point for policy makers could be to reduce dropout rates from teacher education. However, it is a priori unclear whether low dropout rates are associated with a high supply of suitable teachers, as low dropout rates can be a sign of both high quality and low standards. In this chapter, I use Dutch registry data on the population of primary school teacher education students and their labour market outcomes to analyse program level dropout rates and investigate their relationship with these programs' graduates' working careers. This is the first study to consider dropout rates at the teacher training level in explaining differences in labour market outcomes between graduates of different teacher education programs.

The results show that dropout rates are unrelated to the probability of working in the education sector in the first nine years after graduation. Furthermore, dropout rates do not have an influence on graduates' propensity to obtain tenure, nor is there a consistent relationship with the hourly wage. However, graduates from low dropout rate programs work around 2.5% fewer hours per month after several years in the labour market. While non-causal, these results are consistent with a slightly lower teaching suitability for the marginal graduate of these programs leading to a poorer job-match quality, less utility derived from teaching, and ultimately resulting in a willingness to supply fewer hours. This interpretation is corroborated by a lack of association between dropout rates and monthly hours worked outside of the education sector, and inferior labour market outcomes for those who decide to switch to a job outside of education.

For policy makers aiming to increase teaching supply, it is important to know whether the reduction in average hours worked is offset by the increase in the amount of graduates. Calculations investigating this trade-off show that an average sized program with a one standard deviation below average dropout rate supplies 10-15% more teacher FTEs compared to a program with an average dropout rate for the first nine years after graduation. Considering that there are 23 programs that offer teacher training, the potential increase in the supply of teachers through a reduction in dropout rates could be sizable.

The question whether reducing dropout rates through interventions would lead to a similarly sized increase in teacher supply cannot be answered based on these results alone, since the analyses are observational in nature. In order for decreased dropout rates to increase the supply of suitable teachers, the crucial point is that institutions are not able to "game" the system by reducing their standards. If an external quality control body monitors program quality, institutions cannot lower their graduation requirements without risking losing their accreditation, and any reduction in dropout rates has to come from an increase in suitable teachers meeting the labour market requirements. While in a setting where there is no strong accreditation program, programs could reduce dropout rates by decreasing their standards. Policies providing programs with incentives to decrease their dropout rates should therefore always be accompanied by a means through which a certain baseline of program requirements is guaranteed.

3. Teacher-student assortative matching and student achievement

3.1 Introduction

In the Netherlands, as well as in most other developed countries, there is a large achievement gap between students from different migration and socio-economic backgrounds (Inspectorate of Education, 2018; OECD, 2017). There is some international evidence that school segregation is partly responsible for the size of the gap (Card & Rothstein, 2007). One mechanism through which school segregation could harm educational opportunities is through an unequal distribution of teaching resources, as teaching quality has been shown to be strongly related to student achievement (Hanushek & Rivkin, 2006; Hanushek, 2011; Chetty, Friedman, & Rockoff, 2014). When schools serving a disadvantaged student population have more trouble attracting high quality teachers, this could further increase school segregation, as well as its negative impact on equality of educational opportunities (Inspectorate of Education, 2018).

In this chapter, I investigate the extent of positive assortative matching between student and teacher characteristics and its' relationship with student achievement.[19] I use administrative data on all primary school students and teacher assignments in the Netherlands over the period 2008 to 2016 to show that there is strong sorting along both educational lines and migration background. Schools serving a larger proportion of children with university-educated parents employ a larger percentage of teachers holding a master's degree, while schools with a high percentage of students from a non-western migration background employ more teachers with a non-western migration background. These patterns are especially pronounced in urban areas, where both parents and teachers have more options to act on their preferences for school, student, and teacher characteristics. Over time, assortative matching is slightly increasing along educational lines, and slightly decreasing along migration background. Analyses focusing on early career teachers that graduated between 2007 and 2015 show that the sorting patterns of young teachers reinforce the sorting patterns on average.

To investigate the relationship between student educational outcomes and positive assortative matching between student and teacher characteristics, I run OLS and school fixed-effects regressions relating teacher characteristics at the school level to individual student achievement at the end of primary school. The results suggest that assortative matching on

[19] This chapter has benefitted from valuable comments by Tijana Breuer, Bart Golsteyn, Roxanne Korthals, Nienke Ruijs, Mariana Tavares, and Inge de Wolf. Thanks to the Inspectorate of Education, DUO, and Statistics Netherlands for sharing their datasets.

migration background is not negatively related to student performance. Students with a non-western migration background perform slightly better in schools with a larger share of non-western migrant teachers, with no negative effects found for native students. The results are more pronounced for students from a relatively low socio-economic background, and are driven by increased performance on the mathematics part of the test, with no matching effects found for the language part of the test. In contrast, the share of teachers holding a master's degree is unrelated to the performance of students with university-educated parents, nor do students with low educated parents perform worse in schools with a larger percentage of master's degree holding teachers.

Combining the results of positive assortative matching on teacher and student characteristics with the positive interactions between student and teacher migrant background on student achievement, a picture arises in which segregation of teachers may not be detrimental to educational outcomes. However, these results should be interpreted with caution, as the analyses are non-causal in nature. Another caveat is that the data do not allow for students to be linked to their individual teacher in each grade. This implies that there is no certainty that students have been taught by a teacher that matches their background characteristics, but only that this probability increases when the share of teachers with matching characteristics within a school is larger. Therefore, it is likely that the results underestimate the relationship between teacher-student matching and student achievement.

This chapter adds to the literature on teacher sorting and student achievement. In an early paper, Lankford, Loeb, & Wyckoff (2002) examine the extent of teacher sorting for the state of New York. They find that schools serving disadvantaged students employ teachers with fewer qualifications, consistent with the results of this chapter. Other studies on teacher sorting patterns (e.g. Clotfelter, Ladd, & Vigdor, 2006; Goldhaber, Choi, & Cramer, 2007) reach similar conclusions. Analyzing the differences in teacher quality between schools serving advantaged and disadvantaged children, Sass et al. (2012) show that the average value added of teachers in high poverty schools is lower than the average value added of teachers in low poverty schools. However, consistent with the results of this chapter, they find that teacher certification does not explain much of the variation in teacher quality across schools.

This chapter contributes to the literature by using information on the match between individual student characteristics and teacher characteristics at the school level. This information allows me to investigate whether positive assortative matching can potentially affect student performance. Matching effects on student achievement in primary and secondary school have been studied by Dee (2004, 2005). His studies show that an ethnic match between student and teacher improves student outcomes for both majority and minority students, particularly for students from a low socio-economic background. Studies focusing on matching effects in higher education show strong positive interaction effects along ethnic lines on dropout rates and student achievement (Fairlie, Hoffmann, & Oreopoulos, 2014). This chapter corroborates the findings of these previous studies in a different institutional context, while adding evidence on a lack of matching effects between additional teacher certification and students' parental educational background.

While the results suggest that segregation along migration background might not have negative effects from an educational effectiveness point of view, this does not imply that schools and policy makers should stimulate segregation on migration background of their student and teacher force. One concern is that other important functions of the school system such as integration, socialization, and citizenship skills of migrant students may be enhanced more by increased diversity and exposure to teachers and students from different backgrounds. The potential gains in student achievement on standardized tests could be outweighed by the losses in terms of the socialization function of the educational system. However, the same goes for non-migrant students. Their exposure to teachers from different backgrounds may not affect their test scores, but there could be gains in terms of socialization outcomes. Unfortunately, the impact of exposure to teachers with different background characteristics on these socialization outcomes are difficult to quantify and beyond the scope of this chapter.

The remainder of this chapter is structured as follows. In section 3.2, I discuss the data and show some descriptive statistics. Section 3.3 presents the results, and section 3.4 concludes.

3.2 Data

In this section, I introduce the data sources used and provide descriptive statistics. First, the data concerning the school-level averages of teacher and student characteristics are discussed. Second, I describe the subsample of early career teachers in more detail. Finally, I discuss the characteristics of the individual student data used for analyzing the relationship between teacher characteristics and student outcomes.

3.2.1 School-average student and teacher characteristics

The main dataset used for constructing school-level average student characteristics is the registration file that is used to finance schools based on student enrollment data, the 'DUO 1cijferPO' registration file. I use the data for the years 2008-2016. The dataset contains information on all students enrolled in primary education in the Netherlands, their background characteristics, and school characteristics. This dataset is combined with information on parental education using data on highest obtained education from Statistics Netherlands, and collapsed at the *school*year* level (BRIN4) to obtain school-by-year averages of student characteristics.[20]

Information on teacher assignments and teacher characteristics come from the Dutch teacher registration file, the 'DUO Functiemix' file. This file is based on national administrative salary data, and contains yearly information on all teacher assignments in primary and secondary school for all schools in the Netherlands between 2008 and 2016.[21] From this file, I select those teachers working in regular primary education, and add information on the highest obtained level of education, and the municipal administration data for teachers' date of birth, gender, and migrant status from Statistics Netherlands. Finally, information about each teacher's wage, monthly hours worked, and tenure status are added from the salary administration data (the 'Polis administration file') of Statistics Netherlands. These data are then similarly collapsed at the *school*year* level (BRIN4) to obtain *school*year* averages of teacher characteristics and student characteristics.

[20] Ideally, the data would be aggregated at the school-location level (BRIN6), but unfortunately the teacher data are not precise enough to allow identification of teachers at that level of detail. While 98% of primary schools in the dataset have only one location, for the 2% of schools with multiple locations there is some uncertainty with respect to which students are exposed to which group of teachers. The analyses are unaffected by the exclusion of schools for which the exact location of teachers cannot be determined.

[21] The set date for the teacher assignments is October 1 of every calendar year.

Table 3.1: School-level average student and teacher characteristics

	School-level characteristics			
	Netherlands		Urban areas	
Variable	**Mean**	**SD**	**Mean**	**SD**
School characteristics				
Number of schools	56,808		8,621	
Number of students	229.98	138.22	314.20	157.84
Number of teachers	22.77	12.56	31.57	15.66
Student characteristics				
Girls pct.	49.56	3.49	49.72	2.98
Migrant status:				
Non-migrant pct.	76.49	22.28	52.21	27.68
Non-western migrant pct.	16.66	20.67	37.66	28.37
Western migrant pct.	6.83	4.82	10.11	6.40
Parental education:				
University pct.	18.82	15.12	23.93	20.36
HBO pct.	26.49	10.62	19.07	10.29
MBO34 pct.	31.58	9.50	25.58	8.85
Max MBO2 pct.	23.09	17.06	31.41	23.88
Unknown pct.	20.32	9.89	15.39	8.02
Cito-score	535.24	3.81	534.41	4.82
Teacher characteristics				
Female pct.	82.90	7.95	82.74	7.27
Age	43.78	3.80	43.15	3.70
Migrant status:				
Non-migrant pct.	90.97	11.54	80.73	17.51
Non-western migrant pct.	3.73	9.54	11.39	16.56
Western migrant pct.	5.28	5.90	7.86	7.59
Master's degree pct.	19.04	12.01	18.13	10.55
Hourly wage	22.51	2.067	22.19	2.16
Tenured pct.	90.04	11.12	88.04	12.23
Monthly hours worked	120.76	12.42	127.9	11.54
New teacher in school pct.	12.24	11.13	12.12	10.66
Teachers that did not return pct.	12.10	9.05	11.40	7.97

Note: school-level average student (teacher) characteristics are calculated using the number of students (teachers) within a school as analytic weights.
Source: DUO 1cijferPO, DUO Functiemix, Statistics Netherlands POLIS, Statistics Netherlands municipal administration, and Statistics Netherlands highest achieved level of education databases.

Table 3.1 shows the weighted average student and teacher characteristics at the school level for the entire period studied (2008-2016), separately for the whole of the Netherlands and highly urbanized areas. The reason for making the distinction between strongly urbanized areas and the country as a whole is that positive assortative matching is more likely to occur when both students and teachers have more opportunities to sort on their preferences. In non-urbanized areas, there may be just one or two schools for parents to send their children to (as prior research has shown that parents prefer not to travel too far for primary schools (Borghans, Golsteyn, & Zölitz, 2015)). Likewise, teachers living in these rural areas have fewer schools to apply to than their city-dwelling colleagues. There are 56,808 school*year combinations in the dataset in total, i.e. around 6,200 schools per year. Schools in highly urbanized areas tend to serve more students, have a larger percentage of students with a migrant background, and more students whose parents have either a very high or a very low education level. The average test score at the end of grade six is also slightly lower in strongly urbanized areas.

In terms of teacher characteristics, the vast majority of teachers are female (82.9%), and 90% of teachers have no migration background. Around 19% of teachers within each school for whom information on their highest obtained education is available, have obtained a master's degree. The average monthly hours worked within each school (120 hours per month) is significantly less than full time (140-160 hours per month), implying that a large share of teachers work on a part-time contract. Around 90% of teachers within each school are tenured, and yearly teacher turnover comprises around 12% of the total teaching force within each school.

3.2.2 Early career teacher individual level data

To identify early career teachers, I use data from national higher education student registration files (the DUO 1cijferHO database). This administrative dataset includes information on all student registrations between 2002 and 2016 in the Dutch (subsidized) higher education system, both for higher vocational and university level programs. This dataset includes information of the full-time or part-time student status, whether a student has graduated or not, the highest obtained educational level before enrolling in higher education and on the students' grades and track in secondary education. From this file, I select those full-time students[22] that graduated from primary teacher training between the academic years 2007/2008 and 2015/2016.[23] The data on the graduates are then linked to the 'DUO Functiemix teacher registration' file. This creates a panel dataset where each graduate's first observation is the year in which s(he) started their first teaching job. Finally, for all graduates working as a teacher I add the *school*year* average student and school characteristics from the school-average level dataset described earlier.

Table 3.2: Individual early career teacher characteristics per year of graduation

	Early career teacher characteristics								
	Graduation year								
Variables	2007	2008	2009	2010	2011	2012	2013	2014	2015
N	3,734	3,216	2,884	2,846	2,592	2,329	2,231	2,046	1,608
Female pct.	88.9	89.6	87.7	89.6	87.7	87.9	86.7	84.8	85.8
Average age	24.4	24.4	24.6	25.0	25.1	24.8	24.6	24.4	24.0
Non-western migrant pct.	4.47	5.6	3.74	3.72	3.78	4.89	4.35	5.33	5.6
Western migrant pct.	4.31	3.54	3.95	3.55	3.74	3.65	3.54	3.52	3.73
Master's degree pct.	10.28	7.62	7.47	9.23	9.13	9.22	10.48	11.98	.
MBO pct.	34.68	34.61	32.87	33.98	33.83	32.33	32.14	30.89	26.06
Havo pct.	53.83	53.79	54.33	53.09	54.4	53.03	54.1	53.76	55.97
Vwo pct.	8.62	9.11	10.26	11.03	10.11	13.44	12.33	14.03	16.79
Exam grade VMBO graduates	.	.	.	.	6.38	6.53	6.38	6.35	6.44
Exam grade Havo graduates	.	.	.	.	6.19	6.18	6.15	6.16	6.27
Exam grade Vwo graduates	.	.	.	.	6.09	6.05	6.01	6.04	6.02

Note: average exam grades for the cohorts prior to 2011 and information on obtained master's degrees for the 2015 cohort are not available because of data constraints.
Source: DUO 1cijferHO, DUO Functiemix, DUO vakkenanalysebestand, Statistics Netherlands highest achieved level of education, and Statistics Netherlands municipal administration databases.

Table 3.2 shows the average characteristics of the early career primary school teacher subsample by graduation year. As in the full teacher population, the majority of the early career teachers are female (88%) and non-migrants (91%). The share of early career teachers holding a master's degree is relatively low at around 8%, which suggests that teachers holding a master's degree usually obtain one over the course of their working career. There is a slight increase in the amount of non-western migrants that graduate primary teacher education over time. Most early career teachers attended the middle track (Havo) during secondary school, but the share of graduates from the high secondary school track (Vwo) is steadily increasing. The amount of graduates observed in the data is decreasing over time. This is not only because of dwindling enrolment rates into primary education teacher training

[22] Only full-time students are considered because part-time students are likely to already be employed at a primary school well before their graduation date. By excluding part-time students, it is much more likely that the individual level analyses truly capture the sorting behaviour of early career teachers.
[23] The reason for restricting the sample to those graduating from the academic year 2007/2008 onward is that the aim is to follow teachers from the first year of entering the teaching force, and there is no information available on teacher assignments prior to 2008. Those that graduate in the academic year 2007/2008 will be in the labour market by October 2008, plausibly working their first teaching job.

(de Wolf, Vermeulen, & Breuer, 2018), but also because the graduates from later years have had less time in the labour market to find a teaching job. Some graduates only start teaching 3 or 4 years after graduation. As a result, those that graduated relatively recently have a higher chance of not showing up in the teacher registry database yet.

3.2.3 Student level data on educational outcomes

The information on student level educational outcomes is derived from the national register on students in primary education (the '1cijferPO registration' files) from 2008-2016. I keep all students that took the most commonly used grade six end of primary school test (Cito) during the period studied.[24] Table 3.3 shows the descriptive statistics of the individual level student data, as well as the average sixth grade classroom composition. Pooling all cohorts, there are over 1.1 million students that took the Cito end of primary school standardized test, with an average score of 535.25.[25] Furthermore, there is separate information on students' achievement at the math and language sections of the test. The average age at which students take the Cito test is 11.48. Because of the timing at which children start primary school in the Netherlands, students are relatively young within their cohort if they are born between July and September, and relatively old when they are born between October and December. Finally, I calculate the share of their classmates from a certain parental educational background, migration background, and the gender composition and average age in their classroom for all students taking the Cito.

[24] While alternative tests have become somewhat more popular during the last three years, over the period studied the vast majority of students that took any form of standardized test at the end of primary school, took the Cito test. In the first two years of the dataset (2008 and 2009), test scores are not available for relatively many students. All other years see Cito test-score coverage of around 80%. Some concerns about selectivity of the group for whom test scores are available in 2008 and 2009 may arise. However, all analyses are robust to the exclusion of students who took the Cito test in 2008 and 2009.

[25] In the Cito test, the range of possible test scores is 501-550

Table 3.3: Individual student and average sixth grade student characteristics

Cito student characteristics					
Individual characteristics			**Average sixth grade characteristics**		
Variable	**Mean**	**SD**	**Variable**	**Mean**	**SD**
Number of students	1,117,268		Average number of students	39.05	21.35
Girls pct.	50.22		Girls pct.	50.21	10.01
Age	11.48	0.65	Average age	11.58	0.15
Migrant status			Migrant status		
Non-migrant pct.	77.66		Non-migrant pct.	77.63	23.72
Western migrant pct.	6.20		Western migrant pct.	6.21	5.76
Non-western migrant 2nd gen pct.	14.76		Non-western migrant pct.	16.16	22.44
Non-western migrant 1st gen pct.	1.38				
			Parental education:		
Parental education:			University pct.	17.54	15.62
University pct.	13.82		HBO pct.	25.19	13.18
HBO pct.	19.05		MBO34 pct.	33.49	13.62
MBO34 pct.	25.08		Max MBO2 pct.	23.76	18.72
Max MBO2 pct.	18.32		Unknown pct.	23.72	12.79
Unknown pct.	23.72				
Relative age					
Early pct.	26.06				
Average pct.	48.78				
Late pct.	25.15				
Cito-score	535.25	9.79			
Cito-score language	83.92	18.37			
Cito-score math	49.79	14.37			
Predicted Cito-score	535.32	3.26			

Source: DUO 1cijferPO, Statistics Netherlands "kenmerken deelnemers aan de Eindtoets Basisonderwijs van Cito", Statistics Netherlands highest achieved level of education, and Statistics Netherlands municipal administration databases.

3.3 Results

In this section, I present the results of the analyses of teacher sorting and student outcomes. First, the sorting patterns at the school level are discussed. Secondly, results on the sorting behaviour of early career teachers are shown. Finally, I present the results of OLS and school fixed effects regressions showing the association between teacher characteristics and individual student outcomes at the end of primary school.

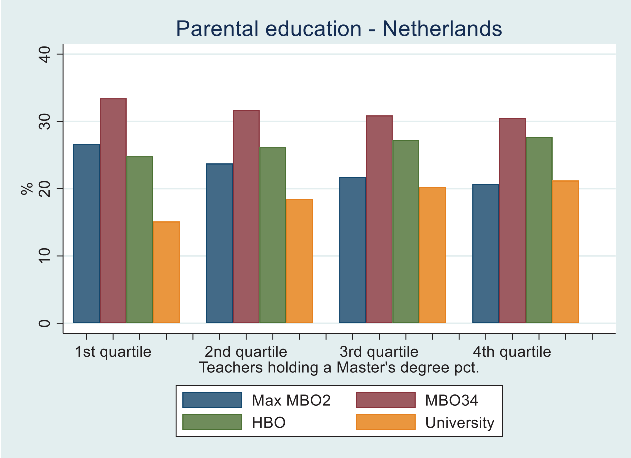
3.3.1 Teacher sorting at the school level

In order to visualize the extent of assortative matching between student and teacher characteristics at the school level, I divide schools into quartiles based on their share of teachers that obtained a master's degree in addition to their initial teaching qualification, and their share of teachers with a non-western migration background separately. For each quartile, the average percentage of students with a certain parental education level, and the average percentage of students from a migration background are calculated. Results are reported for the Netherlands as a whole, as well as for strongly urbanized areas separately.

Figures 3.1A and 3.1B show the average percentage of students from a certain parental educational background against the quartiles of the share of teachers that obtained a master's degree for the whole of the Netherlands, and strongly urbanized areas respectively.[26] In schools with a large share of teachers with a master's degree, the share of students whose parents completed a university degree is higher: from 15% of students for the schools with the lowest share, to around 22% of students for the schools with the highest share. The results are more pronounced in strongly urbanized areas, where the share of students whose parents completed a university degree is 31% on average in schools employing the largest share of teachers holding a master's degree (compared to 17% in schools employing the lowest share of teachers holding a master's degree). These results are indicative of assortative matching on educational levels.

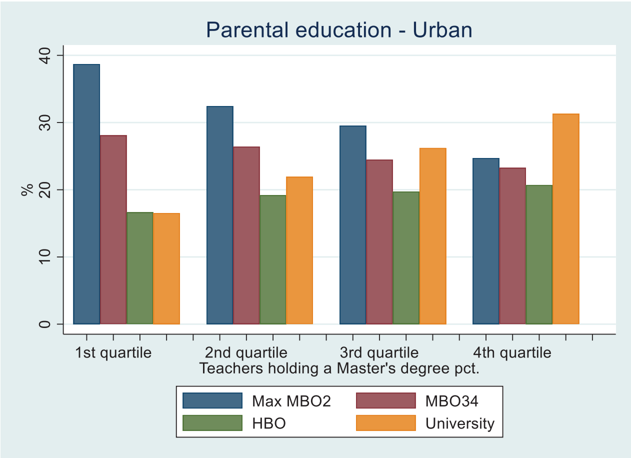
[26] The average shares of teachers holding a master's degree per quartile for the Netherlands are Q1: .05, Q2: .14, Q3: .22, Q4: .35. The shares for strongly urbanized areas are Q1: .06, Q2: .14, Q3: .21, Q4: .33

Figure 3.1A: Average percentage of students from a certain parental educational background per quartiles of the school-level share of master's degree holding teachers - Netherlands



Source: DUO 1cijferPO, DUO Functiemix, and Statistics Netherlands highest achieved level of education databases.

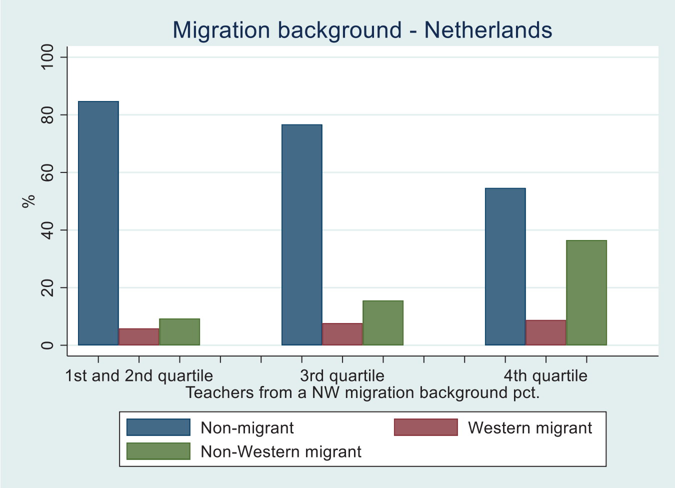
Figure 3.1B: Average percentage of students from a certain parental educational background per quartiles of the school-level share of master's degree holding teachers - Urban



Source: DUO 1cijferPO, DUO Functiemix, and Statistics Netherlands highest achieved level of education databases.

Figures 3.2A and 3.2B follow the same principle, this time dividing the schools into quartiles based on the share of teachers with a non-western migrant background and the average share of students from a certain migrant background. Since there are relatively few non-western migrant teachers in total, the median school in the Netherlands does not employ any teacher with this background. As a result, there is no distinction possible between the first and second quartile. In strongly urbanized areas, there are more teachers with a non-western migration background, and a distinction between the first and the second quartile becomes possible again.[27] Both figures show that schools where the share of teachers with a non-western migrant background is higher tend to serve more students from a non-western migration background. For the whole of the Netherlands, this result is to be expected, since the population of non-western migrants is mostly centered in urban areas. However, zooming in on these particular urban areas, the pattern is even more striking. In these areas, schools that employ the largest share of non-western migrant teachers serve around 75% non-western migrant students, while schools without any teachers from a non-western migrant background serve only 20% non-western migrant students. These results are indicative of strong matching along migration background.

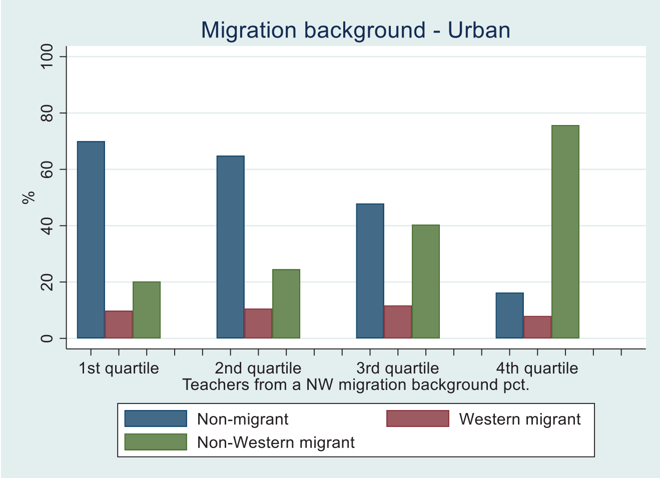
Figure 3.2A: Average percentage of students from a certain migration background per quartiles of the school-level share of non-western migrant teachers - Netherlands



Source: DUO 1cijferPO, DUO Functiemix, and Statistics Netherlands municipal administration databases.

[27] The average shares of teachers from a non-western migration background per quartile for the Netherlands are Q1 & Q2: .00, Q3: .03, Q4: .13. The shares for strongly urbanized areas are Q1: .00, Q2: .04, Q3: .10, Q4: .33

Figure 3.2B: Average percentage of students from a certain migration background per quartiles of the school-level share of non-western migrant teachers - Urban

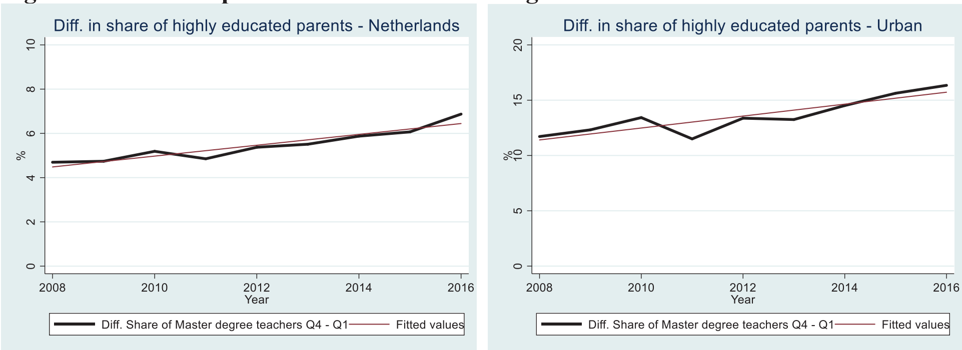


Source: DUO 1cijferPO, DUO Functiemix, and Statistics Netherlands municipal administration databases.

To see how these sorting patterns are developing over time, I plot the difference in the share of students with highly educated parents between schools in the top and bottom quartiles of the share of teachers holding a master's degree for each year separately. For migration background, the difference in the share of students from a non-western migrant background is plotted in the same manner. Figures 3.3A and 3.3B show the results of this exercise for educational background, and migration background respectively.

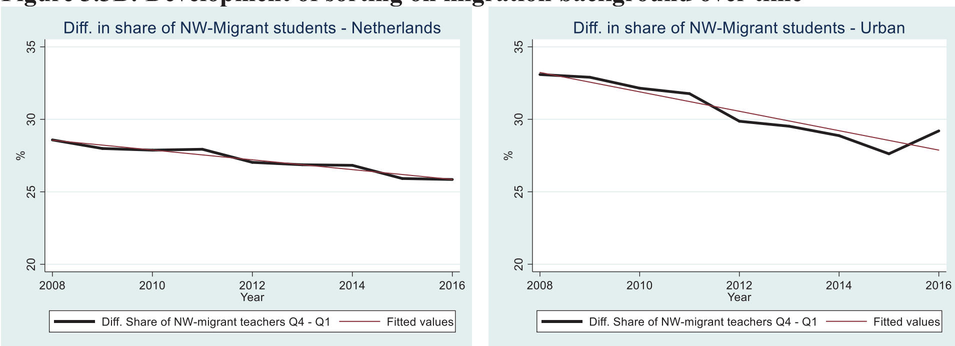
The results show that segregation along educational lines is increasing slightly over time, with the difference in the share of students with highly educated parents increasing from 5% in 2008 to 7% in 2016 between schools in the top and bottom quartiles of teachers holding a master's degree. In contrast, segregation on migration background has decreased from a 28% difference in the share of non-western migrant students in 2008 to a difference of 25.5% in 2016 between the top and bottom quartile of the share of teachers from a non-western migration background. Again, both the increasing educational sorting and decreasing migration background sorting trends are more pronounced in urban areas. These trends in teacher sorting over time mirror the trends in student segregation, where segregation along migration background is decreasing while sorting along educational lines is increasing (Inspectorate of Education, 2018; Boterman, 2018).

Figure 3.3A: Development of educational sorting over time



Source: DUO 1cijferPO, DUO Functiemix, and Statistics Netherlands highest achieved level of education databases.

Figure 3.3B: Development of sorting on migration background over time



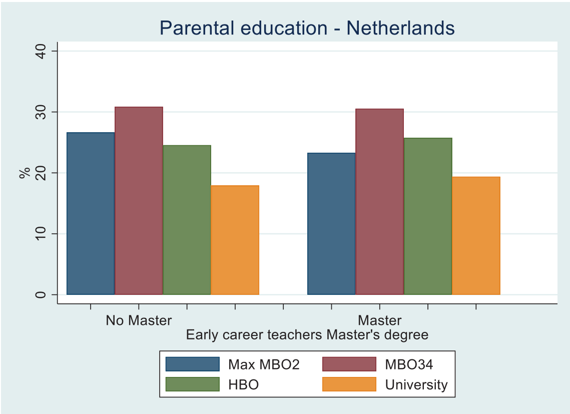
Source: DUO 1cijferPO, DUO Functiemix, Statistics Netherlands highest achieved level of education, and Statistics Netherlands municipal administration databases.

3.3.2 Early career teacher sorting

The previous section showed that there is strong assortative matching between student and teacher characteristics. However, it is unclear whether a particular student composition attracts a particular teacher population, or a certain teacher composition attracts certain students. In this section, I investigate the sorting patterns of early career teachers, relating their characteristics to the characteristics of the student population of the first school they start working at after graduation. Since parents cannot anticipate the characteristics of teachers that have yet to be hired, the sorting pattern of early career teachers is more likely to reflect teacher preferences for a certain student population, or a school's preference for a certain type of teacher, than parental preferences for a certain teaching force.

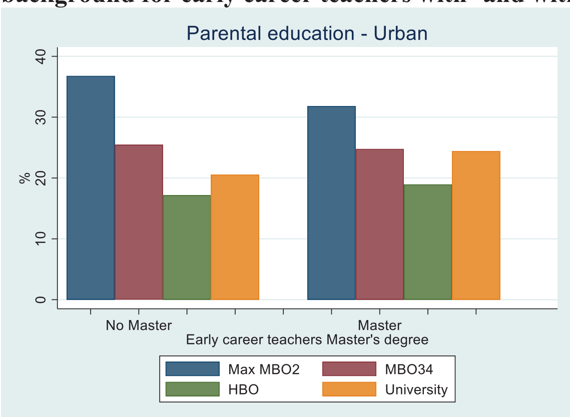
Figures 3.4A and 3.4B show the share of students from a certain parental educational background for early career teachers with and without a master's degree for the Netherlands as a whole and urbanized areas respectively. The share of early career teachers holding a master's degree is relatively low, as many teachers obtain their master's degree over the course of their career. The results show that there is not much evidence of assortative matching for the whole of the Netherlands. However, in strongly urbanized areas, early career teachers holding a master's degree start working at schools serving 4 percentage points more students with university-educated parents on average.

Figure 3.4A: Average percentage of students from a certain parental educational background for early career teachers with- and without a master's degree - Netherlands



Source: DUO 1cijferPO, DUO 1cijferHO, DUO Functiemix, and Statistics Netherlands highest achieved level of education databases.

Figure 3.4B: Average percentage of students from a certain parental educational background for early career teachers with- and without a master's degree - Urban

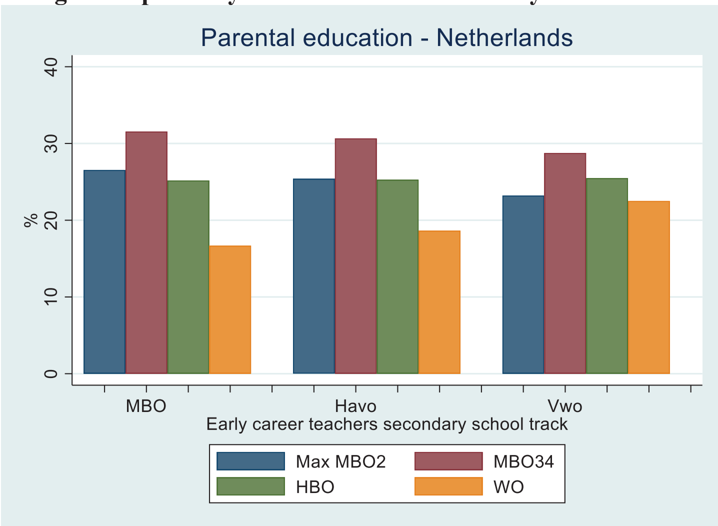


Source: DUO 1cijferPO, DUO 1cijferHO, DUO Functiemix, and Statistics Netherlands highest achieved level of education databases.

For early career teachers, additional information is available on the secondary school track they attended prior to entering teacher training. Since higher secondary school tracks are associated with higher cognitive abilities of students, teacher sorting on this characteristic is analogous to the school-level sorting on the share of teachers holding a master's degree. Figures 3.5A and 3.5B show the share of students from a certain parental educational background against the early career teacher's secondary school track before starting teacher training (for the Netherlands as a whole and urbanized areas respectively). The results show that even though the large majority of early career teachers start with the same level of teaching qualification, there is sorting along teachers' education prior to teacher training. Teachers that graduated from the highest track of secondary school start their careers at schools with a larger percentage of students with highly educated parents. Again, these results are more pronounced in urbanized areas, where teachers from the highest secondary school track start working at schools with a substantially lower percentage of students from low educated households.

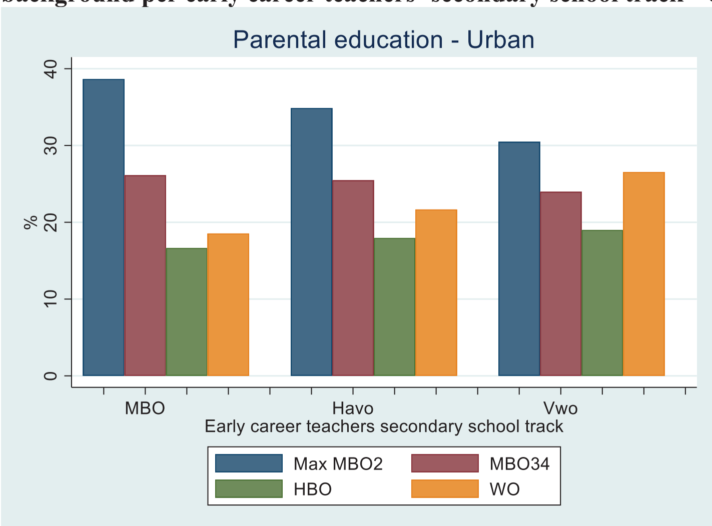
Figures 3.6A and 3.6B show the sorting of early career teachers on migration background for the whole of the Netherlands and urbanized areas respectively. The results show that sorting on migration background is more pronounced in early career teachers than it is on average. Teachers without a migration background work in schools where on average around 75% of the student population does not have a migrant background, while teachers from a non-western migration background start working at schools where 60% of the student population has a non-western migration background. In urbanized areas, the amount of students from a non-migration background is smaller in total, but the same sorting pattern is apparent.

Figure 3.5A: Average percentage of students from a certain parental educational background per early career teachers' secondary school track - Netherlands



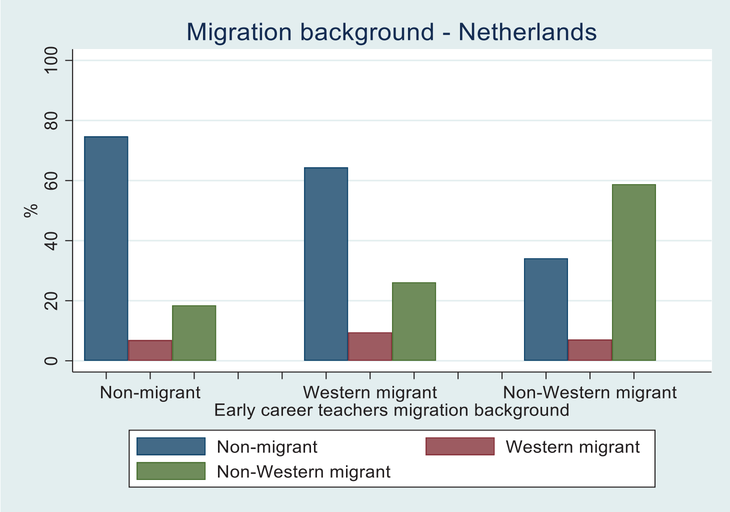
Source: DUO 1cijferPO, DUO 1cijferHO, DUO Functiemix, and Statistics Netherlands highest achieved level of education databases.

Figure 3.5B: Average percentage of students from a certain parental educational background per early career teachers' secondary school track - Urban



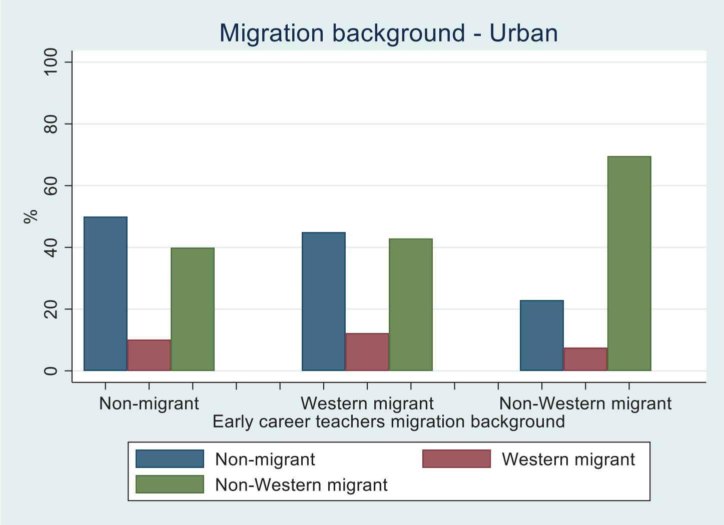
Source: DUO 1cijferPO, DUO 1cijferHO, DUO Functiemix, and Statistics Netherlands highest achieved level of education databases.

Figure 3.6A: Average percentage of students from a certain migration background per early career teachers' migration background - Netherlands



Source: DUO 1cijferPO, DUO 1cijferHO, DUO Functiemix, and Statistics Netherlands municipal administration databases.

Figure 3.6B: Average percentage of students from a certain migration background per early career teachers' migration background - Urban

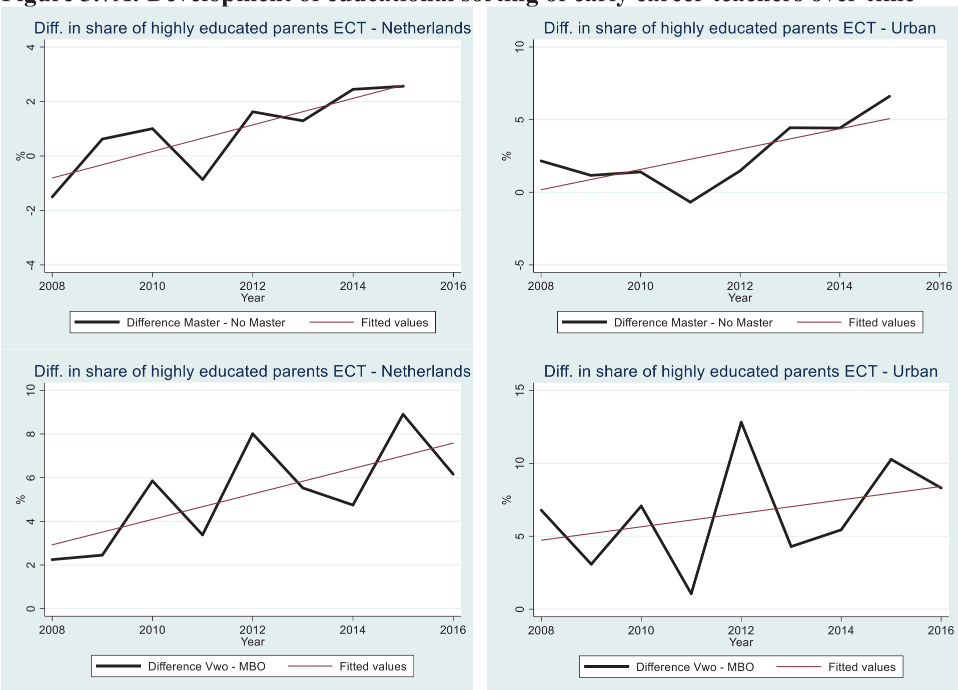


Source: DUO 1cijferPO, DUO 1cijferHO, DUO Functiemix, and Statistics Netherlands municipal administration databases.

Finally, figures 3.7A and 3.7B show the development of early career teacher sorting over time. I plot the difference in the average percentage of students with highly educated parents between schools between teachers with and without a master's degree, and between teachers coming from the lowest and the highest secondary school track for each year. For migration background, I take the difference in the percentage of students with a non-western migration background between teachers with a non-western migration background and teachers without a migration background.

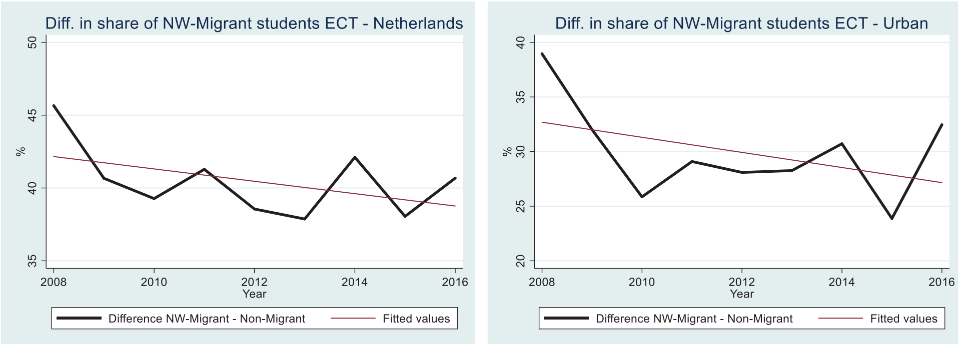
Again, figure 3.7A shows the results for educational sorting, while figure 3.7B looks at migration background. Over time, assortative matching of early career teachers is increasing in educational background, but decreasing in migration background. Together, these results show that the sorting pattern of early career teachers reinforces assortative matching between student and teacher characteristics on average. Furthermore, they show that it is likely that teacher or school preferences, rather than parental preferences, are responsible for the patterns of positive assortative matching.

Figure 3.7A: Development of educational sorting of early career teachers over time



Source: DUO 1cijferPO, DUO 1cijferHO, DUO Functiemix, and Statistics Netherlands highest achieved level of education databases.

Figure 3.7B: Development of sorting on migration background of early career teachers over time



Source: DUO 1cijferPO, DUO 1cijferHO, DUO Functiemix, and Statistics Netherlands municipal administration databases.

3.3.3 Teacher characteristics and student achievement

While the preceding section showed that there is strong positive assortative matching of teacher and student characteristics, the impact of this unequal distribution of teachers across schools on student learning outcomes is unclear. If the characteristics on which teachers are sorted are strongly related to teaching quality, these patterns could reinforce educational inequalities. On the other hand, if student learning is enhanced by being taught by a teacher that shares her background characteristics, it might be optimal to match teachers and students to each other based on exactly these attributes.

There is some prior evidence supporting both of these arguments. Dee (2004, 2005) finds that a match along ethnic lines between teacher and students has a positive impact on student achievement for both black and white students in the United States, particularly for students of low socioeconomic status. For higher education, Fairlie et al. (2014) show strong positive ethnic match effects on the probability of dropping out and student GPA. In this case, positive assortative matching along ethnic lines could increase educational effectiveness.

Conversely, while holding a master's degree does not seem to be related to teacher quality in itself (Harris & Sass, 2011; Coenen et al., 2018), there is some evidence that teachers with high cognitive ability achieve better outcomes for their students (Metzler & Woessmann, 2012; Hanushek, Piopiunik, & Wiederhold, 2018). If education beyond the initial teaching qualification is correlated with cognitive ability, teachers holding such additional master's degrees are expected to be of slightly higher cognitive ability, and therefore quality, on average. In this case, positive assortative matching on educational background would increase educational inequalities. It is therefore an empirical question whether these sorting patterns can be held partly responsible for differences in student performance.

In order to investigate the relationship between teacher characteristics and student achievement, I run OLS and school fixed effects regressions on student performance at the end of primary school exam in grade 6. The main explanatory variables of interest are the primary school-average teacher characteristics. In order to investigate matching effects, I interact the educational and migrant background of teachers with the parental education and migrant background of students.

(1) $$Y_{icst} = X_i + \overline{T_{st}} + (Mig_i * \overline{Tmig_{st}})\beta_1 + (Educ_i * \overline{Teduc_{st}})\beta_2 + \overline{C_{sct}} + \overline{S_{st}} + M_s + \gamma_{st} + \delta_t + \varepsilon_{icst}$$

Equation 1 shows the standard OLS specification. Cito-score Y of student i in classroom c in school s in year t is predicted by a vector of individual student characteristics X, a vector of school average teacher characteristics $\overline{T}$, and two interaction terms: one interaction between a student's migrant background and the school level share of teachers from a certain migration background ($Mig * \overline{Tmig}$), and one between a student's parental education background and the share of teachers holding a master's degree at the school level ($Educ * \overline{Teduc}$). Additionally, I control for classroom-average peer characteristics $\overline{C}$, school-average student characteristics $\overline{S}$, municipality dummies M, observable school characteristics γ, and year dummies δ.

A potential concern with the OLS specification above is that time-invariant unobservable characteristics at the school level that relate both to increased performance of a certain subset of students and to a propensity to employ a certain type of teacher, could bias the results. For example, some schools may focus on offering enrichment programs for their students with high ability, which may simultaneously attract teachers with a master's degree interested in teaching these particular programs. Any association between teachers holding a master's degree and the achievement of high ability students would then by confounded by the availability of the enrichment program.

More generally, any observed interaction between individual student characteristics and teacher characteristics at the school level could potentially be biased due to a time invariant omitted variable at the school level. Therefore, the second specification includes school fixed effects:

$$(2) \qquad Y_{icst} = X_i + \overline{T_{st}} + (Mig_i * \overline{Tmig_{st}})\beta_1 + (Educ_i * \overline{Teduc_{st}})\beta_2 + \overline{C_{sct}} + \overline{S_{st}} + + \gamma_{st} + \varphi_s + \delta_t + \varepsilon_{icst}$$

where φ represents the effect of all school-level time invariant characteristics on students' test scores.

Note that a general limitation of this dataset is that I cannot link individual students to their individual teachers. The standard OLS regression coefficients should therefore be interpreted as the impact of exposure to a certain combination of teachers throughout primary education, while the school fixed effects regressions relate to a change in the composition of teacher characteristics at the school level. An increase in, for example, the share of teachers from a non-migrant background increases the probability that a student is taught by one, but does not make it certain. It is therefore likely that both specifications underestimate the true association between teacher characteristics and student achievement.

Table 3.4 shows the results of the OLS regressions explaining grade six scores on the Cito test. Column 1 includes student and school characteristics, as well as year and municipality dummies. Standard errors are clustered at the school level in all specifications. Column 2 adds teacher characteristics, and column 3 adds the interactions between average teacher characteristics and individual student characteristics. The results show that, while there is no overall relationship between the share of teachers from a certain migration background and test scores, there is a significant positive interaction between the share of non-western migrant teachers and non-western migrant status of the students on student achievement, particularly for first generation migrants. However, the share of non-western migrant teachers relates negatively to student achievement for native students. These results are in line with the results of Dee (2004, 2005), who finds positive match effects along ethnic lines for both majority and minority students. In contrast, the interaction between the share of teachers holding additional qualifications and the parental educational background of students is not significant, nor is there an overall relationship between the share of teachers holding a master's degree and student achievement. These results corroborate earlier literature on the lack of association between additional teacher certification and student outcomes (e.g. Harris & Sass, 2011; Hanushek & Rivkin, 2006; Coenen et al., 2018).

Table 3.4: OLS regressions of Cito-scores on student and teacher characteristics

VARIABLES		(1) Standardized Cito	(2) Standardized Cito	(3) Standardized Cito
Gender (1=F)		-0.028**	-0.028**	-0.028**
Age		-0.271**	-0.271**	-0.271**
Relative age				
	Early	-0.095**	-0.095**	-0.095**
	Late	0.153**	0.153**	0.153**
Migration status:				
	Western migrant	0.031**	0.031***	0.032**
	2nd gen NW-migrant	-0.117**	-0.117***	-0.117**
	1st gen NW-migrant	-0.137**	-0.138***	-0.149**
Parental education:				
	Max MBO2	-0.298**	-0.298**	-0.302**
	HBO	0.310**	0.310**	0.305**
	University	0.575**	0.575**	0.570**
	Unknown	0.004	0.004	-0.005
Mean teacher salary			0.003	0.003
Mean monthly hours worked			0.005*	0.005
Tenured teachers pct.			0.004	0.004
Female teachers pct.			-0.011**	-0.011**
Mean teacher age			0.007*	0.007*
NW-migrant teachers pct.			-0.004	-0.018**
W-migrant teachers pct.			0.000	-0.000
Teachers with master's degree pct.			0.000	-0.002
NW-migrant teachers pct. * Western migrant				0.019*
NW-migrant teachers pct. * 2nd gen NW-migrant				0.017**
NW-migrant teachers pct. * 1st gen NW-migrant				0.030**
W-migrant teachers pct. * Western migrant				0.006
W-migrant teachers pct. * 2nd gen NW-migrant				-0.001
W-migrant teachers pct. * 1st gen NW-migrant				0.008
Teachers with master's degree pct. * Max MBO2				0.004
Teachers with master's degree pct. * HBO				0.003
Teachers with master's degree pct. * University				-0.002
Teachers with master's degree pct. * Unknown				0.006*
Classroom peer characteristics		X	X	X
School peer characteristics		X	X	X
School characteristics		X	X	X
Municipality dummies		X	X	X
Year dummies		X	X	X
Constant		2.146**	4.641**	5.307**
Observations		996.238	994.000	994.000
R-squared		0.174	0.1744	0.1744

Note: school-level teacher characteristics are standardized. The baseline category for relative age is "average". The baseline category for migration status is "non-migrant". The baseline category for parental education is "MBO34".Classroom peer characteristics include share of peers from a certain migration background, share of peers with a certain educational background, share of boys in class, class size, and peer average age. School peer characteristics include school-level share of children from a certain migration background, share of children with a certain educational background, and share of boys. School characteristics include school size, religious denomination, educational philosophy, and school board size. Standard errors are clustered at the school level, and omitted for brevity; the full regression output is available upon request.

** p<0.05; ** p<0.01*

In terms of student characteristics, students' parental education level is the strongest predictor of test scores. Migrant status is related to lower test scores, and there appears to be a relative age effect in that those students that are relatively older within their cohort have higher test scores. Age is negatively related to test scores, as a higher age at the time of test taking is associated with student retention. Finally, girls are slightly outperformed by boys. Regarding teacher characteristics at the school level, a larger share of male teachers is related to higher performance. The average age of teachers is weakly related to student achievement. This is somewhat surprising since age, as a proxy for teacher experience, has been shown to be strongly related to student achievement in many prior studies (e.g. Harris & Sass, 2011; Papay & Kraft, 2015). The result could be explained by the fact that the average age of teachers at the school level is relatively high at 44. While teachers continue to improve across their career (Papay & Kraft, 2015), most of the impact of experience on teacher quality occurs in the first five years of teaching (Hanushek & Rivkin, 2006; Harris & Sass, 2011).

Table 3.5 shows the results of school fixed effects regressions on student performance at the end of primary school. The negative association between the share of non-western migrant teachers and the performance of native students disappears, while the positive interaction effect for non-western migrants stays significant. Again, no significant interaction is found between the share of teachers holding an advanced degree and students' parental educational background. Furthermore, the negative relationship of the share of female teachers as well as the small positive relationship of the average age of teachers at the school level to student test scores is rendered insignificant in the school fixed effects specification.

In terms of the size of the relationship, a one standard deviation increase in the share of non-western migrant teachers is related to a .01 (.03) standard deviation higher Cito score for second (first) generation non-western migrants. This seems like a relatively small association. However, it has to be interpreted in light of the size of teacher effects in general. For example, Papay & Kraft (2015) find that the difference between a novice teacher and one with 5 years of experience is around .08 standard deviations. Considering that the positive return to early career experience is one of the largest effects in the teacher effectiveness literature (Coenen et al., 2018), an association of .03 is not unsubstantial. Furthermore, other studies investigating interaction effects between teacher and student ethnicity find relationships of around .05 standard deviations (Clotfelter, Ladd, & Vigdor, 2010). These prior studies have the benefit of being able to match students to their individual teachers, whereas the results of this chapter are likely underestimated because of uncertainty in the extent to which students were exposed to teachers sharing their migration background.

Table 3.5: School fixed effects regressions of Cito-scores on student and teacher characteristics

VARIABLES		(1) Standardized Cito	(2) Standardized Cito	(3) Standardized Cito
Gender (1=F)		-0.0282**	-0.028**	-0.028**
Age		-0.271**	-0.271**	-0.271**
Relative age				
	Early	-0.096**	-0.096**	-0.0956**
	Late	0.155**	0.155**	0.155**
Migration status:				
	Western migrant	0.0313**	0.031**	0.032**
	2nd gen NW-migrant	-0.117**	-0.117**	-0.115**
	1st gen NW-migrant	-0.132**	-0.132**	-0.146**
Parental education:				
	Max MBO2	-0.298**	-0.298**	-0.302**
	HBO	0.310**	0.310**	0.305**
	University	0. 574**	0.574**	0.568**
	Unknown	0.004	0.004	-0.005
Mean teacher salary			0.000	0.000
Mean monthly hours worked			0.001	0.000
Tenured teachers pct.			0.003	0.0036
Female teachers pct.			-0.000	-0.000
Mean teacher age			-0.001	-0.001
NW-migrant teachers pct.			-0.001	-0.01
W-migrant teachers pct.			-0.003	-0.003
Teachers with master's degree pct.			0.001	-0.001
NW-migrant teachers pct. * Western migrant				0.022**
NW-migrant teachers pct. * 2nd gen NW-migrant				0.012*
NW-migrant teachers pct. * 1st gen NW-migrant				0.030**
W-migrant teachers pct. * Western migrant				0.004
W-migrant teachers pct. * 2nd gen NW-migrant				-0.003
W-migrant teachers pct. * 1st gen NW-migrant				0.011
Teachers with master's degree pct. * Max MBO2				0.002
Teachers with master's degree pct. * HBO				0.005
Teachers with master's degree pct. * University				-0.003
Teachers with master's degree pct. * Unknown				0.0073*
Classroom peer characteristics		X	X	X
School peer characteristics		X	X	X
Year dummies		X	X	X
School fixed effects		X	X	X
Constant		3.299**	3.300**	3.308**
Observations		998.182	995.919	995.919
R-squared		0.164	0.1642	0.1642

Note: school-level teacher characteristics are standardized. The baseline category for relative age is "average". The baseline category for migration status is "non-migrant". The baseline category for parental education is "MBO34". Classroom peer characteristics include share of peers from a certain migration background, share of peers with a certain educational background, share of boys in class, class size, and peer average age. School peer characteristics include school-level share of children from a certain migration background, share of children with a certain educational background, and share of boys Standard errors are clustered at the school level, and omitted for brevity; the full regression output is available upon request.
** p<0.05; ** p<0.01*

3.3.4 Robustness and heterogeneity

A concern with the school fixed effects specification is that a change in the share of teachers from a certain background necessarily coincides with some form of teacher turnover. An increase in the share of migrant teachers implies that either a migrant teacher was hired, or a non-migrant teacher left the school. Since teacher turnover negatively affects student outcomes in itself (Ronfeldt, Loeb, & Wyckoff, 2013), this effect could confound the associations between the changes in teacher characteristics at the school level and test scores. Table 3.6 shows the results of the school fixed effects regressions adding two different measures of teacher turnover to the full model.[28] While teacher turnover does appear to negatively influence student test scores, neither measure of teacher turnover reduces the positive interactions on teacher-student migrant status match.

Because there is information about student performance on the math and language subscales of the Cito-test, it is interesting to see whether the teacher-student match effects are subject specific. Table 3.7 shows school fixed effects regressions on achievement in language (column 1) and math (column 2) separately. The results show that the positive interaction between the share of non-western migrant teachers and students' non-western migration background is only significant for the math part of the test. These results contrast with the findings of Dee (2004), which show gains in both the math and reading domains. A plausible explanation for these discrepant results is that non-western migrant teachers have higher math skills relative to their Dutch language skills than their native colleagues as they have usually been brought up bilingual. In America, however, both white and black teachers have most likely been raised to speak English. Unfortunately, since there are no data available on teacher subject knowledge this interpretation cannot be empirically validated.

[28] Teacher turnover is operationalized as the share of new teachers in a certain school in a certain year (column 1), and the share of teachers that did not return to a certain school after the previous year (column 2).

Table 3.6: School fixed effects regressions of Cito-scores on student and teacher characteristics, and teacher turnover

VARIABLES		(1) Standardized Cito	(2) Standardized Cito
Share of new teachers within a school		-0.0008**	.
Share of teachers that did not return		.	-0.0003
Migration status:			
	Western migrant	0.034**	0.034**
	2nd gen NW-migrant	-0.112**	-0.112**
	1st gen NW-migrant	-0.144**	-0.143**
Parental education:			
	Max MBO2	-0.301**	-0.301**
	HBO	0.306**	0.307**
	University	0.571**	0.571**
	Unknown	-0.005	-0.005
NW-migrant teachers pct.		-0.008	-0.008
W-migrant teachers pct.		-0.003	-0.003
Teachers with master's degree pct.		-0.001	-0.001
NW-migrant teachers pct. * Western migrant		0.023**	0.023**
NW-migrant teachers pct. * 2nd gen NW-migrant		0.013*	0.013*
NW-migrant teachers pct. * 1st gen NW-migrant		0.033**	0.032**
W-migrant teachers pct. * Western migrant		0.004	0.004
W-migrant teachers pct. * 2nd gen NW-migrant		-0.002	-0.002
W-migrant teachers pct. * 1st gen NW-migrant		0.013	0.013
Teachers with master's degree pct. * Max MBO2		0.002	0.002
Teachers with master's degree pct. * HBO		0.005	0.005
Teachers with master's degree pct. * University		-0.003	-0.003
Teachers with master's degree pct. * Unknown		0.007*	0.007*
Student characteristics		X	X
Classroom peer characteristics		X	X
Teacher characteristics		X	X
School peer characteristics		X	X
Year dummies		X	X
School fixed effects		X	X
Constant		3.298**	3.301**
Observations		966.375	963,585
R-squared		0.1644	0.1642

Note: school-level teacher turnover variables are in percentages (scale 0-100). School-level teacher characteristics are standardized. The baseline category for migration status is "non-migrant". The baseline category for parental education is "MBO34". Student characteristics include student gender, and absolute and relative age. Classroom peer characteristics include share of peers from a certain migration background, share of peers with a certain educational background, share of boys in class, class size, and peer average age. Teacher characteristics include the school-level average hourly wage, hours worked, and age, and the percentage of female, and tenured teachers. School peer characteristics include school-level share of children from a certain migration background, share of children with a certain educational background, and share of boys. Standard errors are clustered at the school level, and omitted for brevity; the full regression output is available upon request.
** $p<0.05$; ** $p<0.01$*

Table 3.7: School fixed effects regressions of Cito language and math subscale scores on student and teacher characteristics

VARIABLES	(1) Standardized Cito - Language	(2) Standardized Cito - Math
Migration status:		
Western migrant	0.012*	0.044**
2nd gen NW-migrant	-0.139**	-0.045**
1st gen NW-migrant	-0.162**	-0.046**
Parental education:		
Max MBO2	-0.265**	-0.209**
HBO	0.240**	0.236**
University	0.505**	0.482**
Unknown	-0.013**	0.027**
NW-migrant teachers pct.	0.011	-0.010
W-migrant teachers pct.	0.000	-0.001
Teachers with master's degree pct.	0.001	-0.000
NW-migrant teachers pct. * Western migrant	0.001	0.030**
NW-migrant teachers pct. * 2nd gen NW-migrant	-0.010	0.025**
NW-migrant teachers pct. * 1st gen NW-migrant	0.004	0.025**
W-migrant teachers pct. * Western migrant	0.001	0.010*
W-migrant teachers pct. * 2nd gen NW-migrant	-0.006	-0.001
W-migrant teachers pct. * 1st gen NW-migrant	0.0024	0.007
Teachers with master's degree pct. * Max MBO2	0.007	0.005
Teachers with master's degree pct. * HBO	0.000	0.001
Teachers with master's degree pct. * University	-0.006	-0.005
Teachers with master's degree pct. * Unknown	0.008	0.009**
Student characteristics	X	X
Classroom peer characteristics	X	X
Teacher characteristics	X	X
School peer characteristics	X	X
Year dummies	X	X
School fixed effects	X	X
Constant	2.919**	3.001**
Observations	816.875	816.875
R-squared	0.119	0.097

Note: school-level teacher characteristics are standardized. The baseline category for migration status is "non-migrant". The baseline category for parental education is "MBO34". Student characteristics include student gender, and absolute and relative age. Classroom peer characteristics include share of peers from a certain migration background, share of peers with a certain educational background, share of boys in class, class size, and peer average age. School peer characteristics include school-level share of children from a certain migration background, share of children with a certain educational background, and share of boys Standard errors are clustered at the school level, and omitted for brevity; the full regression output is available upon request.
$* p<0.05$; $** p<0.01$

Finally, Dee (2004) shows that an ethnic student-teacher match is mostly beneficial for students of low socioeconomic status. Starting from the year 2014/2015, Statistics Netherlands calculates a predicted Cito-score based on observable student characteristics for each student which is highly contingent on students' socioeconomic background.[29] To see whether a student-teacher match on migration background is associated with higher test scores specifically for students from a low socioeconomic background, I run the school fixed effects regressions for students with a below average and an above average predicted Cito-score separately. Table 3.8 shows the results. Column 1 (2) shows the results for the subsample of students with a below (above) average predicted Cito-score. The table shows that a student-teacher match on migration background is related to higher student achievement only for the subset of students with a below average predicted Cito-score.

3.4 Conclusions

School segregation across migration background and socio-economic lines is a rising cause of concern for policy makers because of its potential to exacerbate inequality of educational opportunities. One channel through which school segregation could lead to increased inequality is through an unequal distribution of teaching resources. In this chapter, I investigate the extent of teacher sorting using Dutch registry data to find evidence of strong positive assortative matching of students and teachers across migration background and educational lines. Schools serving students with highly educated parents employ a larger share of teachers holding master's degrees. The same holds for migration background: schools with a higher percentage of non-western migrant students employ a larger percentage of teachers with a non-western migration background. The trends over time mirror those in student segregation in Dutch primary education (Inspectorate of Education, 2018; Boterman, 2018): sorting across educational lines is slightly increasing, while sorting on migration background is slightly decreasing. Positive assortative matching is especially pronounced in urbanized areas, and the sorting patterns of early career teachers magnify, rather than mitigate, the extent of teacher sorting.

[29] The characteristics used by Statistics Netherlands to predict Cito-scores are maternal and paternal education level, parental countries of origin, maternal years of residence in the Netherlands, parental gross yearly income, and an indicator for whether a student's parents are currently part of an outstanding debt refinancing program ("schuldsanering") (CBS, 2017)

Table 3.8: School fixed effects regressions of Cito-scores on student and teacher characteristics seperately for students with a below- and above average predicted Cito-score

VARIABLES		(1) Standardized Cito	(2) Standardized Cito
Migration status:			
	Western migrant	0.096**	0.039**
	2nd gen NW-migrant	0.012	-0.046**
	1st gen NW-migrant	0.089**	-0.106**
Parental education:			
	Max MBO2	-0.220**	-0.396**
	HBO	0.179**	0.155**
	University	0.341**	0.401**
	Unknown	0.005	-0.098**
NW-migrant teachers pct.		-0.018	-0.008
W-migrant teachers pct.		-0.007	-0.001
Teachers with master's degree pct.		-0.013	0.018
NW-migrant teachers pct. * Western migrant		0.050**	0.016
NW-migrant teachers pct. * 2nd gen NW-migrant		0.028**	-0.014
NW-migrant teachers pct. * 1st gen NW-migrant		0.047**	0.055
W-migrant teachers pct. * Western migrant		0.012	0.015
W-migrant teachers pct. * 2nd gen NW-migrant		0.013	-0.008
W-migrant teachers pct. * 1st gen NW-migrant		0.017	0.018
Teachers with master's degree pct. * Max MBO2		0.004	0.011
Teachers with master's degree pct. * HBO		0.010	-0.036
Teachers with master's degree pct. * University		0.016	-0.007
Teachers with master's degree pct. * Unknown		-0.013	-0.013
Student characteristics		X	X
Classroom peer characteristics		X	X
Teacher characteristics		X	X
School peer characteristics		X	X
Year dummies		X	X
School fixed effects		X	X
Constant		3.945**	3.481**
Observations		185.068	184.072
R-squared		0.064	0.088

Note: school-level teacher characteristics are standardized. The baseline category for migration status is "non-migrant". The baseline category for parental education is "MBO34". Student characteristics include student gender, and absolute and relative age. Classroom peer characteristics include share of peers from a certain migration background, share of peers with a certain educational background, share of boys in class, class size, and peer average age. Teacher characteristics include the school-level average hourly wage, hours worked, and age, and the percentage of female, and tenured teachers. School peer characteristics include school-level share of children from a certain migration background, share of children with a certain educational background, and share of boys. Standard errors are clustered at the school level, and omitted for brevity; the full regression output is available upon request.
** p<0.05; ** p<0.01*

In terms of the impact of teacher sorting on student achievement, analyses on the link between teacher characteristics and student outcomes reveal no association between test scores and the share of teachers from a migrant background or the share of teachers holding an advanced degree in general. However, a positive interaction between the share of teachers from a non-western migrant background and the share of students with the same background on student test scores is found, providing suggestive evidence of a positive match effect along ethnic lines as previously found by Dee (2004, 2005). The results are more pronounced for students from a low socio-economic background, and are driven by increased performance in mathematics, but not in language. No match effect along educational lines is found. However, the results on student achievement have to be interpreted with care as individual students cannot be matched to individual teachers, and the analyses are non-causal in nature.

In conclusion, considering the positive interaction between the share of non-western migrant teachers and non-western migrant student outcomes, positive assortative matching on migration background may not be negatively related to student achievement and equality of educational opportunities. However, this does not mean that policy makers should design interventions aimed at stimulating student-teacher matches on migration background, as educational effectiveness is only one of several functions of the school system. Potential gains in student achievement from increased assortative matching on teacher and student characteristics may come at the cost of reducing socialization and citizenship skills outcomes, which could be benefited from exposure to teachers from different backgrounds. Therefore, while the extent of positive assortative matching between student and teacher characteristics is sizable, the consequences of this sorting pattern for educational inequalities of opportunities are not yet fully clear.

4. Teacher professional development as a signal: Inefficient sorting and its implications for optimal training provision

4.1 Introduction

The quality of teachers is an essential determinant of educational outcomes in society. According to Hanushek (2011), no other school attribute even comes close to its influence on student achievement. Therefore, providing on-the-job training opportunities for teachers is a key instrument for improving the quality of education (Caena, 2013; OECD, 2014), and policy makers devote many resources to stimulate teachers to enter professional development programs. A recent study by Jacob & McGovern (2015) estimate that around 6 to 9% of school districts' operating budget is being used for teacher professional development in the United States. In the Netherlands, around 10% of the total budget allocated to schools is reserved for this purpose (Ministry of Education, Culture and Science, 2014). Teachers also appear to make use of the professional development opportunities they are given. Results from the OECD's TALIS study show that, across the developed countries, 88% of teachers have participated in at least one professional development program in the past year (OECD, 2014). For the United States, Wei, Darling-Hammond, & Adamson (2010) report a very similar participation rate (87.5% for subject content training) based on the 2008 Schools and Staffing Teacher Survey. However, less is known about how teachers select into available professional development programs. Without an understanding of teacher incentives, it is unclear whether training will be taken by those who would benefit the most from attending the program in terms of ability. This leads to difficulties in evaluating the true added value of the training program, and in designing the most efficient implementation strategy.

In this chapter, we model the decision of teachers to enter on-the-job training.[30] We argue that because teacher performance is imperfectly observable, teachers will incorporate the signalling value of signing up for specific types of training into their decision making process. We show that when training participation is voluntary, sorting into training is inefficient. Programs aimed at improving teachers at the low end of the ability distribution will be underutilized, while programs that aim to improve the top end of the distribution will be overused. Furthermore, we show that in a situation where teachers can choose between a basic and an advanced training program, increasing the attractiveness of the advanced program increases the participation rate for the basic program. The intuition behind this result is that improvements in an advanced training program stimulate more teachers to sign up for

[30] This is a joint work with Olga Meshcheriakova. We thank Timothy Bond, Lex Borghans, Tamás Dávid-Barrett, Mirko Draca, Bart Golsteyn, Susanna Loeb, Arjan Non, Trudie Schils, and seminar participants at Maastricht University, Universität Tübingen, Kiel Institute for the World Economy, as well as participants at EALE 2016, LESE 2017, the Oslo Workshop on Education, Skills, and Labor Market Outcomes 2017, and the AEA ASSA meeting 2018 for their valuable comments.

advanced training. This, in turn, decreases the negative signal associated with signing up for basic training, as the average ability of those who do not take any training decreases. Following the same logic, increasing the attractiveness of the basic program reduces participation in advanced programs. Thus, the availability of different types of training attenuates the signal of each particular training program.

These results have several implications. First, by investing in programs targeting the top of the ability distribution, policy makers can affect the entire pool of teachers and improve average ability overall. In other words, simply providing the option of an advanced training course can be an effective instrument to induce low ability teachers to self-select into training more efficiently. Secondly, evaluations of the effectiveness of single programs should take the potential for spillover effects into account. Even if an advanced course is ineffective on average, its mere existence increases participation in basic courses. Assuming the basic course is effective, this would lead to an increase in average teacher quality. Finally, when designing training programs, policy makers should take the signal the course will give into consideration. Our model predicts that as a course becomes more basic, fewer teachers will sign up for this course unless it is unrealistically effective.

Our model leads to several empirically testable predictions. First, teachers that gain more from sending a positive signal will sort into training based less on gains in objective ability than those for who the signal is of little value. This implies, for example, that teachers that are on a fixed term contract are predicted to sort into advanced training more than their colleagues on permanent contracts. Second, in schools that offer a higher variety of training programs, participation rates among teachers will be higher. Finally, schools that offer and stimulate more advanced training opportunities will have a higher average level of teacher quality, independent of the effectiveness of the advanced training programs.

This chapter contributes to the literature on teacher professional development. Studies on teacher professional development programs tend to focus on evaluating the effectiveness of particular programs (see e.g. Yoon et al. (2008) and Carrillo, Maassen van den Brink, & Groot (2016) for an overview). Not much attention has been given to analysing how teachers choose among the different training options available to them. We show that even when a program is effective in raising teacher quality, it may fail to attract any participants because of the potential negative information it signals about teachers' initial ability. Conversely, we show that programs aimed at high ability teachers improve the efficiency of teacher sorting even if these programs themselves are ineffective on average. Therefore, we argue that evaluations of the effectiveness of single programs should take the potential for these types of spillover effects into account.

More generally, this chapter also contributes to the on-the-job training literature. While there have been many studies on on-the-job training in general (e.g. Becker, 1962; Hashimoto, 1981; Acemoglu, 1997; Acemoglu & Pischke, 1998, 1999a, 1999b; Autor, 2001; Leuven, 2005), less consideration has been given to the decision to participate in training specifically for the public sector. As public sector employees, teachers face weak performance incentives and high levels of wage compression. Additionally, monitoring of performance is difficult in the teaching profession. Therefore, models that predict training participation for private sector employees do not completely capture the considerations of teachers.

The remainder of this chapter is organized as follows. Section 4.2 gives a brief overview of the relevant literature and motivation for our model. Section 4.3 introduces the teacher decision model. Section 4.4 discusses the empirical predictions and Section 4.5 concludes.

4.2 Motivation and Background Literature

In this section, we will discuss some explanations for why employees sign up for on-the-job training, and argue why they do not fully capture teachers' considerations. First, we look at the returns to on-the-job training in terms of wages. Secondly, we consider intrinsic motivation as a reason for training participation. Finally, we introduce career concerns and argue that these, combined with imperfectly observable performance, influence teachers' decision to participate in training because of the signal it confers to their employer.

In general, workers' decision to enter training programs can be thought of from a human capital point of view, in which they decide to enter when the immediate costs of training are outweighed by the net present value of future benefits. These benefits can be either internal (e.g. intrinsic motivation) or external (e.g. increased wages). Regarding external benefits, most empirical studies find that training in the workplace has a positive effect on employee wages (e.g. Frazis & Loewenstein, 2005; Bassanini et al., 2007; Hansson, 2008; Haelermans & Borghans, 2012; Fouarge, Schils, & de Grip, 2013), although some studies do find non-significant returns as well (e.g. Leuven & Oosterbeek, 2004; Leuven & Oosterbeek, 2008). For teachers the future external benefits of training are not that obvious, as in most countries they are paid a fixed salary that is independent of their performance (Lazear, 2003). Teachers themselves also recognize a lack of external incentives to participate in training as a problem (OECD, 2014). One reason that there is no direct monetary benefit of being an excellent teacher, compared to being a poor one, is likely because teacher performance is difficult to quantify. Many factors interact in the educational production function to generate student outcomes, making it hard for schools to reward their teachers based on their objective output.

Training participation in absence of external incentives can occur when performance is intrinsically valued by employees. The increase in performance is a future benefit on its own. Studies on job performance of public sector employees consider intrinsic motivation a main factor in explaining why they choose to exert effort in their job (Wilson, 1989; Dixit, 2002; Prendergast, 2007; Delfgaauw & Dur, 2008). When we think of training participation as a type of job effort that increases future performance, this argument could be an explanation for why teachers participate in on-the-job training. When teachers derive intrinsic value from improving their students' outcomes, and they believe entering professional development programs has a positive influence on their teaching quality, they will enter training even without external incentives.

While we agree that intrinsic motivation plays a role in the training participation decision, we argue that it cannot fully account for the empirical findings in the professional development program effectiveness literature. If intrinsic valuation of one's teaching ability is the main factor in explaining why teachers go into professional development programs, evaluations of these programs are biased towards finding a positive effect if self-selection is not controlled for. Especially considering that there is already a selection bias in which programs are likely to be evaluated. Pritchett (2002) shows that advocates of ineffective programs have an incentive not to subject their program to independent evaluations. Ineffective programs are less likely to be evaluated as a result. In the empirical literature, we therefore expect to find an overestimation of the true effectiveness of the average program.

However, most studies on the effect of specific professional development programs on teacher quality find very little impact on student performance (Yoon et al., 2008; Blank & De Las Alas, 2009; Jacob & McGovern, 2015). Evidence from US studies also point to zero effects of professional development programs on student achievement (Jacob & Lefgren, 2004; Garet et al. 2008; Garet et al. 2010; Garet et al. 2011). Conclusions drawn from these null findings could be that many professional development programs are poorly designed, or that the studies evaluating their effectiveness are flawed. Yoon et al. (2008), for example, argue that the majority of professional development evaluation studies have some methodological issues. There is also a large literature discussing the optimal design of professional development programs that conclude that many programs fall short of meeting best practice standards (see Wei et al. (2009) for an overview).

We argue that the disappointing returns to most professional development programs are partly caused by inefficient selection into training. This inefficient selection into training results from teachers' rational response to their working environment. While teachers' salaries may not depend on their performance, in most countries their salary does increase with tenure (OECD, 2014). Teachers want to stay employed and have career concerns (Gibbons & Murphy, 1992). Because performance is imperfectly observed, teachers have an incentive to strategically divulge information from which their employer will infer they are of high quality. Signing up for particularly advanced forms of professional development programs could be such a signal. Conversely, signing up for a course that improves very elementary skills sends a signal that the teacher does not even master the basics. Advanced courses will be oversubscribed, while courses that improve basic skills will be unpopular. Because program evaluations only look at the objective gains in teaching quality, those who sign up for the signal alone while gaining nothing from the program will bias the observed effect size downwards.

From the preceding, we can conclude that, while there are no direct monetary benefits from *being* a more productive teacher, there are career benefits from *being seen* as a highly productive teacher. This can explain teachers' high participation rate in training programs, in spite of the fact that most programs are evaluated as being ineffective. In the next section, we formalize our argument by developing a model on teachers' decisions to enter professional development programs.

4.3. Teacher training model

In this section, we describe our model. First, we introduce the single program model and state our assumptions. Second, we describe the situation where teachers can choose between different types of available training programs.

4.3.1 Single training program model

We assume that teaching ability (a) is a single united set of skills[31] necessary to improve students' performance, uniformly distributed on $[0,1]$.[32] The assumption of a continuous distribution is used in order to allow for heterogeneity in initial ability.[33] The training program improves all skills in the bundle homogeneously at the same rate. We assume that the total added value of training ($\alpha(a)$) is linear in ability and consists of the fixed benefit (α_0) and marginal returns to training (α_1). Those values are determined by the program characteristics:

(1) $\alpha(a) = \alpha_0 + \alpha_1 a.$

We assume that all the available training programs are advanced or basic by design, depending on which type of teacher benefits more. The main difference between these types of training is the relationship between returns to training and the participants' ability. The added value of basic training is decreasing with teaching ability ($\alpha_1 < 0$), so low ability teachers benefit more from this training. Conversely, for advanced training programs the added value is increasing with ability ($\alpha_1 > 0$), as teachers with higher ability can learn more from these programs. That is to say, added value of basic (advanced) training is higher for teachers with low (high) ability. We restrict the values of the parameters between -1 and 1 to have correspondence in magnitude with the ability variable.

By construction, fixed added value in (1) defines the gains of teachers at the bottom of the ability distribution, since it outweighs marginal added value for low values of a. Analogously, marginal added value determines the gains of teachers with high ability, and we assume that fixed and marginal added values are negatively correlated. An increase in marginal added value makes the training more beneficial for higher ability teachers, and makes it less useful for low ability teachers simultaneously. The same logic holds for an increase in fixed added value: by adjusting to serve the needs of teachers with low ability, the training becomes less useful for high ability teachers.

[31] Teaching ability can also be assumed to be multidimensional (i.e. include more than one skill). In this case, a teacher makes a decision about every specific skill independently, and the initial model can be applied for each dimension separately.
[32] To check for the sensitivity of our results to the choice of a uniform distribution, we simulated the model using several different distributions of initial ability and got qualitatively similar results.
[33] In this, we follow the evidence from previous work (e.g. Aaronson, Barrow & Sander 2007; Staiger & Rockoff, 2010) showing considerable heterogeneity in teacher value added.

Since teacher productivity is not perfectly observable, individual teaching ability and added value of training are private information. The information the market receives about an individual teacher is whether or not he went into training, and general information about the program he took. Based on the information about the program, market agents form beliefs about the level of training. For example, programs that lead to a master's degree will likely be inferred to be an advanced form of training, while a program labeled a 'subject knowledge refreshment course' is likely to be judged as basic. Using the available information, market agents make assumptions about the average level of ability of teachers who did and did not participate in training. Here we analyse the extreme case when training participation is the only available source of information regarding teaching ability.[34] Market agents (e.g. school principals) infer a teacher's ability from the type of training he takes. For example, a teacher who signs up for advanced training, is expected to be of higher initial ability than a basic program participant. Therefore, teachers take into account market beliefs when deciding to take the training.

The decision to participate in the training is a single period voluntary decision. Teachers are rational to the extent that they base their decision on a cost and benefit analysis, i.e. the teacher makes a decision (Y) to participate in the training if the utility of participating is higher than the utility of not participating ($U_p(a) \geq U_{np}(a)$).[35] This decision could be modelled binary:

$$Y = \begin{cases} 1, if\ U_p(a) \geq U_{np}(a) \\ 0, if\ U_p(a) < U_{np}(a) \end{cases}.$$

To distinguish between ability before and after training we refer to ability after the participation decision as knowledge:

$$(2) \qquad K(a) = \begin{cases} \alpha_0 + (1 + \alpha_1)a,\ if\ Y = 1 \\ a,\ if\ Y = 0 \end{cases}.$$

[34] In practice, school principals form beliefs about the productivity of their teachers over time (Rockoff, Staiger, Kane & Taylor, 2012). However, for the results of our model to hold, productivity needs only be partially unobserved, as we show in section 4.3.3.1.

[35] Without loss of generality, we assume that if the teacher is indifferent between taking and not taking the training ($U_p(a) = U_{np}(a)$), he will participate.

This means that for teachers who do not take the training, knowledge is equal to their initial ability.

The utility a teacher gains from the training includes net added value and the signal of the training:

(3) $$U_P(a) = \alpha(a) - c + EK(a^*)_p - K(a),$$

where a^* stands for a participation threshold, and is defined as the ability level for which the utility from training is equal to the utility from not taking the training.

The net added value of training is expressed as the difference between total added value and total costs of training ($\alpha(a) - c$), and represents the teacher's intrinsic valuation of the net gains from training.

The total costs of training (c) include monetary costs and time investment. Following Spence (1973), we can assume that costs decrease with ability, as the smarter individuals require less time to master the same material. However, the decreasing nature of the costs can be incorporated in (3) in the coefficient multiplying the ability, i.e. if $c = c_0 - c_1 a$:

$$U_P(a) = \alpha_0 + a(\alpha_1 + c_1) - c_0 + EK(a^*)_p - K(a).$$

Therefore, without loss of generality we can make a simplifying assumption that the costs are equal for all teachers ($c = c_0$).

We assume that all teachers value their teaching ability in the same way, so there are no differences in the weights they put on the net gains in the utility function.[36]

$EK(a^*)_p - K(a)$ is defined as the signal of the training program. This is the difference between expected knowledge of the group and the individual teacher's knowledge after training. As the best guess the market can make about a person's knowledge is the average knowledge of people (not) taking training, every teacher signing up for training is assumed to be $EK(a^*)_p$. Notice that the signal of the same program can be positive or negative for different teachers depending on their individual ability. Individuals with below average knowledge will have additional (signalling) benefits from the training, since they will be assumed to be more able than they actually are, while for teachers with above average knowledge it will be a penalty. Following the same logic, the decision not to participate in the

[36] Results do not change if valuation of teaching ability is assumed to be heterogeneous.

training can also be treated as a signal of certain knowledge, and these teachers gain (lose) the difference between the expected knowledge of the group not taking training and their individual level of knowledge:

(4) $\quad U_{np}(a) = EK(a^*)_{np} - a.$

Everyone for whom the benefits from training exceed the benefits from not participating will take the training in equilibrium ($U_p(a) \geq U_{np}(a)$). The teacher with marginal ability (a^*), for whom the net benefits from going and not going into training are equal, is indifferent.

Since the solutions differ for different types of training, we look separately at both situations. First we predict participation in the ideal situation when productivity is perfectly observable, and training is only taken by those who benefit from it in terms of ability, then show how the results change once productivity becomes unobservable and signals enter the decision making process.

4.3.1.1 Basic training

We first consider a situation where teacher productivity is perfectly observable. Training cannot be used for signalling and is only valued for the skills increase it provides. In this case only net added value is present in the utility function (3), while utility of not taking the training (4) is zero. Consequently, all teachers who derive positive utility from this training sign up. Using the threshold definition and equating (3) to zero, we find the participation threshold ability (a^*):

$$U_p(a^*) = \alpha(a^*) - c = 0,$$

(5) $\quad a^* = \frac{c - \alpha_0}{\alpha_1}.$

As in this case total added value only decreases with ability, all teachers with ability lower than the threshold derive positive utility from the training and sign up for it. Thus, the participation rate equals the threshold.

Now we look at the situation where ability is not observable.

Proposition 1. In the equilibrium, all teachers are indifferent between taking and not taking the training.

We find the participation threshold by equating utility from training (U_p) to the utility from not signing up (U_{np}) and plugging (2) into (3):

(6) $$EK(a^*)_p - c = EK(a^*)_{np}.$$

Since (6) does not depend on a directly, the benefits from (not) going into training are the same for everyone and boil down to the expected knowledge of each group. This means that, in equilibrium, the utility of taking the training is the same for the entire teacher pool and is equal to the utility of not taking the training. The equilibrium is determined by the number of teachers in each group and their individual abilities. This situation is similar to the perfect competition theory in Microeconomics, where the profits of an individual firm depend on the number of firms in the market. This mechanism regulates the number of operating firms. Similarly to our case, in the equilibrium all active firms receive zero profits (equal to the profits they would receive by not operating).

Proposition 2. Participation rate of a basic training increases with an increase in added value and decreases with an increase in costs.

As $EK(a^*)_p$ ($EK(a^*)_{np}$) is the expected knowledge of the group taking (not taking) the training, $K(0)$ is the knowledge level of the lowest ability teacher, and 1 is the maximum possible ability, we can calculate the average knowledge of each group:

(7) $$EK(a^*)_p = \frac{K(0)+K(a^*)}{2},$$

and

(8) $$EK(a^*)_{np} = \frac{a^*+1}{2}.$$

Taking (2) into account we solve (6) for a^*:

(9) $$a^* = \frac{2c - 2\alpha_0 + 1}{\alpha_1}.$$

Taking into account that $\alpha_1 < 0$, we see from (9) that the participation rate depends positively on added value parameters and negatively on the costs. When α_0 is increasing, participants gain relatively more from the training. The same is true for an increase in α_1 (decrease in absolute value). As the added value curve becomes flatter, on average total gains increase. Therefore, both changes lead to increased participation in the program.

Proposition 3. Basic training should be extremely effective or appealing to a wide audience in order to attract participants.

We also notice that in order for a basic training program to attract participants (a^* to be positive), its net fixed added value should be higher than $\frac{1}{2}$. Signing up for a basic training of this kind will signal low initial ability, for which the training should compensate by being effective enough to bring the lowest ability teacher to an above average level. This extreme result is explained by our assumption that training participation is the only available source of information about teacher quality. Once we allow for more sources of information, the magnitude of the result becomes smaller (see section 4.3.3.1).

The negative signal associated with basic training implies that this type of training will attract few (if any) participants. This poses a problem for policy interventions aimed at low quality teachers. For example, assume that the aim is to improve the teachers at the bottom 10% of the ability distribution and to bring them to the level just above this 10th percentile. Our model shows that nobody will participate in the training due to the negative signal even if the training program is proven to be effective. Policy makers that aim to target the bottom of the distribution will therefore necessarily have to design their program so that it is appealing to a broader audience in order to induce teachers to participate. However, generalizing the program will decrease its efficiency in raising the quality at the low end of the distribution. While somewhat counterintuitive, decreasing the knowledge gained from training for those at the bottom of the distribution increases their willingness to participate. The other, less realistic, option would be to make the training so effective that low ability teachers are better off after the training than not participating at all and being assumed to be average.

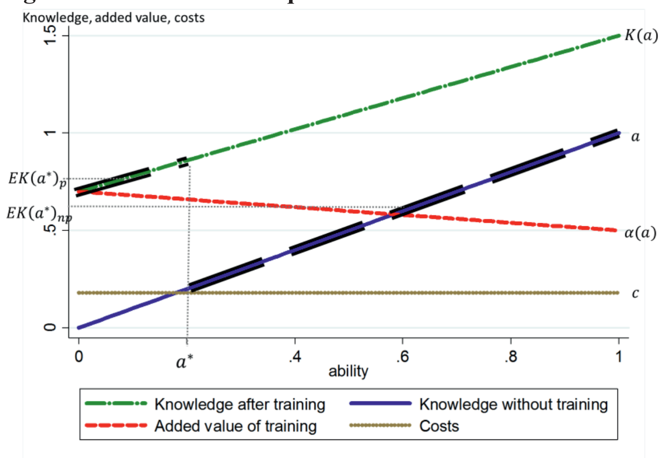
Proposition 4. Imperfect information decreases the participation rate of the basic program.

Now we compare this result with hypothetical situation (5) without signalling discussed earlier. If we rewrite (9) we can treat the second term as a signalling stigma, reducing participation ($c - \alpha_0 > -1$):

$$a^* = \frac{2c - 2\alpha_0 + 1}{\alpha_1} = \frac{c - \alpha_0}{\alpha_1} + \frac{c - \alpha_0 + 1}{\alpha_1}.$$

A graphical example of a basic training program is provided in Figure 4.1. The added value and costs of the training are: $\alpha_0 = 0.7,\ \alpha_1 = -0.2,\ c = 0.18$. We see that the line of gross added value ($\alpha(a)$) is always above the costs (c), therefore in the situation where productivity is observed and the training does not have any signalling value, all teachers sign up for training, and the participation rate is 100%. Everyone improves their initial ability and shifts from a (solid) to $K(a)$ (dash-dot) line. The knowledge after training curve ($K(a)$) is the sum of knowledge without training (a) and added value of training ($\alpha(a)$). If productivity is not observable, using the formulas derived earlier, we calculate that the stigma of the basic training reduces participation to 20% (a^*). Teachers with initial ability below a^* take the training and their knowledge shifts to the $K(a)$, while teachers with initial ability higher than a^* stay on a. The thick dashed black line is used to show the knowledge distribution in the market in equilibrium.

Figure 4.1. Basic course equilibrium



4.3.1.2 Advanced training

Similar to the previous section we first analyse a situation where signalling is not applicable. Since we use the same formulas, the participation threshold is the same as for the basic program:

$$a^* = \frac{c - \alpha_0}{\alpha_1}.$$

In this case teachers with initial ability higher than a^* take the training, therefore the participation rate equals $1 - a^*$.

Now we consider the situation where ability is not observable and teachers can sign up for training to signal their ability. Since the program is assumed to be advanced and highly able teachers benefit from them the most, smart individuals are expected to sign up. Therefore, if for a particular teacher the actual net benefit from the training is negative (added value is lower than the costs), but at the same time his knowledge after training is lower than the average knowledge of teachers attending that training, he can pretend to have higher knowledge just by subscribing to that training. Therefore, because of the positive signal, more teachers are expected to take the training program than just those that benefit from it in terms of increased ability.

Proposition 5. Participation rate of an advanced training increases with increases in added value and decreases with an increase in costs.

Again, the threshold is derived from (6), but this time representations for expected knowledge of the groups are different because the tails of the distribution for which the training is most and least beneficial change places:

$$EK(a^*)_p = \frac{K(1)+K(a^*)}{2} \text{ and } EK(a^*)_{np} = \frac{a^*+0}{2}.$$

$K(1)$ is the knowledge level of the teacher with the highest ability and 0 is the lowest possible ability. Then the threshold is:

$$(10) \quad a^* = \frac{2(c-\alpha_0)-1-\alpha_1}{\alpha_1},$$

and the participation rate is:

$$(11) \quad 1 - a^* = \frac{2(\alpha_1 + \alpha_0 - c) + 1}{\alpha_1}.$$

Here again the participation rate depends positively on added value parameters and negatively on the costs. The underlying intuition is the same as in the previous case. An increase in added value makes more teachers better off taking the training and shifts the threshold to the left, increasing the participation rate.

Proposition 6. The decision to go into training does not have to be related to program effectiveness and can have a pure signalling effect.

We notice that for low or even negative fixed added value (α_0) and low absolute value of marginal returns, in the equilibrium without signalling nobody signs up for training because the net benefits are negative. However, when teachers' productivity is unobservable and signalling is possible, teachers do go into training when the costs are higher than the added value in terms of ability. If the program is useless but assumed to be for highly able individuals, it provides an opportunity for teachers to signal high ability. Therefore, teachers at the margin will sign up for a program from which they gain no knowledge if the program is assumed to be advanced.

Proposition 7. Advanced training programs are oversubscribed.

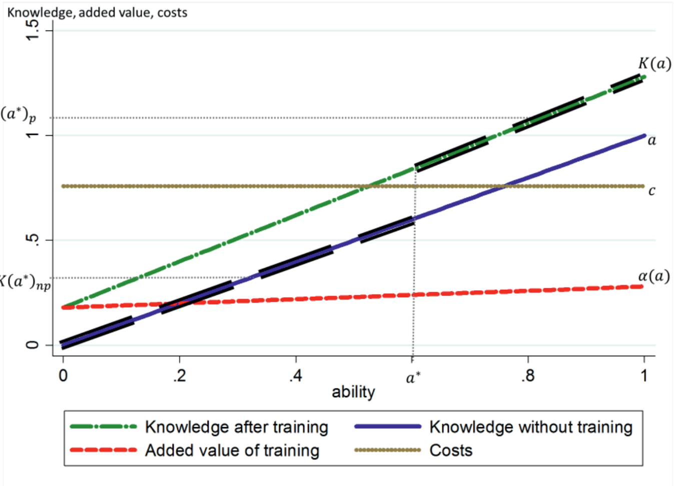
Now we rewrite (10) to compare it with (5) and see what happens to the threshold due to signalling:

$$a^* = \frac{c - \alpha_0}{\alpha_1} + \frac{c - \alpha_0 - 1 - \alpha_1}{\alpha_1}.$$

We see that the threshold is lower and more teachers are participating, including those for whom the training is too costly in addition to those who would take the training without signalling.

Figure 4.2 provides an example of an advanced program. The added value parameters and costs are: $\alpha_0 = 0.18$, $\alpha_1 = 0.1$, and $c = 0.76$. We see that for any ability level the costs are higher than total added value. Therefore in the case that teacher productivity is observable, nobody signs up for the training and the participation threshold $a^* = 1$. In the case that teacher productivity is not observable, participation rate rises to 40% ($a^* = 0.6$), entirely due to the positive signal. Again, the thick dashed black line shows the knowledge distribution in the market.

Figure 4.2. Advanced course equilibrium



4.3.2 Multiple training programs model

So far, we discussed a situation where teachers can only choose one training program. Now we extend the model to a situation where teachers can choose between an advanced and a basic program. We assign additional indices $i = 1$ to the parameters corresponding to a basic training and $i = 2$ to the parameters of an advanced training. Added value and costs of training are now denoted as:

$\alpha_i(a) = \alpha_{0i} + \alpha_{1i}a$ and c_i, $i = 1,2$.

Now teachers can choose between three options: basic training, advanced training, and no training. Each option provides a signal, as described in the previous section. As the nature of the programs does not change, we assume that in equilibrium some teachers from the bottom of the distribution take basic training (with abilities below some a_1^*), and some from the top take an advanced one (with abilities above some a_2^*). Then two cases are possible. First, teachers in the middle of the distribution (between a_1^* and a_2^*) are better off not taking any training and there are two different thresholds a_1^* and a_2^* (Figure 4.3). Second, there is full training participation and only one threshold a^*. Teachers with ability lower than a^* take basic training and teachers with ability above a^* take an advanced one (Figure 4.4).

Figure 4.3. Multiple courses 2-threshold equilibrium

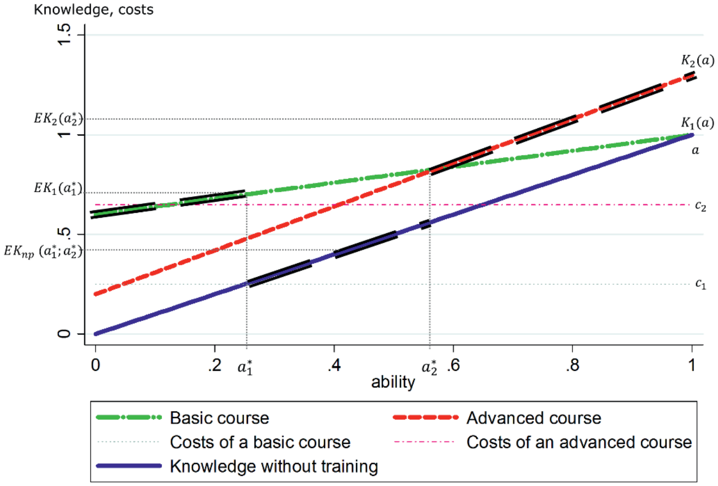
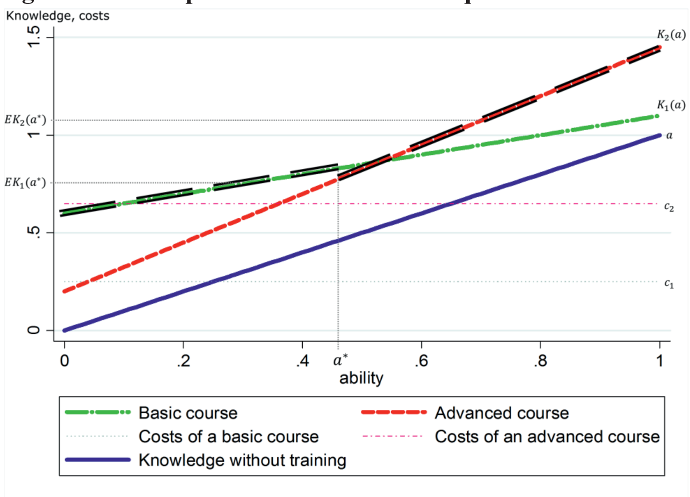


Figure 4.4. Multiple courses 1-threshold equilibrium



Proposition 8. Parameters of different programs affect each other's participation.

We start with the first situation. The threshold of the basic program is strictly lower than the threshold of the advanced program ($a_1^* < a_2^*$). Rewriting (6) for both programs gives a system of equations:

$$(12) \quad \begin{cases} EK_1(a_1^*) - c_1 = EK_{np}(a_1^*; a_2^*) \\ EK_2(a_2^*) - c_2 = EK_{np}(a_1^*; a_2^*) \end{cases}.$$

Solving for a_1^* and a_2^*:

$$(13) \quad \begin{cases} a_1^* = \frac{2\alpha_{02}+\alpha_{12}[2\alpha_{01}+1-2c_1]+1-2c_2}{1-\alpha_{11}\alpha_{12}} \\ a_2^* = \frac{2\alpha_{01}+\alpha_{11}[2\alpha_{02}+\alpha_{12}+1-2c_2]-2c_1}{1-\alpha_{11}\alpha_{12}} \end{cases}.$$

We can see from the parametrical representations of a_1^* and a_2^* that each threshold depends on the specifics of both programs.

Proposition 9. Interventions targeting advanced training have an indirect effect on demand for basic training.

For example, if the advanced training program gets better (increase in fixed or marginal added value), the share of participants of the basic program increases (together with its threshold). It also increases with a decrease in costs of the advanced program. Improvements in design of the advanced program ($\uparrow \alpha_2$) as well as availability to a wider audience ($\downarrow c_2$) boost demand for the basic program without any direct interventions.

The mechanism behind this result is the following. A rise in a_{02} or a_{12} and a drop in c_2 makes the advanced program appealing to more teachers and shifts the participation threshold to the left (a_2^*). As a_2^* is also a right margin of the group not taking the training, its average ability decreases ($\downarrow EK_{np}$). As a result, the benefits from not taking any training go down for all ability levels. Therefore, for teachers with abilities close to the threshold a_1^*, who were better off not taking basic training, the benefits from taking it are now relatively higher. This shifts a_1^* to the right, which increases both the participation rate and the average knowledge of teachers who take the basic training ($\uparrow EK_1$).

Proposition 10. Increased participation of a basic training reduces the positive signal of an advanced training.

Conversely, higher added value of basic training ($\uparrow \alpha_{01}, \alpha_{11}$) and lower costs ($\downarrow c_1$) lead to a higher participation threshold ($\uparrow a_2^*$) and a lower participation rate of the advanced program ($\downarrow 1 - a_2^*$). However, as in the previous case only teachers close to margin (a_2^*) are affected. Therefore, only those teachers who sign up purely because of the signal choose not to participate in the program, and sorting into training becomes more efficient.

Proposition 11. Different professional development programs mutually reduce the absolute value of each other's signal.

Comparing (9) and (11) with (13) we see that both professional development programs mutually reduce the absolute value of each other's signal. Moreover, improvements in basic or advanced training alone affect participants of both types of programs, as well as teachers not taking training at all.

As mentioned above, this holds for the situation where $a_1^* < a_2^*$. Plugging representations for a_1^* and a_2^* from (13) into this inequality we get

(14) $$2[\alpha_{01} - \alpha_{02} + c_2 - c_1] - 1 - \alpha_{12}[2\alpha_{01} + 1 - 2c_1] + \alpha_{11}[2\alpha_{02} + \alpha_{12} + 1 - 2c_2] > 0.$$

(14) can be used as a condition to distinguish between full and partial participation. If (14) holds, we are in the first case and $a_1^* < a_2^*$. Otherwise, if the left-hand side of (14) is less or equal to zero we get the second case and $a_1^* = a_2^* = a^*$. To find out participation threshold, we substitute a_1^* and a_2^* with a^* in (12) and get

(15) $$a^* = \frac{2(\alpha_{01} - \alpha_{02} + c_2 - c_1) - 1 - \alpha_{12}}{\alpha_{12} - \alpha_{11}}.$$

In this situation, all teachers sign up for training: teachers with ability lower than a^*choose the basic training program, and teachers with higher ability sign up for the advanced program. It is clear from (16) that the participation threshold positively (negatively) depends on the added value of the basic (advanced) training and negatively (positively) on the costs of the basic (advanced) one. This is quite intuitive, as due to the full participation, increase in the added value (as well as decrease in costs) of the basic program attracts teachers who were better off signing up for the advanced training. Therefore, the threshold just shifts to the left.

4.3.3 Extensions

4.3.3.1 Multiple signals

So far in this chapter we introduced two simple cases where a teacher's productivity is either perfectly observed or not observed at all, and the training program is the only available source of information about teacher ability. In practice, however, agents usually have additional ways to infer teachers' ability (e.g. personal observations, teacher characteristics such as work experience). Therefore, it is logical to include more sources of teacher quality information into the model.

We assume that in the general case, observed teacher knowledge (OK(a)) is a combination of the signal of a training program ($EK(a^*)$) and information about teacher knowledge ($K(a)$) from other sources:

$$OK(a) = \lambda EK_i(a_i^*) + (1-\lambda)K(a), \;\; i = 1{,}2,$$

where $\lambda \in [0{,}1]$ characterizes information asymmetry in the market. If $\lambda = 1$, we are in the case of perfect asymmetry analysed in detail earlier. With a decrease in λ the importance of the signal of the training program decreases, as information from other sources becomes more reliable. Since λ can be perceived as importance of the signal of the training program, it affects the benefits of teachers from the training and therefore (3) becomes:

$$U_p(a) = \alpha(a) - c + \lambda[EK(a^*)_p - K(a)].$$

Although a decrease in λ attenuates the signal of the course compared to (3), it is clear that, while becoming smaller in magnitude, all the results of the model still hold.

4.3.3.2 Applications

While the model we develop is used solely to analyse teachers' decision to participate in professional development programs, some of the implications of our results can be generalized to other contexts. Particularly, in any situation where productivity is partly unobservable, and employees have an informational advantage regarding their ability, on-the-job training programs have signalling value. Because introducing a new training program attenuates the value of the signal of signing up for other programs, sorting becomes more efficient. Therefore, evaluations of these programs should take the effects of improved sorting into account. Evaluations that do not look at these spillover effects may underestimate the value of offering that particular training program.

Secondly, cost-benefit analyses on the effects of interventions to increase training participation by reducing costs should be based on the expected productivity gains of the marginal, rather than the average, participant. Our model shows that increasing the attractiveness of a training program by reducing its costs (e.g. through subsidizing participation) will induce participation of employees that are close to the participation threshold. However, those that are on the participation threshold are exactly those people for who the training adds relatively little in terms of productivity. Therefore, companies or policy makers that base their cost-benefit analysis on the average increase in productivity per participant will overestimate the benefits of subsidizing training.

4.4 Empirical predictions

The results from our model lead to several testable predictions. In this section, we will suggest some ways to validate our model. We also discuss how the predictions from our model differ from predictions following from models based on intrinsic motivation.

First, we note that the gains from sending a positive signal are greater at certain points in a teacher's career. In our model, we assumed for simplicity that the signal each course sends is valued equally among all teachers. However, those for whom the evaluation by their employer is more important in terms of their future employment, benefit more from signalling high quality. In terms of the model, these teachers will put relatively more weight on the value of the signal ($EK(a^*)_p$) than on the gains in actual knowledge ($K(a)$). In practice, this implies that non-tenured teachers are more likely to sort into advanced courses than teachers that are on permanent contracts. When the signal gains importance relative to actual gains in knowledge, it also means that sorting into training becomes less efficient. At the high participation threshold, teachers will undertake training which adds nothing in terms of knowledge for the positive signal. At the low participation threshold, teachers who would gain knowledge from basic training will not take it because of the negative signal.

A related prediction is that, considering school principals form increasingly accurate beliefs of their teachers' ability (Rockoff et al., 2012), teachers have more influence over their perceived quality early in the employment relationship. Therefore, our model predicts more participation in courses with high positive signalling value for teachers in schools that have come under management of a new school principal. Again, we expect that this effect is stronger for non-tenured teachers.

Importantly, the preceding two predictions diverge from predictions made by a model that considers intrinsic or public sector motivation as the main factor in training participation. If intrinsic motivation would drive training participation, there should not be a difference in training behaviour related to contract status or length of the employment relationship. Each individual teacher would sort into the training program that increases their ability most effectively, regardless of the signal. If anything, assuming basic skills are necessary to benefit from advanced training, we would expect the opposite pattern: teachers should sort into basic training more often early in their career.

Second, because the existence of multiple courses targeted at different parts of the teaching ability distribution reduces the signalling value of each individual course, schools offering a larger variety of training programs will have a higher training participation rate. As we have shown before, adding the option of an advanced course to a choice set consisting only of basic courses reduces the average ability of those teachers who do not sort into any training. This decreases the negative signal associated with signing up for a basic course, increasing participation in said course. A related prediction is that, because the existence of advanced courses increases participation in basic courses, schools that offer advanced training opportunities will have a higher level of average teacher quality. This holds regardless of the effectiveness of the advanced training programs, provided that the basic course is effective.

Again, these predictions differ from predictions made by an intrinsic motivation model. If intrinsic motivation drives training participation, the existence of an advanced course should not influence participation in a basic course. If sorting into training is based only on gains in teaching ability, the existence of a course that is dominated in terms of gains should not influence the rational teacher's sorting decision under the independence of irrelevant alternatives assumption.

4.5 Conclusions

This chapter develops a model on the training participation decision of teachers. We argue that because performance is partly unobserved and teachers face career concerns, teachers can use training participation as a signal of their teaching ability. We show that because teachers incorporate the signalling value of signing up for training into their decision-making process, sorting into training is inefficient. Advanced courses are oversubscribed, and basic courses are underused. Our model can help explain the paradoxical finding that training participation in the teaching profession is high (OECD, 2014), while most programs are evaluated to be ineffective (Yoon et al., 2008).

Schools and policy makers can improve sorting into training by offering a variety of courses. The existence of an advanced course increases participation in basic courses, as the average ability of teachers that do not take any training decreases. The negative signal of signing up for basic courses weakens as a result. Because of these spillover effects, program evaluations that look at the effectiveness of training programs in isolation may miss potential overall positive effects on non-participants whose behaviour is changed through the introduction of the program. Furthermore, policy makers that aim to increase training participation by reducing costs should base their cost-benefit analyses on the gains of the marginal participant. Because the marginal participant's benefits are below average, calculations based on the average increase in productivity per participant will lead to an overestimation of the benefits of decreased costs of training. Finally, policy makers that target the bottom of the distribution will have to design their program to appeal to a wider audience. Generalizing the course will likely decrease its effectiveness for the target population, but is a necessity to induce them to participate. This illustrates the difficulty of improving teaching quality at the bottom of the distribution.

In order to most efficiently raise the quality of education through increasing the quality of teachers, a clear understanding of the decision to enter professional development programs is invaluable. Our model provides an explanation for why sorting into training could be inefficient, and offers suggestions to decrease inefficiencies. However, empirical research should validate whether our model can explain observed training participation patterns before policy makers take up suggestions based on theory alone. For this reason, we make several empirical predictions that diverge from predictions made by models based on intrinsic motivation. Ultimately, when teachers use training as a signal in practice, insights from our model can help policy makers in designing the most efficient implementation strategy for professional development programs.

5. Teacher Literacy and Numeracy Skills: International Evidence from PIAAC and ALL

5.1 Introduction

Teachers are essential for the development of human capital in society. Their skills are formed in teacher training programs, but are also highly influenced by the type and overall quality of the students who enter these programs and become teachers. Understanding which segment of the population is part of the teacher corps is important in order to determine the focus of interventions that can improve the quality of teachers.

This chapter compares the dispersions of literacy and numeracy skills of primary and secondary school teachers relative to the rest of the adult population.[37] We use international data of 15 different countries from the Programme for the International Assessment of Adult Competencies (PIAAC) and the Adult Literacy and Lifeskills Survey (ALL), both conducted by the OECD. These data sets contain representative samples of the adult population in various countries. They include reading and math test scores, as well as detailed information about occupations. For each country, we compare average math and literacy skills between teachers and other respondents, and we investigate differences at the 10th and 90th percentiles of the distributions.

The results show that in virtually all countries, both primary and secondary school teachers score higher on literacy and numeracy tests than the country average. Secondary school teachers outperform primary school teachers on both skill measures. Our analyses of the differences in skill distributions between teachers and others show that the lowest scoring teachers significantly outperform the lowest scoring other respondents both on literacy and numeracy tests. At the top of the distributions, we find that the best secondary school teachers are not strongly outperformed by the best other respondents, and that the best primary school teachers score only slightly lower than the best scoring other respondents. Our results persist when restricting the comparison to the tertiary educated subsample. The results are not driven by age or gender, and are not sensitive to the inclusion of measures for the frequency of skill use.

The extent to which low scoring teachers outperform other low scoring respondents differs substantially across countries. In the Netherlands, primary school teachers at the 10th percentile perform much better than other respondents at the 10th percentile, while in

---

[37] Joint work with Bart Golsteyn and Inge de Wolf. This chapter has been published as Golsteyn, B. H. H., Vermeulen, S., & de Wolf, I. (2016). Teacher literacy and numeracy skills: International evidence from PIAAC and ALL. *De Economist*, *164*(4), 365-389. We thank Lynn Veld for excellent research assistance and seminar participants at Maastricht University, the Inspectorate of Education, KU Leuven, Universität Tübingen, Fontys Institute for Teacher Education Sittard, the Dutch Education Research days 2015, the Dutch Economists Day 2015 and the International PIAAC conference 2015 for their valuable comments.

Denmark for example, primary school teachers at the 10th percentile do not outperform other respondents at the 10th percentile that strongly. In terms of policy, in Denmark it might therefore be effective to focus on the bottom of the distribution (e.g., by raising barriers to enter into teaching, or focusing training on the worst teachers), while in the Netherlands little can be gained in becoming more restrictive at the lower end.

This chapter contributes to the literature on teacher characteristics and teacher quality. Teacher quality has been recognized as one of the most important determinants of educational productivity (Hanushek, 2011; Barber & Mourshed, 2007). Hanushek (1992) finds that being taught by a high quality teacher results on average in 1.5 years' worth of progress in terms of test scores in one academic year, while being taught by a low quality teacher results on average in 0.5 years' worth of progress. Although the exact characteristics of teacher quality are not well-defined (Hanushek & Rivkin, 2006), teachers' skills as measured by scores on achievement tests have been found to be associated with educational productivity (Metzler & Woessmann, 2012; Eide, Goldhaber, & Brewer, 2004). Hanushek, Piopiunik, & Wiederhold (2018) furthermore show that teachers' cognitive skills are an important factor in explaining international differences in student performance.

While earlier studies explore how the relative differences between individuals within the teaching population relate to educational outcomes (e.g., Hanushek, 2003; Rivkin, Hanushek & Kain, 2005) and long-term economic outcomes (e.g., Chetty, Friedman, & Rockoff, 2014; Hanushek, 2011), our study addresses international differences in the *dispersion* of math and language skills of teachers relative to others. Investigating the differences in the distribution rather than the average skill level allows us to pinpoint the focus of potential policy interventions that aim to improve the quality of teachers. Our analyses reveal that the focus of interventions should lie on different parts of the distribution of skills in different countries.

The remainder of this chapter is organized as follows. Section 5.2 explains the relationship between skill dispersions and policy interventions. Section 5.3 describes the data. In section 5.4, the research strategy is discussed. Section 5.5 gives the results. Section 5.6 shows several robustness checks, and section 5.7 concludes.

5.2 Skill dispersions and policy interventions

In the analysis, we investigate math and language skill dispersions of teachers relative to the dispersions of the skills of others in society. These relative dispersions give an indication at which part of the distribution teachers perform relatively well, and where they perform worse than others. Potential policy interventions aiming to increase teacher skills can take different forms, e.g. increasing wages, installing entry barriers, training the current teacher force, or improving the training of new teachers in teacher colleges. A key aspect in each of these examples is that there is an implicit or explicit choice to improve teachers' skills in some part of the distribution.

The costs of these policy measures vary highly. Increasing the wage for all teachers for instance is very costly, especially if the aim is to raise the level of low skilled teachers only. A cheaper and probably more effective solution for countries with a large percentage of low skilled teachers might be to install entry barriers or intensify elementary training programs. In a situation where the lowest skilled teachers already perform relatively well on basic skills, it may make more sense to provide training that adds value to highly skilled teachers. To conclude, getting insight into the dispersion of skills in a country helps to form a decision about which part of the distribution should be targeted and about which tools can be most effective and efficient to reach the target.

In order to show the relative skill dispersions, we compute the score on a measure of a skill (numeracy, literacy) for each percentile of (1) the distribution of the skills for teachers only and (2) the distribution of skills of other respondents, and then subtract these two distributions. The results can be plotted in a figure, such as figure 5.1A. The horizontal axis displays the percentile of the distribution and the vertical axis displays the difference in the score on the measure of the skill between teachers and others. A positive difference at the first percentile implies that teachers in the first percentile of the teacher distribution score higher than other respondents in the first percentile of their distribution.

Figure 5.1A illustrates four hypothetical relative skill distributions.[38] In the first scenario, represented by the horizontal line at zero, teachers and non-teachers are equally skilled across the entire distribution. The lowest scoring teachers are as skilled as the lowest scoring non-teachers, and the highest scoring teachers are as skilled as the highest scoring non-teachers. In this scenario, a possible policy suggestion to increase the average quality of

[38] Note that in these scenarios, we assume the potential teacher supply outweighs demand and schools are not able to perfectly observe differences in skill levels between job candidates.

teachers may be to implement entry barriers to avoid low-skilled teachers from entering. Another policy suggestion is to give teachers better training, either in school or in the labor market.

Figure 5.1A: Hypothetical relative skill distributions: differences at all percentiles

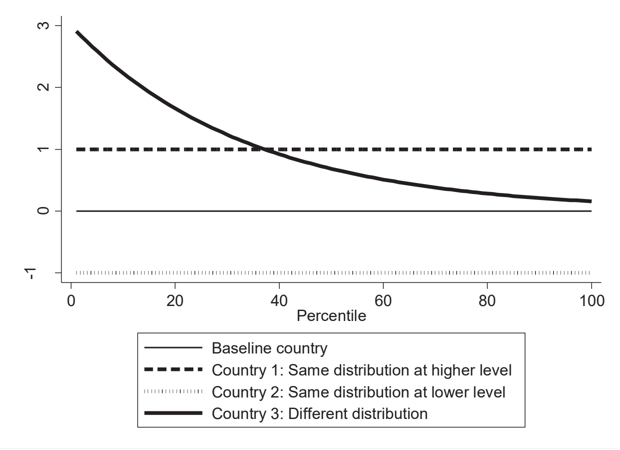
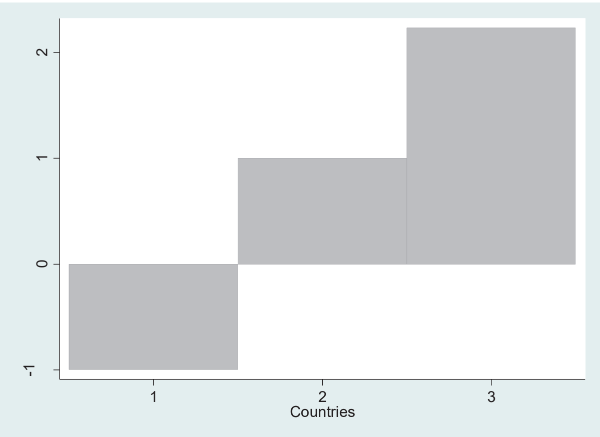


Figure 5.1B: Hypothetical relative skill distributions: differences at the 10th percentile



Increasing barriers to entry might result in a situation such as scenario 2, represented by the uninterrupted convex line. In this situation, teachers outperform the rest of the population mainly in the lower parts of the distribution. Policies aimed at increasing the quality of the lowest skilled teacher by becoming even more selective might result in teacher shortages, and in this case, policies aimed at increasing the attractiveness of the teaching profession for highly skilled individuals might be more effective to increase the quality of the teacher corps.

In scenario 3, represented by the dashed line at one, teachers outperform non-teachers consistently across the entire distribution. Here, teaching appears to be an attractive alternative for people across the full skill distribution. Highly skilled people are sorting into the profession, preferring teaching jobs to non-teaching jobs. An effective policy option may be to install entry barriers for lower skilled people, allowing more highly skilled people to enter the profession and increasing overall quality. In this scenario, however, the teachers are already skilled above average across the board, which raises the question whether the costs associated with intervening would be worth the benefits. Scenario 4 is the opposite of scenario 3 in the sense that non-teachers consistently outperform teachers across the entire distribution. In this case, making the profession more attractive for people across the distribution (e.g., by raising wages) would be an effective policy suggestion.

These examples show the importance of investigating skill distributions beyond analyzing differences in means only: if we would only investigate *average* teacher quality, scenario 2 and scenario 3 would lead to similar conclusions. Looking at the distributions is more informative because it allows policy makers to design interventions aimed at a specific part of the distribution, which will be more effective and efficient than untargeted interventions. The differences in relative skill distributions across countries also give an indication of the result of different standing policies regarding teacher recruitment and teacher training and serve to illustrate why policies to attract better teachers need to differ between countries.

Figure 5.1A shows the differences in scores between teachers and others at all percentiles. In order to reduce complexity, we report the differences in scores in the analyses only at the mean, 10^{th} percentile, and 90^{th} percentile. This is shown in figure 5.1B, which displays the difference between teachers and non-teachers at the 10^{th} percentile only.

5.3 Data

We combine data from the Programme for the International Assessment of Adult Competencies (PIAAC) and the Adult Literacy and Lifeskills Survey (ALL), both conducted by the OECD. These surveys were designed to measure adult skills and competencies across different countries. In this section, we describe the ALL and PIAAC dataset, and provide descriptive statistics.

5.3.1 The ALL data

The ALL measured literacy and numeracy skills of nationally representative samples of 16-65 year olds in participating countries in two rounds. The first round was conducted in 2003 and the second round was conducted between 2006 and 2008. The six countries that took part in the first round were Canada, Italy, Norway, Switzerland, the United States, and Bermuda. In the second round, Australia, Hungary, the Netherlands, and New Zealand participated. The ALL study is the successor of the International Adult Literacy Survey (IALS), which was the first international comparative survey of adult skills undertaken between 1994 and 1998.[39] Measured skills in the ALL survey include prose literacy, document literacy, numeracy, and problem solving. Literacy was defined as "the knowledge and skills needed to understand and use information from texts and other written formats." Numeracy was defined as "the knowledge and skills required to manage mathematical demands of diverse situations."

5.3.2 The PIAAC data

PIAAC measures skills in three domains: literacy, numeracy and problem solving in technology-rich environments. PIAAC has two cycles of assessment: the first cycle is conducted in two rounds, while the second cycle is expected to take place from 2018 to 2023. The first round of the first cycle took place from January 2008 to October 2013. We use the publicly available data from the first round of the first cycle. The countries that took part in the first round of the first cycle were Australia, Austria, Belgium, Canada, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Korea, the Netherlands, Norway, Poland, Russia, the Slovak Republic, Spain, Sweden, the United Kingdom and the United States.

5.3.3 Merging the datasets and descriptive statistics

We identify teachers based on their 4-digit ISCO-88 occupation code.[40] The occupation codes allow us to distinguish between primary and secondary school teachers. For some countries, no teachers are present in the dataset or the occupation code is not detailed enough to

[39] IALS does not contain job codes at a sufficiently detailed level, and could therefore not be included in the analysis.

[40] Approximately 26% of the sample did not report their occupation. 75% of this subgroup reported to be out of the labor force while 15% were unemployed. We investigated the skill level of people not reporting an occupation and found that this group is drawn from the lower part of the skills distribution. Because it is therefore highly unlikely that these individuals are teachers, we consider it best to include these individuals in the control group.

correctly identify them. We exclude these countries from our analyses. The countries that are dropped are Australia, Austria, Canada, Cyprus, Estonia, Finland, Germany, Ireland, Sweden, and the United States.

Table 5.1 shows the summary statistics by country. Italy, Norway and the Netherlands have the most respondents; these are also the countries that were sampled in both PIAAC and ALL. In most countries, there is approximately an equal amount of men and women in the samples. The Russian data are not representative for the entire population as inhabitants of the Moscow municipal area were not included in the PIAAC survey. For this reason, we exclude Russia from our analyses.

The number of tertiary educated people in a country differs substantially. This implies that in our analyses, it is better to relate the skills of teachers to all other respondents in a country than to the tertiary educated sub-sample only. In a robustness check, we nonetheless make the latter comparison.

Table 5.1: Sample characteristics per country

		Age	Gender	Education level	
Country	All respondents	Mean	Pct. Females	Non-tertiary	Tertiary
Switzerland	5,120	43	51.5	66.7	33.2
Bermuda	2,696	42	47.9	68.5	31.5
New Zealand	7,140	40	57.1	66.9	33.0
Hungary	5,575	40	55.9	85.3	14.5
Belgium	5,463	41	50.6	59.0	30.7
Czech Republic	6,102	39	54.6	80.0	19.6
Denmark	7,328	44	50.7	59.4	37.9
France	6,993	42	51.0	68.7	29.3
Italy	11,474	42	52.2	87.5	12.1
Japan	5,278	42	52.3	53.0	44.8
Korea	6,667	41	53.5	62.7	36.9
Netherlands	10,787	43	53.3	65.2	33.6
Norway	10,539	40	48.5	57.3	40.3
Poland	9,366	31	49.5	76.4	23.6
Russia	3,892	36	65.5	35.5	64.4
Slovak Republic	5,723	39	52.7	82.4	17.2
Spain	6,055	40	51.1	70.4	27.6
United Kingdom	8,892	41	58.0	62.0	36.7

Note: pooled PIAAC and ALL data.

Table 5.2 shows the number of teachers and the number of primary and secondary school teachers per country sample. The Netherlands, Denmark, Norway, New Zealand, and the United Kingdom have relatively many teachers in their sample, while in other countries like Bermuda, Switzerland and Japan few teachers were sampled. For these latter countries, our results are less generalizable to the entire teacher population. In our analyses, we will therefore restrict our sample to those countries for which more than 50 primary or secondary school teachers can be identified.[41] Practically, this means that when analyzing primary school teachers Japan, Korea, Bermuda and Switzerland will be excluded. When analyzing secondary school teachers we exclude Japan, Korea, Bermuda, Denmark, Slovakia and the Czech Republic.

Tables 5.1 and 5.2 reveal that the average teacher is somewhat older than the average respondent. They are also more highly educated on average and the proportion of women is higher for teachers. In robustness analyses, we control for these differences.

Both the ALL and PIAAC surveys measure the skill domains on a 0-500-point scale. While ALL and PIAAC both aim to measure the same skills and use the same measurement scale, they do not employ the same tests. In order to be able to pool the two datasets, we standardize the test results based on the full sample, so that they have a mean of 0 and a standard deviation of 1.[42]

Table 5.2: Number of teachers in the sample by country and teacher demographics

					Gender	Education level	
Country	Primary school teachers	Secondary school teachers	All teachers	Mean Age	Pct. Female	Non-tertiary	Tertiary
Switzerland	0	93	138	46	53.6	27.5	72.5
Bermuda	10	16	97	43	77.3	26.8	73.2
New Zealand	157	79	382	44	75.4	11.3	88.7
Hungary	100	51	197	43	79.2	9.1	90.9
Belgium	71	92	264	41	65.5	6.1	93.6
Czech Republic	76	37	184	41	73.9	24.5	74.5
Denmark	292	46	506	47	63.6	13.8	84.2
France	53	106	212	43	60.8	9.9	89.2
Italy	70	163	344	46	70.6	32.0	68.0
Japan	32	42	168	45	58.3	10.1	89.3
Korea	40	43	253	39	73.9	10.7	87.4
Netherlands	224	186	594	45	66.3	12.5	87.2
Norway	194	95	476	43	62.6	7.6	91.6
Poland	104	63	261	36	82.4	8.0	92.0
Russia	64	76	219	39	87.7	9.6	90.0
Slovak Republic	88	16	176	43	80.1	22.7	77.3
Spain	67	52	234	42	66.7	9.4	90.2
United Kingdom	115	91	375	43	72.8	8.0	92.0

Note: pooled PIAAC and ALL data

[41] Including those countries with few observations does not change the conclusions from our findings. The results of these analyses are available upon request.
[42] To further check whether PIAAC and ALL are comparable, we look at the differences in distributions of teachers to non-teachers across the two datasets for those countries sampled twice (The Netherlands, Norway & Italy). The resulting percentile graphs (which can be found in graph A.5.1 of the appendix) show that the distributions are very similar across the two datasets. This can also be interpreted as supporting the stability over time of our main findings.

Table 5.3 shows the mean and standard deviation for the skills in all countries that have teachers in the sample. The results are depicted separately for all respondents and for teachers only. For each country, the mean test score of teachers is higher than the average score of all respondents.

This table reveals the differences in the absolute skill level for teachers between countries. In this chapter, we study the relative differences of skill levels between teachers and others within countries in order to indicate which segment of the population has become a teacher. We choose to focus on relative differences within a country because we are interested in studying the scope for improvement of the skill level of the teacher population. I.e., if the teacher population is relatively low skilled, then it will be more feasible to improve its skill level because there is a large pool of other people in society that may be induced to become a teacher. If this pool is smaller, improving the teachers' skill level is less feasible.

However, the absolute differences of skill levels for teachers between countries also provide an important source of information. Policy to improve teacher skills will partly depend on a minimal required skills level. Hence, the interplay between absolute and relative skill levels is policy relevant. Using the table, we can see that literacy and numeracy levels of teachers are lowest in Italy, while Japanese teachers score highest. Teachers in the Netherlands score very highly both on the literacy and numeracy tests relative to teachers in other countries.

Table 5.3: Mean and standard deviation of test scores by country

	All respondents				Teachers			
Country	Literacy		Numeracy		Literacy		Numeracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Switzerland	.05	.79	.31	.83	.59	.72	.78	.78
Bermuda	.28	1.05	-.02	1.05	.94	.76	.51	.77
New Zealand	.07	.96	-.14	1.07	.68	.66	.43	.82
Hungary	.03	.86	.10	.80	.61	.80	.71	.71
Belgium	.07	1.01	.24	.98	.69	.74	.80	.73
Czech Republic	.09	.85	.16	.84	.63	.77	.63	.71
Denmark	-.11	1.07	.16	1.03	.29	.88	.52	.87
France	-.20	1.06	-.25	1.12	.51	.79	.61	.80
Italy	-.66	1.00	-.58	.93	-.05	.83	-.09	.80
Japan	.56	.82	.43	.83	1.02	.65	.93	.63
Korea	-.01	.89	-.13	.89	.49	.67	.44	.68
Netherlands	.27	.90	.33	.94	.73	.65	.77	.69
Norway	.36	.95	.30	.97	.78	.72	.77	.74
Poland	.03	.97	-.08	.94	.51	.83	.24	.82
Russian Federation	.17	.88	.12	.77	.31	.82	.16	.70
Slovak Republic	.01	.85	.11	.93	.33	.64	.57	.66
Spain	-.51	1.09	-.52	1.05	.41	.71	.26	.67
United Kingdom	-.01	1.01	-.16	1.04	.70	.73	.53	.73

Note: pooled PIAAC and ALL data

5.4 Methodology

To compare the average skills of teachers to those of the other respondents, we respectively regress literacy and numeracy scores on a dummy for being a primary school teacher. We do these analyses for each country separately. We perform a similar set of regressions on a dummy for being a secondary school teacher. This yields the difference in test scores between teachers and other respondents, expressed in standard deviations. We use the full sample weights present in the datasets to account for sampling bias and to ensure our results are representative for the population. We generate bar charts to graphically represent teacher skill levels compared to the country average levels. In the baseline analyses, we do not control for other variables. In robustness analyses, we do control for potential confounders.

Next, we investigate the shape of the distributions. We show this by plotting differences in scores at the 10th percentile and the 90th percentile on the measure of a skill (numeracy, literacy) between non-teachers and primary and secondary school teachers, respectively.

We perform several robustness checks. Because most countries require their teachers to be relatively highly educated, we compare teachers with the college-educated part of the population for each country. These results should be interpreted with caution since the percentage of individuals with a tertiary education differs substantially between countries (see also table 5.1). Still, it is instructive to see whether our results could be explained purely by differences in educational attainment.

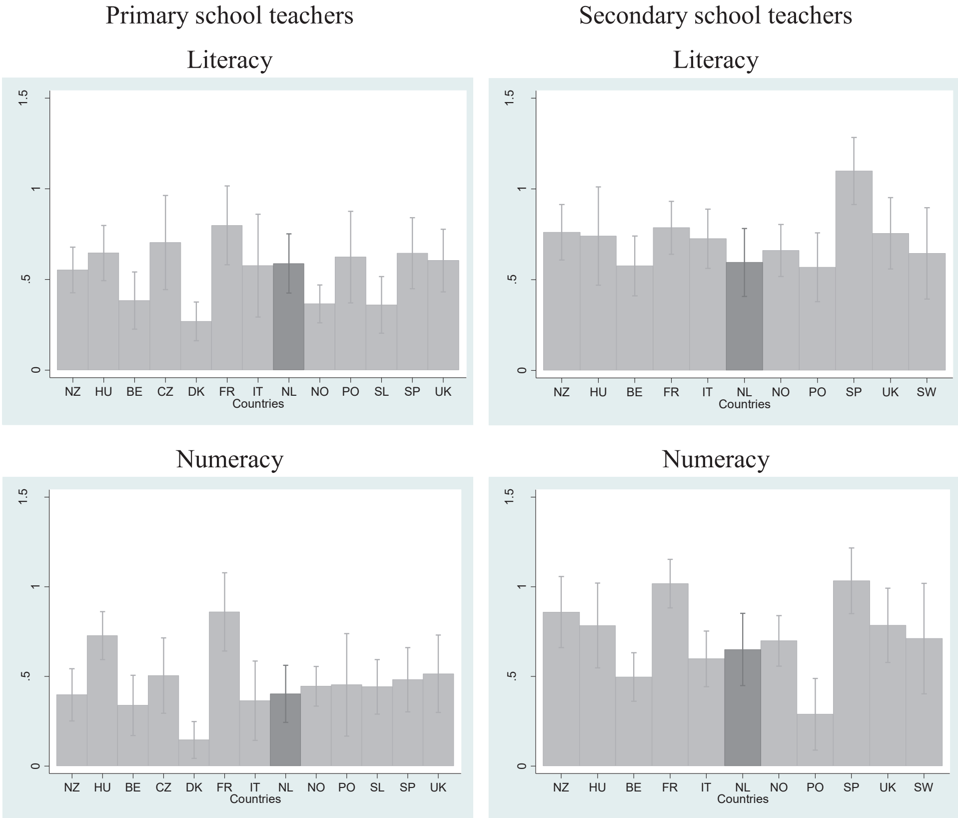
Another interesting question is whether differences in skills are driven by selection or whether these are due to the nature of the profession. Teachers are likely to use their literacy and numeracy skills more than others. Therefore, our results could reflect frequency of use rather than innate ability. To address this possible alternative explanation, we present regression-adjusted graphs where we control for skill use at the workplace.

5.5 Results

5.5.1 Differences in the averages of skills

Figure 5.2 shows how primary and secondary school teachers perform on literacy and numeracy skills compared to the rest of the population per country. In virtually all countries, teachers significantly outperform the average other respondent on both skill measures. The difference is slightly larger for secondary school teachers.[43]

Figure 5.2: Mean Skill Difference: teachers vs. other respondents



Note: bars indicate the mean test score difference in standard deviations between teachers and non-teachers for each country on the horizontal axis. The error bars indicate the 95% confidence interval. Differences were calculated using full sample weights.

[43] Regressions comparing primary school teachers to secondary school teachers show that the difference between the two groups is mostly insignificant, but significant in favor of secondary school teachers in New Zealand, Norway and Spain for both skill measures. Poland is the only country in which the sign of the coefficients is in favor of primary teachers. The graphs depicting these results are available upon request.

The graphs in figure 5.2 suggest that the scores for numeracy and literacy are highly related. To test the relationship, we calculate the correlation between literacy and numeracy skills for non-teachers, primary school teachers and secondary school teachers separately. These results can be found in table 5.4. As expected, literacy and numeracy skills are highly correlated within individuals.

Table 5.4: Correlation between Literacy and Numeracy scores for non-teachers, primary school teachers and secondary school teachers

	Non-teachers Numeracy	Primary school teachers Numeracy	Secondary school teachers Numeracy
Non-teachers Literacy	0.90**	-	-
Primary school teachers Literacy	-	0.83**	-
Secondary school teachers Literacy	-	-	0.81**

*Note: * = p < .05; **= p < .01*

5.5.2 Differences in the distribution of skills

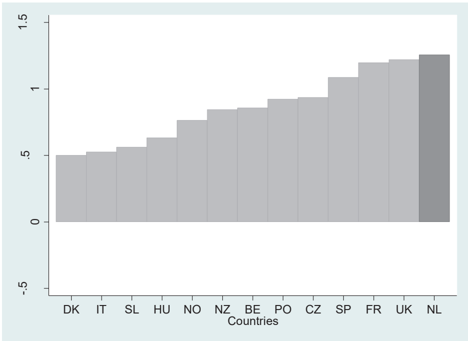
As argued in section 5.2, the shape of the skill distribution can be more informative for policy makers than the average skill level, as different distributions might warrant different interventions. For each available country, we depict the test scores at the 10^{th} and 90^{th} percentiles for teachers and non-teachers.

Figure 5.3 shows the difference between primary school teachers and others in society at the 10^{th} and 90^{th} percentiles for literacy and numeracy skills. For both skills, primary school teachers at the 10^{th} percentile strongly outperform the other respondents at the 10^{th} percentile in most countries. Interestingly, primary school teachers at the 90^{th} percentile are not outperformed by the non-teachers at the 90^{th} percentile, suggesting that there are some very highly skilled teachers. Given that wages for teachers are substantially lower than potential private sector wages at the high end of the skills distribution (Chevalier, Dolton & McIntosh, 2007; Stinebrickner, 2001), these results suggest that non-pecuniary factors (e.g. job security, secondary benefits, intrinsic motivation) may play an important role in the decision to become a teacher for highly skilled people. In figure 5.4, we show the results focusing on secondary school teachers. The results are similar to those found for primary school teachers.

Figure 5.3: Differences in literacy and numeracy scores at the 10th and 90th percentile for primary school teachers

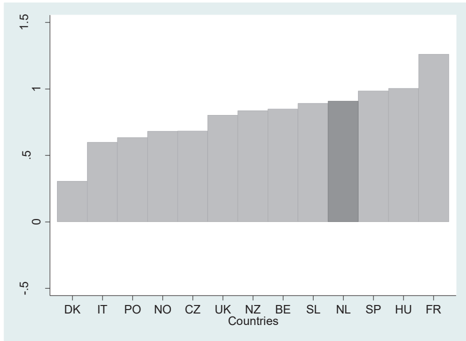
Literacy

10th

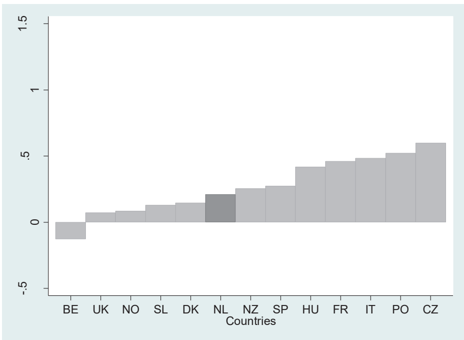


Numeracy

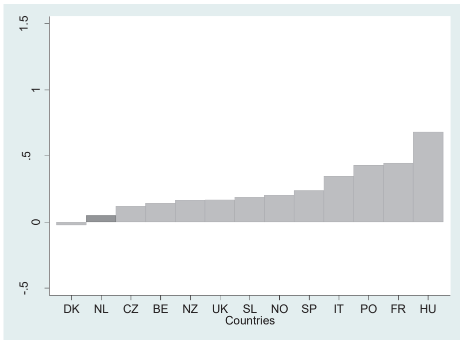
10th



90th



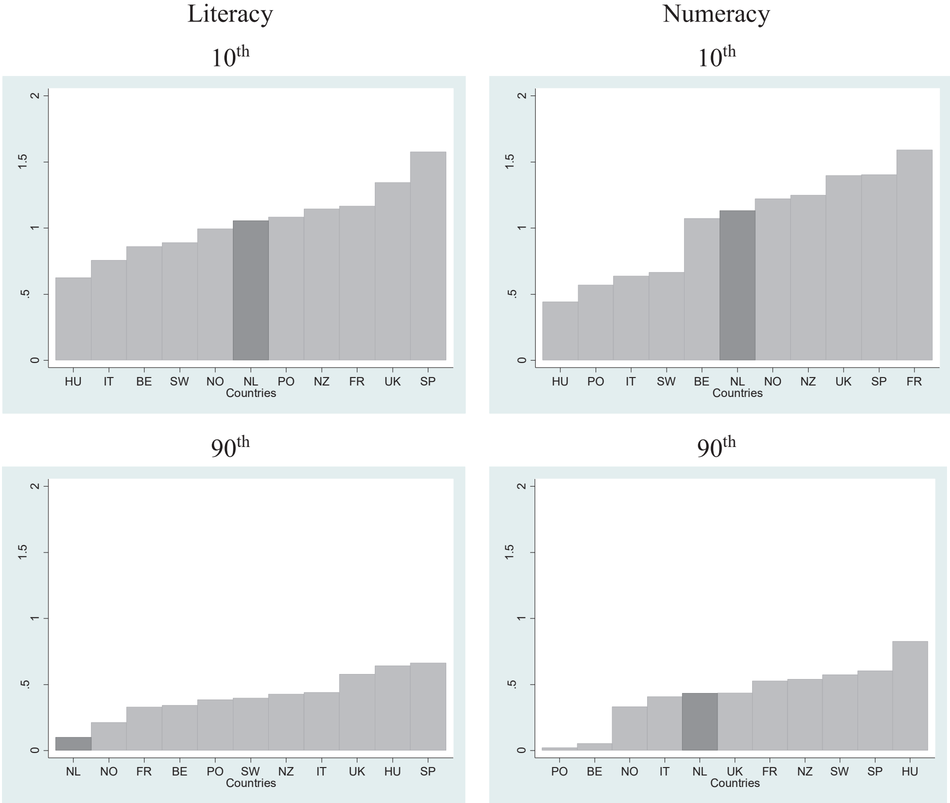
90th



Note: bars indicate the test score difference at the 10th and 90th percentile in standard deviations between primary school teachers and non-teachers for each country on the horizontal axis. Differences were calculated using full sample weights.

These findings provide a guideline for policy makers who design interventions aimed at increasing the average skill level of teachers. For example, if we compare the results from Denmark and the Netherlands in figure 5.2, we see that the average primary school teacher outperforms the average respondent in both countries (this difference is smaller in Denmark than in the Netherlands). Comparing the average difference across countries is informative, but for policy makers it is much more relevant to know at which part of the skill distribution the differences are most apparent. As argued in section 5.2, different shapes of the distribution warrant different kinds of interventions.

Figure 5.4: Differences in literacy and numeracy scores at the 10th and 90th percentile for secondary school teachers



Note: bars indicate the test score difference at the 10th and 90th percentile in standard deviations between secondary school teachers and non-teachers for each country on the horizontal axis. Differences were calculated using full sample weights.

When looking at the differences at the 10th and 90th percentile in figure 5.3, we see that in the Netherlands the primary school teachers at the 10th percentile perform very well compared to the 10th percentile of other respondents, while in Denmark primary school teachers at the 10th percentile are not that much better than the 10th percentile of other respondents. At the 90th percentile, the relative performance of primary school teachers in the two countries is much more similar. By looking at the skills dispersions, we can see that in Denmark an effective policy appears to be to focus on the bottom of the distribution (e.g., by raising barriers to entry into teaching), while in the Netherlands there is little to be gained in becoming more restrictive at the lower end. This example clearly shows the benefit of analyzing differences in distributions relative to differences in means, as the conclusions could not have been drawn if only country means were compared.

It is interesting to investigate whether the order of the countries in the graphs above is similar for numeracy and literacy scores. The order of the scores may be different, for instance, because some countries prioritize numeracy skills over literacy skills in their education system. This preference could be reflected in the relative skills of their teachers. It is important to take into account different orderings of the countries, as policy recommendations become less clear-cut when relative literacy and numeracy scores differ strongly. For example if low literacy scores are accompanied by high numeracy scores it is unclear whether interventions aimed at the bottom of the distribution would be effective in raising teaching quality as much as when low literacy is accompanied by low numeracy. Furthermore, if countries prioritize numeracy, low scores on literacy may not warrant any corrective intervention. In table 5.5, we investigate the correlations of the differences at the 10^{th} or 90^{th} percentile for numeracy and literacy skills. It appears that the differences at the 10^{th} percentile of literacy scores are highly correlated with the 10^{th} percentile differences in numeracy scores, both for primary and secondary school teachers. This implies that at the 10^{th} percentile, the order of the countries is similar for the numeracy and literacy scores. At the 90^{th} percentile, the results for literacy and numeracy skills correlate substantially for primary school teachers but less so for secondary school teachers. This suggests that within countries the relative literacy and numeracy skills of teachers are quite strongly related.

Table 5.5: Correlations of the differences in skills between teachers and non-teachers at the 10^{th} and 90^{th} percentiles

Primary School Teachers	Pct. 10 Literacy	Pct. 90 Literacy	Pct. 10 Numeracy	Pct. 90 Numeracy
Pct. 10 Literacy	1	-	-	-
Pct. 90 Literacy	0.02	1	-	-
Pct. 10 Numeracy	0.54	0.06	1	-
Pct. 90 Numeracy	-0.09	0.51	0.47	1

Secondary School Teachers	Pct. 10 Literacy	Pct. 90 Literacy	Pct. 10 Numeracy	Pct. 90 Numeracy
Pct. 10 Literacy	1	-	-	-
Pct. 90 Literacy	0.20	1	-	-
Pct. 10 Numeracy	0.74**	0.14	1	-
Pct. 90 Numeracy	-0.01	0.45	0.02	1

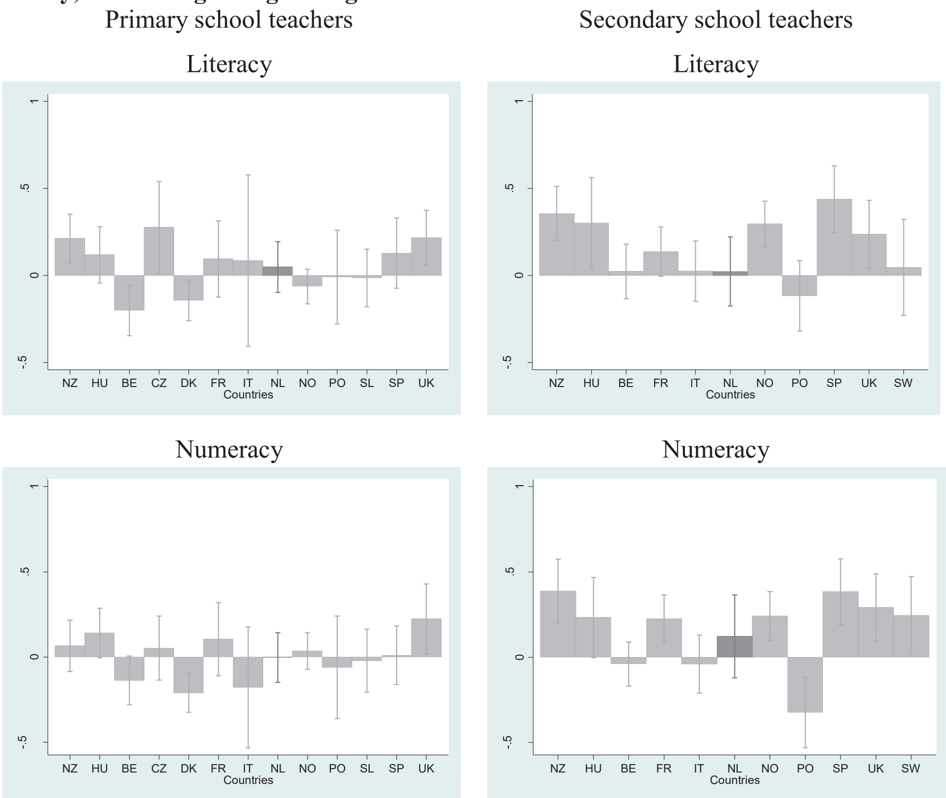
Note: The table shows the correlations of the order of the countries at various percentiles respectively.
* = p < .05; **= p < .01

5.6 Robustness

5.6.1 Comparison with the tertiary educated

In order to become a teacher, one must have a certain level of education. In most countries, only people with a tertiary degree are eligible for teaching jobs (OECD, 2014). One may therefore argue that it could be meaningful to restrict our analyses to that part of the sample that has a tertiary degree. Notice, however, that the percentage of people with a tertiary education differs dramatically between countries. Therefore, in our baseline analyses, we compare teachers' skills to those of all other respondents instead of only tertiary educated respondents. In this robustness check, we restrict the sample to the tertiary educated. Additionally, we add controls for gender and age.

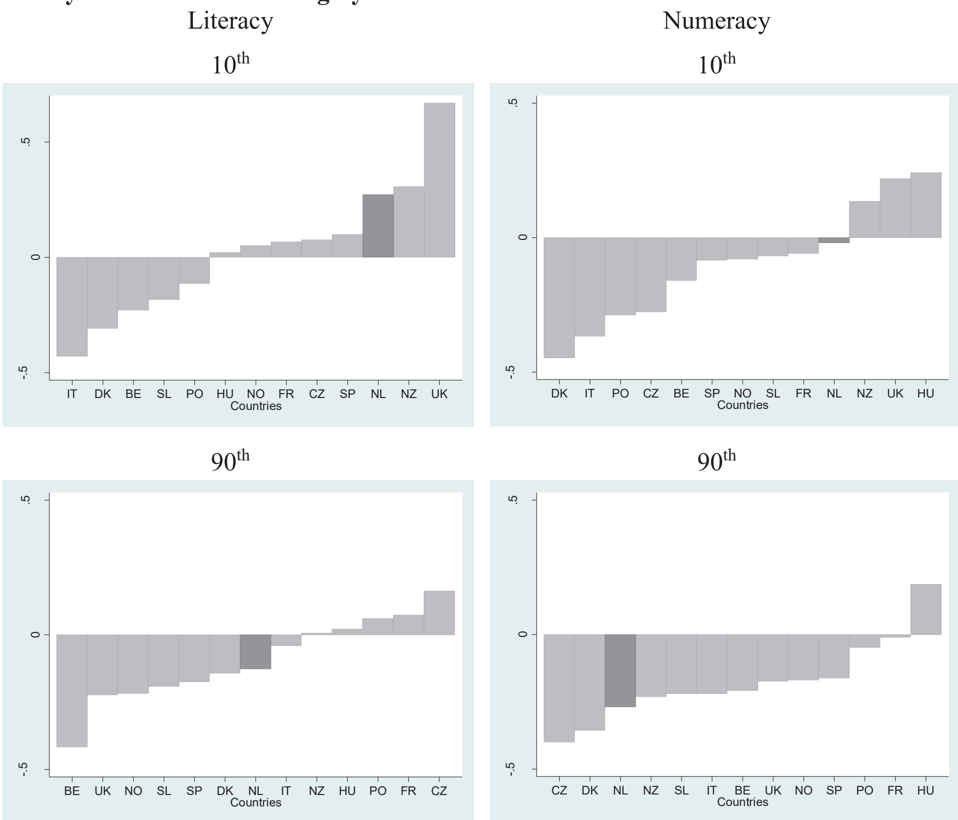
Figure 5.5: Mean skill difference of teachers vs. tertiary educated subsample per country, controlling for age and gender



Note: bars indicate the mean test score difference in standard deviations between teachers and non-teachers for each country on the horizontal axis. The error bars indicate the 95% confidence interval. Differences were calculated using full sample weights using only the tertiary educated subsample and controlling for age and gender.

Figure 5.5 compares the average scores between teachers and tertiary educated other respondents. We find mixed results. Primary school teachers in the UK significantly outperform the average tertiary educated respondent on both skill measures. Only in Denmark do primary school teachers score significantly below average. This suggests that in most countries we observe, teachers are not recruited from the lower part of the college graduate skill distribution. Our findings contrast with earlier results from studies in the United States that showed that the average teacher is less skilled than the average college graduate (e.g., Hanushek & Pace, 1995; Bacolod, 2007). Secondary school teachers in New Zealand, the UK, Norway and Spain outperform the average tertiary educated respondent on both skill measures, and secondary school teachers do not score significantly lower than others on either skill measure in any country.

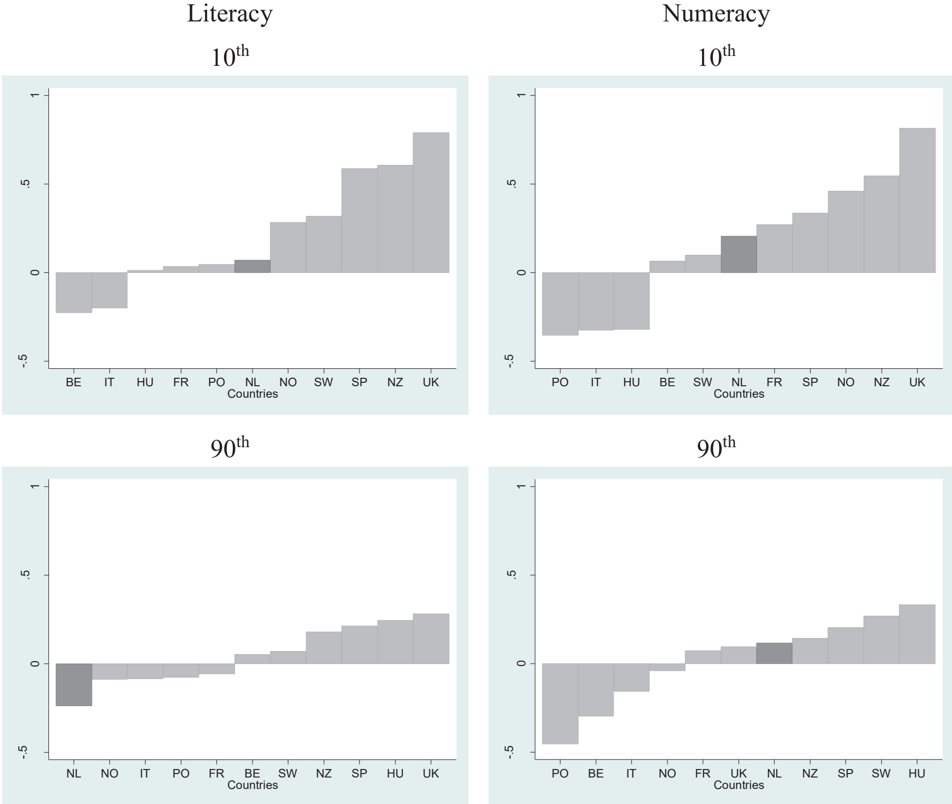
Figure 5.6: Differences in literacy and numeracy scores at the 10th and 90th percentile: primary school teachers vs. highly educated



Note: bars indicate the test score difference at the 10th and 90th percentile in standard deviations between primary school teachers and tertiary educated non-teachers for each country on the horizontal axis. Differences were calculated using full sample weights and controlling for age and gender.

It is also instructive to look at the difference in the distribution of skills between teachers and non-teachers who have finished tertiary education. The graphs depicting these analyses can be found in figures 5.6 and 5.7. The combined findings in figures 5.5, 5.6 and 5.7 suggest that the average teacher is approximately as skilled as the average college graduate. The differences in average test scores we observe in some countries are mainly due to teachers outperforming other graduates in the bottom of the distribution, while in the top of the distribution, (especially secondary school) teachers keep up with the best scoring other graduates.

Figure 5.7: Differences in literacy and numeracy scores at the 10^{th} and 90^{th} percentile: secondary school teachers vs. highly educated



Note: bars indicate the test score difference at the 10^{th} and 90^{th} percentile in standard deviations between secondary school teachers and tertiary educated non-teachers for each country on the horizontal axis. Differences were calculated using full sample weights and controlling for age and gender.

5.6.2 Controlling for differences in skill use

One potential concern with our results is that literacy and numeracy might be skills that sustain or improve because they are used at work. The nature of their profession may make it more likely that teachers use these skills more often than others. This could mean that their higher performance relative to the rest of the population is driven by experience rather than innate ability. PIAAC and ALL allow us to check for this alternative explanation. Both surveys ask respondents a variety of questions regarding the frequency of their literacy and numeracy skill use in the workplace. Table 5.6 shows the correlations of the differences at the 10th and 90th percentile for numeracy and literacy between the original results and the results when controlling for frequency of skill use, age and gender. This allows us to see whether the results are stable across the different specifications.

Table 5.6: Correlations of the differences in skills between teachers and non-teachers at the 10th and 90th percentiles between the original results and the results when controlling for skill use, age and gender.

Primary School Teachers	Pct. 10 Literacy with controls	Pct. 90 Literacy with controls	Pct. 10 Numeracy with controls	Pct. 90 Numeracy with controls
Pct. 10 Literacy	0.57*	-0.06	0.53	-0.07
Pct. 90 Literacy	-0.26	0.78**	-0.07	0.74**
Pct. 10 Numeracy	0.42	-0.01	0.72**	0.07
Pct. 90 Numeracy	-0.03	0.41	0.08	0.68*

Secondary School Teachers	Pct. 10 Literacy with controls	Pct. 90 Literacy with controls	Pct. 10 Numeracy with controls	Pct. 90 Numeracy with controls
Pct. 10 Literacy	0.75**	-0.63*	0.55	-0.27
Pct. 90 Literacy	-0.15	0.59	-0.19	0.23
Pct. 10 Numeracy	0.86**	-0.66*	0.89**	-0.17
Pct. 90 Numeracy	-0.03	0.30	0.28	0.91**

Note: The table shows the correlations of the order of the countries in specifications with and without controls at various percentiles.
* = $p < .05$; **= $p < .01$

The table shows that the differences are highly correlated across both specifications. For primary school teachers, the order of the countries is very similar for the 10th percentile of both skill measures and the 90th percentile of literacy. The 90th percentile for numeracy differs slightly between the two specifications. For secondary school teachers, the order of the countries is very similar for all of our outcomes of interest. In absolute terms, controlling for skill use, age and gender decreases the difference between teachers and the rest of the population mainly in the lower part of the distribution. Teachers still outperform the other respondents across the main part of the distribution. These results can be seen in figures A.5.2, A.5.3 and A.5.4 of the appendix.

Note that skill use and raw ability are most likely not independent. Skilled people probably sort into jobs in which they can utilize their talents. In addition, highly educated people tend to sort into professions where literacy and numeracy skills are more likely to be needed. To show that skill use is not merely a proxy for educational attainment, we also created correlation tables comparing teachers to the tertiary educated subset of the population controlling for frequency of skill use. These results can be found in table 5.7. The results are qualitatively similar as the baseline results: the differences between teachers and other college graduates do not appear to be driven by differences in use of literacy and numeracy skills in the workplace. The absolute differences between teachers and other tertiary educated respondents when controlling for skill use can be seen in figures A.5.5, A.5.6 and A.5.7 of the appendix. Controlling for skill use increases the performance of both primary and secondary school teachers relative to others in numeracy, but decreases it in literacy across the distribution. This could be an indication that teachers use their literacy skills relatively more often than their numeracy skills compared to the rest of the working age population.

Table 5.7: Correlations of the differences in skills between teachers and tertiary educated non-teachers at the 10th and 90th percentiles between the original results and the results when controlling for skill use, age and gender.

Primary School Teachers	Pct. 10 Literacy with controls	Pct. 90 Literacy with controls	Pct. 10 Numeracy with controls	Pct. 90 Numeracy with controls
Pct. 10 Literacy	0.58*	-0.30	0.49	-0.27
Pct. 90 Literacy	-0.10	0.67*	-0.24	0.40
Pct. 10 Numeracy	0.79**	-0.10	0.73**	-0.19
Pct. 90 Numeracy	0.29	0.31	0.16	0.49

Secondary School Teachers	Pct. 10 Literacy with controls	Pct. 90 Literacy with controls	Pct. 10 Numeracy with controls	Pct. 90 Numeracy with controls
Pct. 10 Literacy	0.89**	-0.23	0.70*	0.14
Pct. 90 Literacy	0.60*	0.52	0.38	0.27
Pct. 10 Numeracy	0.81**	-0.49	0.87**	-0.01
Pct. 90 Numeracy	0.48	0.15	0.59	0.91**

Note: the table shows the correlations of the order of the countries in specifications with and without controls at various percentiles.
* = p < .05; **= p < .01

5.7 Conclusions and policy implications

This chapter investigates teachers' relative literacy and numeracy skills in 15 countries, using the OECD's PIAAC and ALL datasets. We analyze differences in average skills, and focus additionally on differences in the distribution of skills. Our main conclusions are that the skill distributions differ between countries and that these distributional differences are more informative for policy makers than average teacher skills.

We find that in virtually all countries, teachers are more highly skilled than the average respondent. Looking at the differences in the skills distributions, we find that in most countries teachers outperform the average respondent mainly in the lower percentiles of the distribution. The lowest scoring teachers perform better than the lowest scoring other respondents. These findings hold when restricting the analyses to the tertiary educated subsample and controlling for skill use, age and gender.

The extent to which teachers outperform other respondents in the lower parts of the distribution varies between countries. This variation is informative for policy makers aiming to increase teacher quality. For instance, interventions focusing on the bottom part of the distribution will be less effective in countries in which the positive difference in the lower parts of the distribution between teachers and others is already large. Such interventions can be more effective in countries where these differences are close to zero or negative.

Considering that earnings for teachers compared to potential earnings in the private sector are relatively low (Stinebrickner, 2001), the finding that the best teachers are not outperformed by the best non-teachers is a result that suggests that non-pecuniary factors may motivate these highly skilled people to enter the teaching profession. For policy makers, it is relevant to investigate what drives these highly skilled individuals to become a teacher. If some of these factors can be identified, attempting to select on these factors may prove a cost-effective way to attract more highly skilled individuals into the teaching corps. This could potentially allow policy makers to increase average teacher quality without having to resort to costly salary increases.

While our policy recommendations are based on teachers' relative skills, other facets need to be taken into account when deciding on a specific intervention. For example, costs and feasibility may differ vastly between interventions. A policy intervention that may be optimal from a skills distribution perspective may not be optimal in terms of other factors. Increasing barriers to entry may for example lead to teacher shortages, while increasing pay for high skilled teachers may create inequality between teachers and decrease overall teacher job satisfaction. Nevertheless, the shape of the current teachers' skills distribution is important information to consider in order to accurately predict the costs and benefits of any policy intervention.

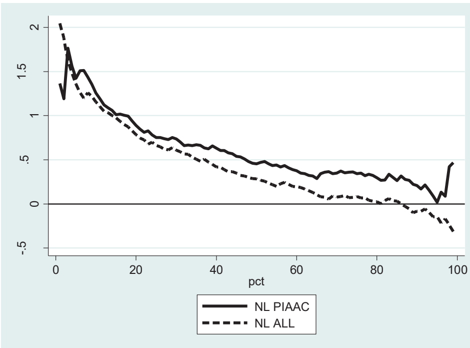
The substantial variation in the skill dispersion across the different countries leads to the conclusion that different interventions are likely to be optimal in different countries. In countries where the teachers with the lowest skills do not outperform the low skilled other college graduates (e.g., Denmark), interventions like increased barriers to entry might be more cost-effective than trying to aim at the top of the distribution. However, in countries where the lowest skilled teachers are already performing relatively well there is less scope for improvement at the bottom and shifting focus to the top end of the distribution might be more efficient. Overall, we conclude that it is important to take the shape of the skills distribution into account when designing policies aimed at improving teacher skills.

5.8 Appendix -- Additional results

Figure A.5.1: Comparing the difference in distributions for PIAAC and ALL for the Netherlands

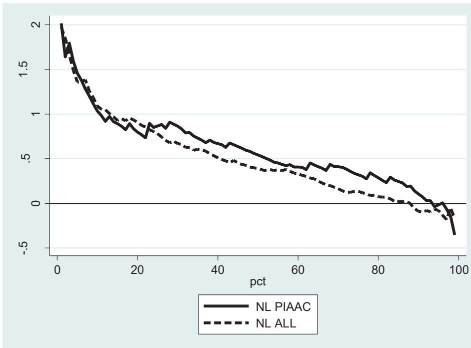
Primary school teachers

Literacy

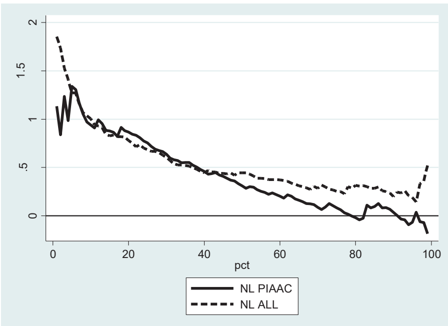


Secondary school teachers

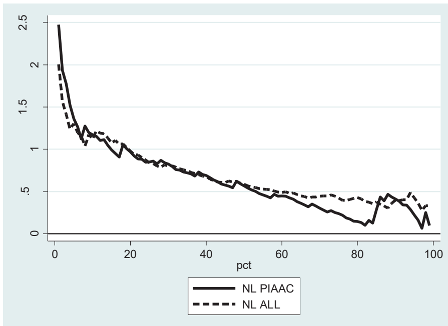
Literacy



Numeracy



Numeracy

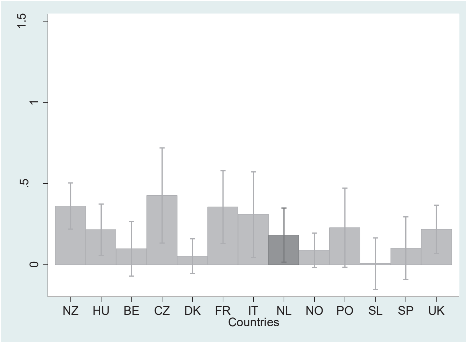


Note: percentile graphs depicting the difference in the distribution of skills of teachers and all non- teachers for the Netherlands (the graphs for Italy and Norway are available upon request). The uninterrupted (interrupted) line depicts the scores in PIAAC (ALL). All differences are in standard deviations from the full sample standardized test scores. Results were calculated using full sample weights.

Figure A.5.2: Regression adjusted mean skill difference of teachers vs. other respondents

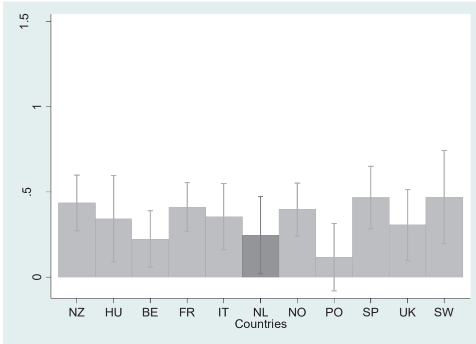
Primary school teachers

Literacy

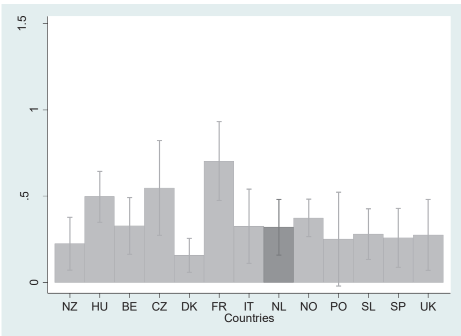


Secondary school teachers

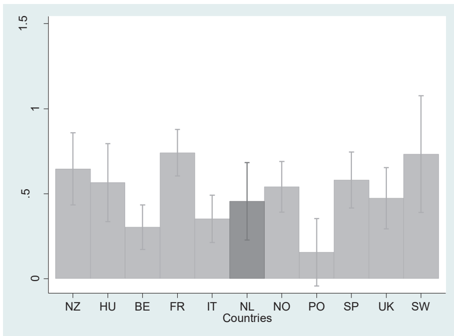
Literacy



Numeracy

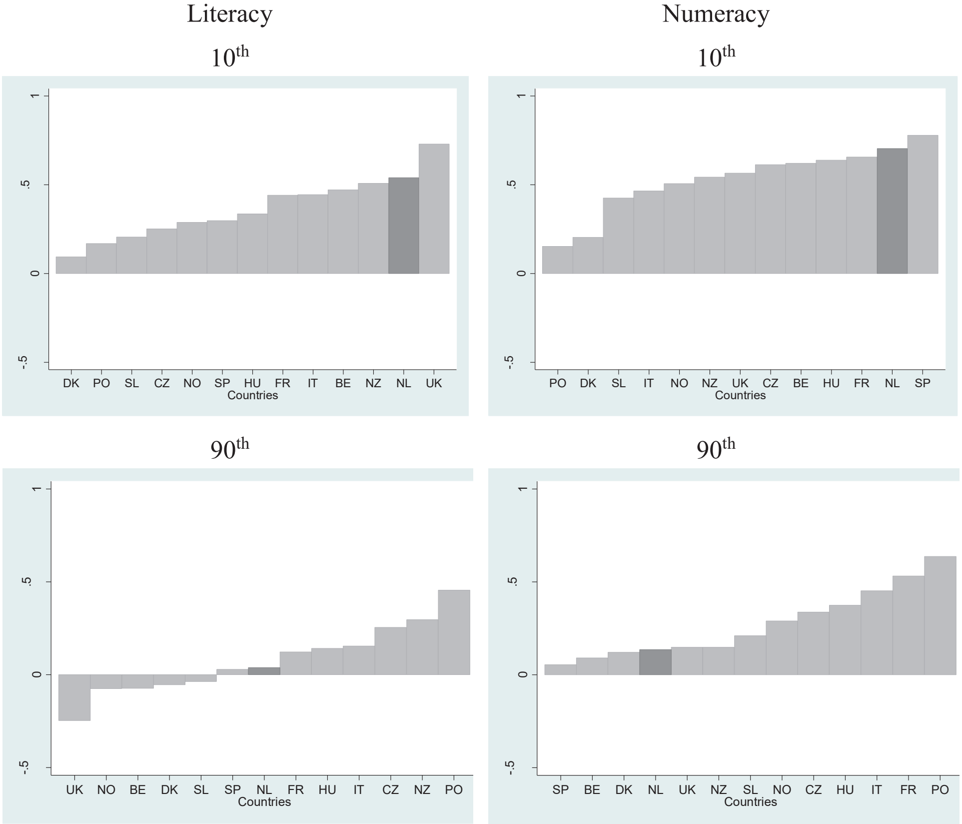


Numeracy



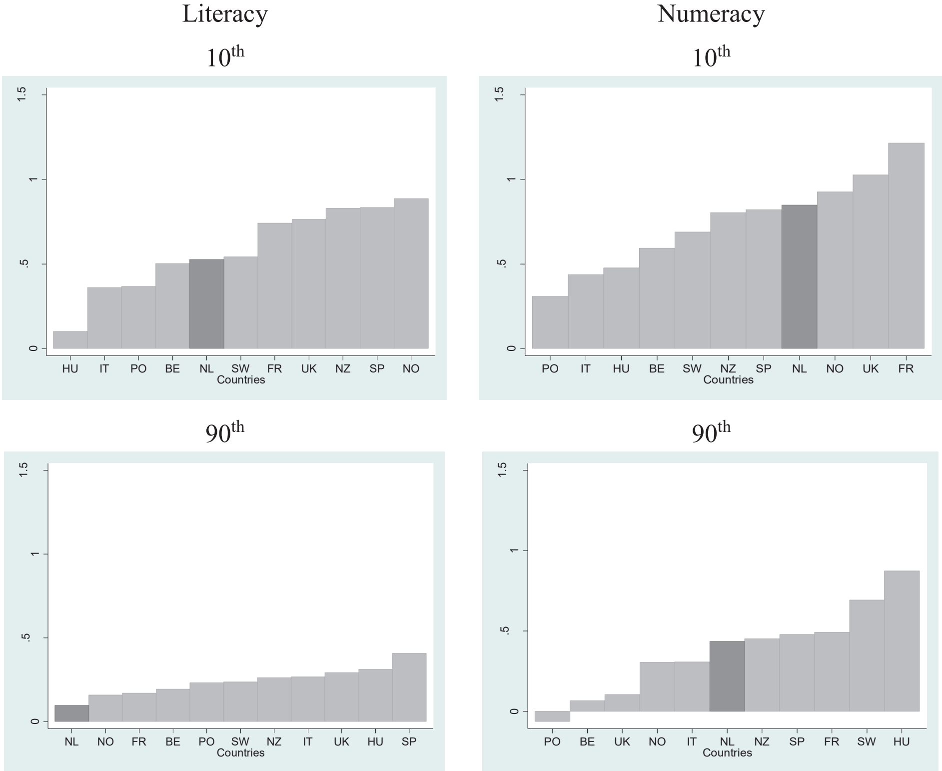
Note: bars indicate the mean test score difference in standard deviations between teachers and non-teachers for each country on the horizontal axis adjusted for frequency of skill use. The error bars indicate the 95% confidence interval. Differences were calculated using full sample weights.

Figure A.5.3: Regression adjusted differences at the 10th and 90th percentile of primary school teachers vs. non-teachers



Note: bars indicate the test score difference at the 10th and 90th percentile in standard deviations between primary school teachers and non-teachers for each country on the horizontal axis adjusted for frequency of skill use in the workplace. Differences were calculated using full sample weights.

Figure A.5.4: Regression adjusted differences at the 10th and 90th percentile of secondary school teachers vs. non-teachers



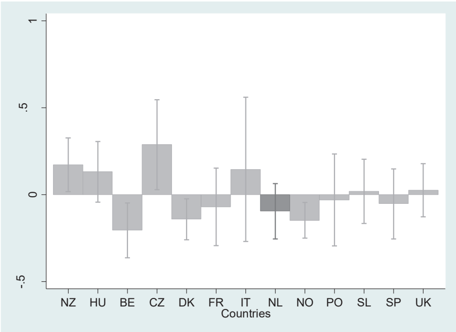
Note: bars indicate the test score difference at the 10th and 90th percentile in standard deviations between secondary school teachers and non-teachers for each country on the horizontal axis adjusted for frequency of skill use in the workplace. Differences were calculated using full sample weights.

Figure A.5.5: Regression adjusted mean skill difference teachers vs. tertiary educated other respondents

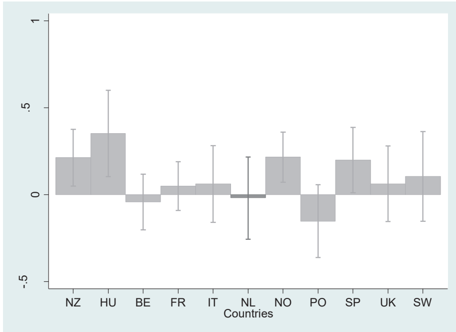
Primary school teachers

Secondary school teachers

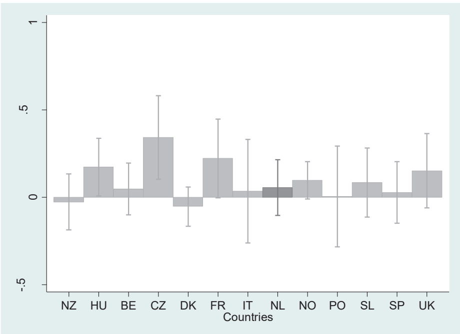
Literacy



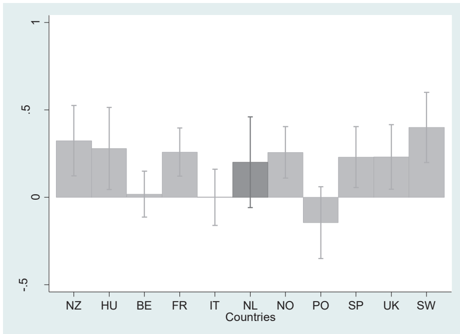
Literacy



Numeracy

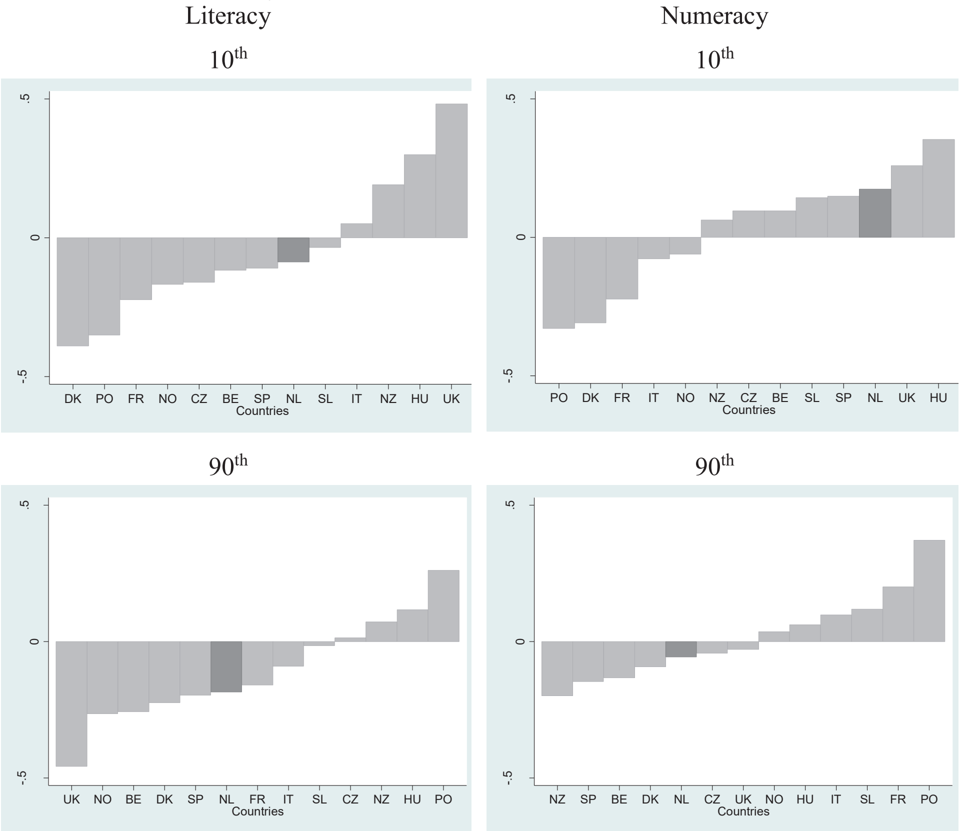


Numeracy



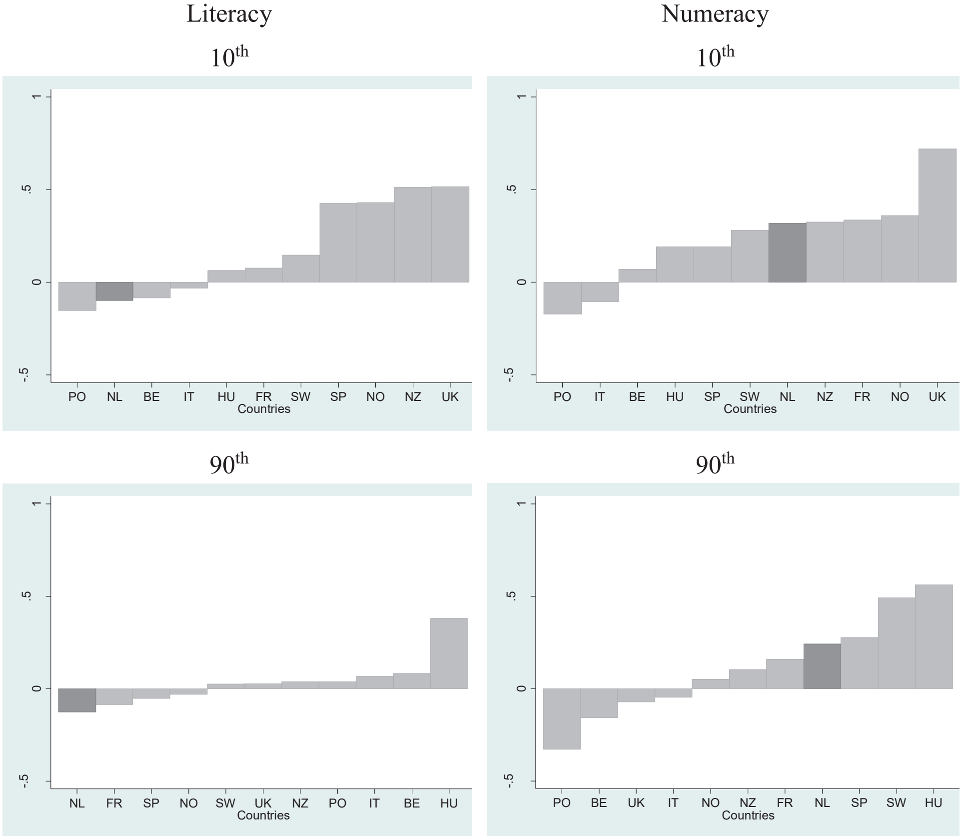
Note: bars indicate the mean test score difference in standard deviations between teachers and tertiary educated non-teachers for each country on the horizontal axis adjusted for frequency of skill use and controlling for age and gender. The error bars indicate the 95% confidence interval. Differences were calculated using full sample weights.

Figure A.5.6: Regression adjusted differences at the 10th and 90th percentile of primary school teachers vs. tertiary educated non-teachers



Note: bars indicate the test score difference at the 10th and 90th percentile in standard deviations between primary school teachers and tertiary educated non-teachers for each country on the horizontal axis adjusted for frequency of skill use in the workplace. Differences were calculated using full sample weights and controlling for age and gender.

Figure A.5.7: Regression adjusted differences at the 10th and 90th percentile of secondary school teachers vs. tertiary educated non-teachers



Note: bars indicate the test score difference at the 10th and 90th percentile in standard deviations between secondary school teachers and tertiary educated non-teachers for each country on the horizontal axis adjusted for frequency of skill use in the workplace. Differences were calculated using full sample weights and controlling for age and gender.

6. Summary, implications and valorization

6.1 Summary of the main findings:

The aim of this thesis has been to investigate the career and skill development of Dutch teachers. Large administrative datasets, theoretical modelling, and international comparative skills surveys were used in four self-contained chapters, each addressing a different facet of Dutch teachers' careers and skills. Insights from the results of this thesis are relevant for policy makers aiming to ensure an adequate supply of high quality teachers, which due to the impending teacher shortage has recently taken a prominent place on the Dutch political agenda. In this section, the main results of each chapter are briefly summarized, followed by a discussion of the implications of these results in informing teacher policies.

Chapter 2 maps the early career labour market outcomes of Dutch primary school teacher training graduates, and investigates the relationship between these outcomes and student dropout rates at the teacher training program level. It finds that graduates from programs with a low dropout rate are equally likely to work in the education sector as those graduating from high dropout rate programs. However, low dropout rate program graduates supply 2.5% fewer hours of teaching on average after several years on the labour market. Nevertheless, the larger proportion of students that graduates from low dropout rate programs more than offsets the lower average amount of hours of teaching supplied by these graduates. A second finding of this chapter is that primary school teacher training graduates face worse labour market outcomes when they switch out of teaching and start working outside of the education sector, at least for the first nine years after graduation. Those that stay in the education sector earn a 10% higher hourly wage, and are 50% more likely to be employed on a permanent contract, while working a similar amount of hours per month as their non-teaching classmates.

Chapter 3 looks at the distribution of Dutch primary school teachers across schools, and investigates how assortative matching between teacher and student characteristics relates to student test scores. It finds evidence of strong sorting on migrant background and moderate sorting on educational background. Schools with a larger share of non-western migrant students employ a larger share of non-western migrant teachers. At schools with a large percentage of students with highly educated parents, the share of teachers holding a master's degree is higher. Assortative matching is more pronounced in highly urbanized areas, and the sorting patterns of early career teachers reinforce, rather than weaken, the existing distribution. Concerning student test scores, this chapter finds no evidence of negative effects of the unequal distribution of teachers. The share of teachers with a master's degree has no impact on student test scores, nor are there differential effects along the dimension of students' parental education. The share of non-western migrant teachers relates positively to test scores for non-western migrant students, with no negative consequences for native students. The gains are concentrated in the math part of the test, and are more pronounced for migrant students from a relatively low socio-economic background.

Chapter 4 models the decision of teachers to sign up for professional development programs. It shows that as teacher quality is (partly) unobservable, training participation can be used to convey a signal about a teacher's initial ability. As a result, courses that are aimed at low quality teachers will be unpopular, while courses aimed at high quality teachers will be oversubscribed. Results from the model show that offering training has spillover effects, and that the availability of different types of courses weakens the signal of each course individually.

Finally, chapter 5 maps the literacy and numeracy skills of primary and secondary school teachers relative to the rest of the population for 15 developed countries. It shows that the 10 percent lowest skilled Dutch teachers outperform the 10 percent lowest skilled college educated population, while the average teacher is about equally skilled as the average college graduate. At the top, the 10 percent highest skilled college graduates outperform the 10 percent highest skilled teachers. Internationally, Dutch teachers perform well on average and are relatively highly skilled at the bottom, while performing relatively poorly at the top.

6.2 Implications and valorization

The results have several implications for policies aimed at reducing teacher shortages and increasing the average quality of teachers in the Netherlands. Chapters 2 and 3 are geared towards addressing the quantitative teacher shortage, while chapters 4 and 5 are aimed at addressing teacher quality.

Chapter 2

First, based on the results of chapter 2 it can be argued that the majority of teaching potential is wasted. Because of student dropout, the fact that a proportion of teaching graduates starts working outside of the education sector, and the propensity of graduates to work part-time, an average sized teacher preparation program entry cohort of 400 students supplies only around 122 FTE of teaching per year. Put another way, every 10 additional students sorting into primary school teacher education reduce teacher shortages by only three FTE. Therefore, in order to eliminate the primary teacher shortage in the Netherlands of 7000 FTE by 2025

(Adriaens et al., 2017) through increased student numbers, the amount of students entering primary education teacher training has to more than double.

Alternatively, if all students that sorted into primary teacher education would end up as full-time teachers, the teacher shortage would be solved within three average sized entry cohorts. The focus of policy makers on increasing entry into teacher education therefore seems inefficient. Rather than convincing prospective students to go into teacher education, additional efforts should be made to reduce student dropout from teacher training programs. By initially signing up for teacher education, these students have already expressed a preference for teaching. This makes interventions aimed at this particular group more likely to be effective in increasing the supply of teachers. However, while chapter 2 shows that low dropout rates at the program level are observationally unrelated to the propensity to work in the education sector, it is important that interventions targeting dropout rates do not come at the expense of graduation quality standards.

Secondly, a promising avenue to increase the supply of teachers is to stimulate full-time employment among teachers. Even if attrition from teacher training cannot be reduced, if all teacher primary education teachers would start working full-time, the amount of FTE supplied would increase by around 30%. This is more than enough to solve the prospective teacher shortage. A potential path to stimulating full-time employment could be to increase teachers' salaries for additional hours worked above a certain part-time threshold, as has been suggested by Kalshoven (2017). This might stimulate teachers to start working additional hours, while being relatively cheaper than raising teacher salaries across the board. Chapter 2 further shows that teachers' labour market opportunities are worse outside of teaching. Therefore, a salary increase across the board is unlikely to result in additional teacher supply from incumbent teachers beyond the increase that could be expected from more highly rewarding additional hours worked above the part-time threshold.

Chapter 3

While interventions to increase teacher supply are needed, one can argue that policies influencing the sorting patterns of teachers are not necessary based on the results from chapter 3. While teachers are unequally distributed across primary schools regarding their level of education and migration background, this does not negatively affect student performance. However, other research has shown that schools that serve a high percentage of students with a non-western migration background have more trouble filling their teaching vacancies (Inspectorate of Education, 2019). Therefore, while the sorting patterns of teachers may not be problematic in itself, the unequal distribution of the teacher shortage warrants further attention. In that vein, it is interesting to note that a larger share of teachers with a non-western migration background appears to be related to increased performance of non-western migrant students. Furthermore, teachers from a non-western migration background tend to work at schools with a larger proportion of non-western migrant students.

In order to reduce the inequality of the distribution of the teacher shortage and increase the performance of non-western migrant students, policy makers should spend more effort in enthusing non-western migrant students for the primary school teaching profession. Note that the assumption underlying this recommendation is that the mechanism through which migrant students benefit from a larger share of non-western migrant teachers is related to attributes of non-western migrant teachers that are unique to them. Examples of such mechanisms include role model effects or a shared cultural belief system. If instead non-western migrant teachers benefit migrant students through a certain imitable teaching approach, this method could in principle be applied by teachers regardless of migration background. Therefore, further research into the mechanism through which a match on migration background positively influences test scores is needed.

If policy makers do want to increase the amount of non-western migrant students entering the primary school teaching profession, current policy measures appear to be counterproductive. After the introduction of the entry tests as a requirement to start primary teacher education in 2015, the amount of students with a non-western migration background entering primary teacher education halved (de Wolf, Vermeulen, & Breuer, 2018), although this finding is not necessarily causal. While the entry tests have been designed to raise the quality of the students sorting into primary teacher education, an unintended consequence appears to have been a reduced number of non-western migrant students entering the primary school teaching profession.

Chapter 4

In terms of improving the quality of the incumbent teaching force through on-the-job training and professional development activities, chapter 4 suggests that the nature of the teaching profession renders effective targeting of training difficult. Because unobservable performance introduces signalling concerns, low ability teachers are likely to avoid signing up for training that is geared towards teachers with their prior ability level. It also suggests that training that is aimed at high quality teachers is likely to be oversubscribed. This might explain why professional development is widely used, but rarely evaluated as being effective. Even if a course is useful in raising the quality of a certain subset of teachers, the signup of teachers for whom the course is of no added value attenuates the estimates of the course's effectiveness.

In order to increase the efficiency of training participation, the model predicts that introducing more courses targeted at different levels of initial teacher ability can reduce the signalling value of signing up for any particular course. This would increase the take up rate of courses aimed at low ability teachers. If policy makers are concerned with increasing the quality of the lowest ability teachers, they could alternatively make certain professional development programs targeting basic skills mandatory for all teachers. This would eliminate the negative signal of basic training participation. However, this option is not ideal as the basic training is wasted on those teachers that already master these skills.

It should be kept in mind that these recommendations are based on results from a theoretical model, the validity of which needs to be tested in an empirical setting. However, the point that selection into on-the-job training needs to be taken into account when designing, offering, and evaluating professional development programs stands, regardless of whether signalling is the main mechanism explaining the nature of selection.

Chapter 5

Taking stock of the literacy and numeracy skills of Dutch teachers relative to the rest of the population and compared to other countries, chapter 5 shows that Dutch teachers are relatively highly skilled at the bottom of the skills distribution. This means that the 10 percent lowest skilled teachers perform better than the same group in other countries relative to the rest of the population. The 10 percent highest skilled teachers perform lower than the top teachers in other countries, again relative to the rest of the population. What this implies in terms of policies aimed at increasing the skills of the Dutch teacher corps is that interventions aimed at raising the floor of teachers' cognitive skills might not be most effective. Relatively low skilled teachers already outperform other low cognitively skilled college graduates. Therefore, becoming more restrictive by raising skills barriers to entry into the profession is likely to result in marginal gains in the average quality at the expense of reducing the quantity of teachers. Considering that the most highly skilled Dutch teachers are outperformed by high skilled other college graduates, interventions that aim to raise the ceiling of the skills of the Dutch teacher corps seem more promising. Attracting highly skilled students to enter teaching could not only increase the average quality of the Dutch teacher corps directly, but also help in increasing the status of the profession in the long term (Cörvers et al., 2017).

In aiming to recruit teachers from the high end of the ability distribution, it should be kept in mind that while cognitive skills are somewhat predictive of teacher quality (Coenen et al., 2018; Hanushek et al, 2018; Metzler & Woessmann, 2012), the relationship is unlikely to be linear. It makes sense that a teacher who is uncomfortable performing long division is unlikely to competently teach long division. However, it is unclear whether much can be gained from employing a Nobel Prize winning mathematician instead of a regular teacher holding a math's major. It is still an open question how desirable it is for teachers to be drawn from the absolute top of the skills distribution, and whether this is the most efficient allocation of skills from a societal point of view. Therefore, further research should address the shape of the relationship between teacher subject knowledge and student performance.

In conclusion, this thesis has aimed to provide a better understanding of the development of Dutch teacher careers and skills. This new information suggests that several avenues for policies that are currently being pursued may be suboptimal. Focussing on increasing the number of students in teacher education is inefficient if student attrition is not addressed simultaneously. Furthermore, the policies aimed to raise the floor of the cognitive skills of the primary teacher corps are likely to result in larger quantitative shortages with small gains in average teacher quality. Additionally, an unintended consequence of these policies seems to be a reduced amount of non-western migrant students entering the teaching profession. Instead, policies aimed at increasing the amount of highly skilled teachers seem more promising in the Dutch context. Secondly, it has shown that policies that are designed to target incumbent low quality teachers through professional development programs should take the potential signalling value of professional development programs into account when offering courses. Finally, the results of this thesis have suggested focal points for policy interventions in reducing student attrition from teacher training, increased efforts to enthuse non-western migrants for the teaching profession, and stimulating full-time employment of incumbent teachers.

Bibliography

Aa, van der, R., Cörvers, F., & Schoon, R. (2017), *Met een blik op het verleden, Een inventarisatie van eerdere projecten en maatregelen ter vermindering van lerarentekorten in het po, vo en mbo*, CAOP, Den Haag.

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

Acemoglu, D. (1997). Training and innovation in an imperfect labour market. *Review of Economic Studies*, *64,* 445–464.

Acemoglu, D. and Pischke, J.-S. (1998). Why do firms train? Theory and evidence. *Quarterly Journal of Economics*, *113*(1), 79–119.

Acemoglu, D. and Pischke, J.-S. (1999a). Beyond Becker: Training in imperfect labor markets. *Economic Journal*, 109, F112–F142.

Acemoglu, D. and Pischke, J.-S. (1999b). The structure of wages and investment in general training. *Journal of Political Economy*, *107*(3), 539–572.

Adriaens, H., Fontein, P., Uijl, den, M., Vos, de, K. (2017). *De toekomstige arbeidsmarkt voor onderwijspersoneel po, vo en mbo 2017-2027.* Tilburg: CentERdata

Altonji, J. G., & Pierret, C. R. (2001). Employer learning and statistical discrimination. *The Quarterly Journal of Economics*, *116*(1), 313-350.

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

Autor, D. H. (2001). Why do temporary help firms provide free general skills training?. *Quarterly Journal of Economics*, 1409-1448.

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science (New York, NY)*, *344*(6186), 843-851.

Bacolod, M. (2007). Who teaches and where they choose to teach: college graduates of the 1990s. *Educational Evaluation and Policy Analysis* 29 (3), 155-168.

Barber, M., & Mourshed, M. (2007). *How the world's best performing systems came out on top.* London: McKinsey & Co.

Bassanini, A., Booth, A., Brunello, G., De Paola, M. & Leuven, E. (2007). Workplace training in Europe, in *Education and Training in Europe* (Eds) G. Brunello, P. Garibaldi and E. Wasmer, Oxford University Press, Oxford, pp. 143–309.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, *70*(5, Part 2), 9-49.

Blank, R. K., & De Las Alas, N. (2009). *The effects of teacher professional development on gains in student achievement: How meta analysis provides scientific evidence useful to education leaders*. Council of Chief State School Officers. Washington, DC.

Borghans, L., Golsteyn, B. H.H., & Zölitz, U. (2015). Parental preferences for primary school characteristics. *The BE Journal of Economic Analysis & Policy*, *15*(1), 85-117.

Borman, G. D., & Maritza-Dowling, N. (2008). Teacher attrition and retention: A meta-analysis and narrative review of the research. *Review of Educational Research, 78*, 367–409.

Boterman, W. (2018*). Segregatie in het Nederlandse onderwijs*. (Unpublished Manuscript)

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management 27*(4): 793–818.

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

Brok, den, P., Wubbels, T., & Tartwijk, van, J. (2017). Exploring beginning teachers' attrition in the Netherlands. *Teachers and Teaching*, *23*(8), 881-895.

Caena, F. (2013). Supporting teacher competence development for better learning outcomes. *Education & Training, European Commission*, 5-59.

Card, D., & Rothstein, J. (2007). Racial segregation and the black–white test score gap. *Journal of Public Economics*, *91*(11-12), 2158-2184.

Carrillo, C., Maassen van den Brink, H., & Groot, W. (2016). *Professional development programs and their effects on student achievement: A systematic review of evidence* (No. 16/03). TIER Working Paper.

Centraal Planbureau (2017). *Paden naar leraarschap*. Den Haag: Centraal Planbureau

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633-79.

Chevalier, A., Dolton, P., & McIntosh, S. (2007). Recruiting and retaining teachers in the UK: an analysis of graduate occupation choice from the 1960s to the 1990s. *Economica* 74 (293), 69-96.

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2006) Teacher-student matching and the assessment of teacher effectiveness. The Journal of Human Resources *41*(4): 778–820.

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

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2010) Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources 45*(3): 655– 681

Coenen, J., Cornelisz, I., Groot, W., Maassen van den Brink, H., & Klaveren, van, C. (2018). Teacher Characteristics And Their Effects On Student Test Scores: A Systematic Review. *Journal of Economic Surveys*.

Coenen, J. & Klaveren, van, C. (2016) Better test scores with a same-gender teacher. *European Sociological Review, 32*(3), 452–464.

Cörvers, F., Mommers, A., Ploeg, van der, S., & Sapulete, S. (2017). *Status en imago van de leraar in de 21ste eeuw*. Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA), Maastricht en Ecorys, Rotterdam

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, *86*(1), 195-210. 121

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter?. *American Economic Review*, *95*(2), 158-165.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human resources*, *42*(3), 528-554.

Delfgaauw, J., & Dur, R. (2008). Incentives and workers' motivation in the public sector. *The Economic Journal*, *118*(525), 171-191.

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*, 696-727.

Education Council (2013). *Kiezen voor kwalitatief sterke leraren*. Den Haag: Onderwijsraad.

Ehrenberg, R.G., Brewer, D., & Goldhaber, D. (1995) Do teachers' race, gender, and ethnicity matter? Evidence from the NELS. *Industrial and Labor Relations Review 48*(3): 547–561.

Eide, E., Goldhaber, D., & Brewer, D. (2004). The teacher labour market and teacher quality. *Oxford Review of Economic Policy* 20 (2), 230-244.

Evers, A. T., Heijden, van der, B. I., Kreijns, K., & Gerrichhauzen, J. T. (2011). Organisational factors and teachers' professional development in Dutch secondary schools. *Journal of European Industrial Training*, *35*(1), 24-44.

Fairlie, R. W., Hoffmann, F., & Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review*, *104*(8), 2567-91.

Fouarge, D., Schils, T., & Grip, A. de (2013). Why do low-educated workers invest less in further training? *Applied Economics*, *45*(18), 2587-2601.

Fontein, P., Kieruj, N., Uijl, den, M., Vos, de, K., Boom, van der, E., & Vrielink, S. (2016). *Loopbaanmonitor onderwijs*. Tilburg: CentERdata

Frazis, H., & Loewenstein, M. A. (2005). Reexamining the returns to training functional form, magnitude, and interpretation. *Journal of Human Resources*, *40*(2), 453-476.

Fredriksson, P., Öckert, B., & Oosterbeek, H. (2012). Long-term effects of class size. *The Quarterly Journal of Economics*, *128*(1), 249-285.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., ... & Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement.* NCEE 2008-4030.National Center for Education Evaluation and Regional Assistance.

Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Walters, K., Song, M., ... & Doolittle, F. (2010). *Middle school mathematics professional development impact study: Findings after the first year of implementation.* NCEE 2010-4009. National Center for Education Evaluation and Regional Assistance.

Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., ... & Sepanik, S. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation.* NCEE 2011-4024. National Center for Education Evaluation and Regional Assistance.

Gershenson, S., Hart, C., Hyman, J., Lindsay, C., & Papageorge, N. W. (2018). *The long-run impacts of same-race teachers* (No. w25254). National Bureau of Economic Research.

Gibbons, R., & Murphy, K. J. (1992). Optimal incentive contracts in the presence of career concerns: Theory and evidence. *The Journal of Political Economy*, *100*(3), 468-505.

Goldhaber, D., Choi, H. J., & Cramer, L. (2007). A descriptive analysis of the distribution of NBPTS-certified teachers in North Carolina. *Economics of Education Review*, *26*(2), 160-172.

Goldhaber, D., & Cowan, J. (2014). Excavating the teacher pipeline: Teacher preparation programs and teacher attrition. *Journal of Teacher Education*, *65*(5), 449-462.

Goldin, C., & Katz, L. F. (2007). *The race between education and technology: the evolution of US educational wage differentials, 1890 to 2005* (No. w12984). National Bureau of Economic Research.

Gong, J., Lu, Y., & Song, H. (2018). The Effect of Teacher Gender on Students' Academic and Noncognitive Outcomes. *Journal of Labor Economics*, *36*(3), 743-778.

Haelermans, C., & Borghans, L. (2012). Wage Effects of On-the-Job Training: A Meta-Analysis. *British Journal of Industrial Relations*, *50*(3), 502-528.

Hansson, B. (2008). Job-Related Training and Benefits for Individuals: A Review of Evidence and Explanations. OECD Education Working Papers, No. 19. OECD Publishing (NJ1).

Hanushek, E. A. (1992). The Trade-off between Child Quantity and Quality. *The Journal of Political Economy,* 100(1), 84-11.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal, 113* (485), F64-F98.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*(3), 466-479.

Hanushek, E. A., & Pace, R. R. (1995). Who chooses to teach (and why)? *Economics of Education Review, 14*(2), 101-117.

Hanushek, E. A., Piopiunik, M., & Wiederhold, S. (2018). The Value of Smarter Teachers: International Evidence on Teacher Cognitive Skills and Student Performance. *Journal of Human Resources*.

Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. *Handbook of the Economics of Education* 2, 1051-1078.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, *95*(7-8), 798-812.

Hashimoto, M. (1981). Firm-specific human capital as a shared investment. *American Economic Review*, 71(3), 475–482.

Hong, J. Y. (2010). Pre-service and beginning teachers' professional identity and its relation to dropping out of the profession. *Teaching and Teacher Education, 26*, 1530–1543.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal, 38*, 499–534.

Inspectorate of Education (2017*). De staat van het onderwijs. Onderwijsverslag 2015/2016*. Utrecht: Inspectie van het Onderwijs.

Inspectorate of Education (2018*). De staat van het onderwijs. Onderwijsverslag 2016/2017*. Utrecht: Inspectie van het Onderwijs.

Inspectorate of Education (2019*). De staat van het onderwijs. Onderwijsverslag 2017/2018*. Utrecht: Inspectie van het Onderwijs.

Jacob, B. A., & Lefgren, L. (2004). The impact of teacher training on student achievement quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources*, *39*(1), 50-79.

Jacob, A., & McGovern, K. (2015). *The mirage: Confronting the hard truth about our quest for teacher development*. TNTP.

Kalshoven, F. (2017, December 2). Waarom een algehele loonsverhoging de problemen in het onderwijs niet oplost. *De Volkskrant,* retrieved from: https://www.volkskrant.nl/nieuws-achtergrond/waarom-een-algehele-loonsverhoging-de-problemen-in-het-onderwijs-niet-oplost~babb4399/

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615-631.

Kim, L. E., Dar-Nimrod, I., & MacCann, C. (2018). Teacher personality and teacher effectiveness in secondary school: Personality predicts teacher support and student self-efficacy but not academic achievement. *Journal of Educational Psychology*, *110*(3), 309.

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, *113*(485), F34-F63.

Lange, F. (2007). The speed of employer learning. *Journal of Labor Economics*, *25*(1), 1-35.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37-62.

Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review*,*10*(2), 179-214.

Leuven, E. (2005). The economics of private sector training: A survey of the literature. *Journal of Economic Surveys, 19*: 91–111.

Leuven, E., & Oosterbeek, H. (2004). Evaluating the effect of tax deductions on training. *Journal of Labor Economics*, *22*(2), 461-488.

Leuven, E., & Oosterbeek, H. (2008). An alternative approach to estimate the wage returns to private-sector training. *Journal of Applied Econometrics*, *23*(4), 423-434.

Lim, J., & Meer, J. (2017). The impact of teacher-student gender matches: Random assignment evidence from South Korea. *Journal of Human Resources*, 1215-7585R1.

Lim, J., & Meer, J. (2019). Persistent effects of teacher-student gender matches. *Journal of Human Resources*, 0218-9314R4.

Metzler, J., & Woessmann, L. (2012). The impact of teacher subject knowledge on student achievement: evidence from within-teacher within-student variation. *Journal of Development Economics* 99(2), 486-496.

Ministry of Education, Culture and Science (2014). *Rijksbegroting 2015 VIII Onderwijs, Cultuur en Wetenschap* (kst-34000-VIII-2 ISSN 0921 – 7371). 's-Gravenhage: Sdu Uitgevers.

Munoz, M.A., & Chang, F.C. (2007).The elusive relationship between teacher characteristics and student ˜ academic growth: a longitudinal multilevel model for change. *Journal of Personnel Evaluation in Education 20*(3–4): 147–164

NAO (2016). *Training New Teachers.* National Audit Office.

OECD (2014), *Education at a Glance 2014: OECD Indicators*, OECD Publishing, Paris.

OECD (2014), *TALIS 2013 Results: An international perspective on teaching and learning*, OECD Publishing, Paris.

OECD (2017), *Educational Opportunity for All: Overcoming Inequality throughout the Life Course*, Educational Research and Innovation, OECD Publishing, Paris

Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, *130*, 105-119.

Prendergast, C. (2007). The motivation and bias of bureaucrats, *American Economic Review*, vol. *97*(1), pp.180-96.

Primary Education Council (2017, September 5). *Stijging aantal pabo-inschrijvingen druppel op gloeiende plaat*.. retrieved from: https://www.poraad.nl/nieuws-en-achtergronden/stijging-aantal-pabo-inschrijvingen-druppel-op-gloeiende-plaat

Pritchett, L. (2002). It pays to be ignorant: a simple political economy of rigorous program evaluation. *The Journal of Policy Reform*, *5*(4), 251-269.

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

Rockoff, J. E., Staiger, D. O., Kane, T. J., & Taylor, E. S. (2012). Information and employee evaluation: Evidence from a randomized intervention in public schools. *The American Economic Review*, *102*(7), 3184-3213.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4-36.

Ruijven, van, K., (2016). The Effects of Entry-Tests during Primary-School Teacher Education Netherlands. Thesis Erasmus Universiteit Rotterdam.

Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, *72*(2-3), 104-122.

Scafidi, B., Sjoquist, D. L., & Stinebrickner, T. R. (2006). Do teachers really leave for higher paying jobs in alternative occupations? *Advances in Economic Analysis & Policy*, *6*(1).

Schanzenbach, D.W., 2014. *Does class size matter?* In: Policy Briefs, National Education Policy Center, School of Education, University of Colorado, Boulder.

Sokal, L., Katz, H., Chaszewski, L., & Wojcik, C. (2007) Good-bye, Mr. Chips: male teacher shortages and boys reading achievement. *Sex Roles 56*(9–10): 651–659.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355-374.

Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *The Journal of Economic Perspectives, 24*(3), 97-117.

Steeg, van der, M., & Gerritsen, S. (2016). Teacher evaluations and pupil achievement gains: Evidence from classroom observations. *De Economist*, *164*(4), 419-443.

Stinebrickner, T. R. (2001). Compensation policies and teacher decisions. *International Economic Review* 42 (3), 751-780.

Stinebrickner, T. R. (2002). An analysis of occupational change and departure from the labor force: Evidence of the reasons that teachers leave. *Journal of Human Resources*, 192-216.

Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2016). *A coming crisis in teaching? Teacher supply, demand, and shortages in the U.S.*. Palo Alto, CA: Learning Policy Institute

Venhorst, V., Dijk, van, J., & Wissen, van, L. (2011). An analysis of trends in spatial mobility of Dutch graduates. *Spatial Economic Analysis*, *6*(1), 57-82.

Wei, R. C., Darling-Hammond, L., & Adamson, F. (2010). *Professional development in the United States: Trends and challenges* (Vol. 28). Dallas, TX: National Staff Development Council.

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the US and abroad. Technical Report*. National Staff Development Council.

Wilson, J. Q. (1989). *Bureaucracy: What government agencies do and why they do it*. Basic Books.

Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics. 100*, 61-78.

Wolf, de, I., Vermeulen, C.S.C., & Breuer, T. (2018). De lerarenopleidingen: Wie beginnen eraan en wie halen de eindstreep?. In F. Cörvers & M. van der Meer, *Onderwijs aan het werk – 2018: Analyses, feiten en visies over werken in het onderwijs* (pp. 75-89). Den Haag: CAOP

Yarnell, L. M., & Bohrnstedt, G. W. (2018). Student-Teacher Racial Match and Its Association With Black Student Achievement: An Exploration Using Multilevel Structural Equation Modeling. *American Educational Research Journal*, *55*(2), 287-324.

Yoon, K. S., Duncan, T., Lee, S. & Shapley, K. (2008). The effects of teachers professional development on student achievement: Findings from a systematic review of evidence. Paper presented at the Annual Meeting of the American Educational Research Association, March 24-28, 2008, New York

Nederlandse samenvatting (Dutch summary)

Een onderwijssysteem is zo goed als zijn leraren. Uit eerder onderzoek blijkt dat de leraar de meest invloedrijke factor op schoolniveau is voor de prestaties van leerlingen (Hanushek, 2011). De impact van leraren is groot: goede leraren kunnen in één schooljaar anderhalf jaar aan leerwinst boeken voor hun leerlingen, terwijl slechte leraren slechts een half jaar leerwinst voor hun leerlingen kunnen bereiken (Hanushek, 1992). Recent onderzoek uit Nederland laat zien dat het verschil tussen het hebben van een goede of een slechte leraar in groep acht wel een half adviesniveau kan schelen (van der Steeg & Gerritsen, 2016). Om het niveau van het onderwijs te verbeteren zijn er dus voldoende goede leraren voor de klas nodig. Het verzorgen van dit aanbod staat dankzij het lerarentekort (Adriaens et al., 2017), en zorgen over de kwaliteit en ongelijke verdeling van leraren (Inspectie van het Onderwijs, 2019; Onderwijsraad, 2013) hoog op de Nederlandse politieke agenda.

Dit proefschrift onderzoekt hoe de carrières en vaardigheden van Nederlandse leraren zich ontwikkelen. Deze inzichten zijn relevant voor beleidsmakers die een voldoende groot aanbod van goede leraren nastreven. Het proefschrift richt zich op leraren in het basisonderwijs en voortgezet onderwijs. De carrières worden beschreven voor basisschoolleraren, terwijl de hoofdstukken over vaardigheden zowel het basisonderwijs als het voortgezet onderwijs betreffen. De bijdrage van dit proefschrift is tweeledig. Ten eerste levert het proefschrift nieuwe inzichten in de loopbanen van leraren door het gebruik van grote administratieve datasets. Deze maken het mogelijk alle beginnende leraren te volgen vanaf de middelbare school tot aan het negende jaar op de arbeidsmarkt. Ook kan hiermee gekeken worden naar relaties met de school en leerlingkenmerken van de school waarop de leraren werken. Met deze schat aan data kunnen de carrièrepatronen van Nederlandse basisschoolleraren worden bestudeerd op een detailniveau dat tot nu toe niet mogelijk was. Ten tweede draagt dit proefschrift bij aan onze kennis over hoe leraren ervoor kiezen hun vaardigheden te ontwikkelen door middel van bijscholingsprogramma’s, en over hoe de taal en rekenvaardigheden van Nederlandse leraren in het basisonderwijs en voortgezet onderwijs zich verhouden tot die van de rest van bevolking, en hoe deze zich verhouden ten opzichte van leraren in andere landen.

Hoofdstuk 2 brengt de arbeidsmarktuitkomsten van beginnende leraren in het Nederlandse basisonderwijs in kaart en onderzoekt de samenhang tussen deze uitkomsten en de uitvalpercentages van de lerarenopleiding waaraan ze hebben gestudeerd. Het laat zien dat afgestudeerden van lerarenopleidingen met een laag uitvalpercentage, net zo vaak in het onderwijs terecht komen als afgestudeerden die van lerarenopleidingen komen met hogere uitvalpercentages. Afgestudeerden van lerarenopleidingen met lage uitvalpercentages werken echter wel 2.5% minder uren per maand na enkele jaren op de arbeidsmarkt. Desalniettemin compenseert het hogere aantal afgestudeerden ruimschoots voor het gemiddeld lagere aantal gewerkte uren van de afgestudeerden van lerarenopleidingen met lagere uitstroompercentages. Een tweede bevinding van dit hoofdstuk is dat pabo-afgestudeerden slechtere arbeidsmarktuitkomsten hebben wanneer ze ervoor kiezen om buiten het onderwijs aan het werk te gaan. Degenen die in het onderwijs werken verdienen een 10% hoger uurloon en hebben een 50% hogere kans op een vast contract. Het gemiddeld aantal gewerkte uren per maand verschilt niet tussen degenen die binnen en buiten het onderwijs werkzaam zijn.

Hoofdstuk 3 kijkt naar de basisscholen waar de Nederlandse leraren werken, en naar de mate waarin de match tussen leraren en leerlingpopulatie samenhangt met leerlingprestaties op de scholen. Het laat zien dat er een sterke match tussen leraren en leerlingpopulatie is als het gaat om migratieachtergrond. Op scholen met een hoger aandeel leerlingen met een migratieachtergrond werken relatief veel leraren met een migratieachtergrond. De match naar opleidingsniveau is gemiddeld: op scholen waar een hoog percentage leerlingen hoogopgeleide ouders heeft werken meer leraren met een masterdiploma. Deze vorm van matching is vooral zichtbaar in sterk verstedelijkte gebieden, en de sortering van beginnende leraren volgt hetzelfde patroon. Voor de prestaties van de leerlingen zien we geen bewijs voor negatieve effecten van deze sortering. Het aandeel leraren met een masterdiploma op een school heeft geen impact op de prestaties van leerlingen, noch is er sprake van heterogeniteit in de effecten naar het opleidingsniveau van de ouders van leerlingen. Het aandeel leraren met een niet-westerse migratieachtergrond is positief gerelateerd aan de toetsscores van leerlingen met een migratieachtergrond, terwijl er geen negatieve effecten zijn voor leerlingen zonder migratieachtergrond. De positieve resultaten zijn gedreven door betere prestaties bij rekenen, en zijn met name zichtbaar voor leerlingen een relatief lage socio-economische achtergrond.

Hoofdstuk 4 gaat over nascholing van leraren. Het modelleert de beslissing van leraren om zich aan te melden voor professionele ontwikkeling. Het model laat zien dat leraren bijscholingsparticipatie strategisch kunnen gebruiken als signaal voor hun onderliggende kwaliteit, omdat lerarenkwaliteit deels onobserveerbaar is. Als gevolg hiervan zullen bijscholingsprogramma's gericht op basisvaardigheden van leraren impopulair zijn, terwijl leraren zich inefficiënt vaak zullen inschrijven voor programma's gericht op de meer geavanceerde vaardigheden van leraren. De resultaten van het model laten verder zien dat het aanbieden van een programma spillover effecten heeft. Dit houdt in dat het aanbieden van een bepaalde cursus ook de participatie van leraren in andere cursussen kan beïnvloeden. Een gevarieerd aanbod van bijscholingsprogramma's kan er dus voor zorgen dat leraren zich op een efficiëntere manier gaan aanmelden voor professionele ontwikkeling.

Ten slotte brengt hoofdstuk 5 de taal en rekenvaardigheden van leraren in het primair en secundair onderwijs ten opzichte van de rest van de bevolking in kaart voor 15 ontwikkelde landen. Het laat zien dat er weinig laag presterende leraren zijn in Nederland: de 10% Nederlandse leraren met de minste vaardigheden scoren hoger dan de 10% laagst scorende tertiair geschoolden in andere landen. De gemiddelde leraar is ongeveer even vaardig als de gemiddelde tertiair geschoolde persoon. Aan de top doen leraren het minder goed dan andere hoger opgeleide personen. De 10% meest vaardige hoogopgeleiden presteren beter dan de 10% meest vaardige leraren. Internationaal gezien presteren Nederlandse leraren gemiddeld en aan de onderkant van de verdeling goed, maar zijn ze aan de bovenkant relatief zwakker. Er zijn vergeleken met andere landen weinig excellent presterende leraren.

Al met al levert dit proefschrift een bijdrage aan de bestaande kennis over de ontwikkeling van de carrières en vaardigheden van Nederlandse leraren. De resultaten van dit proefschrift suggereren dat de huidige beleidsmaatregelen gericht op het aanpakken van het lerarentekort voor verbetering vatbaar zijn. De focus op het verhogen van de instroom in de lerarenopleidingen is niet efficiënt wanneer niet tegelijkertijd de hoge uitval uit de opleiding wordt aangepakt. Daarnaast hebben de huidige maatregelen met als doel de minimale kwaliteit van de instroom te verhogen (zoals de entreetoetsen en de reken/taaltoetsen) waarschijnlijk een groter negatief effect op de kwantiteit, dan een positief effect op de kwaliteit van leraren. Daarbovenop lijkt een onbedoeld bijeffect van deze maatregelen te zijn dat nog minder studenten met een niet-westerse migratieachtergrond instromen in het lerarenberoep. In plaats van beleid te richten op de minimale kwaliteit, lijkt beleid gericht op het aantrekken van leraren met hoge vaardigheden kansrijker in de Nederlandse context.

Ten tweede toont dit proefschrift dat maatregelen gericht op het verbeteren van basisvaardigheden van leraren door middel van nascholing, het negatieve signaal dat uitgaat van het meedoen aan deze vorm van professionele ontwikkeling in acht moeten nemen.

Ten slotte bieden de resultaten van dit proefschrift aangrijpingspunten voor toekomstig beleid op het gebied van leraren. Met name het tegengaan van uitval tijdens de lerarenopleiding, het enthousiasmeren van studenten met een niet-westerse migratieachtergrond voor het lerarenberoep, en het stimuleren van fulltime werk onder de huidige lerarenpopulatie zijn kansrijke richtpunten voor beleidsmaatregelen in de strijd tegen het lerarentekort.

Curriculum Vitae

Christiaan S.C. Vermeulen was born on September 15, 1989, in Breda, The Netherlands. He obtained his Bachelor's degree in International Business at Tilburg University in 2011, and graduated from the Human Decision Science Master's program at Maastricht University in 2014. Between November 2014 and November 2018, he was a PhD-candidate at the Department of Economics in the Economics of Education group at Maastricht University, working on the "Academische Werkplaats Onderwijs" project in cooperation with the Dutch Inspectorate of Education. Since November 2018, he is employed as a postdoctoral researcher at the Research Centre for Education and the Labour Market (ROA) and the Department of Economics (AE2) at Maastricht University, where he works on the "Academische Werkplaats Onderwijskwaliteit" project, again in collaboration with the Dutch Inspectorate of Education.