ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

Ministerie van Onderwijs, Cultuur en Wetenschap

The Effects of Entry-Tests during Primary-School Teacher Education in the Netherlands

Koen van Ruijven

Supervisor: Prof. dr. Dinand Webbink

Master Thesis Policy Economics

Erasmus University Rotterdam, Erasmus School of Economics, Department of Economics

October 2016

Abstract

This thesis investigates the effects of the implementation of entry-tests in the Dutch teacher education for primary-schools. During the first year of the study, students were required to pass tests in mathematics and the Dutch language in order to proceed towards the second year. Using the synthetic control method originally coined by Abadie and Gardeazabal (2003), we find a gradual but severe decline in change of enrollments-year-by-year developing to about 40% points in 2009. These effects are roughly similar per subgroup of highest previously attained education. More importantly, our results document a severe deterioration in the rate of students being granted permission to enter the second year (6.5%). In turn, the magnitude of these effects are in line with the presumed set of cognitive standards per highest previously attained education. About two-third of the retention decline correspondingly also develop into reduced graduation rates. Lastly, we find no evidence for the notion that the cognitive skills of entering pabo students have improved over time.

Keywords: Primary-school teacher education, selection after the gate, quantity-quality tradeoff in the labor markets for teachers

Table of Contents

Acknowledgements

The beauty of economics is that it makes apparent several patterns one never thought about before. Yet, once familiar with the concept, phenomena come to life, which sometimes are so blatantly obvious. Books like '*Why Nations Fail'*, *'Thinking, Fast and Slow*' and *'House of Debt'* are great examples of this. But also concepts like 'skills beget skills' and the 'dynamic complementarity' of skills formation. The latter constitute one of the many excellent topics professor Dinand Webbink has introduced me to during last year. I would like to thank him for his supervision, the constructive feedback given to this paper and relatedly, for being a wonderful teacher.

Teachers in general are the major theme in this paper. More specifically, primary-school teachers in the Netherlands, many of whom will retire in the upcoming decade, inducing a severe and growing shortage of primary-school teachers. I would like to express my appreciative gratitude to Ib Waterreus and Marc van der Steeg at the Ministry of Education, Culture and Science for providing the opportunity to explore the effects of specific entry-tests at the gateway of becoming a primary-school teacher, the pabo – Ib for his enthusiasm, knowledge and expressed notions in our meetings which motivated me to proceed with the project, and Marc for his excellent sparking ideas and knowledge, both contributions were incredibly valuable to me.

As a general matter of fact, the quality of research hinges on its available data. To this end, I am indebted to Arrian Rutten and Erik Fleur of DUO for having made available a dataset with detailed information about Dutch higher vocational education students and Maud Huygen of Studiekeuze123 for presenting specific study-related data. Additionally, I would like thank Annelise Sprenger of the Ministry for the information provided about the policy changes at the pabo.

My fellow interns – Peter and Vishand- were also of great help to me during our internship. My discussions with Peter about the synthetic control method were essential in acquiring the knowledge to produce accurate results.

Lastly, all the people of the Ministry within the Direction Kennis, and most notably the cluster Kennisverbindingen en Kenniseconomie are gratefully acknowledged for their shared enthusiasm and ambition to enhance the quality of the Dutch education system.

I. Introduction

'*The quality of an education system cannot exceed the quality of its teachers.*' A now notorious statement by Barber and Mourshed (2007) that has become common knowledge among policymakers during the last couple of years. At first glance, this observation is not particularly remarkable. It is generally recognized that schools possess no other attributes that have a more vigorous influence in student achievements as the quality of teachers (e.g. Hanushek, 2011)[1]. Students and parents alike follow this impression in their ambition to ensure placement in classes of specific teachers. And indeed, one should expect so, given the substantial estimates of gains in students learning being assigned to specific teachers. For instance, the estimates of Rivkin *et al.* (2005) and Aaronson *et al.* (2007) reveal an one standard deviation improvement in teacher quality enhances student performances in the order of 0.1 to 0.2 standard deviations. Relatedly, the pupils' cognitive tests scores seem to translate towards the future -both in terms of college completion and earnings (e.g. Murnane *et al.*, 1995). The policy implications of these findings were profound and have revitalized the attention given to the quality of teachers.

The Netherlands used to be an epitome of a well-functioning education system. A feature that has slowly been blurring since the early 2000's however. Dutch test scores are increasingly losing ground compared to the top nations (van der Steeg *et al.*, 2011), fueling policy debates to reverse such tendencies[2]. In Dutch policy circles it was generally perceived that the relative deterioration in international test scores was attributable to the average decline in the quality of primary school teachers. As such, it was not surprising that stakeholders pinpointed the education to become a primary school teacher, the pabo, as the dominant element to reverse these tendencies. Also pabo-principals acknowledged the, at the time, prevalent pabo-education was unable to deliver an uniform and satisfactory quality of graduated students and mainly aimed at increasing the quality of entrance students.

In 2005 Cito developed a calibrated mathematics test to juxtapose the math scores of primary-school pupils in the eighth grade with the scores of pabo students in

---

[1] The literature regarding class size reduction policies suggest positive effects in terms of student achievements see e.g. Krueger (1999) in Project STAR. Yet, despite the fact that the effects are worthwhile, class size policies are generally deemed to be very costly.

[2] Concerns about the educational quality of primary schools were driven by inspection reports, conveying a mere quarter left primary education with insufficient literacy skills. Relatedly, there were surges in the complaints of secondary and tertiary schools arguing that entering students possessed insufficient numeracy skills (Inspectie van het Onderwijs, 2007). Within an international context however, van der Steeg et al. (2011) note that the Dutch test scores of the weakest students are among the best of the world. In contrast, the top-students in the Netherlands have lost ground compared to their international counterparts.

their first year. Using a representative sample of both groups, the conducted tests were seized as a proxy to portray the problematic scores of pabo students. The latter were at least deemed to score at the 80^{th} percentile of primary school pupils in order to show they possessed 'sufficient mathematics skills'. The results however did not come as a surprise. Only a mere 47% of all entrance pabo students were able to meet the succeeding criteria, where the variation in success rates was mainly dependent on the highest previously attained education of pabo students (Straetmans & van Eggen, 2005).

In effect, earlier concerns about the quality of primary-school teachers became validated, which eventually induced experts to widely embrace the ambition to increase the quality standards of the pabo's. During the course of 2006, stakeholders proposed to implement mandatory mathematics and language 'entry'-tests for students in the first year of the pabo. Since the enrollment cohorts of 2006 and 2007 respectively, pabo freshmen are deemed to show they possess at least the sufficient set of cognitive standards in terms of math and Dutch language skills, or otherwise students are obliged to leave the pabo after the first year. It was conjectured that this, once and for all, would ensure pabo students had at least the required cognitive aptitudes to become primary-school teachers. Following this, the image of primary-school teachers would gradually improve, rendering the occupation more attractive for students with high cognitive abilities. The latter reflected the idea that the entry-tests would additionally generate a better composition of students, thereby contributing to a higher graduation rate efficiency of pabo's.

Earlier literature regarding the effects of the implementation of the entry-tests at the pabo's is scarce. A study by van den Enden (2012) solely focused on the enrollment change effects, and found that the implementation led to an enrollment decline of about 15%-points. In this paper, we aim to sketch a complete image of the effects in a chronological manner– particularly diverting our focus on the change in retention rates and graduation rate efficiency. Our identification strategy relies on an application of the synthetic control method for comparative case studies coined by Abadie and Gardeazabal (2003). This method extents the standard differences-in-differences methodology by reducing the ambiguity on which control units are chosen. In the light of our research, a data driven method constructs a 'synthetic' pabo by assigning weights to studies in the donor pool of control units dependent on how well they resemble the outcome variable of the actual pabo. Correspondingly, we simply measure the difference between the outcome variables of the actual pabo and the synthetic pabo during the post-intervention period to obtain causal inferences of the entry-tests.

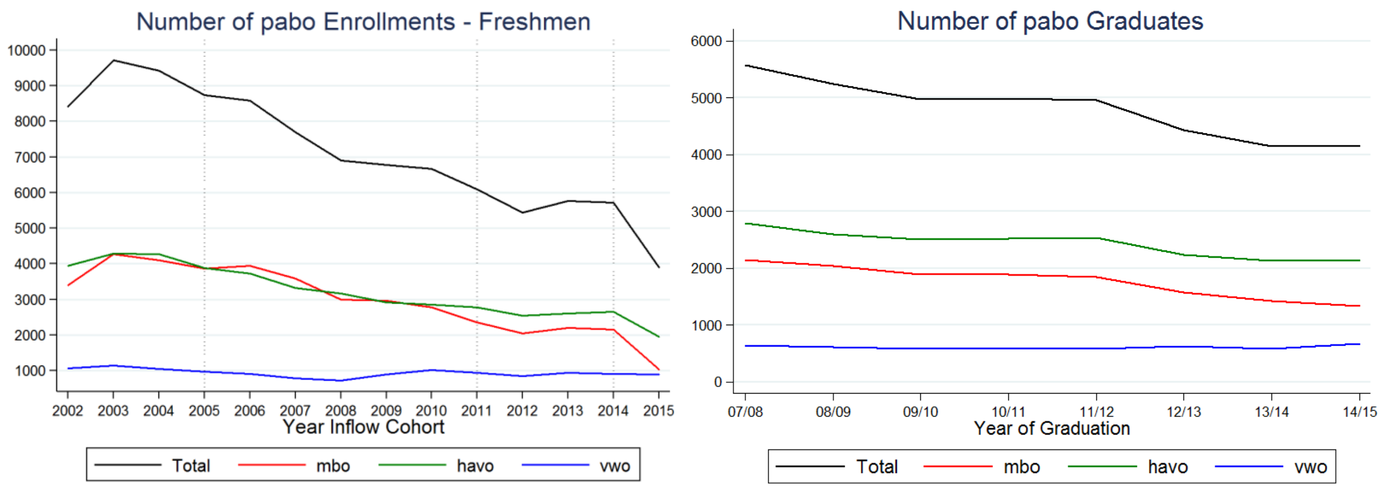
We find the entry-tests induced a gradual adverse effect on enrollments – year-by-year developing to about 40% points in 2009. These effects are roughly similar per subgroup of highest previously attained education. Moreover, the prevalent notion that students *de facto* entering the pabo perform better since the entry-tests were implemented cannot be backed by empirical evidence. In contrast, our results document a severe deterioration in the rate of students being granted permission to enter the second year (6.5%). The magnitude of the effects are now in line with the presumed cognitive abilities of students with different previously attained educations; most notably those students with a highest pre-education of mbo are negatively affected, as illustrated by a drop in retention rates of about 9% points. Our estimates reveal the drop in retention rates correspondingly pass-through imperfectly to a decline in the graduation rate efficiencies of pabo's -about two-third of the retention decline eventually settle in a reduction of graduation rates.

The former effects are consistent with the observation that *on average* the cognitive abilities of students entering the pabo did not improve since the entry-tests were implemented, which is confirmed by our analyses. The main function of the entry-tests must therefore be perceived as a *selection mechanism*, precluding pabo students the entrance towards the second year with insufficient cognitive abilities.

At the backdrop of our results are large predicted shortages of primary-school teachers. In the upcoming years, a substantial amount of primary-school teachers approaching their retirement age have to be replaced – mostly by students graduating from the pabo. Current conservative estimates predict a gradual growth in the shortage of primary-school teachers, reaching its peak in 2025 with an amount of 10000 fte's (CentERdata, 2016). The predicted shortages resemble the consequence of the quality-quantity tradeoff in the labor market for teachers. Stakeholders of the pabo have increasingly wanted to ensure the cognitive element of the former, while meanwhile the amount of pabo graduates have declined annually (see figure 1). Our results suggest a current disbalance in the tradeoff, certainly considering the evidence that certain observable cognitive elements of teachers constitute to higher quality teachers is scarce.

The remainder of this paper is organized as follows. The next section provides a historical synopsis of the policy changes at the pabo, followed by section three –a broad

Figure 1 – The Decreasing Amount of Enrollments and Graduates at the pabo



Notes: The figure on the left-hand side displays the amount of students *de facto* entering first year of the pabo. Only students with a main subscription at the pabo are considered. In the figure on the right-hand side the amount of graduates are visualized per year of graduation. In both figures, the students without an explicit highest previously attained educational level are still labelled as students with a highest previously attained education of mbo.

overview of the related literature, which eventually gives rise to a discussion about the quality-quantity tradeoff in the Dutch primary-teacher labor market. Section four gives a detailed description of our identification strategy – the synthetic control method. This section also presents a brief description of the used data. The results will be discussed in section five, following the path of gradual more specific settings. The last section will wrap up the main conclusions of our analyses, and set up a discussion of its implications. This section also contains an overview of the enrollment effects of the newly imposed policies at the pabo since the entrance enrollment cohort of 2015.

II. Background Policy Changes at the pabo

Since the early 2000's, the top-performing pupils in the Netherlands are losing ground relative to their international counterparts during primary school. Specifically, there are large, and diverging, differences in test scores of the best scoring 9-year olds in the Netherlands compared to the best nations in the world (van der Steeg *et al.*, 2011). The current situation even reflects that the Dutch top-scoring primary school pupils perform below the world's average in the subjects of mathematics and sciences[3], which is the backdrop of an even larger phenomenon. Not only the top-scoring students in the Netherlands have experienced a relative decline, also the *average* Dutch scores have deteriorated, both relatively and absolutely, which in context of the findings of Hanushek and Woessman (2012) is highly unfavorable. These authors show that

[3] This may not come as a surprise since primary schools in the Netherlands on average devote the least time on their science education. Table 1 in Appendix C illustrates the decreasing performance of Dutch students.

average math scores resemble the highest relation to economic growth. The reduction in average math scores of Dutch students could therefore have large effects on the size of the Dutch economy.

From another perspective is the average decay of Dutch math performances comparable to the effect of replacing average teachers by low-quality teachers (Hanushek and Rivkin, 2010). In Dutch policy circles, stakeholders mainly identified the declining average quality of primary-school teachers as reason for the deterioration in Dutch pupils' test scores. The dominant element to reverse these tendencies, so was argued, was the education to become a primary-school teacher, the pabo.[4]

Within the curriculum of the pabo's, there is generally a trade-off between specific-cognitific competence subjects and the focus on didactic and pedagogic elements. In earlier reports the notion was conveyed that this tradeoff has increasingly moved towards the latter elements during the last decades (Onderwijsraad, 2005). Pabo's were simply not able to devote more time on closing deficiency gaps of its students within the dense curricula[5]. Following this notion, the commission Meijer, Vermeulen-Kerstens, Schellings and Meijden (2006) argued that the capabilities the enrolling entrance students were deemed to have were too low. The main culprit of the declining quality of graduated pabo teachers was thus the below par quality of entering students. Also pabo-principals acknowledged the, at the time, prevalent pabo-education was unable to deliver an uniform and satisfactory quality of graduated students and mainly aimed at increasing the quality of entrance students.

This notion is, in a sense, not new. In the decades building up to the early 2000's, the deficiency of entrance students in the pabo in terms of both math- and language skills was subject to fierce debates within the Dutch policy circles. Already in 1987 the Dutch Education Inspection recommended the Ministry to select the entrance students more thoroughly. In the following year, it was the '*wiskunde-maatregel'*, the obligation for each entrance student to have at least mastered math level A within its pre-education profile, which had to ensure each entrance student had sufficient math capabilities. However, the obligation was dropped in 1990, probably induced by a growing shortage in primary school teachers. Only eight years later in 1998, the policy once again was reversed with the implementation of the 'Studiehuis'; havo-students

---

[4] See Appendix B figure 1 for an outline.

[5] In 2003, the Dutch and Flemish Accreditation Organization (NVAO) argued the envisaged basic set of skills pabo students were deemed to have after completing the pabo, were almost impossible to realize within a duration of four years.
(Entrance enrollment is defined as students enrolling in the first year of a specific study.)

were required to accomplish mathematics in their end-exam. Yet, with all preceding flawed policies in mind, the implementation did not induce the desired consequences.

Earlier concerns about the deficiencies in mathematics skills of pabo students became validated in 2005, at the time when Cito developed a calibrated math tests to compare the math skills of pupils in the eighth grade of primary-schools with those of pabo students. Pabo students were at least deemed to score at the 80th percentile of primary school pupils in order to show they possessed 'sufficient mathematics skills'. The results were not startling[6]. Only a mere 47% of all entrance pabo students were able to meet the succeeding criteria. The variation in success rates was mainly dependent on the highest previously attained education of pabo students. Where 76% of all vwo students succeeded the test, only 53% of the havo students were able. The success rates of mbo students were even more striking: merely 28% of all mbo students succeeded (Straetmans & van Eggen, 2005).

In effect, the minister concluded the precluding policies did not induce the desired consequences and cancelled the mathematics end-exam requirement for havo students with a C&M profile. Other policies were deemed more effective by the Minister of Education. The newly proposed policies were two-sided. On the one hand, it was proposed to offer stand-alone math courses in the pre-education leading up to the pabo – later on what formally became the so-called *'rekentoets'* during secondary education. While on the other hand it was proposed to implement a mandatory mathematics test during the first year of the pabo. The latter proposition was well received by the pabo's themselves who embraced the ambition to enhance its quality standards. In a step-by-step integration, the quality standards at the pabo have been gradually scaled up. The main policy interventions have been summarized in table 1.[7]

The quality enhancements started during the course year 2006-07 with the implementation of a standardized mandatory math test for pabo students in their first year. During the course of 2006 there was some controversy among stakeholders who argued that the possession of sufficient math skills is only part of the skills primary school teachers are deemed to possess. The command of the Dutch language is of similar

[6] The results by Warps *et al.* (2010) indicate that pabo students perceive their study as less difficult before starting their studies (relative to other vocational-education students). In contrast, they expect to incur study problems with numeracy skills and critical thinking. Pabo-students attribute this due to the flawed connection between their acquired pre-education skills and the skills they are required to possess during the pabo.

[7] *History repeats itself*, the ambition to increase the quality standards at teacher educations must always be examined at the backdrop of the prevalent conditions at the labor market. For instance, during 2014 there was a vast surplus of primary school teachers. As such, stakeholders thought they could 'afford' itself to be relatively more strict. – more generally showing that the effects of imposed quality increments during teacher education programs can analyzed in an quality-quantity tradeoff. See related literature section B3.

Table 1: Evolution policy changes PABO

Entrance cohort	*Admission Changes*	*Required Level*
2006-07	Implementation of mathematics entry-test for first-year pabo students, in combination with an academic dismissal policy (BSA). Relatedly, the language entry-test was also implemented, however here only an advisable standard was set.	Minimum required math skills of at least the 20 percent best –scoring pupils of the eight primary-school grade. Maximum of three opportunities. Anyone with at least a MBO-4 certificate is able to subscribe to the pabo.
2007-08	The language entry-test for first-year pabo students also mandatory, in combination with an academic dismissal (BSA).	Minimum required language skills of at least havo-4 level Maximum of three opportunities. Anyone with at least a MBO-4 certificate is able to subscribe to the pabo.
2008-09	Establishment of academic pabo's, comprising a combination of the pabo education itself and either an university bachelor of pedagogic sciences or a pre-master pedagogic science. Possible increased attractiveness of pabo's for VWO'ers	VWO
2011-12	Implementation of 14 mandatory 'knowledge-bases'; these consists of required end-levels of specific-pabo subjects. These have to be mastered by the student in order to guarantee each graduated pabo-student has sufficient capabilities to succeed as a teacher. Partially initiated by pabo's themselves.	In practice, students have to successfully complete numeracy and literacy tests during the third-year of their study. Pabo's themselves develop these tests.
2012-13	Implementation of mandatory entry-tests of 'Mens & Wereld' for first-year pabo-students, in combination with an academic dismissal (BSA). These concern three subjects; history, geography and natuur&techniek.	Minimum required geography, history and natuur & techniek skills of at least the 20 percent best–scoring pupils of the eight primary-school grade. Maximum of three opportunities. Anyone with at least a MBO-4 certificate is able to subscribe to the pabo, independent of the pre-educational background.
2015-16	Concerns about the difficulty and quality-content of HBO teacher education studies remain[8]. The objective was, once and for all, to increase the quality of outflowing pabo-students by selecting students before they start the pabo. In this way, pabo's can guarantee a similar quality-level of outflowing pabo-students[9] and do not have to adhere their curricula to the quality of inflowing students. All prospective students have to show -either due to satisfactory secondary school grades or by passed test tailored for prospective pabo-students- that their cognitive abilities for the subjects English, history, geography and natuur&techniek are sufficient to enroll at the pabo.	Selection before the gate as constituted within Dutch law; minimum required entry level in the subjects geography history and natuur&techniek of at least havo-3/vmbo-4 level. There are two ways the student can show it has sufficient cognitive capabilities (i) by showing satisfactory secondary school grades (vmbo and havo) within the subjects. An unsatisfactory grade is allowed in only one of the subjects. A sufficient grade for the secondary-school subjects of either physics, biology or NLT can be used for the subject natuur&techniek or (ii) by passing test specifically tailored for prospective pabo students in these subjects. If the student fails in only one of these test it can still be provided with an assessment to enter. The math entry-tests during the first year of the pabo remain. The standardized language tests will end after the cohort 2015-2016.

utmost importance. Therefore, a mandatory standardized language test was implemented in the ensuing course year 2007-2008. Since then, students have the opportunity to successfully complete both tests during the first semester of the pabo[10]. If students do not successfully complete the tests in their first opportunity, two remaining opportunities are provided to the student. In case the student has still not successfully

[8] Kwaliteitsagenda Krachtig Meesterschap
[9] Een goede basis, commissie Meijerink.
[10] In commission of the Ministry of education, culture and science CITO has developed the standardized mathematical entry-test; the Wiscattoets. The skills of students are tested using an adaptive computer test; the precedent answer given by the student determines whether she or he gets a more (less) difficult question afterwards. After the last question has been completed, the acquired level within the calibrated 'difficulty' scale reflects the test score of the student. More detailed information available at Straetmans and van Eggen (2011). Also the language test has been developed by CITO and is based on the same principle; an adaptive computer test to inquire whether the student possesses a sufficient command of the Dutch language.

completed either one of the tests after the first year, an academic dismissal (i.e. negatief bindend studieadvies) is given to the student, implying it is obliged to leave the pabo.

One may wonder –if the perceived math and language abilities of pabo students were so weak- why pabo's did not conduct these entry-tests in preceding years to the actual implementation? This could be answered quite simply: The majority of the pabo's *already did* conduct math and language tests during the (first year of) the pabo. Yet, these tests were not standardized and not directly linked to an academic dismissal policy[11]. In practice, this meant pabo students could advance their studies, without *de facto* having the required skills during the second year of the pabo. Relatedly, the 'kennisbases' were not yet implemented, which implied pabo students with insufficient skills had the opportunity to obtain a degree.

In sum, the 'entry-tests' should be perceived as an objective, grounded selection mechanism within the first year of the pabo, such can be guaranteed that students entering the second year of the pabo have a relatively uniform level of cognitive skills – skills which primary school teachers are at least deemed to possess[12]. In this paper, we specifically focus to present a complete view of the effects of these entry-tests. A comprehensive overview of the effects of the newly imposed changes in admission requirements since the course year 2015-16 will be analyzed in a separate report. Yet, with the currently available data, we *are* able to present the enrollment effects of the changed admission requirements, these are incorporated in the discussion of this paper.

III. Related Literature

A. The Importance of Teachers

Over the course of the past decades, academic research has developed an impressive amount of literature stressing the importance of teachers. While speaking of a term like '*the quality of a teacher*', researchers should develop an understanding what factors actually constitute toward it. The surge in standardized student testing has fulfilled this gap by providing an accountability measure to stakeholders- commonly termed as the 'value-added' approach. The main findings of this 'economic' approach suggest that

[11] The same could be noted for the Mens & Wereld tests, which were finally standardized in 2012-13. These tests concern the subjects geography history, and biology/nature sciences. The wish to standardize tests at the pabo came forward in the report *beleidsagenda lerarenopleidingen 2005-2008 "Meer kwaliteit en differentiatie: de lerarenopleidingen aan zet"*, which reflected the goal among pabo's to develop a singular quality standard at pabo's.

[12] The least set of cognitive skills of each subject were more formally instituted after advice of the Inspection of Education (2008) with the development of the so-called '*kennisbases'*. These described the required end-levels which graduated pabo students had to possess in order to succeed as a teacher.

there are large differences in average test-score gains by students across classrooms. Even classrooms within schools portray substantial variation among average test-score gains of students (Hanushek & Rivkin, 2010). Some teachers are consistently able to produce larger achievement gains in student test scores than others. At first sight these findings are not startling, stakeholders may point out that endogeneity may be at stake; families may select themselves into neighborhoods and schools based on their unobserved preferences and resources (Tiebout, 1956). Yet, even studies that aimed to circumvent the potential bias that may arise, have found significant differences in student test-scores attributable to teachers. For instance, Rivkin *et al.* (2005) use repeated observations of students to cancel out time-invariant effects of student characteristics affecting its achievement gains[13]. These authors specifically focus on the within-school variation in quality of teachers, exploiting the between-cohort average test-score gain per specific grade. Their results indicate that teachers have significant impact on the variation in test-score gains in mathematics and reading[14]. The findings by Aaronson *et al.* (2007) are in line with the magnitude of the math score effects. They also employ a fixed-effects strategy, but extent it with corrections for sampling error due to their relatively small amount of data (Chicago). Interestingly, higher-quality teachers have the largest impacts on groups with relatively disadvantaged backgrounds, most notably African-American students.

Even though the former studies devote a great deal in trying to circumvent potential biases, the assumption that there are no confounding factors cannot be ruled out completely (Rothstein, 2010). The holy grail of identifying causal effects in this sense would be a random assignment of teachers and students to classes- exactly what was done in the in Project STAR in Tenessee. Nye *et al.* (2004) seized this data to exploit the random variation in between-teacher quality within schools and again found effects that were fairly similar to the earlier mentioned studies. [15]

Research has thus confirmed folk wisdom that the quality of teachers can have significant impacts on the achievements of students. From a policy point of view it would correspondingly be attractive to link these findings with two questions; (i) whether

[13] The authors use matched panel data from the Texas Schools Project. The authors included fixed-effects on the on both a student level and on a school level. The former capture a myriad of family factors, e.g. parental education, parental permanent income, whereas the latter capture time-invariant school characteristics, e.g. school resources and curricula.
[14] One SD improvement in teacher quality leads to 0.11 and 0.095 SD student-improvement, respectively.
[15] Another study that uses random assignment of teachers is Kane and Staiger (2008) of an experiment in Los Angeles. Interestingly, they were able to find that pre-experimental value-added measures of teachers were able to predict the value-added of student achievement during the experiment.

student achievement gains translate towards academic success and income variation in later life and (ii) whether specific attributes of teachers contribute to more effective teachers (see outline below). Research by Chetty *et al.* (2014) has focused on the first question and found both using a cross-sectional design and a quasi-experimental design that an one standard deviation improvement in teacher quality enhances the probability of college attendance by almost one percent at age 20. After their college, students also earn more and have steeper earning trajectories – an one standard deviation improvement in teacher quality raises yearly earnings by more than one percent (see also Murnane *et al.*, 1995)[16]. The evidence thus points out that the quality of teachers matters for student outcomes and has substantial influence for them during their occupational career.

Concerning the second question, placing a floor on teacher quality should require some knowledge that some characteristics of (prospective) teachers are systematically related to improved student outcomes. In contrast, there is no conclusive evidence that particular attributes of teachers can be ascribed to higher quality teachers. In general observable characteristics explain only little in terms of teacher quality deviation. Despite this, Boyd *et al.* (2008) argue that screening policies of principals based on observable characteristics could lead to an improvement in student learning- something other economists have disputed since the quality of teachers being explained by observable characteristics is dwarfed by the unexplained attributes of teachers. [17]

For instance, there is substantial controversy in the literature whether the experience of teachers has any impact past the initial years of profession. The fact that inexperienced teachers –being less than three years in profession- do adversely impact teacher performances is fairly well established in the literature (Kane *et al.*, 2008; Staiger & Rockoff, 2010). But whether experience matters beyond this cutoff is

[16] From a bird perspective, Hanushek and Woesmann (2012) show that international variation in tests math and science scores also translate towards deviations in growth trajectories among countries.

[17] It must be noted however that the link between test scores and teacher quality become more pronounced within an international context. Hanushek *et al.* (2014) use data of an international assessment of the cognitive skills of adults (PIAAC) to link the variation in teacher quality to the international variation in student performances, and show that an one SD improvement in teachers scores increases student performances by 0.2-0.3 SD. Teachers in the best performing student nations score above people with a master's or PhD degree, whereas teachers in poor performing countries score similar to the skills of adults with just a post-secondary education. See also Borghans *et al.* (2012).

debatable. More recently, some high quality studies have shown that experience is associated with higher test-scores of elementary pupils in math and language; Harris and Sass (2011) using a three-way fixed effects model and Gerritsen *et al.* (2014) using the random assignment of twins in classrooms and Chetty *et al.* (2011) using data from Project STAR.

In this paper though, it would be more interesting to infer whether pre-education characteristics of prospective teachers are related to more effective teachers. In the US literature most studies find that initial education of teachers is unrelated to student performances (Aaronson *et al.* 2007; Fryer & Levitt, 2004). And when a significant effect of initial education is found, these tend to be small (Clotfelter *et al.*, 2007). An optimistic view thus postulates that the possession of a master degree is weakly related to the quality of a teacher. More often a significant effect is found of test scores of students in high-schools as a predictor of the value-added of teachers (Hanushek, 2003). For instance, Clotfelter *et al.* (2007) indicate that the effects of initial test scores of teachers are particularly large for students' achievements in math. The effects on language achievements are much less pronounced[18], which is in line with the results of Boyd *et al.* (2008). US research though, does not uniformly convey that test scores are important predictors of teacher quality. Angrist and Guryan (2007) for example find that placing a cognitive aptitude floor by specific licensure exams induces a surge in wages rather than it improves quality, i.e. it decreases the pool of potential applicants. Also Buddin and Zamarro (2009) find that the value-added of teachers in the state Los Angeles is largely unrelated to licensure tests measuring general knowledge, multiple subjects and reading pedagogy.

The effects of racial/ethnic characteristics of teachers however are mostly ignored in the former studies. Dee (2005) has shown that such characteristics can already by seized by school boards as determinants of teacher quality. If students share teachers of a similar racial or ethnic designation, they are less likely to show disruptive and inattentive behavior, which enhances their test-score performances. Once again, students of low socioeconomic backgrounds seem to benefit most from teachers as 'role model'. [19]

Recent advancements in the literature have gained increased access to more specific nontraditional attributes of teachers. In this way, school principals do not have

[18] These authors have made use of longitudinal data across grades 3 to 5 from North Carolina data and use student fixed effects to deal with potential non-random matching of students and teachers

[19] In the Netherlands, the ethnic and racial diversity among prospective primary teachers at the pabo has undergone a steady decline over the past years. See NRC article '*De pabo wordt steeds witter*'.

to focus on the traditional observable characteristics of teachers. Instead, principals have more opportunities to identify the quality of a teacher by extending the demand of teachers credentials in order to make more effective recruiting decisions- simply because being a good teacher involves more than achieving high grades in cognitive tests. For example, Rockoff *et al.* (2011) investigate, next to cognitive measures, a broad set of non-cognitive skills to probe which factors affect the effectiveness of teachers at the times after their first hire. In relation to the cognitive elements, factors like the SAT score, the teacher certification exam score and the number of courses students took in the subject matter are taken into account. These all are positively -albeit marginally significant- associated with teachers effectiveness during the first year. The 'Big Five' non-cognitive personality traits comprising factors as extraversion, conscientiousness, agreeableness, emotional stability and openness to experience convey a similar pattern of positive effects. Especially conscientiousness and personal efficacy seem of teachers seem to benefit student performances. [20]

B. Analysis of Relationship Between the Quantity and Quality of Primary-school Teachers

As argued above, evidence has spread that the value-added variation of teachers can have large long-term effects on individuals and the economy as a whole, yet much less is known about which attributes constitute to teacher quality. Still, there is a prevalent notion among economists and the general public alike that the average 'quality' of teachers has declined over the past decades. Quality here being referred to as the pre-educational background of teachers entering the profession. Before analyzing the quantity-quality tradeoff of imposed quality increments, we explore what effects determine the (declining) quality of inflowing teachers. These factors will be considered in the following paragraph. Thereafter, a paragraph will be devoted on what factors induce students to self-select themselves into teacher education. These paragraphs will be combined to construct a framework to analyze the effects of increased quality standards in a quantity-quality tradeoff.

[20] See also Van der Steeg & Gerritsen (2016) and the literature review by van der Rijst *et al.* (2015). The latter combine a set of studies having investigated the predictive power of observable characteristics of students as determinant of study success during teacher education. Interestingly, most studies indicate a mix of non-standard personal characteristics is able to (weakly) predict the success of students in didactic courses. Yet, the predominance of evidence suggests that mainly poor scores on a variety of indicators is able to predict the poor performance of students in teacher education. For instance, Andrew *et al.* (2005) indicate that critical thinking and (low) verbal skills are important predictors of the GPA of students during teacher education. Whereas Bieri and Schuler (2011) find that the combination of low assessments on communication, cooperation, assertiveness, motivation and factual research skills is correspondingly able to predict deficient performances of students in teacher education.

B_1. Explanations for declining quality of teachers

According to Lakdawalla (2006), the contraction in the average quality of teachers in the US can primarily be attributed to technological changes, which have enlarged the productivity of skilled workers outside teaching. At the same time, technical progress in teaching education has been slow, inciting a relative wage dispersion of high-skilled teachers compared to their high-skilled counterparts[21]. The unfortunate consequence is that the attractiveness of becoming a teacher has deteriorated, which is further aggravated by the occupational specificity of teacher education. For instance, a study of economics or law provides a student to alter between future job opportunities, whereas this opportunity is largely limited for teachers after their education. In other words, the latter group has low substitution opportunities, exemplified by the fact that the unemployment rate of recently graduated teachers is highly dependent on the demand and supply of teachers. [22]

While seizing the skill-biased technological change narrative of Lakdawalla, Bacolod (2007) further specifies it by recognizing the developments of women and disadvantaged minorities in (Western) labor markets over the past decades. In historic sense, high-quality females had less outside opportunities besides the opportunity to become a teacher. For example, Becker (1985) conjectured that women choose for teaching because it is more compatible with tasks at home. The last decades have shown remarkable alterations in this notion; females and disadvantaged minority groups have seen large improvements in terms of professional opportunities[23]. Simply put, it has increasingly become less attractive for high-ability females to opt for a career as teacher, rendering a relative decline in the quality pool of teachers.[24]

Yet Hoxby and Leigh (2004) observe that alternative opportunities for women improved in all occupations, which makes the narrative of improved labor market opportunities for solely high-aptitude women incomplete. These authors decompose the decline in average teacher aptitude into an alternative opportunities effect and an

---

[21] Lakdawalla indicates that the relative change in wages of Dutch primary school teachers has declined by 23% within a period of three decades attributable due to productivity divergences. He argues that this induced school principals to substitute away from teacher skill to teacher quantity. In this study, a specific type of intelligence test is used as a proxy for quality of teachers.

[22] Cörvers (2014). In Appendix A, we have outlined an analysis of the functioning of the Dutch primary-school teacher market.

[23] In the Dutch context this is the main reason put forward why students refrain from enrolling in a teacher education. Warps *et al.* (2010) results indicate that those students that were interested in choosing a teacher education, but eventually chose another study, primarily attributed this due to the perceived labor market opportunities and its related wages. For instance, the lack of career- and income-growth opportunities are perceived as major elements in the choice to waive a teacher education. It must be noted that these findings are likely to be related to the cognitive ability of individuals.

[24] The study by Bacolod (2007) concerns the US labor market, seizing IQ and AFQT test as proxy for the quality of the teachers. She estimates a multinomial logit model to determine whether teaching relative to alternative opportunities has caused the earnings of an individual, while allowing the estimates to alter over different quintiles of the ability distribution. To capture possible endogeneity, cohort- and labor market area fixed effects are included.

unionization effect. In fashion of Roy's (1951) model of occupational choice, it is hypothesized that compressing the pay-for-aptitude will drive the high-aptitude women out of teaching. Relatedly, increasing teacher wages for all aptitudes will attract more students to opt for a career as teacher, but not necessarily change the distribution of aptitude between occupations. Hoxby and Leigh find it is exactly the unionization effect that has been the principal determinant of the relative decline of high-aptitude women in teaching (in the US). This is the group of whom the relative pay loss has been highest.

In consideration of both alternative opportunities and compression of wages, Leigh (2012) empirically investigates whether changes in (relative) wages do have an effect on the quality distribution of inflowing teachers. His identification strategy hinges on within-state variation of teacher salaries over time in Australia, while combining it with the quality of inflowing students into teaching education. Also Leigh uses a proxy of test-scores of students ranging over a 15-year period for the quality of prospective teachers. As a broad measure of outside opportunities, the findings suggest that a 1% rise in teacher pay (relative to other occupations requiring a college degree) is associated with a 0.6% increase in the average cognitive percentile rank of prospective teachers. In addition, a higher wage-dispersion in alternative occupations is associated with a decline in teacher aptitude, as it is likely to increase the attractiveness for high-aptitude individuals[25].

It must be noted however, that the academic literature on average does not convey an uniform pattern of the responsiveness of individuals with higher cognitive aptitudes towards wages (e.g. Gilpin, 2012). But on average, those countries who pay relatively well in the relative wage distribution of teachers compared to other higher educated individuals experience less difficulty in alluring students to start studies to become a teacher and cope less/not at all with shortages of teachers (Waterreus, 2003).

B_2. Self-selection into teacher educations

Students deciding upon entering the Dutch primary-teacher education – the pabo- have traditionally been recognized by some dominant characteristics; being female, havo as highest previously attained education, and their relatively low-initial standards of math. The latter is an imperative peculiarity of prospective teachers, being recognized as less

[25] In a political economy setting, it is easy to see why increases in teacher wages tend to be hard to defend to the public. Since both ineffective and effective teachers get similar wage increases due to wage compression, there may be no clear evidence that salary increases are related to improved student outcomes.

capable students. In contrast, the more suitable students tend to self-select themselves into other specific vocational (HBO) educations (Webbink, 2000). Despite the fact these designated traits of pabo students reflect a grain of truth, they still miss the most imperative one to explain the logic behind these self-selection patterns; *risk-aversion.* Essential in this sense is the presence of organizational literature, which has directed its attention to the prevalence of incentives within organizations. As a general matter of fact, incentive structures largely determine the type of people being attracted to the organization. And accordingly, organizations can alter occupational characteristics in order to change the 'type' of their workforce.

Within the Dutch primary-school educational system, one element directly becomes clear; *the lack of incentives.* Primary-school teachers have (almost) no tools to differentiate themselves from others, earn wages dependent on their performances or career prospects to advance within a primary-school. Rather, primary-school teachers earn relatively fixed salaries –salaries which are severely compressed, making the presence of wage differentials very limited. Moreover, the risk of unemployment is fairly low[26]. All these specific traits of the occupation of a primary-school teacher induces a self-selection mechanism of a specific 'type' of individuals. Namely, directly related to the self-selecting patterns of individuals into occupations, is the decision of individuals to choose for specific education. To this end, researchers have aimed to relate the characteristics of students to the education they choose. As an example, Dohmen and Falk (2010) seized data of HBO-monitor to link the preferences, behavior and personality traits of Dutch students to the education of they have selected. Using a probit model, the results indicate that females are significantly more probable of becoming a teacher than males. Moreover, teachers are more probable of being careful, altruistic and open to new experiences. More importantly, there is a strong negative relationship between the probability of becoming a teacher and the *inclination to take risk*, being the strongest for students coming from the pabo. Prospective teachers seem to prefer fixed salaries rather than incentive systems with a variable wage scheme based on their productivity. Nielsen and Vissing-Jorgensen (2006) consociate this with an individual's aptitude to incur labor income risk. They show that the decision of Danish students to enroll in a certain education is partly determined by the (un)certainty of its associated

[26] Another factor for which the labor market for teachers is characterized is its large degree of part-time employees. In the Dutch primary-labor market almost two-third work part-time- a feature associated with the feminization of the labor market. Moreover, primary teachers can largely operate independently and have relatively favorable opportunities to combine labor- and private time (e.g. during school-holidays).

future income[27]. Obviously, this is also dependent on the risk of unemployment, i.e. people with a low aptitude for risk work less in occupations with a high change of becoming unemployed (see Fouarge *et al.*, 2011)[28]. Note that the so-called 'non-pecuniary' elements are relatively strong in the labor market for (primary-school) teachers and all have a tendency to substantially *mitigate (income) risks for individuals*.

Teachers are deemed to attach substantial value to the non-pecuniary job characteristics. Their intrinsic motivation is relatively large compared to their extrinsic motivation and are therefore inclined to incur a financial concession (Vogels & Bronneman-Helmers, 2006; Appendix B figure 7). At the start of their education, teacher education students already indicate they choose their study less so for the perceived wage opportunities relative to other vocational students (de Graaf & Heyma, 2014). [29]

In essence, one should note all former specific treats of the Dutch labor market for primary-school teachers propagates the *self-selection mechanism* of *risk-averse* students in teacher educations (see Dohmen (2010) for an excellent overview).

Even though pabo-students are in more risk-averse in general relative to their average counterparts, the educational backgrounds of individuals flowing into the pabo are not uniform. In this light it is important to ascribe the risk-aversion of individuals to their cognitive ability – in detail studied by Dohmen *et al.* (2010). These authors demonstrate that individuals with lower cognitive abilities tend to be associated with a larger degree of risk-aversion[30]. Yet even students who are different in their cognitive abilities but possess similar risk-aversion characteristics, are able to make different study choices. De Poala and Gioia (2011) show that the pool of relatively capable risk-averse students choose to enroll in teacher education systems due to relatively stable labor market prospects, whereas the relatively weak, risk-averse students rather decide to enroll in relatively less difficult education[31]. For the former group, positive labor-

[27] The authors use choice models to document the effect of labor income risk on the educational decisions.

[28] In addition, uncertainties concerning technological and demographic changes, (bad) luck and a person's state of health also constitute to labor income risk. The findings by Warps *et al.* (2010) may be surprising in this context. They show that Dutch (vocational) students choosing for alternative education other than teaching have a larger probability of choosing their education due to its perceived opportunity of finding a job.

[29] Estimates by Heyma *et al.* (2006) point out that the 'soft' labor conditions of teachers may compensate for at least 5% of the gross wage differences relative to the market sector. Also note that Dutch students are fairly well able to predict their starting salaries (Webbink & Hartog, 2004), which exemplifies teacher education students do 'rationality' choose to become a teacher.

[30] The authors employed a dataset of 1000 randomly selected individuals in Germany and conducted choice experiments accordingly to relate their cognitive ability to the risk-aversion. The cognitive ability is based on sub-modules of IQ-tests. Their estimates control for personal characteristics.

[31] The study focuses on Italian students, using multinomial logit models. Relatively less difficult education is here defined as the students' choice of a major during an economics study; i.e. the success rate students in majors represents the perceived difficulty.

market prospects and stable predictable incomes outweigh the possibility of dropping out during education- the reverse being true for the latter group. [32]

B3. The Trade-Off between the Quality and the Quantity of Teachers

The policy disposition to raise the quality of the education systems to become a primary school teacher has been gradually implemented since the course of 2006. There is a strand of economic analyses that have been invoked to probe the effects of these quality increments. As a general rule, economic theory resembles the tradeoff between quality standards and quantity of teachers (Gilpin & Kaganovich, 2012). In particular, raising quality standards induces a quantity decline of teachers, given the presumption of unchanging relative wages and alternative labor market opportunities. Economists have used both the former and the latter element interchangeably to examine the tradeoff.

Based on the evidence and notions above, teacher quality depends on the average wage in teaching relative to alternative occupations, the wage dispersion between high- and low-ability individuals opting for similar jobs, the non-wage characteristics in teacher jobs relative to alternative occupations, and the quantity constraints imposed by the government both in teacher education as in alternative education (Leigh, 2012). The quantity-constraints must be perceived as the required (quality) conditions which the pool of students must satisfy in order to graduate (or proceed during education). Following the former elements respectively, and particularly focusing on the mean quality of students in the teacher education, the average student quality is given by[33]:

$$\overline{TQ}^{TCH} = F(\overline{W}^{TCH}, \overline{W}^{ALT}, \frac{W_{High}^{TCH}}{W_{Low}^{TCH}}, \frac{W_{High}^{ALT}}{W_{Low}^{ALT}}, \overline{NW}^{TCH}, \overline{NW}^{ALT}, Q^{TCH}, Q^{ALT})$$

Where the superscript TCH represents the conditions of teachers, and ALT the alternative non-teaching conditions. First, considering a situation absent any changes in the quantity standards. The model by Leigh (2012) shows the partial derivative of student quality with respect to the enhancements in the average wage of alternative occupations will induce a decline in the average quality of (prospective) teachers $\frac{\delta \overline{TQ}^{TCH}}{\delta \overline{W}^{ALT}} < 0$. In quality-quantity terms; those individuals opting for alternative occupations are being replaced by individuals of a relatively lower ability, portraying the demand for

[32] In the context of the entry test at the pabo however, van den Enden and van der Wiel (2012) provided evidence that the discouragement effect was not especially strong for students with relatively weak pre-educational backgrounds.
[33] This formula is based on Leigh (2012).

teachers can be fulfilled simply when there are no quality standards deterring low-ability individuals to opt for a teacher education program.

Second, when diverting our focus to the effects of imposing quantity constraints on teacher education, the partial derivative is expected to be negative $\frac{\delta \overline{\mathrm{TQ}}^{\mathrm{TCH}}}{\delta \mathrm{Q}^{\mathrm{TCH}}} < 0$. Since the teacher quality is a function of the teacher quantity however, and entry-tests of the pabo are aimed to increase the quality standards of primary-school teachers, the model only implicitly conveys the idea that quality conditions will render a decrease in the amount of graduates. Whether the amount of enrollments at teacher educations also experience a drop remains ambiguous. For this reason, we have developed *a simple model in Appendix A,* which formalizes a students' decision to enroll at the pabo or not, and its ability to eventually obtain a degree. [34]

C. Student Dropout, Student Graduation and Selection 'after the gate'

During the past few decades, stakeholders of higher educational institutes have augmented their emphasis on student graduation rates as it provides a benchmark of institutional effectiveness (Fike & Fike, 2008). Mainly induced by the implementation of performance-based funding, education institutes have large incentives to encourage students to graduate within nominal study time. The general thought is that this will both benefit students -who have lower forgone earnings- and the institute, as nominal graduation leads to earlier funding (Sneyers & de Witte, 2015). From a policy perspective it is correspondingly natural to inquire what factors and conditions actually constitute to the dropout of students. The body of research has inferred many conditions which boil down to two perspectives.

First, at an institutional level factors like student-staff ratios and institutional size have been put forward as having an adverse impact on retention rates. As an example of the former, Tinto (2002) argues that a more frequent interaction between students and staff enhances the academic aptitude and persistence of students. Relatedly, if institutions expect more of their students, their involvement increases such that students value their learning more. The institutional size plays a crucial role in this

[34] Angrist and Guryan (2007) examine the effects of the imposition of high-stakes testing in a theoretical model. In the US, teachers can obtain a teacher certificate when they pass a high-stake test. They mainly show that within the potential pool of prospective teachers, the imposition of a test will enable education institutes to deter people to enter the labor market for teachers with insufficient skills. On the other hand, the most-skilled prospective teachers will be less willing to enter the market since the costs of effort is enlarged. When the image of teachers also increases though, the latter effect could potentially be outweighed such that more high-skilled are willing to become a teacher.

‘social engagement’ story as it is easier to generate an engaged environment within a smaller sized institution. The results by Calagno *et al.* (2008) confirm this notion.

Second, student characteristics are perceived as major determinants of an students academic success. For example, Scott *et al.* (2006) stress the gender gap in explaining why some (vocational) institutions have higher graduation rates. Women are referred to as to possessing psychological and biological factors that are related to attainment (McNabb *et al.*, 2002). It is suggested that females, due to their earlier puberty, make more conscious study choices in sooner stages, and therefore on average perform better than males. Moreover, in a study by Warps *et al.* (2010) who focus on factors determining the dropout of students in the first year of vocational education, the results indicate that the age of an individual is significantly associated with the change on dropout. For example, if students are older they are more probable of having doubled a class during high school. Another reason could be that older students already have completed an earlier vocational education of intermediate level (MBO), which enables them to search for jobs at a relatively lower level. [35]

In addition, the (initial) skill level of a student - exposed by average (math) scores in high school- is observed to be important for success of freshmans during college (Leuven *et al.*, 2010). Students with higher cognitive abilities seem to perform especially well when incentivized with extrinsic (financial) awards, whereas the opposite effect is true at students with lower cognitive abilities. Camerer and Hogarth (1999) ascribed these contrasting results to the interplay of (i) an incentive effect, which dependent on the cognitive ability of the student can result into a binding participation constraint, and (ii) an intrinsic motivation effect, which can be crowded out in case students do not achieve the incentive performance threshold. The consequences are students with higher abilities who were induced to perform better during the first year, do not slow down but benefit from the high-powered incentives during subsequent years. This is in sharp contrast to students with lower abilities, who have become discouraged and obtain less credits compared to the situation without any incentives.[36]

When specifically focusing on selective criteria -next to cognitive skills- which teacher education principals can seize as predictors of a student’s study success, mainly

[35] Other potential reasons for student dropout are for example the less intensive thinking about the study choice, and related the later choice of students. Moreover, teacher education students that dropout during their first year tend to have ascribed less value (utility) to their social responsibility and the tool as knowledge transfer, already in advance of their decision to enroll in teacher education (Warps *et al.*, 2010). Due data unavailability, it is not possible to control for these determinants.

[36] Leuven *et al.* (2010) study financial incentive effects of first year economics students of the cohort 2001-02. At the time, there were no (binding) dismissal policies, i.e. students were not obliged to obtain a certain credit threshold.

the importance of non-cognitive skills is exemplified in assessing the quality of (prospective) teachers. In a variety of countries, selection into teacher education is based on a number of interventions including; intake conversations, the necessity to write essays on questions and sometimes even demonstration lessons so the student can show its aptitude for the teacher occupation. In a study of Finnish students, Valli and Johnson (2007) show that the performance in the latter have the highest predictive power for students success during internships. The verbal competence of students thus seems to be imperative in practical courses during teacher education. In this paper, these kind of non-cognitive skills are not available in advance of the study, simply because the Dutch education principals do not select on these criteria. [37]

In contrast to the former situation, most (vocational) institutes and universities *have* introduced a selection mechanism 'after the gate'. Over the course of the 2000's an increasing amount of institutes aimed to enhance their effectiveness using academic dismissals (in Dutch 'Bindend Studie Advies'). Students are required to display a satisfactory progress during the first year of their studies, which is measured in terms of the obtained credits. Every course students take is linked to these credits, dependent on their workload. In case performance in a course is unsatisfactory, no credits are earned by the student. And if the total acquired amount of credits does not exceed a certain threshold, an academic dismissal (AD) is given to the student. In this situation, academic dismissals can thus be perceived as a high-stake incentive without any financial rewards.

In most studies investigating the effects of academic dismissals, researchers generally find that despite high attrition rates of college freshman's, the institutions effectiveness in terms of students graduating in time is larger (Lau, 2003). Arnold (2014) notes the AD policy has proven to be an effective tool to select the least capable students and prevent them from lingering through their studies. In other words, on the one hand the AD policy has increased the dropout rate during the first year, while on the other hand it has induced a rise in the completion rate of students graduating within 4 years. This thus shows that selection during the first year of college works as a guidance to retain a motivated group of students.

[37] The literature review by van der Rijst *et al.* (2015) indicate only a mix of disparate non-cognitive and cognitive characteristics of prospective teacher education students is able to predict a students' success during the study. It must be noted though that the evidence surrounding predictors of the success in teacher education studies is very limited.

IV. Identification Strategy

In an attempt to obtain causal inferences of the implementation of the mandatory entrance tests during the first year of the pabo the synthetic control method for comparative case studies will be used. Originally coined by Abadie and Gardeazabal (2003), and later further popularized by Abadie *et al.* (2010), this method provides a design to select comparison 'units' to construct a counterfactual in case an event or intervention at the unit of interest would not have occurred. At first glance, this is in similar vein as the importance of selecting 'valid' comparison units in differences-in-differences settings. The latter which premise mainly relies on finding controls units with a similar pre-intervention trend, building on the assumption that in absence of the intervention the treated unit would have followed the same trend akin to its comparison unit(s). According to Abadie (2005) even then comparison units have to be sufficiently similar to the unit of interest, otherwise obtained differences during the intervention may simply reflect disparities in their characteristics[38]. In this line of thought, the synthetic control method is based on a formal, data driven design to construct one combined 'synthetic' control unit, using a weighted combination of control units ($\sum 1$)- where the weights are designated to control units dependent on how well they resemble the treatment group during the pre-intervention period. That is, the more the control unit resembles the treatment unit during the pre-intervention period, the higher the weights are given to this control unit. In this way, the method is typically better able to reproduce similar characteristics as the unit of interest relative to single comparison units – this is thought to render a more 'reliable' counterfactual and to reduce the ambiguity about how control units are chosen. During the post-intervention period, we simply compare the 'synthetic' counterfactual to the actual outcome variable, which gives us the estimate of the treatment effect. In the following, a formal overview will be provided to explain this method in the case of the pabo intervention.

A. The Synthetic Control Method

Suppose there are $J + 1$ studies in higher vocational education in the Netherlands; the treated unit (pabo) represents unit $j = 1$ and studies $j = 2$ to $J + 1$ are units that constitute the donor pool of potential controls. Moreover, suppose the studies were in place in the pre-intervention period T_0 as well as the post-intervention period T_1. The

[38] This notion is merely based on unobserved differences that change over time. Relatedly, selection for treatment may be influenced by transitory (income) shocks of an individual, referred to as Ashenfelter's dip (see also Geddes (2003).

only unit exposed to the intervention is the pabo and the intervention does not induce any anticipation effects affecting outcomes in pre-intervention periods. In addition, It is assumed the units do not interfere with one another. As stated above, the synthetic control paradigm can be defined as the construction of a counterfactual of the unit of interest based on relevant characteristics. Let X_1 define the vector of characteristics specific to the pabo *ex-ante* the intervention. Additionally, let X_0 define a $k \times J$ matrix of characteristics of studies that constitute the donor pool. In order to most closely resemble the main features of the pabo (X_1), the X_0 matrix will be seized using a weighted average of all the studies in the donor pool. Formally, the 'synthetic pabo' can be defined by a $J \times 1$ vector of weights designated to each higher vocational education study $W = (w_2 \dots, w_{J+1})'$, where the weights represent nonnegative numbers $0 \leq w_j \leq 1$ and add up to one $w_2 + \cdots + w_{J+1} = 1$. In general sense, the vector of weights can be determined in such a manner that specific features of the pabo can be resembled closely. One may note however that within the X_1 vector and X_0 matrix, not all pre-intervention characteristics are of equal importance as determinant of the outcome of interest. For this reason, the discrepancy vector $X_1 - X_0W$, measuring the differences in pre-intervention characteristics, is minimized with respect to W in such a way it reflects the predictive power of each particular characteristic. In practice, let m represent $1, \dots, k$ characteristics and X_{1m} respectively X_{0m} reflect the value of the m-th variable for the pabo and the study in the donor pool (the latter being a $1 \times J$ vector). In resonance to Abadie and Gardeazabal (2003), the relative importance assigned to the m-th variable is displayed by v_m. The discrepancy vector is operationalized in the following manner:

$$\sum_{m=1}^{k} v_m (X_{1m} - X_{0m}W)^2$$

which is minimized by selecting the optimal weight for each study in the donor pool $W^* = (w_2^*, \dots, w_{J+1}^*)$ and the optimal predictive power of each characteristic $v_m{}^*$. The procedure thus searches among all matrices of V and sets of W-weights in order to minimize the distance in pre-intervention outcomes variables between the pabo and the 'synthetic' control group – coined by Abadie and Gardeazabal (2003) as the *mean squared prediction error (MSPE)* of the synthetic control estimator. [39]

[39] In technical sense, the vector V is chosen among positive definite and diagonal matrices such that the mean squared prediction error of the pre-intervention outcomes is minimized (see Abadie and Gardeazabal (2003; appendix) for further technical details.) In all specifications, the option nested is employed in the selection of control units. This is an optimization procedure that searches among all positive semidefinite and diagonal matrices of V and sets of W-weights in order to approximate the 'best' fitting *convex* combination of control units.

Each of the characteristics are meant to have predictive power on post-intervention outcomes at the pabo. It is important to note that the pre-intervention 'matching' characteristics may include values of the pre-intervention outcomes of the pabo. That is, one is able to construct a 'synthesized' outcome variable using a weighted combination of other studies in order to (virtually) perfectly be able to replicate the actual outcome variables of the pabo in the pre-intervention period. For this reason, to obtain causal inference of the intervention on the outcome of interest, simply the difference between the post-intervention outcomes of the pabo and the 'synthetic' pabo' is taken[40]. Let Y_1 represent a $(T_1 \times 1)$ vector of the pabo on the outcome of interest. Relatedly, let Y_0 define a $(T_1 \times J)$ matrix of post-intervention outcomes of 'synthetic' pabo. The synthetic control estimator can be described as follows:

$$Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$$

where t represents the post-intervention period ($(t \geq T_0)$.

As with most econometric methods, an illustration provides the reader the most intuitive way of understanding. In this case, the peculiarity of the pabo in terms of characteristics of entering students is shown in Table 2. Suppose we are interested in the effect of the intervention on the evolution of enrollments at the pabo. Column (1) reports the average pre-intervention characteristics of the pabo. Using the procedure described above, column (4) reports the average pre-intervention characteristics of the "synthetic" pabo. The synthesized unit has (approximately) the same evolution in pre-intervention enrollments as the pabo (figure 2), yet remarkably differs in a number areas from the pabo. Although not part of the procedure, column (7) illustrates the magnitude of these pre-intervention differences. Notable are the discrepancies in gender rates, the age, the degree of non-natives and perceived labor market opportunities. This indicates that, given the other covariates, these variables do not have substantial predictive power in determining the evolution in enrollments of the pabo before the intervention was implemented *within this specific procedure*. If we still build on the conjecture that these characteristics have a (small) predictive power in determining the

[40] One of the main conditions of the synthetic control method is the availability of a balanced panel during the entire evaluation period. Due to the unavailability in institutional-study -and labor market data in the year 2002, we use a balanced panel of 2003 to 2009 as input in all specifications to construct a synthesized control unit. In practice, the outcome variables of the pabo in the years 2003, 2004 and 2005 were included as predictors for the software to construct a synthesized control unit that is virtually perfectly able to replicate the pabo. In addition, the outcome variable of the year 2002 was included as a separate covariate for each study. That is, throughout the entire panel duration, this covariate reflects the outcome of 2002 for each study. In this way, while including the covariate in the matching procedure, we are able to additionally 'synthetically' replicate the outcome variable of the pabo in 2002. Note that the discrepancy between the pabo and the synthetic unit thus does not show up in the *mean squared prediction error*.

enrollment paths of studies, then the discrepancy between the pabo and the 'synthetic' pabo should not be much of a concern as long the synthesized group does not differ noteworthy over time. Column (6) displays the differences between the pre- and post-intervention covariates of the synthetic unit. Although most of the covariates differ over time, these are often very limited. Therefore, as the observable characteristics of the

Table 2 - Descriptive Statistics

	pabo			*"Synthetic" pabo*			(1) - (4)	(2) - (5)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pre-Mean	Post-Mean	Mean Dif	Pre-Mean	Post-Mean	Mean Dif	Pre- Dif	Post-Dif
Panel a) Total sample								
Sex	0.850	0.837	-0.0126**	0.326	0.343	0.0174	0.524***	0.494***
	(0.00468)	(0.00503)	(0.00401)	(0.0647)	(0.0902)	(0.0376)	(0.0649)	(0.0903)
Vwo	0.114	0.111	-0.00240	0.108	0.0987	0.00896	0.00599	0.0133
	(0.00584)	(0.00791)	(0.00466)	(0.0150)	(0.0186)	(0.00817)	(0.0161)	(0.0203)
Mbo	0.439	0.451	0.0113	0.585	0.584	-0.000590	-0.145**	-0.134*
	(0.0111)	(0.0158)	(0.0111)	(0.0484)	(0.0610)	(0.0310)	(0.0497)	(0.0630)
Havo	0.447	0.438	-0.00893	0.308	0.317	0.00955	0.139***	0.121**
	(0.0112)	(0.0137)	(0.0104)	(0.0348)	(0.0435)	(0.0237)	(0.0365)	(0.0456)
Age	22.04	20.78	-1.265***	23.15	22.88	-0.264	-1.129	-2.107**
	(0.343)	(0.299)	(0.288)	(0.900)	(0.734)	(0.271)	(0.961)	(0.791)
Native Dutch	0.870	0.882	0.0115	0.789	0.787	-0.00205	0.0817***	0.0953***
	(0.0162)	(0.0126)	(0.0107)	(0.0171)	(0.0229)	(0.0140)	(0.0234)	(0.0260)
Western Immigrant	0.0534	0.0516	-0.00181	0.103	0.116	0.0131	-0.0495***	-0.0644**
	(0.00350)	(0.00520)	(0.00507)	(0.0126)	(0.0191)	(0.0117)	(0.0131)	(0.0198)
Non-Western Immigrant	0.0765	0.0668	-0.00965	0.109	0.0976	-0.0110*	-0.0321	-0.0309*
	(0.0139)	(0.0108)	(0.00837)	(0.0101)	(0.00885)	(0.00480)	(0.0171)	(0.0138)
Direct Entrant	0.735	0.807	0.0714***	0.640	0.650	0.00983	0.0973*	0.156***
	(0.0192)	(0.0164)	(0.0139)	(0.0350)	(0.0265)	(0.0196)	(0.0397)	(0.0311)
Graduate (hbo or wo)	0.0819	0.0811	-0.000855	0.0784	0.0691	0.00929	0.00335	0.0123
	(0.0105)	(0.0103)	(0.00681)	(0.0103)	(0.0118)	(0.0139)	(0.0146)	(0.0156)
Academic Dismissal	34.37	28.82	-5.548***	35.62	41.79	6.166***	-1.080	-13.87***
	(0.392)	(0.175)	(0.379)	(0.948)	(1.261)	(0.616)	(1.020)	(1.271)
Numerus Fixus	0.00652	0.0307	0.0242***	0.0346	0.0301	0.00454	-0.0277**	0.00144
	(0.000293)	(0.000199)	(0.000398)	(0.00983)	(0.00591)	(0.00384)	(0.00984)	(0.0102)
Labor Market:								
Outlook	3.687	3.713	0.0269	2.959	3.002	0.0428	0.708***	0.748***
	(0.0141)	(0.0104)	(0.0171)	(0.0925)	(0.192)	(0.152)	(0.0936)	(0.192)
Alternate Occupations	1	1	-	3.207	3.496	0.288	-2.207***	-2.496***
	(.)	(.)		(0.0906)	(0.269)	(0.198)	(0.0906)	(0.269)
Cyclicality	1	1	-	3.095	3.354	0.259*	-2.095***	-2.354***
	(.)	(.)		(0.127)	(0.224)	(0.121)	(0.127)	(0.224)
Student-Staff Ratio	13.73	13.78	0.0488**	12.94	12.93	-0.00416	0.796	0.818*
	(0.00956)	(0.00592)	(0.0241)	(0.467)	(0.389)	(0.167)	(0.467)	(0.389)
N	27873	29955	57828	200452	288755	489207	228325	318710

Continues on next page

Table 2 Continued - Descriptive Statistics

	pabo			*"Synthetic" pabo*			(1) - (4)	(2) - (5)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pre-Mean	Post-Mean	Mean Dif	Pre-Mean	Post-Mean	Mean Dif	Pre- Dif	Post-Dif
Panel b) havo grades								
Final Average	6.564	6.553	-0.0111*	65.79	65.85	0.0554	-0.150	-0.328**
	(0.00920)	(0.0101)	(0.00540)	(0.105)	(0.0566)	(0.117)	(0.138)	(0.117)
Dutch	6.620	6.513	-0.107***	6.454	6.417	-0.0371	0.163***	0.0971***
	(0.0203)	(0.0147)	(0.00939)	(0.0163)	(0.0250)	(0.0271)	(0.0260)	(0.0289)
English	6.293	6.253	-0.0400**	6.396	6.426	0.0299	-0.104	-0.171*
	(0.0348)	(0.0368)	(0.0136)	(0.0638)	(0.0652)	(0.0228)	(0.0725)	(0.0746)
Mathematics	6.604	6.656	0.0520***	6.438	6.486	0.0484*	0.167***	0.168***
	(0.0184)	(0.0192)	(0.0116)	(0.0201)	(0.0113)	(0.0224)	(0.0271)	(0.0222)
Panel c) vwo grades								
Final Average	67.07	67.19	0.121	66.45	66.42	-0.0294	0.635*	0.777**
	(0.217)	(0.158)	(0.145)	(0.209)	(0.150)	(0288)	(0.298)	(0.217)
Dutch	6.836	6.752	-0.0838***	6.655	6.511	-0.144	0.180**	0.241***
	(0.0273)	(0.0202)	(0.0211)	(0.0561)	(0.0386)	(0.0905)	(0.0620)	(0.0436)
English	6.438	6.394	-0.0444	6.403	6.143	-0.260	0.0377	0.250*
	(0.0301)	(0.0394)	(0.0304)	(0.0750)	(0.119)	(0.161)	(0.0805)	(0.125)
Mathematics	6.496	6.527	0.0312	6.339	6.348	0.00891	0.158**	0.179**
	(0.0355)	(0.0250)	(0.0297)	(0.0486)	(0.0602)	(0.0553)	(0.0597)	(0.0649)

Notes: Descriptive statistics as denoted in panel a) based on most complete synthetic matching procedure, including controls on the individual-, study-, institutional- and city-level. Panel b) and panel c) hinge on the most complete procedure in the subsample of individuals with respectively havo and vwo as highest completed pre-education. Only the most relevant characteristics are shown. Individuals without an explicit previously completed education certificate are labelled as having mbo as highest previously attained education. Mean differences are based on regressions of characteristics as dependent variable and an time dummy indicating the post-intervention period as explanatory variable. Pre- (post-)intervention differences based on simple regressions with characteristics as dependent variable and a pabo dummy as explanatory variable. The means of characteristics of the synthetic unit are based on the assigned study weights, these are reassigned to the individual level. Descriptives are reported with standard errors in parenthesis, clustered at the institutional-study cohort level. Significance at the 10, 5 and 1 percent levels are respectively denoted by * / ** / ***.

synthetic unit endure limited changes, we could deduce that the *unobserved* differences are limited as well. The only real concern could be the large upheaval in the imposed ECTS standard of academic dismissals and the improvement in labor market opportunities in the post-intervention period. Relatedly, the discrepancy of pabo pre- and post-intervention characteristics are shown in column (3) – discrepancies which may be expected as the treatment could induce an alteration in the characteristics of people being attracted to the pabo. After the intervention, pabo entrants become younger, more often directly enter the pabo (instead of taking a break between), and have relatively lower (higher) final average grades dependent whether they completed havo (vwo) as highest pre-education.

It is imperative to note that the descriptive statistics as denoted in panel a) of table 2 are based on the most complete specification in the synthetic matching procedure considering the indexed change in enrollments outcome variable. For each specific outcome variable and each separate specification, we use the synthetic control procedure to construct a different ‘synthetic‘ pabo in order to best approximate the pabo outcome variable during the pre-intervention period[41]. The most complete specification include covariates on the (aggegrated) individual level, the study-level and the institutional (and city)-level. We label this specification as specification 4. The remainder of specifications (1, 2 and 3) respectively build up to this specification, where specification 1 does not include any covariates in the synthetic matching procedure, specification 2 includes covariates on the aggregated individual level and specification 3 includes controls on the aggregated individual and study-level. The former specification procedure has also been used in subsets of aggregated individuals with a similar pre-education (i.e. mbo, havo or vwo). In practice, this implies that the covariates inherent to the subgroup were used as an input in the synthetic control method. Furthermore, the relative proportions of each subgroup (e.g. 45% havo students) within a study were used in specifications (2-4) to take account of potential peer effects. In this way, the 'synthesized havo subsample' pabo also hinges on a 45% composition of havo students entering the study in the post-intervention period.

B. Assumptions of the Synthetic Control Method

The question whether the main assumptions of the synthetic control method can be maintained is essential to the extent how much we can trust our identified results. The synthetic control method relies on three main assumptions. We will provide in depth information about the validity of each of them.

i) The only unit exposed to the intervention is the pabo and the intervention does not induce any anticipation effects affecting outcomes in pre-intervention periods. Anticipation effects could arise when students in the years precluding to the implementation of the entry-tests strategically recognize the increased cost of effort and therefore, instead of taking a break of one (or multiple) year(s), directly enter the pabo, such they prevent the necessity of having to pass the entry-tests during the first year. The data however do not support this hypothesis; in the precedent intervention years the amount of direct entrants was lower than in the post-intervention years. This is likely due to the fact the students could simply not anticipate the implementation of the tests, since only at the beginning of 2006 the decision was made to conduct a standardized math test. Therefore, anticipation effects are plausibly absent. The decrease in pabo enrollments before the tests were implemented could primarily be

[41] The descriptive statistics of the 'synthetic' pabo of other outcome variables and/or other specifications are available upon request.

attributed to the deterioration of the image of primary school teachers, which in general were perceived as insufficiently competent. (The labor market was not yet of an issue of the years around 2006.)

ii) It is assumed the units do not interfere with one another. Strictly spoken, the assumption of no interference between units would imply only studies where individuals choose to enroll at the particular study without an alternate study choice would constitute the donor pool of studies. Obviously, there is always some sort of ordering in study preferences of students, violating this assumption. The assumption therefore must be seen as more implicit; it assumes the outcomes of untreated units are unaffected by the intervention. This is the prime reason why only the following five studies are removed from the sample of potential donor units; (i) Pedagogiek (ii) Sociaal pedagogische hulpverlening (iii) Maatschappelijk werk en dienstverlening (iv) Social work and (v) Verpleegkunde (hbo). These are the traditional alternate studies that students choose who waive a pabo education (see e.g. Warps *et al.*, 2010). Moreover, pabo students that switch after one year mostly choose to enroll in one of these studies (apart from the pabo at other institutions).

iii) In the original synthetic control method for comparative case studies, the pre-intervention characteristics are supposed to have predictive power on post-intervention outcomes. In light of the pabo, this implies the assumption of time-invariant characteristics of both students, the study and even the future labor market. Critics may point out whether this assumption can be maintained – and rightfully so as shown by columns (3) and (6). For instance, studies in higher vocational education in the Netherlands have gradually implemented academic dismissal policies, possibly inducing a surge in switch and dropout rates. Furthermore, the allurement of starting a study is usually dependent on its perceived labor market opportunities. Changes in the latter might influence the inflow into a study.

Controlling for Potential Biases - If one fails to correct for these variations over time, it is conceivable there is a bias in the synthetic control estimator. Consequently, while employing the weights obtained in the synthetic control method, standard differences-in-differences models have been estimated to overcome possible biases. Note that these biases would be primarily induced when the synthetic unit endures unusual changes in terms of (un)observable characteristics. Then, sudden developments in the counterfactual could simply reflect differences in these characteristics (i.e.

(un)observables affecting outcomes). As we saw in column (6), observable changes are limited, and we could assume the unobservable differences are limited as well.[42]

In the remainder of all equations, the weights designated to each study j have been reassigned to respectively the level of the study j at a specific institution s, and the level of an individual i who enrolled in a study j at institution s, dependent on its respective size within the assigned study weight. In this way, the original weights designated to each study once again sum up to one for both the pabo as the 'synthetic' pabo within each enrollment cohort.

In the following, the estimated equations will be provided in the natural order of occurrence. To start off, the response variable is the logged enrollment of institution s at study j in the inflow cohort year t. To measure the enrollment effects of the intervention, the following functional form will be estimated:

$$\log\ (enrollment_{jst}) = \alpha_j + \delta_t + \beta D_{jt} + X'_{jst}\gamma + Z'_{jt}\eta + Q'_{st}\theta + \varepsilon_{jst}$$

where D_{jt} is the treatment variable, indicating an interaction variable which equals zero *ex-ante* the intervention period and reverts to one only at the pabo in the post-intervention period ($T_0 = 2003, \dots, 2005$), ($T_1 = 2006, \dots, 2009$). On a related note, X'_{jst} represents a vector of aggregated personal characteristics of enrolled students, Z'_{jt} reflects a vector of characteristics specific to the study, whereas $Q'_{st}\theta$ symbolizes a vector of characteristics specific to the institution and its city. Study fixed effects α_j are included to control for time-invariant effects of the study and year fixed-effects denoted by δ_t are included to take account of shocks occurring at each year. The standard errors ε_{jst} will be clustered at the institutional level to take account of student choices influencing one another.[43]

When individuals have enrolled at a specific study j at institution s, there are three possibilities that can arise at the end of the first year. First, the individual is able accomplish all required conditions to be granted permission to enroll at the second year of the study. Second, the individual fails to accomplish all conditions and chooses to

[42] Despite the fact that we know other studies did not implement a standardized tests of similar magnitude during the first year of the study, we cannot guarantee that studies in the donor pool gradually increased their quality standards during the first year. This would likely render a decrease in the retention rate of studies, thereby inducing a downward bias in the estimates of the treatment effects. The sets of figures later on in this paper however, seem to indicate that the 'synthesized' retention rate and graduation rate efficiency of the pabo are relatively constant during the post-intervention period, therefore making it unlikely that other studies have introduced sudden measures that affect these outcome variables. All in all, if such interventions would not be captured by quality standard proxies (e.g. #ECTS for academic dismissal), then still the estimate is more likely to be a conservative estimate of the treatment effect rather than an overstated treatment effect.

[43] Warps *et al.* (2010) indicate the enrollment decision of pabo-students relative to their HBO fellows is peculiar in the sense that their choice to enroll at specific pabo-institutions is consciously determined by the institutions' reputation, its perceived quality and atmosphere, and less so by the attractiveness of its city and student life.

enroll at another study in the subsequent year. And third, the individual fails to accomplish all conditions and chooses to completely dropout of higher vocational education[44]. The first possibility will be estimated in a probit model, where the binary variable is respectively defined as one in case the individual enters the second year, and zero otherwise[45]. When estimating binary outcome variables, the estimated error term is dependent on its chosen functional form in terms of a cumulative distribution function (cdf). The probit model induces a normally distributed cumulative function, i.e. $p = P[y_{ijst} = 1 | x_{ijst}] \sim N(0,1)$. It merely treats individuals who switch respectively dropout as successes, and individuals who do not as failures. The estimated probit model will encompass the same covariates as equation x, yet will be estimated at the individual level i:

$$P[y_{ijst} = 1 | x_{ijst}] = \Phi(x_{ijst}\beta)$$

where Φ denotes the standard normal cumulative distribution function. Since the estimation outputs of probit models cannot be directly interpreted, average marginal effects (AMEs) will be estimated subsequently to obtain causal inferences, defined as:

$$AME = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial E[y_i|x_i]}{\partial x_i}$$

which describes the populations' averaged partial derivative of the conditional mean of y with respect to a predictor covariate. Put differently, it measures for each individual the change in the probability of switching respectively dropping out due to a one unit change in the predictor variable, while controlling for all other covariates. And correspondingly averages the sum of marginal effects. Moreover, since the assumption that the error term is independently distributed can easily be rejected, the error term is clustered at a class cohort level (a particular study j at institution s at time period t).

Following the chronological order of occurrence and taking a giant leap forward, students that do not dropout or switch during the first year are able to enroll at the second year of the study. If these students do not dropout or switch in the remaining

[44] Note that dropout is denoted as the individual not enrolling in the subsequent year.

[45] The retention rate of a study within an institution is the mirror image of the sum of the switch and dropout behavior of individuals. The switch and dropout behavior of individuals have also been separately estimated in probit models, where individuals are respectively labelled as one whey they switch to another study or another institution, and as one when they completely dropout of higher vocational education, and zero otherwise. *These probit models are available upon request.* Note that within these probit models, in the group labelled as zero, there is also a group that *actually* represents a third category. In other words, this is a multidisciplinary category. Therefore, the covariates (e.g. grades, nationality) are unlikely to display the valid effects on dropout (switch) when individuals that switch (drop out) are present in the dataset of individuals. For this reason, one may suggest a multinomial logit estimation, however unfortunately, this is not possible with the synthetic control method since each outcome variable procedure produces its own weights of studies. Averaging these weights would produce a counterfactual with non-similar trends on each of the outcome variables. Also the weights of the retention counterfactual alone do not produce similar trends on dropout and switch, which would make a multinomial logit severely biased. The direction of the bias of the estimated average treatment effects of switch and dropout depend on its relation with the covariates. In the current specifications, the switch and dropout estimates must be perceived as a rough decomposition of the retention estimates.

(three) years of the study, they will eventually be able to obtain their degree[46]. In order to estimate the marginal causal effects of the intervention on the percentage students obtaining a degree within the nominal plus one year study duration a similar approach will be conducted as the switch and dropout probit models. Now let $P[y_{ijst} = 1|x_{ijst}] = \Phi(x_{ijst}\beta)$ define the probability of an individual obtaining its degree within nominal plus one study time. The computed marginal effects will indicate the change in the percentage students obtaining a degree within nominal plus one study duration due to a one unit change in a predictor covariate.

C. Data Description

In this paper a combination of two main databases has been used. First, we use student-level data of each individual ever having been enrolled since 2002 at a specific bachelor study in higher vocational education (hbo) in the Netherlands. Specifically, in the remainder of this paper an enrollment is defined as a student which also enters the first year of the particular study[47]. The so-called '1cijferho' database was made available by Dienst Uitvoering Onderwijs (DUO), and contains information about the gender, age, ethnicity, migration generation, nationality, enrollment form (full/part-time or dual), first-year of enrollment (inflow) and characteristics about the highest previously attained educational level of each entity. The latter could comprise three distinct education forms; mbo, havo or vwo. In case of the first, the individual already has completed a study of the intermediate-level vocational education form. In addition, individuals that are not in the possession of an explicit havo or vwo diploma, are labelled as having mbo as highest completed pre-education -even in cases when data is missing about the mbo study type inherent to the individual. For mbo students with explicit mbo degrees, we are only able to control for the mbo study and its type. [48]

More specific information about the cognitive ability is available for students coming from the second and third track. For these groups, we are able to use the average (final) grade as a proxy for the cognitive ability of individuals. In addition, the (final)

[46] During the time period of interest, the vast majority of studies had a nominal duration of four years; students were required to acquire 240 credits to obtain a degree.
[47] Strictly speaking, students are able to subscribe to multiple studies before the choice is made to (*de facto*) also enter one of the subscribed studies. We do not observe the amount of subscriptions of a student, only *de facto* enrollments are considered.
[48] The study type has been classified in specified fields dependent on their attached future labor market. In total there are nine of possible fields to which mbo studies belong (e.g. economics, technique or healthcare). Note that this labelling induces the mbo subgroup to comprise of a substantial amount of individuals without an explicit mbo degree (about 20%), within the subgroup of mbo students, these individuals are labelled as a remainder study type.

grades of different subjects (the Dutch language, mathematics and English[49]), and the profile type[50] are employed as separate cognitive ability proxies. With regards to the (havo) students that first choose to complete a mbo study before enrolling in higher vocational education (about 4.2%), these are labelled as possessing havo as highest pre-education, while controlling for the fact the student, actually, has a double degree.

For each individual, we only observe the first enrollment in the first bachelor year (propedeuse) of a study in higher vocational education. More specifically, for each individual we observe which study it started, at what year, and after that one year, whether it switched to another study or completely dropped out of higher vocational education. There are two options to be labelled as a switcher i) the student enrolls at a study different from the study chosen in the precluding year and ii) the student enrolls at the same study (first year) but has changed from institution. An individual is indicated to be a dropout when it does not chooses to enroll in higher vocational education in the subsequent year. Additionally, if a student eventually was able to obtain a degree, the year of graduation and study code was specified. Only students of whom the enrollment is principal are considered in this paper. Note that we do not observe whether a pabo student switched respectively dropped out due to its failure at the entry-test. Implicit to our identification strategy therefore, is that these individuals are not able to enter the second year of the pabo, *due to their lack of cognitive abilities*, and not due to a sudden surge in pabo individuals that find out that their match with their future job is poor, (but still are able to pass the entry-tests). There are two factors that validate our approach: i) There were no alternative policy changes at the pabo during the intervention period 2006-2009, apart from the introduction of the academic pabo in 2009. The only change during the intervention period therefore, is the necessity to pass the entry-tests in order to enter the second year. ii) The overall pass rates of the math- and language tests respectively were fairly constant throughout time (see Appendix B figure 2). Since the retention rates of pabo students fluctuate between enrollment cohorts, we have to explicitly assume that the 'unobserved' part (e.g. flawed match) of pabo students not passing through the second year also fluctuates over time. The results we later obtain therefore, must be conform this pattern, i.e. a quite constant treatment effect over

[49] Unfortunately, there were a substantial amount of missing data concerning the grades of the indicated subjects for individuals. In case a grade was missing, the average subject grade of other students with available grades *within the specific enrollment* cohort was designated to the individual. Then also a dummy variable indicating the grade was missing was added as a control in the equations.
[50] After the third year in secondary education, students are obliged to choose a profile. Generally there are four profiles students can choose, however individuals are able to complete a double profile in case they successfully graduate in more subjects.

distinct cohorts. More specifically, the estimations should be in line with the observed failure rates per pre-educational subsample for both the math- and language tests.

The second main dataset is the so-called Nationale Studenten Enquête (NSE) and has been made available by Studiekeuze123 ranging over the time period 2003 to 2015. It comprises information on a wide range of factors both on the study-level and the study-institution level. Before 2008, the data was gathered by Centrum Hoger Onderwijs Informatie (CHOI), thereafter the data collection was powered up and conducted by Studiekeuze123[51]. The data on the study-level are used to proxy for the allurement of starting a particular study. Since this is generally linked to the perceived opportunities on the labor market, we control for the labor market outlook (i.e. perceived opportunities), the cyclical dependency of the occupation, and whether the study focus' is on a single occupation or the graduate is able to practice a wide set of occupations. All former variables range from one to five, where one respectively reflects the worst outlook, the least dependency of the business cycle and the most narrow set of possible occupations to practice.

On the study-institutional level controls are included to proxy for the (perceived) difficulty of a study at an institution. These controls entail the number of credits deficit at which an academic dismissal is given to the student, the availability of a numerus fixus, the availability of decentralized selection, the student-staff ratio and number of employees of the institution. On a less formal note, also controls have been included to check for a possible disparate attractiveness of the cities, these include student evaluations (1-10) on both culture and hospitality (i.e. bars and restaurants) and the amount of students relative to the total amount of inhabitants.

As outlined in the identification strategy the relevant sample of studies should hinge on two main requirements i) *balanced panel:* studies both comprise a complete enrollment duration and a complete set of available covariates during the sample period (2003-2009)[52], (ii) *no interference:* studies do not interfere with one another[53]. In case a study did not conform to either of these criteria, the study was removed from the sample. The end result is a sample of 562.466 students over the identification period.

[51] Since 2008, the survey is conducted multiple times a year. In this paper, only data of the last survey of each year is used to construct a panel data of higher vocational education studies throughout the years. Later these were merged to the individual level for each enrollment cohort per year.

[52] Since the enrollment specifications were estimated at the institutional-study level, here the additional condition was imposed that the institution provided the study program during the relevant sample period. Also the condition that the study had at least 100 aggregate entrance enrollments in each enrollment cohort was imposed. In the robustness estimates, more strict conditions were imposed on the institutional level. That is, the institution should have at least 20 entrance enrollments in each enrollment cohort.

[53] The studies as mentioned in assumptions of the synthetic control method are removed from the sample of potential donor units.

V. Results

A. Total Sample

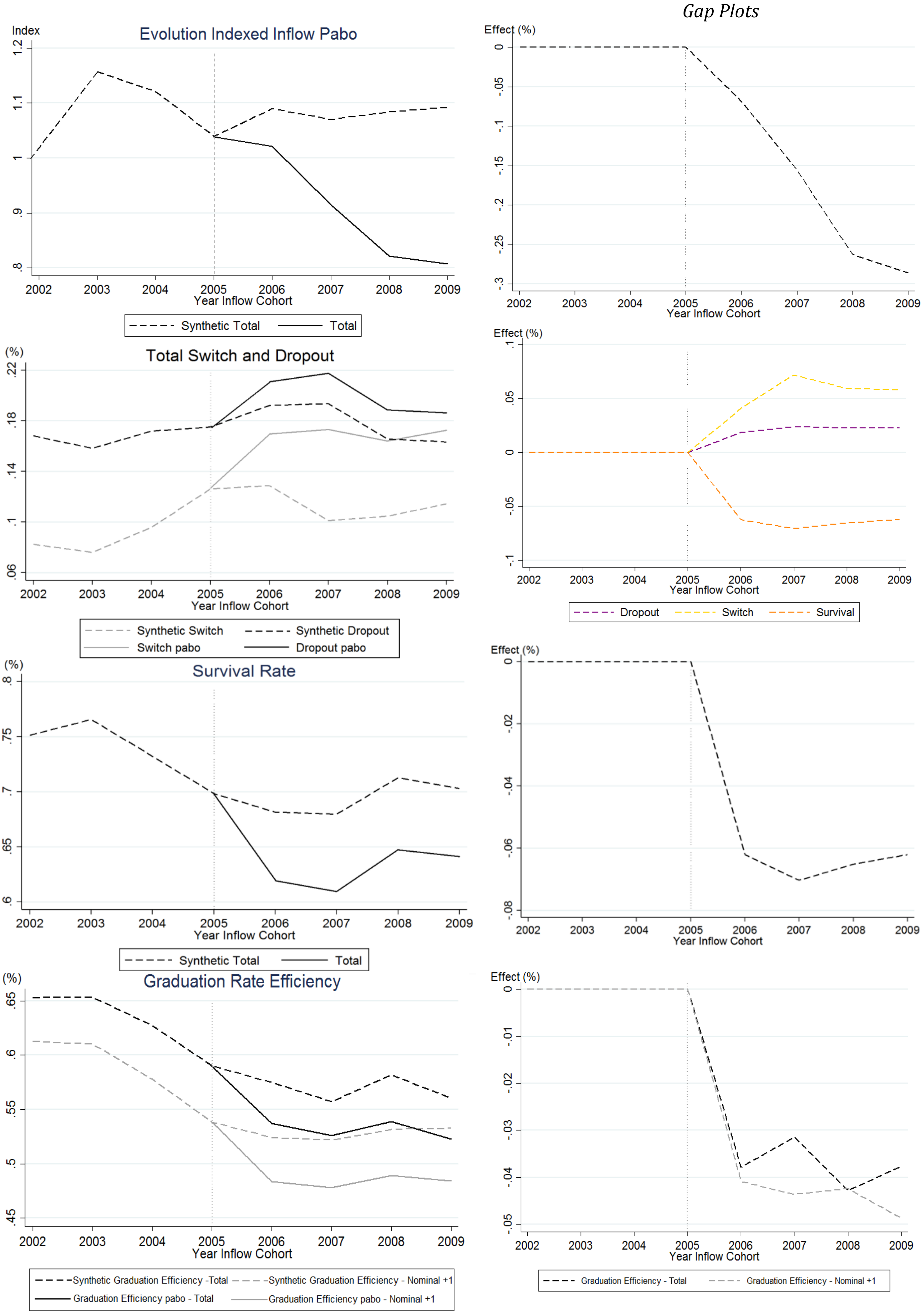
Plots of the most complete synthetic control procedures on the indexed evolution of pabo enrollments, switch and dropout rate, retention rate and graduation rate efficiency are respectively presented on the left-hand side of the subfigures in figure 2. The dashed lines show the virtually perfect replication of the outcome variables before the entry-tests were implemented during the pre-intervention period of 2002-2005[54]. The solid lines exemplify the *de facto* outcome variables of the pabo after the entry-tests were implemented. In contrast, the continuation of the dashed line after 2005 illustrates the counterfactual of the pabo with regard to the specific outcome variables. In other words; what would have happened at the pabo if the intervention *would not* have occurred? The simple differences between the two -the actual outcomes at the pabo and the 'synthesized' outcomes of the counterfactual- are visualized on the right-hand side of the subfigures in figure 2. All of the figures represent the entire relevant sample. Later on, the outcomes of the synthetic control procedures will be exhibited for the subsamples of individuals with respectively a highest attained previous education of mbo, havo or vwo.

The top left figure clearly shows a gradual surge in the indexed amount of entrant enrollments at the pabo. As the years passed by, an effect what was earlier coined by van den Enden (2012) as the '*discouragement*' effect, progressively developed into a 30% point decrease in enrollments in 2009. The pattern is thus evoked that students have augmented their unwillingness to start a study at the pabo due to the perceived increased costs of effort. Obviously one may inquire whether the intervention solely comprises a discouragement effect, or whether other effects may be subject to the intervention. For instance, the image of the pabo already was already in decline in the years prior to the intervention (see e.g. Researchned, 2012)[55]. The implementation of the entry-tests strengthened this as an implicit signaling device. That is, the implementation of the entry-test worked as a signal to potential pabo students that the image of the occupation was poor at the time, which induced potential pabo entrants to waive the pabo. We coin the former as the '*image'* effect. The magnitude of the effect as displayed on the right hand-side can be perceived as the joint discouragement- and image effect.

[54] The *root mean squared prediction error* (RMSPE) of each specification are available upon request, but in general we can note over all specifications it never exceeded $3.57*10^{-7}$. (Recall that the RMSPE is minimized and represents the fit of the synthetic control unit with respect to the pabo.)

[55] The perceived quality standards of the pabo were fairly poor already before the entry-tests were implemented. At the time, the pabo was seen as 'easy' and typically for women. These are the principal reasons why students refrain from choosing the pabo – next to the poor career opportunities and lowly perceived wages (Researchned, 2012). Most primarily males and students with different ethnicities indicated the deteriorating quality of the pabo as reason why they did not choose to enroll at the pabo.

Figure 2 – Total Sample: The Effects of Entry-Tests



Notes: Each of the (lines within) the subfigures are based on the most complete specifications of the synthetic control procedure inherent to the particular outcome variable. The mean squared prediction error is minimized over the period 2003-2005. The year 2002 is used in the matching procedure as covariate intrinsic to each study. Labels on the vertical axis represent percentage points within a 0-1 range (e.g 0.1 reflects 10 %).

In a way, the implementation of the entry-tests rendered a decline in pabo enrollments *just* as could be expected of the policy. Namely, the policy not only aimed at guaranteeing pabo students possess the at least sufficient capabilities to become a primary-school teacher, it also aimed to reverse the tendency within society where pabo graduates were labelled as being not competent enough. Specifically, pabo's wanted to resemble the image of a more selective study – it was conjectured that in this way more and more students of higher cognitive abilities were gradually persuaded to choose for the pabo. The gradual reverse of the image of the pabo was ensured by the intervention as students with insufficient math (and language) capabilities were not able to enter the second year. The magnitude of this proposition is more formally investigated in the twin middle figures, which respectively plot the yearly estimates of the impact of the intervention on switch and dropout patterns and the interrelated *'survival'* rate (or retention rate) pattern of the pabo relative to its synthetic counterparts. The plot of the switch rate suggests that the entry-tests had a large positive effect on the switch behavior of pabo students, an effect which increased after 2006 to 5.5% points per enrollment cohort. The intervention effects on dropout switch behavior of pabo students were less severe as suggested by the dropout plot, evoking a pattern of increased dropout behavior of around 2% points a year of the pabo relative to its synthetic counterpart. *En masse,* the increased dropout and switch pattern have induced a contraction in the survival rate of about 6.5% points, which is more intuitively plot together with dropout and switch patterns at the right-hand side. The figure suggests either (or both) of the synthetic control procedures on dropout respectively are relatively spacious estimates, since they jointly do not add up to the drop in the survival rate[56]. Put differently, we can conservatively say the intervention has caused a 6.5% point decline in retention, consisting of a surge in the switch and dropout of pabo students relative to its synthetic counterfactual.

Although the pabo's had hoped the intervention would have induced an inflow of students with a higher cognitive ability, we saw that this conjecture, on average, was not supported by the data (see descriptive statistics table 2, columns 1-3). The drop in retention rates therefore did not come as a surprise at the pabo's. Yet, *ex-ante* the entry-tests were implemented, policymakers hypothesized a decrease in the inflow to the pabo

[56] Note that the synthetic control procedures are all based on singular estimates with respect to the outcome variable. Only the most complete specification of the procedures are shown in the subfigures. The switch and dropout estimates of the synthetic control method must be perceived as a rough decomposition of which proportions could be attributed to the decrease in retention rates. These estimates however, must not be taken as causal inferences.

would not necessary imply a similar decrease in the outflow of pabo students. This reflected the idea of an imperfect pass-through relationship of inflow relative to outflow in the pabo, driven by an incremental graduation rate efficiency.

The bottom left-figure plots the graduation rate efficiency trajectories of the pabo and its synthetic counterparts for both students that graduate within nominal study duration plus one year, and students that *ever* graduate irrespective of the duration. The figure clearly makes apparent that the policymakers' hypothesis can be falsified; as demonstrated on the right-hand side, the intervention not only generated a decrease in the retention rate of the pabo, it also induced a reduction in the graduation rate efficiency of about 4% points. The imperfect pass-through relationship thus does not exist *between* the inflow and outflow in the pabo, yet it exists *within* the pabo. More specifically, the rate of students that obtain their degree *conditional* on entering the second year of the pabo, has increased due to the intervention (see figure 5, Appendix B). The disparity between the retention rate- and graduation rate efficiency plots (≈ 2%) suggests that about one-third of the students that would have entered the second year in absence of the intervention, would still never have obtained their degree. For this reason, the intervention can also be perceived as a commitment mechanism where students are kept from lingering throughout the pabo without ever obtaining a degree.

Statistical Inference

In order to gauge statistical interferences of the synthetic control procedures, while at the same time controlling for potential biases, the weights designated to each study in the donor pool were reassigned to respectively the institutional- and individual level such that regression analyses were possible[57]. Table 3 presents the estimation results of the average treatment effects relating the introduction of the entry-tests to the outcome variables. Intuitively, one would imagine the entry-tests to have the most direct (and therefore significant) effect on the outcome variables during the first year of the pabo and less so in the years subsequent to the tests. This pattern is largely confirmed in the table: the combination of the surge in switch and dropout behavior build up to the decrease in retention behavior of about 6.7% points in the most complete specification. These estimates are all statistically significant at the one-percent level, independent of the specification. For the outcome variables on which the entry-tests have a less direct

[57] The probit model estimates of the synthetic control procedures on respectively the switch, and dropout behavior of students are available upon request. In general, these estimates were fairly well able to reproduce the estimates of the synthetic control method.

Table 3 - The Average Treatment Effects of Entry Tests: Total Sample Estimates

	ln (Enrollment)				*Retention*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Average Treatment Effect	-0.178**	-0.208***	-0.206***	-0.193**	-0.0509***	-0.0513***	-0.0675***	-0.0673***
	(-2.60)	(-3.31)	(-3.32)	(-2.63)	(-3.11)	(-3.35)	(-5.50)	(-5.01)
Observations	3948	3948	3948	3948	562466	562466	562466	562466
(Pseudo) R^2	0.722	0.776	0.783	0.784	0.038	0.053	0.054	0.054
	Graduation Rate Efficiency - Nominal +1				*Graduation Rate Efficiency - Total*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Average Treatment Effect	-0.0393*	-0.0374*	-0.0368	-0.0429*	-0.0308	-0.0319	-0.0353*	-0.0348**
	(-1.82)	(-1.80)	(-1.54)	(-1.74)	(-1.39)	(-1.55)	(-1.90)	(-2.08)
Observations	562466	562466	562466	562466	562466	562466	562466	562466
Pseudo R^2	0.039	0.074	0.074	0.074	0.035	0.059	0.057	0.061
Year and Study Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Individual Controls	N	Y	Y	Y	N	Y	Y	Y
Study Controls	N	N	Y	Y	N	N	Y	Y
Institution and City Controls	N	N	N	Y	N	N	N	Y

Notes: Dependent variables in italics. All coefficients reflect the average marginal change of the variable of interest with respect to the treatment for unique enrollment cohorts, based on individually estimated probit models. Four different specifications are estimated for each of the outcome variables, the build-up being presented at the lower region of the table. The t-statistics of standard errors clustered at the institutional-study cohort level are presented between parenthesis. The only exception are the coefficients of the change in enrollment outcome variable, being estimated using OLS regressions at the institutional-study level. Here standard errors are clustered at the institutional level. Significance at the 10, 5 and 1 percent levels are respectively denoted by * / ** / ***.

influence, the statistical significance, in turn, is also less immersive. The graduation rate efficiency models are just marginally significant, whereas the estimation outputs of the change in enrollments do themselves not necessarily transmit useful information. In an earlier stage, we saw that the joint discouragement and image effect induced a growing negative enrollment effect. In this case, it would be more interesting to measure the entry-tests effects per enrollment cohort rather than the average treatment effect. Following this, we are also able to control for enrollment-specific effects that are possibly related to the outcome variables. That is, when covariates vary per enrollment cohort they are likely to affect the corresponding treatment effect estimates per enrollment cohort.

This proposition is more formally investigated in table 4, which again presents a chronicled look of the entry-tests effects, now per enrollment cohort. The estimates are fairly well able to reproduce the magnitude of the most complete specification effects as displayed on the right-hand side of the subfigures in figure 2. Slight but notable discrepancies between the estimates of the synthetic control method and the probit model estimates can be identified in the retention estimates[58]. Relative to the synthetic control procedures (3) and (4), the retention rate effects as estimated by

[58] The inclusion of year-specific covariates affect the magnitude of the estimated treatment effects dependent on its respective relationships with both the treatment effect estimator and the outcome variable. When the estimation of the treatment effect increases in magnitude, the (changes in) covariates are *on average* negatively (positively) related to the outcome variable and have a positive (negative) relationship with the treatment effect estimator. The vice versa reasoning applies to the case when the estimated treatment effect declines in magnitude.

Table 4 - The Effects of Entry Tests per Enrollment Cohort: Total Sample Estimates

	ln (Enrollment)				*Retention*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Treatment (2006)	-0.0283	-0.0712	-0.0746	-0.0879	-0.0536***	-0.0654***	-0.0639***	-0.0697***
	(-0.26)	(-1.23)	(-0.87)	(-0.95)	(-2.87)	(-3.32)	(-4.84)	(-4.74)
Treatment (2007)	-0.155	-0.141**	-0.144	-0.124	-0.0604***	-0.0699***	-0.0659***	-0.0729***
	(-1.53)	(-1.72)	(-1.47)	(-1.71)	(-3.10)	(-3.18)	(-4.07)	(-4.00)
Treatment (2008)	-0.268**	-0.240**	-0.261**	-0.280**	-0.0566***	-0.0453***	-0.0659***	-0.0677***
	(-2.85)	(-3.09)	(-2.67)	(-2.71)	(-2.77)	(-2.74)	(-2.96)	(-3.10)
Treatment (2009)	-0.262*	-0.329***	-0.293**	-0.288**	-0.0627*	-0.0544**	-0.0667***	-0.0645***
	(-2.55)	(-3.45)	(-2.79)	(-2.75)	(-1.92)	(-2.12)	(-2.88)	(-3.04)
Observations	3948	3948	3948	3948	562466	562466	562466	562466
(Pseudo) R^2	0.731	0.784	0.784	0.784	0.038	0.053	0.054	0.054
	Graduation Rate Efficiency - Nominal +1				*Graduation Rate Efficiency - Total*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Treatment (2006)	-0.0268	-0.0518*	-0.0496**	-0.0393**	-0.0383*	-0.0375*	-0.0373*	-0.0431**
	(-1.12)	(1.85)	(-2.13)	(-2.48)	(-1.86)	(-2.01)	(-1.96)	(-2.13)
Treatment (2007)	-0.0255	-0.0431**	-0.0303	-0.0354**	-0.0345**	-0.0340*	-0.0367*	-0.0330**
	(-1.00)	(2.42)	(-1.15)	(-2.19)	(-2.14)	(-1.78)	(-1.85)	(-2.23)
Treatment (2008)	-0.0271	-0.0420**	-0.0395***	-0.0395***	-0.0386*	-0.0361*	-0.0359**	-0.0385**
	(-1.10)	(2.57)	(-2.82)	(-2.98)	(-1.93)	(-1.72)	(-2.25)	(-2.03)
Treatment (2009)	-0.00585	-0.0316	-0.0355*	-0.0364*	-0.0388**	-0.0327*	-0.0323*	-0.0394*
	(-0.24)	(1.52)	(-1.72)	(-1.86)	(-2.31)	(-1.69)	(-1.70)	(-1.88)
Observations	573258	525600	572185	572185	573661	573661	573519	573661
Pseudo R^2	0.045	0.075	0.088	0.089	0.035	0.040	0.057	0.061
Year and Study Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Individual Controls	N	Y	Y	Y	N	Y	Y	Y
Study Controls	N	N	Y	Y	N	N	Y	Y
Institution and City Controls	N	N	N	Y	N	N	N	Y

Notes: Dependent variables in italics. All coefficients reflect the average marginal change of the variable of interest with respect to the treatment for unique enrollment cohorts, based on individually estimated probit models. Four different specifications are estimated for each of the outcome variables, the build-up being presented at the lower region of the table. The t-statistics of standard errors clustered at the institutional-study cohort level are presented between parenthesis. The only exception are the coefficients of the change in enrollment outcome variable, being estimated using OLS regressions at the institutional-study level. Here standard errors are clustered at the institutional level. Significance at the 10, 5 and 1 percent levels are respectively denoted by * / ** / ***.

the probit models are about one percentage point larger than as suggested. Other discrepancies are generally in the order of tenths of a percentage point. Yet overall, it must be emphasized that the discrepancies in estimated effects between the two methods is limited, making the synthetic control procedure a promising strategy to evaluate treatment effects. Moreover, within each specification, the statistical inferences behave as one would imagine they would do. For instance, one would imagine that in the change in enrollment specifications the statistical significance of the results gradually become larger over time, which is confirmed by the estimation outputs.

Placebo Effects Evaluation

In contrast to former regression models, the inferential method in the original synthetic control procedure does not rely on regression analyses. Abadie *et al.* (2012) propose an evaluation method based on placebo tests to probe the statistical significance of their estimated treatment effect. In this case, the placebo tests hinge on iteratively applying the synthetic control method on studies within the donor pool of potential control units, which did not experience the implementation of the entry-tests, but are treated as if they did[59]. Correspondingly, one evaluates the magnitude of the estimated treatment effect of each study in the donor pool. In other words, to what extent are our obtained pabo results driven by chance? If the real treatment effect of the pabo stands out within the plot of placebo effects, then we can deduce the estimated treatment has a significant effect on the particular outcome variable.

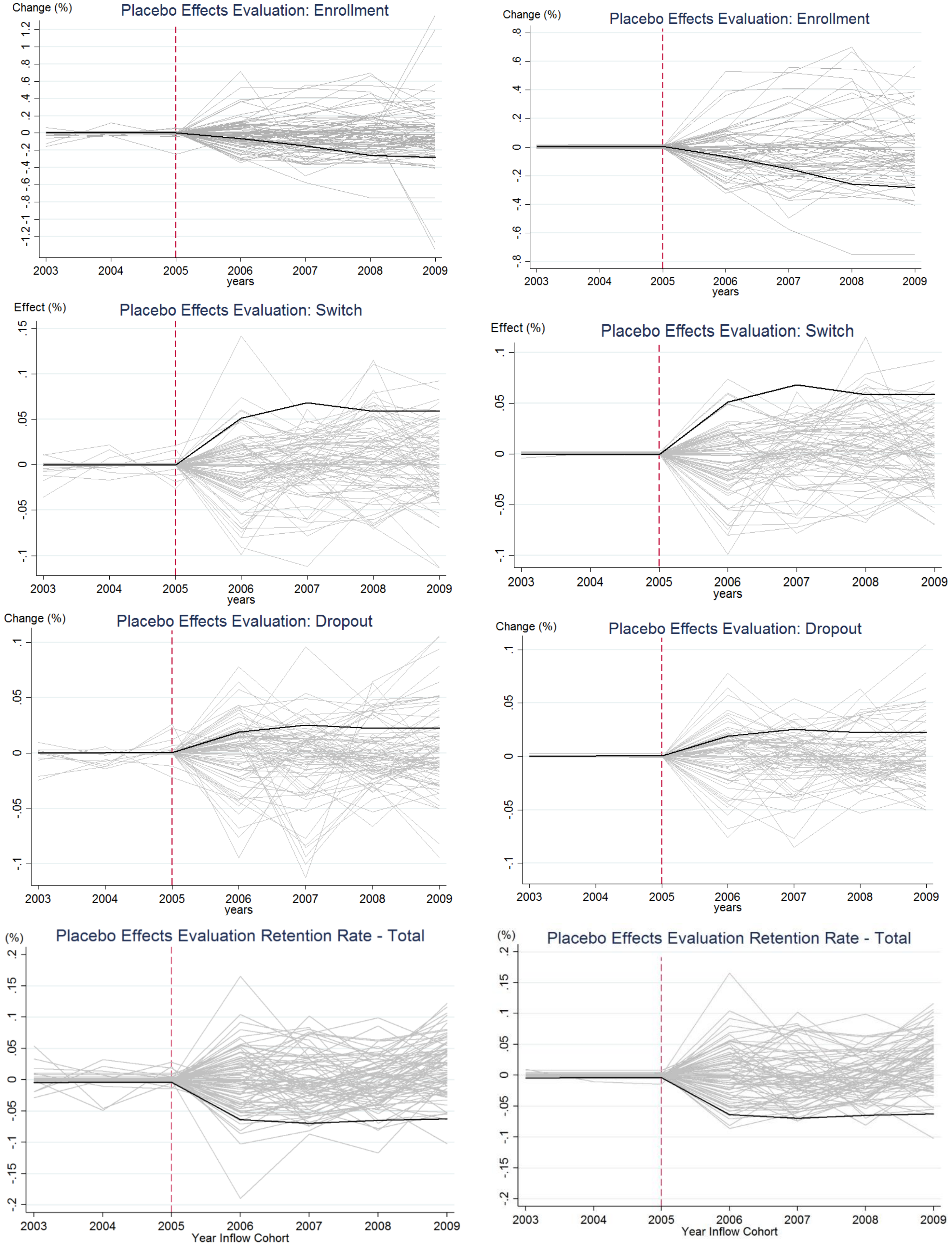
Before mentioning the results, it must be noted that the assumption of no interference between units is likely to be violated in most placebo procedures. That is, studies where at least its enrollment is related to the enrollments of another study, are present in the donor pool of potential control units. For this reason, especially the enrollment placebo procedures should not be taken as causal inference.

Plots of the placebo procedures on each of the respective outcomes are visualized on the subfigures in figure 3. The light gray lines represent the effects in gap plots associated with the placebo procedures of 109 studies, whereas the black line symbolizes the 'true' effect of the pabo. It immediately becomes apparent most of the plotted graphs consist of a vast amount of substantial year peaks. There are two possible explanations that can justify why these are so prominent: i) Several studies have an extreme difficulty in

[59] We use the most complete (data) procedure, where the pabo is kept in de donor pool of units.

Figure 3 –The Placebo Effects of Entry-Tests

Exclusion 20% worst MSPE



Notes: Each of the (lines within) the subfigures are based on the most complete specifications of the synthetic control procedure inherent to the particular outcome variable. The mean squared prediction error is minimized over the period 2003-2005. The year 2002 is used in the matching procedure as covariate intrinsic to each study. Labels on the vertical axis represent percentage points within a 0-1 range (e.g 0.1 reflects 10 %). The enrollment plots are based on 87 studies, including the pabo. The remainder of the placebo runs on each of the outcome variables entail the complete donor pool of studies, i.e. 110 with the pabo included. Placebo plots on graduation rate efficiency outlined in Appendix B.

producing a reasonable synthetic version of themselves with a similar pre-trend (i.e. large MSPE's). ii) The violation of the assumption of no interference between units. Concerning statement i), we could argue that the extensive gaps of these studies are likely to be artificially generated. That is, due to a lack of fit. This is the main reason why Abadie et al. (2012), iteratively exclude units with the largest MSPE's in their placebo plot. Due to space constraints, here only the studies with the 20% worst MSPE's are excluded at the right-hand side of each of the subfigures.

The simple juxtaposition of the figures on the left-hand side with those on the right-hand side, indicates that the worst fitted studies indeed experienced large shocks, but that this can likely be attributed to the procedure artificially generating a synthetic control group and thereby the large gap(s) in certain years, rather than plausibly generating the placebo effects. If we take a look at the right-hand side figures, the problem of no interference between units remains. Still, keeping this potential problem in mind, we choose to assess the significance of the pabo estimates. Table 5 presents the p-values for each of the outcome variables per enrollment cohort. It aims to answer the subsequent question; what is the probability of obtaining a similar effect as the pabo, when one were to assign the intervention at random to one of the studies in the donor pool? As example: the probability of finding a dropout estimate in 2006 as large as the pabo is 15/110=13.6%.

If we compare the statistical inferences as obtained by the regression analyses with the placebo procedures, it becomes clear that the statistical significance in the placebo procedures is about five percentage points lower. This statement is remarkably coherent over each of the outcome variables. One interpretation is to ascribe these differences to the power of the analyses, since the regression models were based on individual data,

Table 5 – To What Extent Are the Results Driven by Chance?

					Graduation Rate Efficiency	
	ln(Enrollment)	*Dropout*	*Switch*	*Retention*	*Nominal +1*	*Total*
Treatment (2006)	0.322	0.136	0.036	0.036	0.055	0.064
Treatment (2007)	0.172	0.082	0.009	0.027	0.109	0.127
Treatment (2008)	0.080	0.100	0.073	0.027	0.073	0.055
Treatment (2009)	0.057	0.145	0.045	0.027	0.073	0.136

Notes: Dependent variables in italics. P-values are based on one-sided t-tests and correspond to the most complete synthetic control procedures on each of the outcome variables.

and the placebo plots are based on aggregated study-level data. Another interpretation is the violation of the assumption of no interference between studies. [60]

B. The Effects of Entry-Tests per Highest Previous Educational level Attained

So far, the analyses have focused on the effects of entry-tests on the entire relevant sample of individuals. However, the entry-tests aim to select those with a sufficient set of cognitive abilities in mathematics and the Dutch language, making the presence of heterogeneous effects likely per subsample of individuals with different highest previous educational levels attained (i.e. mbo, havo and vwo). For this reason, we conduct a similar set of synthetic control method procedures, now per subgroup to probe whether students with presumably lower cognitive abilities are treated more intensely by the implementation of the entry-tests.

Plots of the most complete synthetic control procedures on each of the outcome variables per pre-educational subsample are respectively presented on the left-hand side of figure 4. Again, the dashed lines show the virtually perfect replication of the outcome variables before the entry-tests were implemented. The continuation of the dashed lines during the post-intervention period illustrate the counterfactual of the pabo with regard to the specific outcome variables. For example, what would have happened with the enrollment behavior of pabo students with respectively a highest attained educational level of mbo, havo or vwo if the intervention *would not* have occurred? The counterfactuals are contrasted to the solid lines, exemplifying the actual outcome variables of the pabo after the entry-tests were implemented. In resonance with figure 2, the simple differences between the actual outcome variables of the pabo and the 'synthesized' outcomes of the counterfactual are visualized on the right-hand side of the subfigures in figure 4. These differences reflect the estimated effects for each of the outcome variables. In the havo and vwo procedures, we are now able to include covariates proxying for the (study-level aggegrated) cognitive ability of students.

When we compare the estimated treatment effects of the change in enrollments per subsample of individuals to the estimates of the entire relevant sample, it seems that the inclusion of more specific (cognitive) covariates for the former group correspondingly also generates a more specific counterfactual. That is, where *for the*

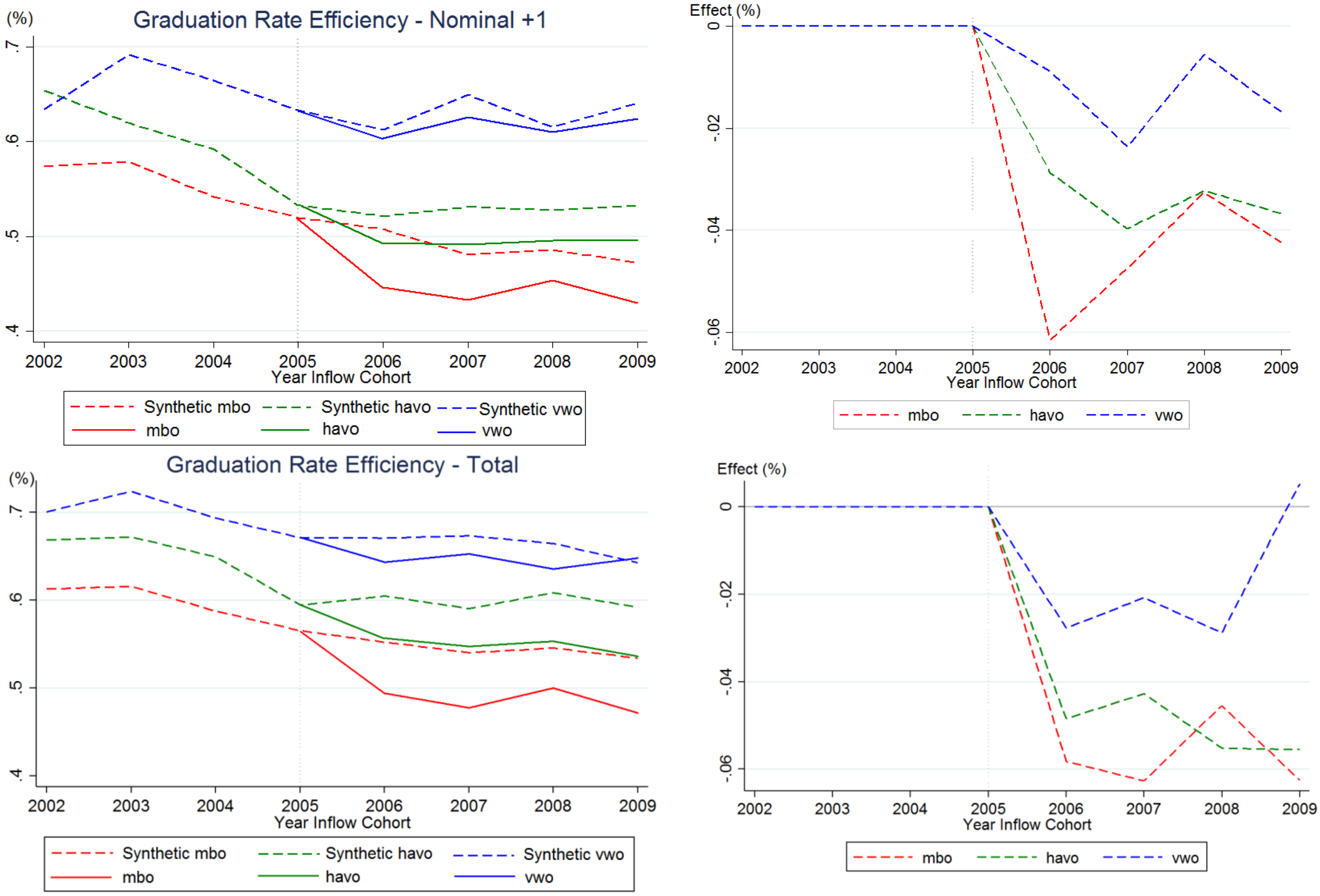
[60] An alternative way to evaluate the significance of the average treatment effect estimates is to look at the distribution of the pre- and post MSPE's of each placebo run. The post-MSPE's simply correspond to the size of the areas below or above the generated placebo effect lines relative to the y-axis of zero. In this way, it becomes even clearer that within the switch and retention rate procedures, the pabo stands out as an effect that is very unlikely to be randomly found within the pool of donor units.

Figure 4 –The Effects of Entry-Tests per Pre-Education

Gap Plots

Figure 4 Continued –The Effects of Entry-Tests per Pre-Education

Gap Plots



Notes: Each of the (lines within) the subfigures are based on the most complete specifications of the synthetic control procedure inherent to the particular outcome variable. The mean squared prediction error is minimized over the period 2003-2005. The year 2002 is used in the matching procedure as covariate intrinsic to each study. Labels on the vertical axis represent percentage points within a 0-1 range (e.g 0.1 reflects 10 %).

entire relevant sample the indexed amount of entrant enrollments at the pabo gradually built up to a 30% point decrease in enrollments in 2009, the indexed amount of enrollments *per subgroup* each build up to a decline of roughly 40% points in 2009. The total sample estimate of the change in enrollment effect must therefore be perceived as a conservative estimate, which does not take into account specific information about subgroups of students. Moreover, if we take a look at the magnitude of the estimates at the right-hand side, the pattern is conveyed that the subgroup of students which presumably possesses a relatively low set of cognitive skills (i.e. mbo) were not more discouraged to start the pabo compared to the other subgroups. In contrast, there are hardly any differences in the patterns of the declining enrollment change effects, except for the subgroup of vwo students, which experienced a relatively larger decline in 2007, and catch-up in 2009. [61]

[61] Part of the declining vwo change in enrollment effect in 2009 is due to the implementation of the academic pabo's. By then, more institutions started to provide an academic pabo program, which increased the attractiveness for vwo students to start the pabo.

The relatively homogenous per subgroup effects cannot be similarly detected in the outcome variable estimates that occur at the end of the first year of enrollment. Intuitively, one would imagine the entry-tests to have a particular negative effect on the retention rates of the groups with the lowest set of cognitive skills. This pattern is closely reproduced as could be seen in the bottom figures. Pabo students with a highest previously attained educational level mbo were most negatively affected by the intervention, as perceived from a decline of about 10% points. The magnitude of the latter effect sharply contrasts to the obtained effect for the group of students with presumably the highest set of cognitive skills (vwo). Here students were only mildly negatively affected by the entry-tests, whereas the students with the presumed intermediate cognitive skills range (havo) were correspondingly also moderately affected with a contraction in the retention rate of about 6 to 7% points.

For each subgroup, the drop in retention rates can be decomposed into a dropout factor and a switch factor. Only the students with a highest previously attained mbo education experience both a substantial surge in dropout behavior ($\approx$ 5%) as well as a rise in switch behavior ($\approx$ 4%). The fact that the surge in dropout behavior is so rampant for this group in particular can most probably be ascribed to these individuals already being in the possession of a qualification for a job. For the remainder of the students with respectively a highest previously attained educational level havo and vwo this is not the case. It should come as no surprise that in case pabo students of these groups do not successfully complete the first year, their only option is to switch to another study in order to obtain a qualification for the labor market.

For each subgroup, the decrease in retention rates again pass-through imperfectly to the decrease in graduation rate efficiency. In absence of the entry-tests about one-third of the pabo students that enter the second year would not have completed the pabo anyways. Since the implementation of the entry-tests, these individuals are now urged to either switch or dropout after the first year of the pabo. The discrepancy between the plot of students that *ever* graduate (total) and graduate within nominal study duration plus one year, additionally indicates that this *'frontloading'* effect is stronger for the seemingly weakest students entering the pabo.

In sum, if we compare the total sample estimates to the estimates per subgroup, we can note that the increase of dropout behavior of about 2% points can almost be entirely explained by the rise in dropout behavior of students with mbo as highest previously attained education. The switch patterns are more homogenously per subgroup and together constitute fairly well to the growth in switch behavior of the total sample (≈5.5%). This premise cannot be similarly concluded in the per subsample retention estimates, which together constitute to a decline in retention rates of about 7.5% points. Earlier we argued that the total sample estimates are therefore are likely to be rather conservative estimates. The contrast is true for the suggested declines in per subgroup graduation rate efficiency, which again fairly well jointly reproduce the total sample estimates.

Statistical Inference

In order to obtain statistical inferences of the estimates per subgroup of individuals with different levels of highest previously attained education, while controlling for potential covariate changes in the treatment and control group, the weights designated to each study in the donor pool were reassigned to weights on respectively the institutional and the individual level. The weights now reflect the respective size of the institution or the individual within the assigned study weight per subgroup (i.e. mbo, havo or vwo). Table 6 presents the average treatment effects relating the introduction of the entry-tests to each of the outcome variables per subgroup of individuals.

To start off with the group of individuals with a highest previously attained educational level of mbo. This group appears to be the most intensely affected by the introduction of the entry-tests. All of the estimated treatment effects on the outcome variables are either significant at the five percent or the one percent level, which should not come as a surprise since the magnitude of the estimated treatment effects are generally among the largest. When we juxtapose the former statistical inferences to those with the subgroup of individual with a highest previously attained education level of vwo, it becomes apparent that the latter are only moderately (directly) affected. Still it must be noted that the combined surge in dropout and switch behavior, induces a statistically significant decreased retention rate of about 4% points. Those students with a highest previously attained education level of havo again fall within the middling category, with statistical significances at either the ten percent or five percent level. There seem to be no notable discrepancies between the estimated magnitudes in the synthetic control method and the regression analyses.

To confirm the latter proposition, we additionally check whether cohort-specific covariates have an effect on the estimates of the synthetic control method. In the

Table 6 – The Average Treatment Effects of Entry Tests per Highest Previously Attained Education

	ln (Enrollment)				*Retention*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Average Treatment Effect mbo	-0.332**	- 0.319***	-0.290***	-0.309***	-0.0938***	-0.0922***	-0.0868***	-0.0941***
	(-3.24)	(4.00)	(-3.74)	(-3.86)	(-4.68)	(-4.98)	(-3.46)	(-3.68)
Observations	3944	1938	3944	3944	248074	248074	248074	248074
(Pseudo) R^2	0.708	0.837	0.780	0.785	0.022	0.039	0.042	0.044
Average Treatment Effect havo	-0.263*	- 0.220*	-0.235*	-0.261**	-0.0598**	-0.0618**	-0.0699*	-0.0623**
	(-2.53)	(2.32)	(-2.39)	(-2.93)	(-2.36)	(-2.25)	(-1.96)	(-2.31)
Observations	3944	1938	3944	3944	249664	249664	249664	249664
(Pseudo) R^2	0.708	0.837	0.780	0.785	0.081	0.123	0.123	0.123
Average Treatment Effect vwo	-0.287**	-0.290**	-0.290**	-0.283**	-0.0350**	-0.0400***	-0.0376**	-0.0417***
	(-3.02)	(-3.24)	(-3.20)	(-3.15)	(-2.38)	(-2.98)	(-2.53)	(-2.74)
Observations	3828	3616	3828	3800	64728	64728	64728	64728
(Pseudo) R^2	0.727	0.805	0.803	0.804	0.014	0.037	0.037	0.037
	Graduation Rate Efficiency - Nominal +1				*Graduation Rate Efficiency - Total*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Average Treatment Effect mbo	-0.0384**	-0.0469***	-0.0540***	-0.0584***	-0.0368**	-0.0443***	-0.0546***	-0.0566***
	(-2.11)	(-2.80)	(-3.14)	(-4.03)	(-2.27)	(-2.89)	(-3.30)	(-4.13)
Observations	248074	248074	248074	248074	248074	248074	248074	248074
Pseudo R^2	0.026	0.055	0.056	0.056	0.021	0.048	0.049	0.049
Average Treatment Effect havo	-0.0447*	-0.0397	-0.0364***	-0.0369***	-0.0359**	-0.0359	-0.0390*	-0.0372*
	(-1.86)	(-1.61)	(-3.38)	(-3.23)	(-2.01)	(-1.54)	(-1.76)	(-1.71)
Observations	248784	248784	248784	248784	249598	249598	249598	249598
Pseudo R^2	0.052	0.116	0.119	0.123	0.059	0.112	0.112	0.114
Average Treatment Effect vwo	-0.0103	-0.0180	-0.0113	-0.0192	-0.00828	-0.0146	-0.0125	-0.0187
	(-0.49)	(-0.88)	(-0.35)	(-0.74)	(-0.29)	(-0.76)	(-0.65)	(-0.83)
Observations	64728	64728	64728	64728	64728	64728	64728	64728
Pseudo R^2	0.025	0.073	0.073	0.074	0.025	0.073	0.073	0.074
Year and Study Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Individual Controls	N	Y	Y	Y	N	Y	Y	Y
Study Controls	N	N	Y	Y	N	N	Y	Y
Institution and City Controls	N	N	N	Y	N	N	N	Y

Notes: Dependent variables in italics, estimated per pre-education. All coefficients reflect the average marginal change of the variable of interest with respect to the treatment for all enrollment cohorts subject to the treatment, based on individually estimated probit models. Four different specifications are estimated for each of the outcome variables, the build-up being presented at the lower region of the table. The t-statistics of standard errors clustered at the institutional-study cohort level are presented between parenthesis. The only exception are the coefficients of the change in enrollment outcome variable, being estimated using OLS regressions at the institutional-study level. Here standard errors are clustered at the institutional level. Significance at the 10, 5 and 1 percent levels are respectively denoted by * / ** / ***.

appendix, tables 3 and 4 present the treatment effects per enrollment cohort for respectively the subgroups with a highest previously attained education mbo, havo and vwo. A similar pattern emerges as within the total sample estimates. Despite the fact that the inclusion of year-specific covariates *often* either enlarges or lessens the effects as shown in the subfigures of figure 4 with about 1% point, there are no predominant biases to mention. That is, it does not occur that the inclusion of control covariates in the regression models suddenly causes a reverse in sign in the treatment estimate, or even a de(in)crease of the estimated treatment effect by more than 2% points.

Placebo Effects Evaluation

To check whether the statistical inferences as obtained using regression models remain robust to placebo tests procedures, we again iteratively apply the synthetic method on studies within the donor pool of potential control units. These studies are treated as if they experienced the intervention of the entry-tests, whereas in reality they did not[62]. Using this procedure, we can evaluate to what extent our synthetically obtained results are driven by chance.

Plots of the placebo procedures on each of the respective outcome variables per subgroup are presented in figures 9-14 of the Appendix. The light gray lines represent the effects in gap plots associated with the placebo procedures of 109 studies, whereas the red, green and blue line respectively symbolize the 'true' mbo, havo an vwo pre-education effect of the pabo. The associated p-values are demonstrated in table 7.[63]

When we compare the statistical inferences of the two methods, again a clear pattern emerges as being evoked in the table. In congruence with the placebo effects evaluation of the total sample, the p-values decrease about 5% points in terms of statistical significance compared to the regression estimates. This only holds within the estimates of the havo and vwo samples. Apparently, it seems that the inclusion of specific covariates relating the cognitive skills of individuals (e.g. grades and profiles) to the outcome variables, provides more information to the synthetic control procedure, which correspondingly is able to construct a more reliable counterfactual. The reason why this proposition is put forward is due to the discrepancy in p-values between the

[62] We use the most complete data procedure on each of the outcome variables, implying that the specific variables inherent to the subgroup of pre-educational mbo/havo or vwo students is used in each of the procedures. Additionally, the relative size of each subgroup (e.g. 12% vwo) of the students entering the pabo is used as a matching variable to take account of potential peer effects.

[63] Recall that the p-values aims to answer the following question; what is the probability of obtaining a similar effect as the pabo, when one were to assign the intervention at random to one of the studies in the donor pool?

Table 7 – To What Extent are the Results per Subsample Driven by Chance?

					Graduation Rate Efficiency	
	ln(Enrollment)	*Dropout*	*Switch*	*Retention*	*Nominal +1*	*Total*
mbo						
Treatment (2006)	0.540	0.118	0.127	0.127	0.136	0.127
Treatment (2007)	0.287	0.055	0.064	0.018	0.064	0.082
Treatment (2008)	0.069	0.036	0.155	0.027	0.082	0.136
Treatment (2009)	0.103	0.036	0.173	0.027	0.118	0.173
havo						
Treatment (2006)	0.184	0.155	0.136	0.082	0.255	0.164
Treatment (2007)	0.080	0.145	0.045	0.064	0.136	0.191
Treatment (2008)	0.069	0.309	0.036	0.027	0.073	0.136
Treatment (2009)	0.069	0.345	0.064	0.118	0.127	0.118
vwo						
Treatment (2006)	0.287	0.336	0.255	0.218	0.118	0.155
Treatment (2007)	0.069	0.282	0.218	0.255	0.236	0.282
Treatment (2008)	0.069	0.345	0.209	0.236	0.291	0.336
Treatment (2009)	0.138	0.327	0.282	0.391	0.345	0.391

Notes: Dependent variables in italics. P-values are based on one-sided t-tests and correspond to the most complete synthetic control procedures on each of the outcome variables. Placebo plots outlined in the Appendix

synthetic control method and the regression estimates for the pre-education mbo individuals. For this group, the discrepancy in p-values is generally in the order of 10% points. Another reason could be that the within the subsample of mbo individuals, the violation of the interference assumption induces a more prevalent bias within the mbo sample estimates.

C. The Heterogeneous Effects of Entry-Tests

So far, we have aimed to sketch the average treatment effects of the entry-tests gradually in more specific settings. We started off by presenting the entire picture, then followed by zooming in to the level per highest previously attained education. Yet, we did not allow for a disparate responsiveness to the entry-tests for distinct subgroups within these settings. In order to continue the path of zooming in, we gauge whether specific subgroups were affected differently by the implementation of the entry-tests. To this end, we (separately) include interaction variables of the average treatment effects with dummies for respectively females, a non-Western immigrant, a Western-immigrant and a direct entrant. In addition, we test whether the age of individuals has a differential responsiveness to the entry-tests. Table 8 presents the results.

Table 8 - Heterogeneous Effects of Entry Tests – Which Subgroups are most Affected?

	ln(Enrollment)				*Retention*			
	Total	mbo	havo	vwo	Total	mbo	havo	vwo
ATE	-2.234**	-1.886*	-2.221**	-2.103	-0.0679***	-0.0775***	-0.0381*	-0.0855***
	(-2.15)	(-1.97)	(-2.18)	(-1.39)	(-4.28)	(-4.13)	(1.82)	(-3.67)
ATE* Female	2.482**	1.979*	2.432**	2.588	-0.0124	-0.0133	-0.0350**	0.0585**
	(1.97)	(1.86)	(1.97)	(1.41)	(-1.23)	(-0.70)	(-2.22)	(2.56)
(Pseudo) R^2	0.857	0.842	0.863	0.658	0.058	0.044	0.132	0.041
ATE	1.141	0.319	1.908**	-0.227	-0.0582***	-0.124***	-0.0181	-0.0659*
	(1.31)	(0.31)	(2.05)	(-0.25)	(-2.73)	(-3.23)	(-0.58)	(-1.72)
ATE* Age	-0.0652	-0.0221	-0.102**	0.0140	-0.000952	0.00149	-0.00212	0.000810
	(-1.52)	(-0.43)	(2.20)	(0.31)	(-1.11)	(0.88)	(-0.97)	(0.52)
(Pseudo) R^2	0.860	0.848	0.864	0.645	0.057	0.045	0.100	0.041
ATE	-0.209***	-0.136*	-0.190***	-0.242**	-0.0715***	-0.0838***	-0.0474**	-0.0446***
	(-2.59)	(-1.68)	(-3.13)	(-2.41)	(-4.68)	(-4.09)	(-2.32)	(-2.89)
ATE* Non-WI	-0.380	-0.398	-0.425	0.102	-0.0650***	-0.0456*	-0.0304*	0.0480
	(-0.65)	(-0.41)	(-0.72)	(0.13)	(-3.79)	(-1.74)	(-1.87)	(0.99)
(Pseudo) R^2	0.856	0.844	0.863	0.648	0.057	0.046	0.097	0.039
ATE	-0.125	-0.127	-0.109*	-0.149*	-0.0710***	-0.0817***	-0.0558**	-0.0387**
	(-1.47)	(-1.42)	(-1.69)	(-1.81)	(-4.84)	(-3.33)	(-2.51)	(-2.50)
ATE* WI	-2.108*	-0.701	-2.537***	-0.695	-0.0447**	-0.0807*	-0.0429	-0.0445
	(-1.93)	(-0.58)	(-3.33)	(-1.47)	(-2.27)	(-1.65)	(-1.47)	(-1.11)
(Pseudo) R^2	0.855	0.843	0.861	0.643	0.054	0.045	0.097	0.039
ATE	-1.425***	-1.073	-1.744***	-0.606	-0.0919***	-0.0752**	-0.0574**	-0.0518**
	(-2.68)	(-1.63)	(-3.89)	(-1.02)	(-5.21)	(-2.36)	(-2.31)	(-2.47)
ATE* DE	1.458**	1.143	1.844***	-0.686	0.0194*	-0.0214	-0.0108	0.0216
	(2.35)	(1.48)	(3.49)	(-0.96)	(1.68)	(-0.91)	(-0.32)	(0.74)
(Pseudo) R^2	0.859	0.845	0.870	0.669	0.056	0.045	0.098	0.040
Observations	4235	4205	4106	4018	562466	248074	249664	64728

Notes: Dependent variables in italics, estimated for the total sample and per pre-educational subsample. All coefficients reflect the average marginal change of the variable of interest with respect to the treatment, based on individually estimated probit models. All estimations include controls on the individual, study and institutional/city level complemented with year and study fixed effects. In addition, for each specification of interest, interaction variables with all other covariates were included to control for potential disparate responses of the variable. For example, females may react different to labor market outlooks. The t-statistics of standard errors clustered at the institutional-study cohort level are presented between parenthesis. The only exception are the coefficients of the change in enrollment outcome variable, being estimated using OLS regressions at the institutional-study level. Here standard errors are clustered at the institutional level. Significance at the 10, 5 and 1 percent levels are respectively denoted by * / ** / ***. Abbreviations as denoted on the left-hand side (Non-WI, WI, DE) respectively indicate Non-Western Immigrant, Western Immigrant and Direct Entrant.

Table 8 Continued - Heterogeneous Effects of Entry Tests – Which Subgroups are most Affected?

	Graduation Rate Efficiency - Nominal +1				*Graduation Rate Efficiency - Total*			
	Total	mbo	havo	vwo	Total	mbo	havo	vwo
ATE	-0.0517***	-0.103***	-0.0337**	-0.0807**	-0.0538***	-0.0940***	-0.0381*	-0.0915***
	(-2.62)	(-4.94)	(-2.41)	(-2.10)	(-3.14)	(-4.58)	(-2.13)	(-3.29)
ATE* Female	0.0125	0.0418**	0.00905	0.0790***	0.0148	0.0301	-0.0105	0.0786***
	(0.92)	(2.09)	(0.47)	(2.64)	(0.95)	(1.52)	(-0.62)	(2.82)
(Pseudo) R^2	0.079	0.056	0.128	0.077	0.065	0.049	0.119	0.065
ATE	0.0135	-0.0554	0.175*	-0.0170	0.0194	-0.0587*	-0.0201	-0.0242
	(0.43)	(-1.54)	(1.95)	(-0.38)	(0.73)	(-1.79)	(-0.44)	(-1.01)
ATE* Age	-0.00261***	-0.000645	-0.00991***	-0.000185	-0.00279**	-0.000417	-0.00552**	0.000558
	(-2.61)	(-0.46)	(-2.81)	(-0.11)	(-2.57)	(-0.32)	(-2.38)	(0.33)
(Pseudo) R^2	0.077	0.058	0.131	0.076	0.063	0.051	0.119	0.063
ATE	-0.0390**	-0.0597***	-0.0341*	-0.0179	-0.0347**	-0.0588***	-0.0258**	-0.0123
	(-1.99)	(-3.52)	(-1.60)	(-0.72)	(-2.46)	(-3.71)	(-2.20)	(-0.51)
ATE* Non-WI	-0.0457**	-0.0848***	-0.0207	-0.0378	-0.0764***	-0.0822**	-0.0326*	-0.000491
	(-2.48)	(-2.78)	(-1.44)	(-0.52)	(-3.49)	(-2.53)	(-1.82)	(-0.01)
(Pseudo) R^2	0.076	0.058	0.128	0.074	0.063	0.050	0.118	0.061
ATE	-0.0376*	-0.0672***	-0.0385**	-0.0206	-0.0356**	-0.0669***	-0.0321**	-0.0226
	(-1.89)	(-3.93)	(-2.33)	(-0.77)	(-2.39)	(-4.12)	(-2.48)	(-0.76)
ATE* WI	-0.0137	-0.0430	0.0472	0.00831	-0.0354	-0.0427	-0.00518	-0.0102
	(-0.53)	(-1.36)	(1.53)	(0.13)	(-1.51)	(-1.57)	(-0.14)	(-0.23)
(Pseudo) R^2	0.077	0.057	0.129	0.074	0.062	0.050	0.118	0.062
ATE	-0.0665***	-0.0861***	-0.136**	-0.0270	-0.0717***	-0.0918***	-0.0982***	-0.0220**
	(-3.35)	(-3.66)	(-2.55)	(-0.86)	(-4.07)	(-4.13)	(-3.89)	(-2.13)
ATE* DE	0.0317**	0.0195	0.124***	0.0195	0.0407***	0.0301*	0.0813**	0.0200
	(1.97)	(1.05)	(2.71)	(0.70)	(2.79)	(1.76)	(2.53)	(0.77)
(Pseudo) R^2	0.076	0.058	0.132	0.077	0.063	0.051	0.119	0.064
Observations	562466	248074	248784	64728	562466	248074	249598	64728

Notes: Dependent variables in italics, estimated for the total sample and per pre-educational subsample. All coefficients reflect the average marginal change of the variable of interest with respect to the treatment, based on individually estimated probit models. All estimations include controls on the individual, study and institutional/city level complemented with year and study fixed effects. In addition, for each specification of interest, interaction variables with all other covariates were included to control for potential disparate responses of the variable. For example, females may react different to labor market outlooks. The t-statistics of standard errors clustered at the institutional-study cohort level are presented between parenthesis. The only exception are the coefficients of the change in enrollment outcome variable, being estimated using OLS regressions at the institutional-study level. Here standard errors are clustered at the institutional level. Significance at the 10, 5 and 1 percent levels are respectively denoted by * / ** / ***. Abbreviations as denoted on the left-hand side (Non-WI, WI, DE) respectively indicate Non-Western Immigrant, Western Immigrant and Direct Entrant.

First of all, it must be noted that the outcome variable measuring the change in enrollments cannot be interpreted in a similar manner as the remainder of the outcome variables. The reason is that the former models are estimated on the institutional level, whereas the latter on an individual level. In practice, this implies for example that the interaction variable of the average treatment effect with the female dummy should be interpreted as follows; when a pabo institution experienced a very high percentage of females as composition of the entrance into the pabo during the intervention period, it could confine its decline in enrollments. An analogous statement could be made on the percentage of direct entrants. In contrast, institutions experienced a larger decline in enrollments when the degree of both Western- and non-Western immigrants was larger during the post-intervention period. Notice however, that on grounds of these results, we cannot conclude that the combined discouragement/image effect is larger for a certain *group* of (potential) pabo students.

Second, on an individual level especially pabo females with a highest previously attained educational level of vwo have a much higher performance rate relative to their male counterparts since the entry-tests were implemented- the difference being in the order of 8% points in terms of graduation rate efficiency. Or to put more it correctly: these females are less negatively affected by the implementation of the entry-tests. Moreover, since the intervention young(er) pabo students with a highest previously attained havo education perform better compared to their older counterparts, which in line with the observation that also direct (havo) entrants to the pabo were less intensely affected in terms of graduation rate efficiency.

Third, if we divert our focus to the entry-tests effects on pabo students with different ethnic backgrounds, it becomes apparent that most prominently non-Western immigrants with a mbo pre-educational background are affected adversely by the entry-tests -individuals of this group perform about 6% points worse relative to their native Dutch counterparts in terms of retention rates, which culminates to even worse levels in terms of graduation rate efficiency. On a similar note, Western-immigrants with a mbo background were also more negatively affected by the entry tests in terms of retention relative to their native Dutch counterparts, however this only partially propagates in a deterioration in graduation rates.

The Importance of Cognitive Skills

In a way, the former effects are not particularly surprising to policymakers - these groups are generally recognized to possess a relatively lower set of cognitive skills. Yet, whilst speaking of cognitive skills, so far we do not even know whether these matter as determinant of passing the entry-tests (i.e. retention) and proceeding towards the end of the pabo. Table 9 investigates this formally, relating the average grade of the group of students with a highest previously attained education of respectively havo, and vwo to each of the outcome variables.

Table 9 - Do Cognitive Skills Matter?

					Graduation Rate Efficiency			
	ln(Enrollment)		*Retention*		*Nominal +1*		*Total*	
	havo	vwo	havo	vwo	havo	vwo	havo	vwo
Average Treatment Effect	-0.803	-4.984	-0.0945***	-0.0726**	-0.245*	-0.600***	-0.253*	-0.455***
	(-0.23)	(-1.44)	(-2.89)	(-2.05)	(-1.71)	(-3.88)	(-1.77)	(-3.11)
ATE* Final Average	0.0901	0.750	0.0113	0.00818**	0.0318*	0.0872***	0.0347**	0.0664***
	(0.17)	(1.44)	(1.48)	(2.14)	(1.94)	(3.59)	(2.00)	(2.87)
(Pseudo) R^2	0.863	0.667	0.096	0.038	0.125	0.077	0.118	0.064
Observations	4106	4018	249664	64728	248784	64728	249598	64728

Notes: Dependent variables in italics, estimated for the total sample and per pre-educational subsample. All coefficients reflect the average marginal change of the variable of interest with respect to the treatment, based on individually estimated probit models. All estimations include controls on the individual, study and institutional/city level complemented with year and study fixed effects. In addition, for each specification of interest, interaction variables with all other covariates were included to control for potential disparate responses of the variable. The t-statistics of standard errors clustered at the institutional-study cohort level are presented between parenthesis. The only exception are the coefficients of the change in enrollment outcome variable, being estimated using OLS regressions at the institutional-study level. Here standard errors are clustered at the institutional level. Significance at the 10, 5 and 1 percent levels are respectively denoted by * / ** / ***.

A clear pattern emerges from the table. It tends to confirm the conjecture that cognitive skills determine whether pabo students are either mildly or greatly affected by the entry-tests. Most primarily a shift can be observed within the group of students with a pre-educational vwo background. A (vast) difference of more than two points on the average grade list determines whether or not a pabo student is affected in terms of retention rates. But perhaps more interesting, the final average grade has become more important since the entry-tests were implemented as predictor of eventually obtaining a degree. This concerns both students with a havo and vwo pre-educational background. To gauge the magnitude with an example: if a student with a pre-education of respectively havo and vwo has a final average of a seven instead of a six, it experiences *no* decline in the probability of obtaining a degree as opposed to a decline of about 7 to 8% points for vwo students and about 3% points for havo students. Therefore, the pabo has become more selective in 'singling out' students with a proper set of cognitive skills.

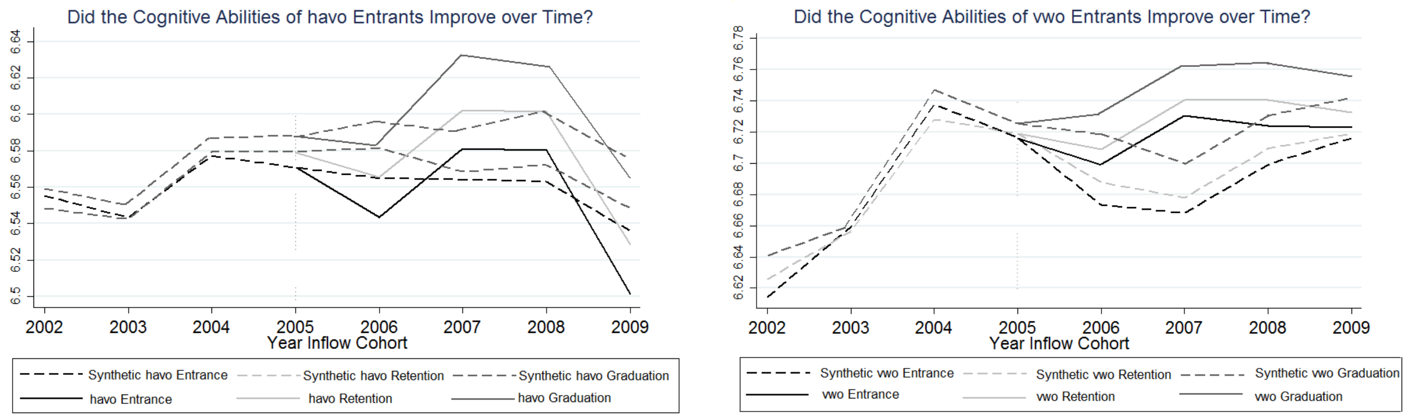
Still, as we saw earlier in the descriptive statistics table, the alleged image improvement of the pabo's –that is, being more selective- did *on average* not induce students with higher average grades to enter the pabo. To this end, we lastly investigate whether this also holds when we use the synthetic control method relating the final average grades to the implementation of the entry-tests.

Plots of the most complete synthetic control procedures of havo an vwo entrants are respectively presented on the left- and right hand side of figure 5. The black lines illustrate the average final grades of students entering the pabo, whereas the grey lines display the average grades of pabo students *conditional* upon entering the second year/ upon obtaining a degree. The joint gapsplot is shown in Appendix figure 4.

There are a couple statements conveyed by these figures. On a first note, the discrepancy between the lines respectively depicting the average grades of pabo students upon entrance ánd the average grades *conditional* upon entering the second year (obtaining a degree) during the pre-intervention period are almost nonexistent. Before the entry-tests were implemented therefore, there was no selection mechanism electing mainly those with high cognitive abilities to be granted permission to enter the second year. Intuitively, one would expect that this would reverse since the selection mechanism was put into place with the entry-tests. And indeed, the solid lines during the post-intervention period exemplify that the average grades improve when students enter the second year. Still, the gain is very limited - for havo entrants about 0.02 points and for vwo entrants 0.01 point.

On a second note, if we compare the counterfactuals demonstrating the average grade for both subgroups with those of the actual average grades of the pabo, a small

Figure 5 –The Allurement and Selection Effects of Entry-tests



Notes: Each of the (lines within) the figures are based on the most complete specifications of the synthetic control procedure inherent to the particular outcome variable. The cognitive abilities are proxied by the final averages of students with a highest obtained pre-education of havo or vwo. The retention respectively graduation estimates concern the average cognitive abilities for the subset of students who *de facto* enter the second year/ ever obtain their degree.

overall increase can be observed within the group of vwo entrants. This somewhat contrasts to the group of havo entrants of whom the change is less uniform and only positive for the enrollment cohorts of 2007 and 2008[64]. If we investigate this formally on grounds of statistical inference, none of the specifications turn out to be statistically significant. For this reason, if we use the average grades of entering students (cognitive abilities) as a proxy for the image of the pabo it becomes apparent that the 'allurement' effect has not induced students in possession of higher cognitive abilities to enter the pabo[65]. In other words, we cannot argue that the entry-tests rendered the pabo more attractive for students with higher cognitive abilities.

The latter observation, on a certain level, confirms our estimation results regarding the decline in retention rates and graduation rate efficiency. At the time the entry-tests were implemented, the prevalent belief was the pabo would attract 'better' students, which due to their higher cognitive abilities would be much more able to eventually obtain a degree. However the opposite became true, we just saw the cognitive abilities remained relatively constant, whereas the retention rates dramatically dropped eventually propagating to a partial decline in the graduation rate efficiency. The entry-tests therefore must simply be perceived as a selection mechanism, precluding the entrance towards the second year of the pabo for those students who possess insufficient cognitive skills.

D. Limitations

One should of course be aware that some caveats have to be mentioned in the interpretation of the main results of our paper. These caveats are mostly inherent to our identification strategy and to the available data. First, we do not observe whether pabo students *de facto* fail to attain the second year of the pabo *due to the entry-tests themselves.* To this end, we referred earlier to the presence of one main factor that helps to validate the belief in our obtained results – next to the primary argument that there were simply no alternative policy changes at the pabo during the intervention period. This factor relies on the pass-rates of pabo students respectively passing the entry-test of math and the Dutch language. If one compares the failure rates of the entry-tests per

[64] A potential reason for the drop in final averages of havo students in 2009 could be that the standardized exams were relatively difficult for the enrollment cohort of 2009. The counterfactual indeed shows a partial decline.

[65] Obviously, employing the average final grade of a student as a proxy for its cognitive abilities is imperfect. For instance, some profiles students choose during secondary education are generally recognized as being more challenging. Therefore an equal average grade for students with different profile types chosen does not imply these students have a similar set of cognitive skills. In addition, even similar grades of students with similar profiles may be imperfectly comparable in case these students are from different enrollment cohort backgrounds.

pre-educational subgroup to the obtained deterioration rates in respectively retention- and graduation rate efficiency behavior for each respective subgroup, one can observe a close connection between the two. Those students that generally experience the most difficulty of passing cognitive ability tests –the ones with a pre-educational level of mbo- were correspondingly the ones with the highest failure rates at the entry-tests. Whereas the groups with presumably higher cognitive abilities (havo/vwo) instinctively coincide with the observed failure-rate results[66]. For this reason, our results concerning the deterioration in retention rates for each pre-educational subgroup closely reproduced the pattern as conveyed by the failure rates of the entry-tests; *selection on cognitive skills*. This confirmation strengthens our belief in the obtained results.

A second caveat of this study is that we have not shown the results were robust to more strict conditions imposed on the donor pool of studies. Intuitively, more strict conditions could primarily affect our enrollment estimates. When we conducted a few robustness tests using the synthetic control method however, the results were fairly similar to the base outcomes. For example, when we imposed the condition that each institution should have at least 20 enrollments for a particular study, we only lost three studies in the donor pool. Yet, this condition only marginally changed the estimated enrollment change effects (these estimations are available upon request). In addition, the estimates of the remainder of the outcome variables only changed in the order of less than 0.5% points for certain enrollment cohorts.

The third and last main limitation is that were unable to control for the prevalent labor market conditions. Prima-facie analyses by De Graaf and Heyma (2014) show a close association between on the one hand, the growth in employment levels for primary school teachers and at the other hand, the growth in graduated teachers. According to their results, there seems to be a four year gap between the two – exactly the time to complete the pabo (Appendix figure 6). This indicates that enrollment growth rates at the pabo are strongly related to the *prevalent* labor market conditions, and not necessarily labor market outlooks (which we did control for). Ideally, one would then include prevalent labor market conditions within specific regions, enabling one to discriminate among pabo's in employment growth- and decline regions, respectively. It must be noted however, that the inclusion of prevalent labor market conditions are

[66] The discrepancies between the per pre-educational subgroup failure rates at the entry-tests and the estimated decline in retention rates for each pre-educational subgroup suggests that most primarily students with a pre-educational mbo background that failed the tests would have also stopped the pabo in absence of the entry-tests (about three-quarters for mbo students).

unlikely to substantially change our enrollment results. During the evaluation period, the total employment level (in fte) of primary-school teachers remained relatively constant (Appendix figure 3), and nationwide the unemployment level of graduates was fairly low. Data per region was not yet available by then, but graduated pabo students in declining employment regions are generally found to be quite mobile (OCW, 2016). We therefore state that during the evaluation period, the prevalent labor market conditions were very unlikely to be the principal reason to waive the pabo.

VI. Conclusions

After long-term concerns about the quality of pabo students, mandatory entry-tests in mathematics and the Dutch language were implemented over the course years of 2006 and 2007, respectively. Since the introduction, pabo students have to pass both tests during the first year of their study in order to ensure they possess at least the sufficient set of cognitive abilities to become a primary school teacher. This paper has examined the effects of the implementation of these entry-tests for pabo students. We obtain causal inferences by capitalizing the synthetic control method (e.g. Abadie and Gardeazabal, 2003) on a chronicled array of outcomes. This method resembles a formalization of differences-in-differences models, where -instead of multiple control units- a data driven design constructs one combined 'synthetic' control unit, by designating weights to units in the control group, dependent on how well they resemble the treatment group during the pre-intervention period. Correspondingly, the estimates of the treatment effects are simply generated by comparing the 'synthetic' counterfactual to the actual outcome variable of the pabo.

We have sketched the average treatment effects of the entry-tests gradually in more specific settings. The results being conveyed by the entire picture indicate that the entry-tests induced a progressive joint discouragement and image effect – step-by-step developing to a 30% point enrollments decline in 2009. A closer look per highest previously attained educational level indicated the magnitude of the former effect is relatively conservative. The availability of more specific covariates inherent to the group of students allowed us to discover that the three pre-educational groups all experienced a roughly similar drop in enrollments – gradually developing in a decline of about 40% points in 2009. We therefore cannot state the group with presumably the lowest set of cognitive skills were more discouraged to start the pabo compared to other subgroups.

In a way, the decline in enrollments was just as partially hypothesized by policymakers. It was conjectured the decrease in enrollments would not necessarily imply a similar decrease in the amount of graduates – reflecting the thought that the entry-tests would render an increment in the graduation rate efficiency of the students that *de facto* did enroll at the pabo. The results we obtained however, illustrate exactly the reverse occurred: The entry-tests not only induced a decline in the amount of enrollments, it also caused the 'survival' rate of the students being granted permission to enter the second year of the pabo to decline by about 6.5% points. Relatedly, when we scrutinized the magnitude of the effects per subgroup, intuitively, one would imagine the effects to be in line with the presumed cognitive standards of each subgroup; heterogeneous. And indeed, the results conform this notion. Pabo students with a pre-educational level of mbo were most adversely affected by the implementation of the entry-tests, as perceived from a decline of about 9% points in the survival rate. The drop in the survival rate was much less severe for those with the presumed highest cognitive abilities (4%), whereas the students within the intermediate cognitive range were correspondingly moderately negatively affected (6%).

At the time the entry-tests were implemented, another prevalent belief among policymakers was that pabo's would be able to attract 'better' students, which due to their higher cognitive abilities would have a higher probability of eventually obtaining a degree. The former conjecture however, could not be supported by empirical evidence. We found the cognitive abilities of pabo students remained relatively constant, which – on a certain level- is exactly in line with our estimation results occurring at the end of the first year of the pabo. Simply put, we can draw the conclusion that the introduction of the entry-tests functions as a *selection mechanism*, precluding the entrance towards the second year of the pabo for those students who possess insufficient cognitive skills, and not as an *alluring mechanism*, attracting students in possession of higher cognitive abilities due to an alleged image improvement of pabo's.

According to our results, the deterioration in the retention rates of the pabo's pass-through imperfectly to a decrease in the graduation rate efficiency. For all subgroups, about two-third of the declines in retention rates also finally settle in a reduction in graduation rates. Therefore, the entry-tests can additionally be perceived as a *commitment mechanism*: about one-third of the students that would have entered the second year in absence of the entry-tests would still have never obtained their degree.

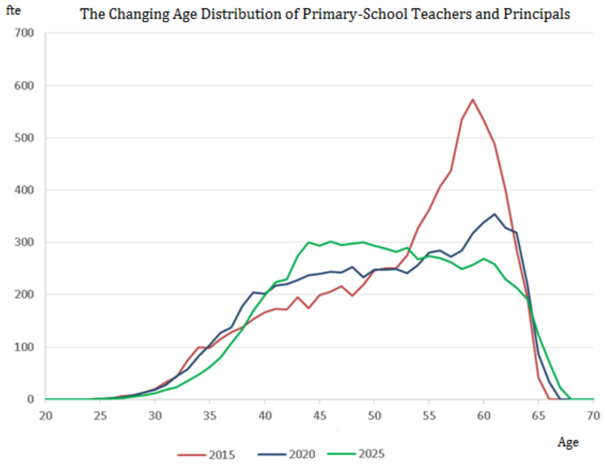
Lastly, one may inquire whether some groups were more affected by the entry-tests within the specific subgroups, we therefore continued the path of zooming in by allowing for a disparate responsiveness to the entry-tests within the distinct subgroups. The obtained results were not particularly surprising; females with a havo pre-educational background perform better relative to their male counterparts and students with disparate ethnic backgrounds (non-Western and Western immigrants with a pre-educational level of mbo) were hit more adversely relative to native Dutch students of the pabo since the entry-tests were implemented.

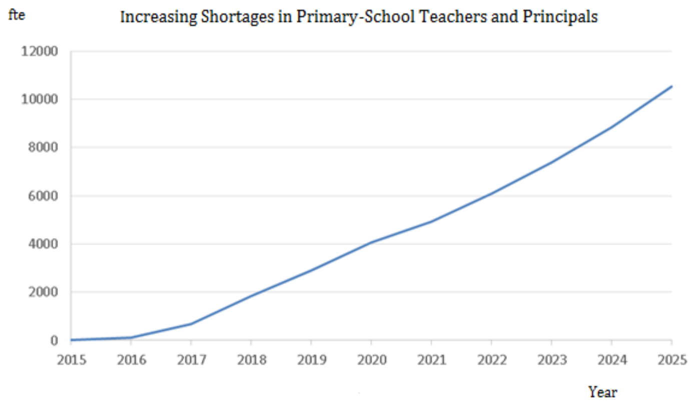
A few questions remain open after this paper and demand further research. Most importantly, we do not know to what extent the effectiveness (value-added) of *ex-post* intervention pabo graduates has changed relative to *ex-ante* pabo graduates. And if this is the case, did the (higher) cognitive set of skills induce the a change in effectiveness or can this be designated to non-cognitive factors? Moreover, does the extent to which cognitive skills of teachers matter differ for particular grades within primary-school? In other words, do *all* pabo graduates really need to acquire the imposed minimum cognitive set of skills, or can this been differentiated to the required skills for particular grades?

Discussion

The main conclusions of this paper come at the backdrop of large predicted shortages of primary-school teachers. Certainly, policy changes can only render the desired outcomes if one understands how it effects people's decisions in the labor market for teachers. The latter is a peculiar one. Its supply and demand never seem to be in full balance. And policymakers generally have a hard time to implement the right incentives for students to begin the pabo at the right time. This mainly has to do with the fact that labor market predictions get reacted to with lagged responses. As such, the surpluses of primary-school teachers of the last couple of years can quickly reverse into large shortages (figure 6). One of the primary reasons of the predicted shortages relies on the current age composition of primary school teachers (left-hand side figure 6). In the upcoming years, the teachers approaching their retirement age have to be replaced by novel ones – mainly those graduating from the pabo. In the current situation however, there are too few graduates in order to mitigate the shortages. The upcoming shortages are

Figure 6 –Increasing Shortages Primary-School Teachers





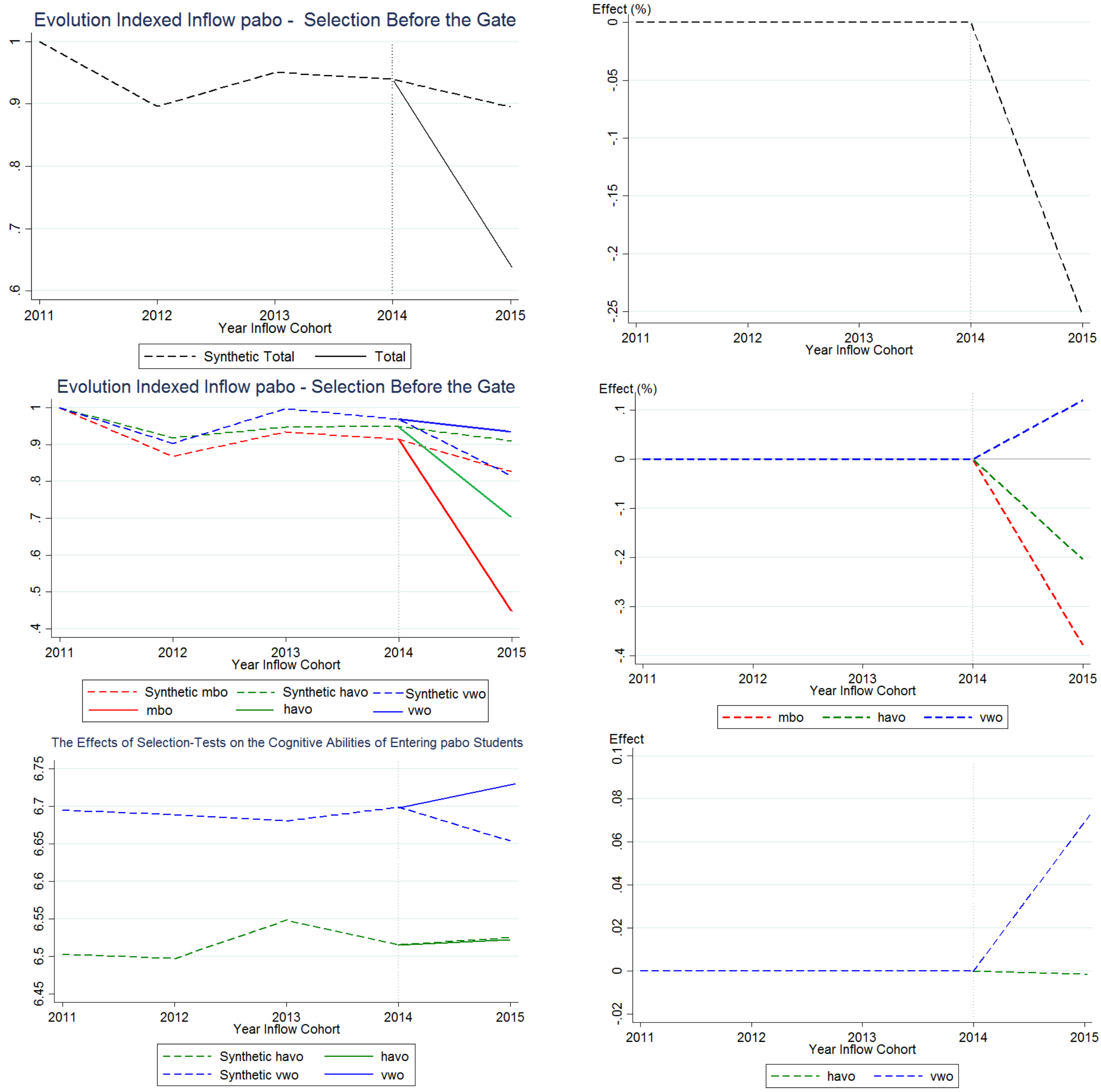
Notes: The figure on the left-hand side indicates the current cumulative age distribution (2015) and related the estimated cumulative age distributions in the year 2020 and the year 2025 respectively. The estimations hinge on a neutral outlook on the in-and outflow of primary-school teachers and principals, without any policy changes. The right-hand side figure indicates its consequences: An increasing discrepancy between the in-and outflow will induce a shortage. The predicted shortage is even an optimistic one, considering the standards used for graduation rate efficiency.
Source: CentERdata (2016)

essentially being aggravated by the imposed policy changes since the entrance cohort of 2006, i.e. the entry-tests since 2006 and the newly imposed selection tests since 2015. *Concerning the entry-tests:* despite the notion of policymakers that the graduation rate efficiency of pabo's would increase due to a potential increased allurement for students with higher cognitive abilities, we concluded earlier that the graduation rate efficiency declined due to the entry-tests, whereas the cognitive standards remained constant.

In addition, when we again use the synthetic control method to probe the magnitude of *the newly imposed policies* at the pabo since the entrance cohort of 2015 we can observe a severe enrollment decline - being shown in the subfigures of figure 7. The figures convey the admission tests *before one can enter the pabo,* additionally induced a decline of about 25% points in enrollments. Especially the subgroup with a highest previously attained mbo educational level shows to be most unable to pass *the tests in order to be allowed to enter the pabo*, as seen from a decline in enrollments of about 40% for this group.

If we analyze whether the cognitive skills of the students that were able to enter the pabo actually increased since these additional conditions were imposed (2015), we – perhaps surprisingly- found that this was not the case for havo entrants. In contrast, for the vwo entrants, we found a marginal increase in the cognitive abilities of entering students (i.e. average grade vwo). Since the apparent shortcomings of cognitive abilities are most prominent for the former group, the havo entrants of the enrollment cohort of 2015 are unlikely to substantially increase their performance rates at the entry-tests. In other words, the retention rates per subgroup are unlikely to improve significantly due

Figure 7 –The Effects of Selection Tests before the gate



Notes: Each of the (lines within) the figures are based on the most complete specifications of the synthetic control procedure inherent to the particular outcome variable. The mean squared prediction error is minimized over the period 2003-2005. The year 2002 is used in the matching procedure as covariate intrinsic to each study. The bottom figures display the cognitive abilities of respectively havo and vwo students over time. The cognitive abilities of both subgroups are proxied by their final average during secondary education.

to their inability to pass the entry-tests during the first year of the pabo. If the retention rates per subgroup of the enrollment cohort of 2015 were to increase, it is rather likely to be due to less students dropping out of the pabo because of a mismatch with the study. The composition change however, will still render an overall improvement in the retention rate of the pabo, simply because less students with a pre-educational mbo background are able to enter the pabo.[67]

[67] Moreover, the students that additionally dropped out due the Mens&Wereld tests in the first year, do not even enter the pabo.

The Implications

Despite the likelihood that at times of growing shortages the allurement to start the pabo will attract more students to choose for the pabo (i.e. job security), the current cognitive demands of the pabo's will remain a step too far for a lot of students potentially choosing the pabo. In essence, this describes the quality-quantity tradeoff, of which the government has increasingly wanted to ensure the former, now leading to a disbalance in the tradeoff. To this end, one must recognize students – and people in general- have a natural tendency to respond to incentives. Why would someone choose to enroll at the pabo in order to eventually become a primary school teacher? The occupation which has experienced a deterioration in terms of financial rewards relative to other tertiary education related occupations, the occupation where career opportunities are limited, where wages are severely compressed and largely unrelated to performance, and the occupation where the probability of getting a job was poor before during the years 2012-2015? Taking these factors into consideration, it should not be surprising that higher cognitive aptitude students in secondary education generally waive the pabo, other related occupations simply have surpassed the primary-school occupation – both in pecuniary and non-pecuniary factors[68]. As such, the ambition to attract higher aptitude students can only be met when the government introduces a severe raise pay[69], increasingly develops career opportunities and in general enhances the status of primary-school teachers, where one aims to reverse the image of a typical women-related occupation, making it more attractive for males to enter the pabo.

All in all, since the latter two factors have been aimed to have been addressed, and the former factor is not realistic due to insufficient government resources, one should move away from focusing explicitly on ensuring all new primary school teachers have excellent cognitive abilities. Certainly considering the evidence that (above-)high cognitive abilities of teachers translate towards a higher value-added in terms of students test-score gains is particularly scarce. This is not to say that a decent floor of cognitive abilities of teachers is not an essential feature of teachers- in contrast, primary school teachers should at least have the basic set of skills, but above average cognitive skills are deemed not to be an essential feature of primary-school teachers by the

[68] Recall the formula of Leigh (2012), page 17.

[69] This corresponds to a general notion*: A country gets what it deserves*. If it pays within the top 20 percentile of the earnings distribution, a country is correspondingly much more able to attract higher aptitude students. Despite most evidence is not causal, research by for example Dolton (2011) suggests -using an international panel of countries- the higher the teachers relative pay is within the earnings distribution, the higher correspondingly the average value added is of its teachers. It must be noted though, that Dolton's study considers the relative pay of secondary school teachers and relates it to the test scores of secondary school students.

predominance of the evidence in the academic literature. And while the graduation rate efficiency of students with higher cognitive aptitudes is higher, the current cognitive standards to enter the pabo demand a dramatic increase in the graduation rate efficiency of the students that de facto did enter the pabo[70], a dramatic increase in the order of this magnitude is not *realistic*. Explicitly stated: we recommend to drop the newly additionally imposed entrance conditions since 2015 for the subgroups of students with a pre-educational mbo or havo background. To this end, one must recognize that students entering teacher education self-select themselves into an occupation without a lot of risks and incentives for performances (Dohmen, 2010). Imposing strong incentives during the pabo therefore, should be compensated with an enhanced attractiveness of (non)-pecuniary factors of primary-school teachers. Since the latter did barely happen, the entry-tests as high-stake incentives caused a substantial amount of potential pabo students to waive the pabo (see also Appendix A).

For getting the disbalance in the quality-quantity tradeoff right, one should revert its focus to the quantity side. That is, one should ensure more students of adequate cognitive abilities to enter the pabo, rather than only those with above average cognitive aptitudes. One may ask how this works in practice? The government should exemplify its strong alluring features and guarantee that students with decent capabilities have the opportunity to enter the pabo. We will provide a couple examples.

First, the strongest upcoming feature of the pabo is that paves the gateway to an occupation in very high demand. That is, students currently in secondary education should know the pabo almost certainly guarantees a job in the future. Historical patterns show communication is key in this sense. Normally students study choices respond to prevalent situations in labor markets rather than its outlooks (at least in teacher-related occupations). Hence, by communicating its outlooks students will become more future-oriented and are more likely to enter the pabo. Ideally, the communication should be differentiated per region, particular focusing on shortages in the Randstad.

Second, the pool of recently graduated pabo students which were not able to get a job due to the surplus of graduates, should be incentivized to take up the occupation of a primary-school teacher. Not utilizing this pool of potential teachers would imply a severe waste of government resources and waste of potential talent.

---

[70] If one wants to ensure a similar amount of students entering the second year of the pabo in the enrollment cohort of 2015 relative to the enrollment cohort of 2014. The retention rate of the pabo has to increase with a vast 30% points. Back-on-the-envelope estimations based on the composition shift (2%), changes in characteristics (0.5%), and the drop of the necessity of having to make Mens & Wereld tests during the first year of the pabo (2.5%) constitute to an expected enhanced retention rate of about 5% points.

Third, pabo's should consider to implement a transition year for those students with insufficient cognitive abilities to directly enter the pabo. During this year, students are given the opportunity to upgrade their cognitive standards to a sufficient level in order to be able to graduate within a similar duration as students with higher cognitive abilities. Students that were previously unable to enter the pabo due to the obligation to pass the cognitive tests before entering are now still able to enter, albeit with a sideway. Moreover, the amount of students being discouraged to even try to enter are likely to decrease, simply because these students have more time to prepare themselves.

Lastly, a potential effective -albeit costly- measure to mitigate the upcoming primary-teacher shortages is to financially compensate pabo students already during their education. One must recognize (implicitly) it has become increasingly less attractive to start the pabo since the implementation of the so-called *'sociale leenstelsel'*. Students with high-levels of risk-aversion (i.e. generally women and individuals with low socioeconomic backgrounds) should be expected to most adversely react to such a scheme- even to the extent of not even enrolling at higher vocational education studies. In this light of thought, mainly teacher education studies are among the first to experience a loss in the amount of students enrolling. A surge in the financial costs of teacher education studies should be compensated by an even more substantial gain in benefits[71], otherwise one will find it very hard to attract sufficient students to start the pabo in order to make sure the upcoming shortages are adequately mitigated (Dohmen, 2010). Ideally, policymakers should consider making the financial compensation conditional on *de facto* entering the occupation as primary-school teacher.

[71] Developments in the increasing knowledge of the functioning of the human psyche suggest that individuals respond differently to direct versus delayed gains and direct versus delayed losses. If we start from the point of reference without a 'leenstelsel' , and we coin the implementation of the 'sociale leenstelsel' as a direct costs, whereas the delayed gains are primarily seen as an image improvement, then even without consideration of any risk-aversion, people on average tend to discount the delayed gains more than they would do so for direct gains, whereas the reverse is true for losses (Thaler (1981); Weber *et al.* (2007)). Therefore, increasing the current gains for students will attract more students- even to the extent future they are willing to impair future losses.

VII. References

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.

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1), 1-19.

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *The American Economic Review*, 93(1), 113-132.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490), 493-505.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2), 495-510.

Angrist, J. D., & Guryan, J. (2008). Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review*, 27(5), 483-503.

Arnold, I. J. (2015). The effectiveness of academic dismissal policies in Dutch university education: an empirical investigation. *Studies in Higher Education*, 40(6), 1068-1084.

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

Barber, M., & Mourshed, M. (2007). *How the world's best-performing schools systems come out on top.* McKinsey & Company.

Becker, G. (1985). Human Capital, Effort, and the Sexual Division of Labor. *Journal of Labor Economics*, 3(1), 33-58.

Bieri, C., & Schuler, P. (2011). Cross-curricular competencies of student teachers: a selection model based on assessment centre admission tests and study success after the first year of teacher training. *Assessment & Evaluation in Higher Education*, 36(4), 399-415.

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.

Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, 66(2), 103-115.

Calcagno, J. C., Bailey, T., Jenkins, D., Kienzl, G., & Leinbach, T. (2008). Community college student success: What institutional characteristics make a difference? *Economics of Education review*, 27(6), 632-645.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of risk and uncertainty*, 19(1-3), 7-42.

CentERdata. (2013). *De toekomstige arbeidsmarkt voor onderwijspersoneel 2013-2025.* Tilburg.

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.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics*, 126(4), 1593-1660.

Clotfelter, C. L., 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.

Cörvers, J. G. (2014). *De arbeidsmarkt voor leraren: theorie, beleid en werkelijkheid.* Tilburg University.

CPB. (2013). *Arbeidsmarkt leraren: aanpassingsmechanismen en aangrijpingspunten voor beleid.* CPB Notitie.

De Paola, M., & Gioia, F. (2011). Risk aversion and major choice: Evidence from Italian students. (No. 201107).

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

Dohmen, T. J. (2010). Waarom sommigen liever leraar worden en anderen beter geen bankier kunnen worden. *Inaugural speech.* Maastricht University.

Dohmen, T., & Falk, A. (2010). You get what you pay for: Incentives and selection in the education system. *The Economic Journal*, 120(546), F256-F271.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are Risk Aversion and Impatience Related to Cognitive Ability? *American Economic Review*, 100, 1238-1260.

Enden, S. v. (2012). *An econometric analysis on inflow and outflow of future primary school teachers.* Master Thesis.

Fike, D. S., & Fike, R. (2008). Predictors of First-Year Student Retention in the Community College. *Community College Review*, 36(2), 68-88.

Fouarge, D., Kriechel, B., & Dohmen, T. (2014). Occupational sorting of school graduates: The role of economic preferences. *Journal of Economic Behavior & Organization*, 106, 335-351.

Fryer Jr, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, 86(2), 447-464.

Geddes, B. (2003). *Paradigms and sand castles: Theory building and research design in comparative politics.* University of Michigan Press.

Gerritsen, S., Plug, E., & Webbink, D. (2014). *Teacher quality and student achievement: Evidence from a Dutch sample of twins.* (No. 294). CPB Netherlands Bureau for Economic Policy Analysis.

Gilpin, G., & Kaganovich, M. (2012). The quantity and quality of teachers: Dynamics of the trade-off. *Journal of Public Economics*, 96(3), 417-429.

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

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *The American Economic Review*, 100(2), 267-271.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of economic literature*, 607-668.

Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4), 267-321.

Hanushek, E. A., Piopiunik, M., & Wiederhold, S. (2014). The value of smarter teachers: International evidence on teacher cognitive skills and student performance. (No. w20727). National Bureau of Economic Research.

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

Heyma, A. O., de Graaf, D., & Klaveren, C. P. (2006). *Exploratie van beloningsverschillen in het onderwijs 2001-2004.* Ministerie van Onderwijs, Cultuur en Wetenschap.

Hoxby, C. M., & Leigh, A. (2004). Pulled away or pushed out? Explaining the decline of teacher aptitude in the United States. *The American Economic Review*, 94(2), 236-240.

Inspectie van het Onderwijs. (2007). De staat van het onderwijs. Onderwijsverslag 2005-2006. *Utrecht: Inspectie van het Onderwijs*.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American psychologist*, 39(4), 341.

Kane, T. J., & Staiger, D. O. (2008). Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation. *NBER Working Paper No. 14607*.

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.

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, 114(2), 497-532.

Lakdawalla, D. (2006). The Economics of Teacher Quality*. *Journal of Law and Economics*, 49(1), 285-329.

Lau, L. K. (2003). Institutional factors affecting student retention. *Education*, 124(1), 126.

Leigh, A. (2012). Teacher pay and teacher aptitude. *Economics of education review*, 31(3), 41-53.
Leuven, E., Oosterbeek, H., & Klaauw, B. (2010). The Effect of Financial Rewards on Students' Achievement: Evidence From a Randomized Experiment. *Journal of the European Economic Association*, 8(6), 1243-1265.
McNabb, R., Pal, S., & Sloane, P. (2002). Gender differences in educational attainment: The case of university students in England and Wales. *Economica*, 69(275), 481-503.
Meijer, J., Vermeulen-Kerstens, L. M., Schellings, G. L., & Meijden, A. J. (2006). Reken-en taalvaardigheid van instromers Lerarenopleiding Basisonderwijs.
Meijerink. (2012). *Een goede basis: Advies van de Commissie Kennisbasis pabo.* HBO-raad.
Ministerie van Onderwijs, Cultuur en Wetenschap. (2005). *Meer kwaliteit en differentiatie: de lerarenopleidingen aan zet. Beleidsagenda lerarenopleidingen 2005–2008.* Den Haag: OCW.
Ministerie van Onderwijs, Cultuur en Wetenschap. (2008). *Krachtig meesterschap, kwaliteitsagenda voor het opleiden van leraren 2008-2011.* Den Haag: OCW.
Minne, B., & Webbink, D. (2008). Raising teacher supply: An assessment of three options increasing wages. *CPB Memorandum 194*.
Murnane, R., Willett, J. B., & Levy, F. (1995). The Growing Importance of Cognitive Skills in Wage Determination. *The Review of Economics and Statistics*, 77(2), 251-66.
Nielsen, H. S., & Vissing-Jorgensen, A. (2006). *The impact of labor income risk on educational choices: Estimates and implied risk aversion.* Unpublished Manuscript, Department of Economics, University of Aarhus.
NRC. (2016, february). De pabo wordt steeds witter. *NRC*.
Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational evaluation and policy analysis*, 26(3), 237-257.
Onderwijsraad . (2005). Kwaliteit en inrichting van de lerarenopleidingen.
Researchned. (2012). *Imago PABO en lerarenopleidingen onder mannelijke en allochtone havisten.*
Rijst, R. v., Tigelaar, D., Driel, J. v., Snoek, M., & Verseveld, M. v. (2015). Effecten van selectie ten behoeve van de lerarenopleidingen. Een literatuurreview in opdracht van NRO. *Leiden/Amsterdam: ICLON, Universiteit Leiden/Kenniscentrum Onderwijs en Opvoeding, HvA*.
Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education*, 6(1), 43-74.
Rothstein, J. (2010). Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement. *The Quarterly Journal of Economics*, 125(1), 175-214.
Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2), 135-146.
Scott, M., Bailey, T., & Kienzl, G. (2006). Relative success? Determinants of college graduation rates in public and private colleges in the US. *Research in higher education*, 47(3), 249-279.
Sneyers, E., & De Witte, K. (2015). The effect of an academic dismissal policy on dropout, graduation rates and student satisfaction. Evidence from the Netherlands. *Studies in Higher Education*, 1-36.
Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *The Journal of Economic Perspectives*, 24(3), 97-117.
Steeg, M. v., & Gerritsen, S. (2016). Teacher evaluations and pupil achievement gains: Evidence from classrooms observations. *De Economist*, 1-25.
Steeg, M. v., Lanser, D., & Vermeer, N. (2011). Niveau onderwijs daalt. *CPB Policy Brief 2011/05 | 6 06 2011*.
Straetmans, G. J., & Eggen, T. J. (2005). Afrekenen op rekenen: over de rekenvaardigheid van pabo-studenten en de toetsing daarvan. *Tijdschrift voor hoger onderwijs*, 23(3), 123-139.
Straetmans, G. J., & Eggen, T. J. (2011). *WISCAT-pabo: Ontwerp, kwaliteit en resultaten van een geruchtmakende toets.* Schramade ea (Red.) Handboek Effectief Opleiden, sectie, 9-2.

Tiebout, C. M. (1956). A pure theory of local expenditures. *The journal of political economy*, 416-424.
Tinto, V. (2002, June). Promoting student retention: Lessons learned from the United States. *In 11th Annual Conference of the European Access Network, Prato, Italy (Vol. 19).*
Valli, R., & Johnson, P. (2007). Entrance examinations as gatekeepers. *Scandinavian Journal of Educational Research*, 51(5), 493-510.
van den Enden, S., & van der Wiel, K. M. (2012). Entreetoets schrikt pabo-studenten af. *Economisch Statistische Berichten*, 97(4646), 630-632.
Vogels, H. M., & Bronneman-Helmers, H. M. (2006). *Wie werken er in het onderwijs?: op zoek naar het 'eigene' van de onderwijsprofessional.* SCP Sociaal en Cultureel Planbureau.
Warps, J., Wartenbergh, F., Kurver, B., Muskens, M., Hogeling, L., & Pass, J. (2010). *Studiesucces en studie-uitval bij eerstejaars in de hbo-onderwijsopleidingen. Rapportage op basis van de Startmonitor 2008-2009.* ResearchNed: Nijmegen, februari.
Waterreus, I. (2003). *Lessons in Teacher Pay; Studies on Incentives and the Labor Market for Teachers.* Amsterdam: Thela Thesis: Proefschrift, Universiteit van Amsterdam.
Webbink, D. (2000). Wie wil er nog leraar worden?(Who wants to be a teacher?). *Economisch-Statistische Berichten*, 85(4257), 444-446.
Webbink, D., & Hartog, J. (2004). Can students predict starting salaries? Yes! *Economics of Education Review*, 23(2), 103-113.
Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric discounting in intertemporal choice a query-theory account. *Psychological science*, 18(6), 516-523.

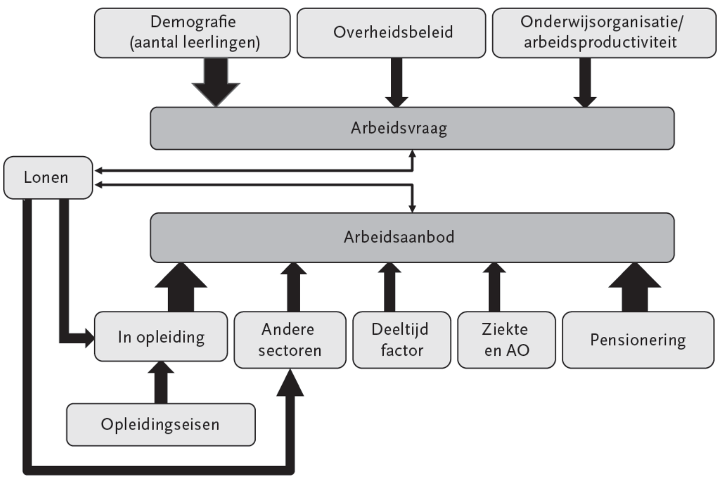
Appendix A: Concepts

Analysis of the Dutch labor market for primary-school teachers

The *demand for teachers* is primarily determined by demographic factors, i.e. the inflow of new students into primary education determines whether more (or less) teachers are required. Relatedly, governmental policies concerning class-size, workload and retrenchments on primary-school resources could attribute to an alteration in the demand for teachers. Regulation of the quantitative (#hours standard) and qualitative (standards on outflow quality pabo-students) content of education are essential herein. It is simple to imagine this could directly translate into the supply of students enrolling in teacher educations. For instance, retrenchments at primary-schools and rigid wages decrease the attractiveness to enroll at teacher education institutions.
The *supply of teachers* is primarily dependent on the outflow of students of the teacher education programs, of whom the inflow into teacher education is partly determined by its financial attractiveness. Waterreus (2003) shows that in those countries where teachers are paid relatively well, the education institutions have the ability to select the best applicants due to the availability of more applicants opting for a teacher education. Unfortunately, students that have completed a pabo education have relatively low outside opportunities, which declines the attractiveness of starting a pabo program. Another determinant of the attractiveness of teacher education is the employment perspective of inflowing students. De Graaf and Heyma (2014) show that employment surpluses or shortages on the teacher labor market tend to be followed by changes in outflow of graduated teachers about four years later – exactly the required time to complete a teacher education. Next to the inflow of graduated students into the teacher labor market

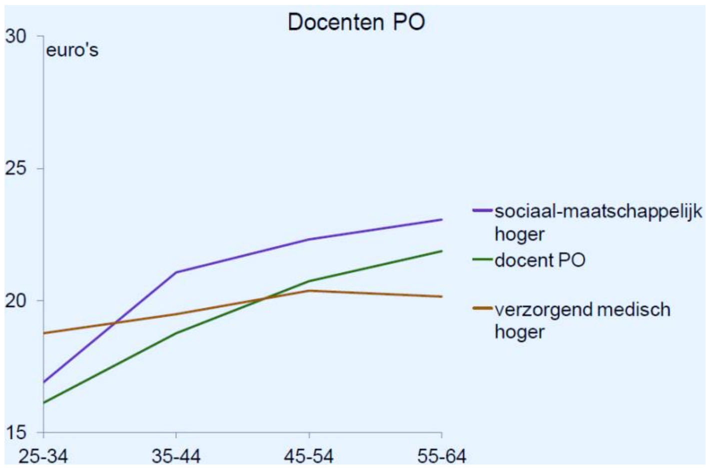
is the outflow of teachers due to retirement options, either due to favorable options to retire earlier (BAPO) or due to obligatory government provisions (AOW).
In general terms, the preferences of people largely determine the attractiveness of being a teacher. Dohmen and Falk (2010) show the relatively favorable non-pecuniary job characteristics of teachers closely relate to the preferences of women, relative to men who value the pecuniary factors generally more. The *average* elasticity of supply with respect to wages is therefore deemed to be fairly low (about 0.2-0.4) (Minne & Webbink, 2008), yet when the government aims to attract more males to become teacher in primary-education, paying higher wages is required relative to non-pecuniary factors.
The wages of (primary) teachers in the Netherlands are characterized by the low amount of differentiation. For instance, teachers are scaled based on their function 'mix', in case of primary teachers this mainly consist of the LA-scale. The remaining quarter of the primary-school teachers are in the higher LB-scale. These are largely predetermined by cao settlements, which are partly in correspondence with the wage tendencies in the market sectors. Despite the correspondence, the development in teacher wages have lost ground compared to the developments in the market sector. This can be attributed due to the teacher wages moving in correspondence with the total market sector, and not the sector for high-educated individuals. The latter have especially gained in the last decades. Relatedly, the wages of teachers do not depend on tendencies in bonus schemes in the private sector, of which have especially surged in the private sector relative to teachers.
An analysis of *future developments in the Dutch labor market for primary-school teachers* has been provided by CentERdata. Their main ingredients for future demand are demographic shifts, whereas the supply is dependent on net inflow of primary-teachers. The analysis includes three business cycle scenarios, and intuitively, under favorable circumstances the supply of teachers declines. Under the latter scenario the future primary-teacher deficit is largest, and will peak in the year 2025.
The determinants of the demand and supply in the Dutch labor market for primary-school teachers are summarized in left-hand side figure below. (For detailed Dutch labor market for teachers analyses see CPB (2013) and De Graaf & Heyma (2014)).

Determinants of demand and supply in labor markets for teachers



Source: Wat is goed onderwijs? Chapter 12, De Graaf & Heyma

Outside options: *Gross average hourly wages of teachers with pabo credentials and gross wages of private sector employees of related occupations (ex-post decision teacher).*



Source: CPB (2013)

Gross hourly wages on vertical axis, age intervals on horizontal axis.

Conceptual Framework

To what extent will teacher quality will improve by student testing? Essentially, individuals of both a low(er) and a high cognitive ability are affected by the effects of high-stake testing. Suppose the utility function of risk-averse students is defined by:

$$U_a = \sqrt{x+\alpha} - \left(\frac{e-1}{\theta}\right)$$

Where x refers to the perceived future wage- and non-pecuniary benefits of the agent within the field of study choice and α reflects an intrinsic motivation component. Moreover, agents dislike to exert effort on their study, but find it increasingly less strenous when their cognitive ability θ is higher. To this end, it is assumed that the cognitive ability of agents is positively related to their outside option, in this case, the perceived future net benefits of another (vocational) study. Agents that want to enter a study of becoming a primary teacher face the new higher required level of effort due to the introduction of entrance tests, yet on the other side also the image of primary teacher increases x_{1new}. The agent prefers the primary education teacher study above its outside option study when the following *participation constraint* is satisfied:

$$(PC) \qquad U_a(e_{new}) \geq U^{out} \Rightarrow \sqrt{x_{1new}+\alpha_1} - \left(\frac{e_{1new}-1}{\theta}\right) \geq \sqrt{(x_2(1+\theta)+\alpha_2)} - \left(\frac{e_2-1}{\theta}\right)$$

$$\Rightarrow \sqrt{(x_{1new}-x_2(1+\theta))+(\alpha_1-\alpha_2)} \geq \left(\frac{e_{1new}-e_2}{\theta}\right)$$

The formula above conveys two main forces opposing one another compared to the situation without any entry-tests. The right-hand side indicates the additional cost of effort costs of effort students have to endure when entering the pabo. This effect induces less students to opt for the pabo. In turn, the left-hand side indicates the potential enhanced image of the pabo, which mitigates the former effect and makes it more attractive for students to enter the pabo. Note that if we could introduce a time component in the model we could explain factors such as a deterioration in pay of teachers relative to alternate occupations, and a decline in the image of primary-school teachers. Moreover, we could allow for students differing in their degrees of risk-aversion. For the sake of presentation, the current formulation is easier.

Translating the Model to our Obtained Results

Our results revealed a gradual but severe decline in the change in enrollments of the pabo. This must be seen as the outcome of the formula above, implying if both forces are combined, the perceived additional cost of effort outweigh the potential image improvement by a landslide. Furthermore, we found that on average the cognitive skills of pabo entrance did not improve over time. This shows not necessarily students with a lower set of cognitive skills are discouraged to enter the pabo. In contrast, we should refer to differences in *risk-aversion* of students to explain why certain students do enter, and others do not. Another potential reason elucidating the decline in enrollments could simply be due to the relative deteriration in (non-) pecuniary factors of primary-school teachers over time.

Considering the group that did in fact enter the study should be able to exert the new level of effort in order to succeed the entrance tests. It is assumed that the effort one exerts is related to the perceived future (non-)pecuniary benefits. In *incentive compability* terms:

$$(ICC) \qquad U_a(e_{new}) \geq U_a(e_{old}) \Rightarrow \sqrt{x_{1new}+\alpha_1} - \left(\frac{e_{1new}-1}{\theta}\right) \geq \sqrt{(x_{1old}+\alpha_1)} - \left(\frac{e_{1old}-1}{\theta}\right)$$

$$\Rightarrow \sqrt{(x_{1new}-x_{1old})} \geq \left(\frac{e_{1new}-e_{1old}}{\theta}\right)$$

If we interpret the formula on an individual level then we could state the enhancement in perceived future (non)-pecuniary wage benefits should outweight the increase in costs of effort, the latter which is mitigated by the cognitive set of skills by a student. On an aggegrated level however, one should interpret the formula as follows: The students that do enter the study but have a relative lack of cognitive ability will not be able to exert enough effort to succeed the entrance tests, which compels them to stop the study. This group will either switch or completely dropout of higher vocational education. On the other hand, if the cognitive skills of entrants (and the potential increase in (non)-pecuniary factors) is larger than the surge in effort costs, we should observe an increase in retention rates. Our obtained results reveal that *on average* the cognitive skills of pabo entrants did not change, whereas the (non)- pecuniary factors of primary-school teachers barely improved over time. Therefore, the hike in the costs of effort simply is a bridge too far for students failing at the entry-tests.

The former elements are summarized below.

Student	Do not enter	Enter
Fail	PC & ICC both bind	PC satisfied, ICC binds
Succeed	Meet ICC but PC binds	Both PC and ICC satisfied

The principal is in this case the government, which minimizes the costs of the (non-) pecuniary factors x_{1new}. Combining the PC and the ICC yields;

Rewriting (ICC) $\sqrt{(x_{1new} - x_{1old})} \geq \left(\frac{e_{1new} - e_{1old}}{\theta}\right) \Rightarrow \sqrt{(x_{1new} - x_{1old})} + \left(\frac{e_{1old}}{\theta}\right) \geq \left(\frac{e_{1new}}{\theta}\right)$

Plugging into (PCC) $\sqrt{(x_{1new} - x_2(1+\theta)) + (\alpha_1 - \alpha_2)} \geq \left(\frac{\sqrt{(x_{1new} - x_{1old})} + \left(\frac{e_{1old}}{\theta}\right) - e_2}{\theta}\right)$

Multiplying by θ $\sqrt{\left(x_{1new} - x_2(1+\theta)) + (\alpha_1 - \alpha_2)\right)\theta} - \sqrt{(x_{1new} - x_{1old})} \geq \left(\frac{e_{1old}}{\theta}\right) - e_2$

Quadrating yields $\left(x_{1new} - x_2(1+\theta)) + (\alpha_1 - \alpha_2)\right)\theta^2 - (x_{1new} - x_{1old}) \geq \left(\left(\frac{e_{1old}}{\theta}\right) - e_2\right)^2$

Solving for x_{1new} gives $(x_{1new}\theta^2 - x_2(\theta + \theta^2)) + (\alpha_1 - \alpha_2)\theta^2) - (x_{1new} - x_{1old}) \geq \left(\left(\frac{e_{1old}}{\theta}\right) - e_2\right)^2$

$$x_{1new}\theta^2 - x_{1new} \geq \left(\left(\frac{e_{1old}}{\theta}\right) - e_2\right)^2 + x_2(\theta + \theta^2)) + (\alpha_1 - \alpha_2)\theta^2 - x_{1old}$$

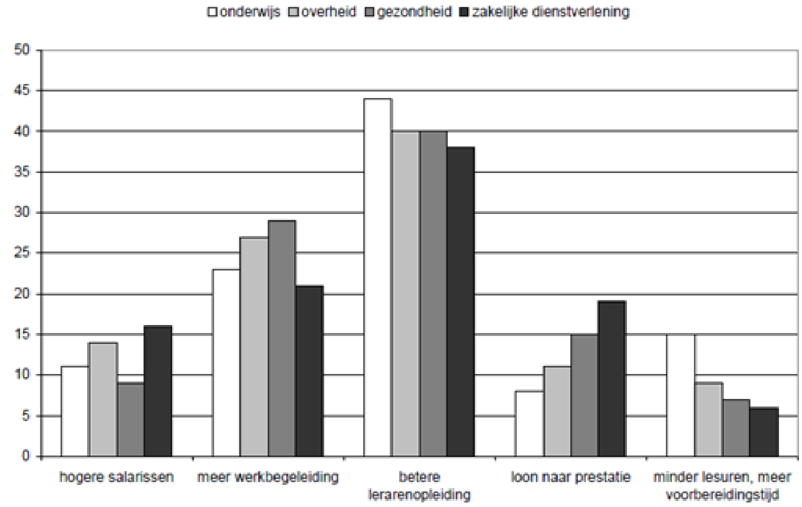
$$x_{1new}(\theta^2 - 1) \geq \left(\left(\frac{e_{1old}}{\theta}\right) - e_2\right)^2 + x_2(\theta + \theta^2)) + (\alpha_1 - \alpha_2)\theta^2 - x_{1old}$$

$$x_{1new} \geq \frac{\left(\left(\frac{e_{1old}}{\theta}\right) - e_2\right)^2 + x_2(\theta + \theta^2) + (\alpha_1 - \alpha_2)\theta^2 - x_{1old}}{(\theta^2 - 1)}$$

Calibrating the model using new data would generate the optimal wages for primary-school teachers.

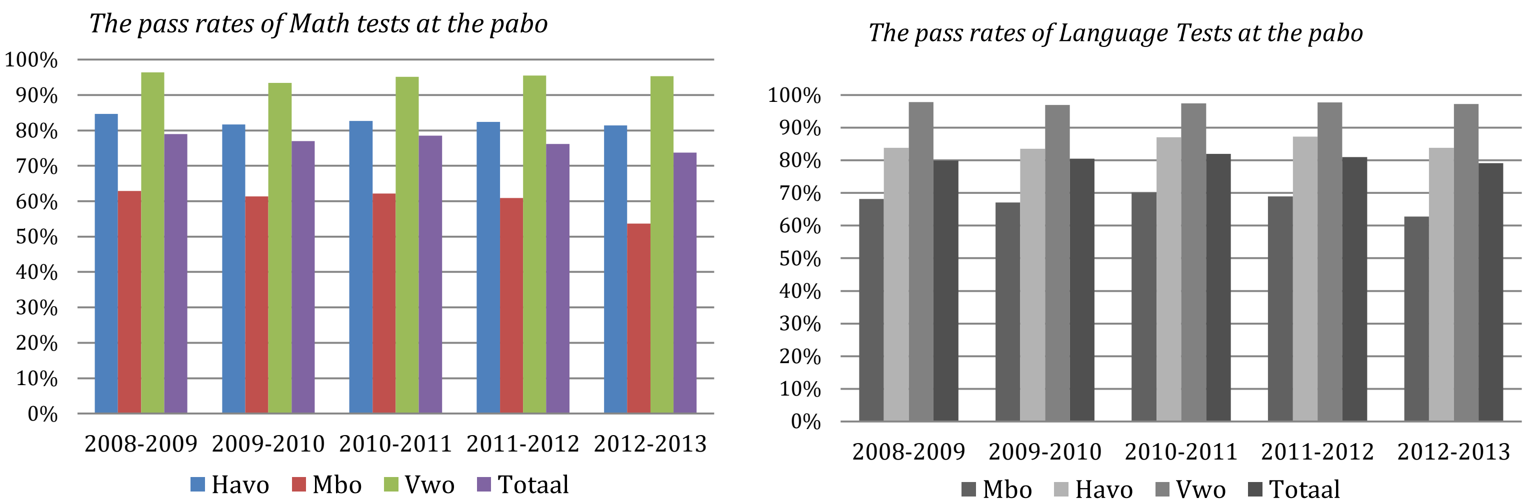
Appendix B:. Figures

Appendix figure 1: Required Measures to Improve the Quality of Primary-School Teachers



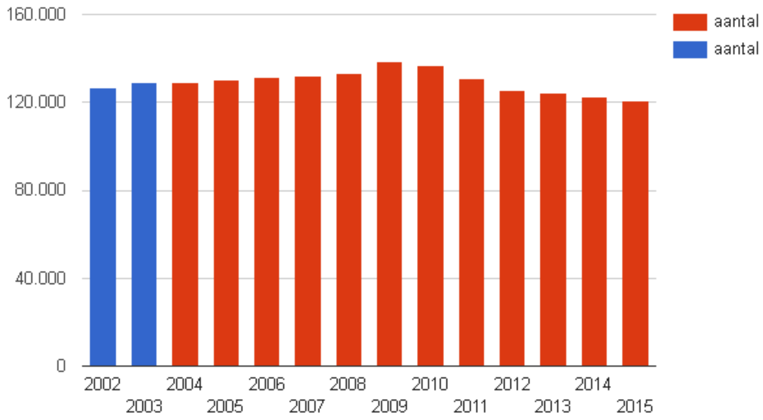
Notes: Responses differentiated per labor market level, 2005. Percentages on vertical axis.
Source: McKinsey and Co. (2005)

Appendix figure 2: Pass Rates of Entry-Tests at the pabo



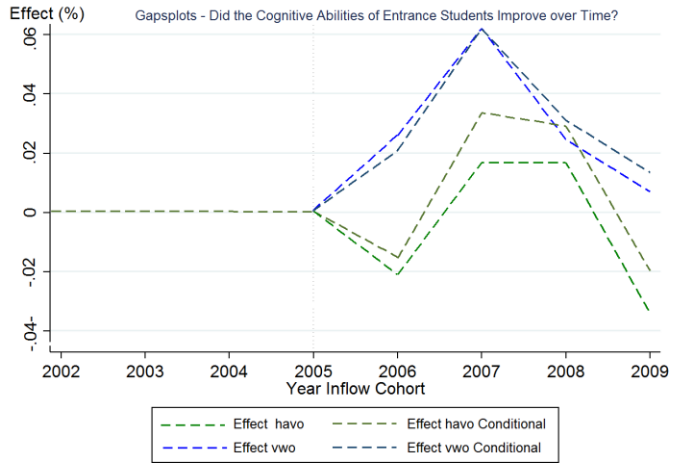
Notes: Non-reported but available data from the cohort 2006-2007 is in line with the pass rates as indicated in the graphs. *Source:* Cito annual reports

Appendix figure 3: Employment levels in primary-education (in fte's)



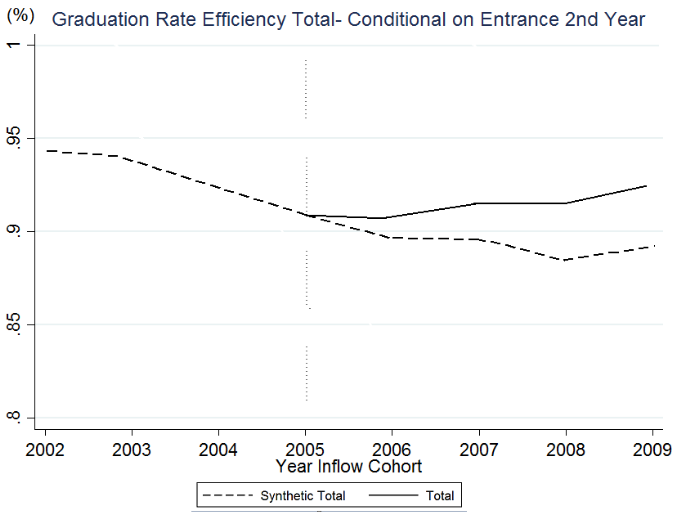
Notes: Amounts until 2003 may be hard to compare between years, due to other selection criteria being used.
Source: Stamos.

Appendix figure 4: Gapsplot Cognitive skills of pabo Entrants



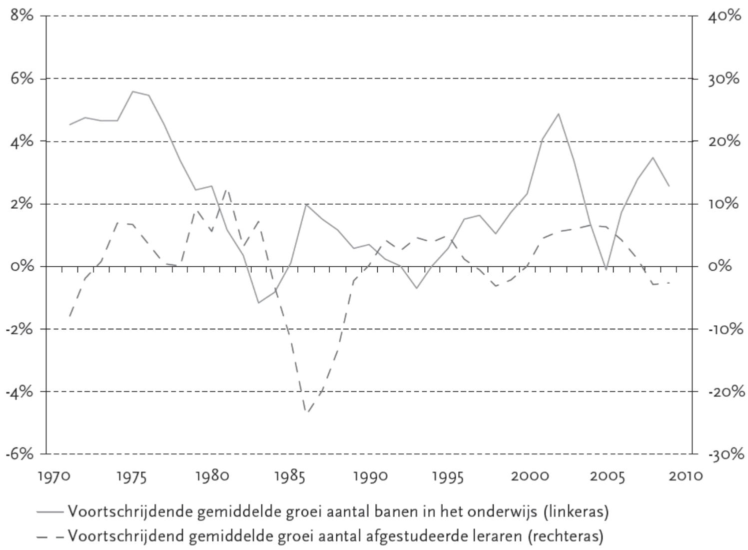
Notes: Conditional lines reflect average final grades conditional upon entering the second year. For clarity, not shown are the effects conditional upon obtaining a degree. Yet its magnitudes are roughly similar as shown above.

Appendix figure 5: Conditional lines reflect average final grades conditional upon entering the second year.



Notes: The lines of this figure is based on the most complete specification of the synthetic control procedure. Labels on the vertical axis represent percentage points within a 0-1 range (e.g 0.1 reflects 10 %).

Appendix figure 6: Lagged Relationship between Growth in Primary-School Jobs and Amount of Entrants in pabo



Source: De Graaf & Heyma , Wat is goed onderwijs? Chapter 12

Appendix figure 7: Stated motives of high-educated people to choose a job in the government sector (in %)

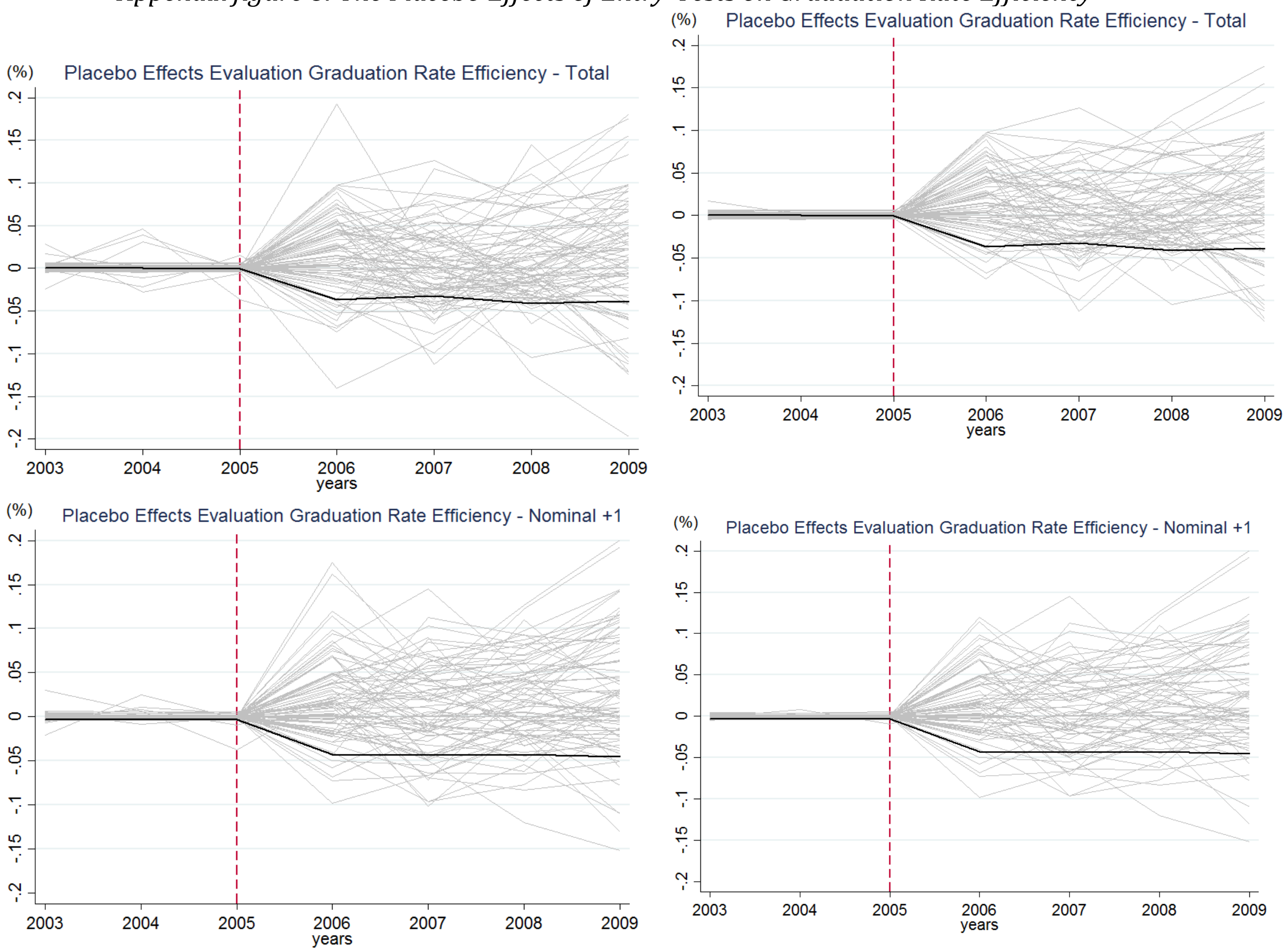
	inhoud werk	zelfstandigheid en/of verant-woordelijkheid	sfeer	combineren werk/zorg	leiding organisatie	direct-leiding-gevende	loopbaan-mogelijk-heden	beloning	resultaat-gerichtheid organisatie	secundaire arbeids-voorwaarden	werkplek (fysiek)	werkdruk	informatie/ communicatie	promotie
overheid totaal[b]	86	79	65	48	44	47	51	46	42	43	34	34	31	24
openbaar bestuur	89	83	62	47	42	49	62	54	44	55	30	32	29	31
veiligheid	93	84	67	44	45	49	68	54	44	56	35	31	32	26
onderwijs[c]	86	77	66	50	47	46	44	42	41	38	37	35	33	20
primair onderwijs	83	76	72	50	54	51	40	43	44	37	40	37	38	16
voortgezet onderwijs	83	74	67	55	51	43	41	46	38	42	43	33	35	19
bve	86	80	63	54	42	45	48	41	41	44	34	38	36	22
hbo	91	81	54	51	31	34	44	43	35	36	29	34	28	18
universiteiten	92	84	58	40	31	47	61	35	37	28	25	29	18	35

a Percentage ingestroomde ambtenaren dat het functieaspect (heel) belangrijk vond bij de keuze van de nieuwe baan.
b Totaal overheid is inclusief onderzoeksinstellingen en academische ziekenhuizen.
c Totaal onderwijs is exclusief onderzoeksinstellingen en academische ziekenhuizen.

Bron: BZK (Personeels en mobiliteitsonderzoek'04)

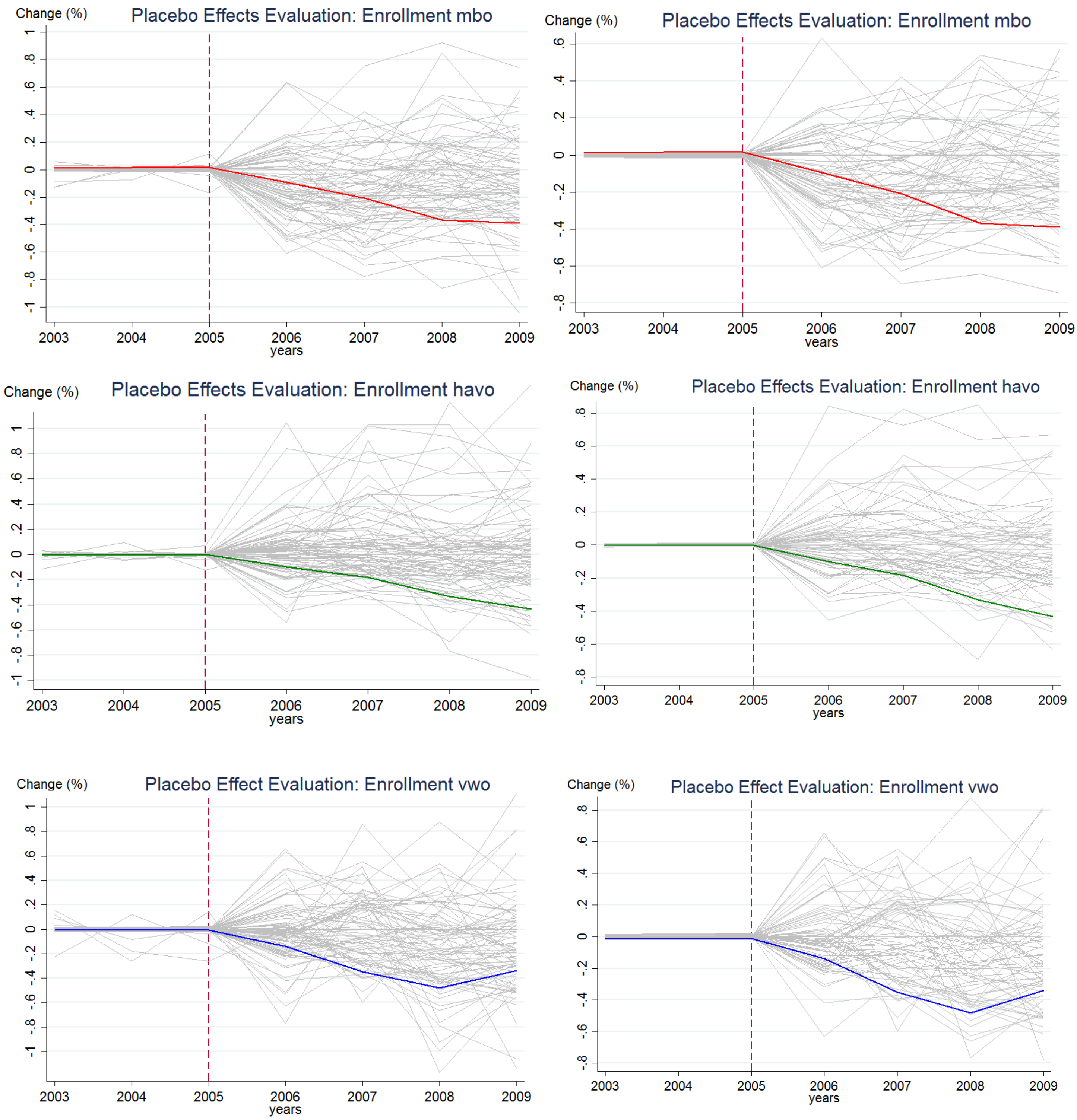
Notes: Based on survey data of 2003, primary-school teachers mainly self-select into the occupation due to the work content, its atmosphere and being an independent professional
Source: Vogels, H. M., & Bronneman-Helmers, H. M. (2006)

Appendix figure 8: The Placebo Effects of Entry-Tests on Graduation Rate Efficiency



Notes: Each of the (lines within) the subfigures are based on the most complete specifications of the synthetic control procedure inherent to the particular outcome variable. The mean squared prediction error is minimized over the period 2003-2005. The year 2002 is used in the matching procedure as covariate intrinsic to each study. Labels on the vertical axis represent percentage points within a 0-1 range (e.g 0.1 reflects 10 %). The plots are based on 110 studies, including the pabo.

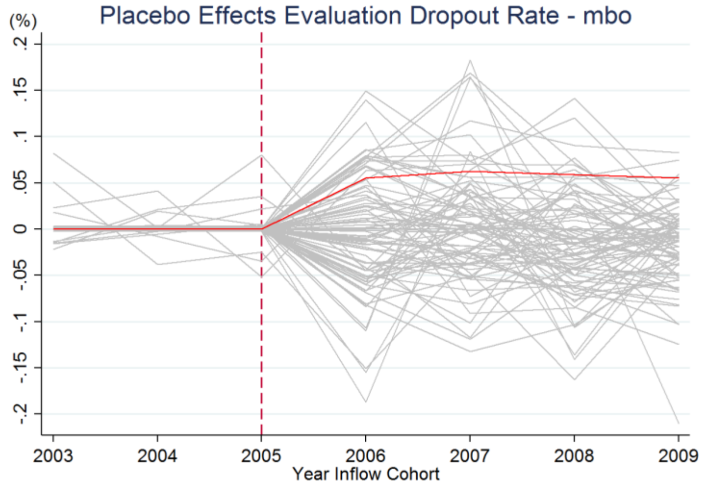
Appendix Figure 9–The Placebo Effects of Entry-Tests per Highest Previously Attained Education
Change in Enrollment
Exclusion 10% worst MSPE

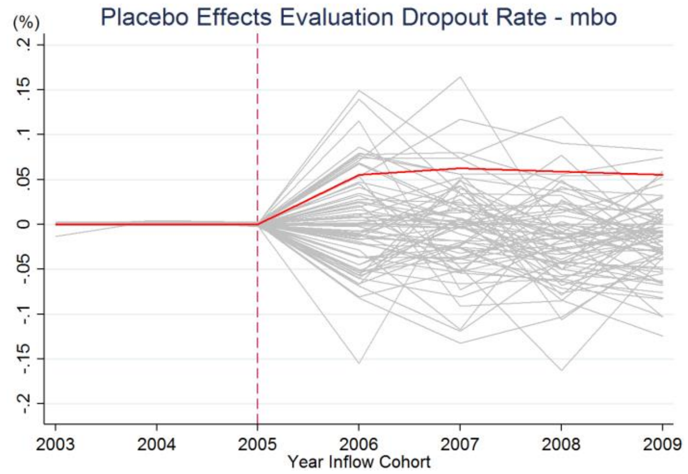


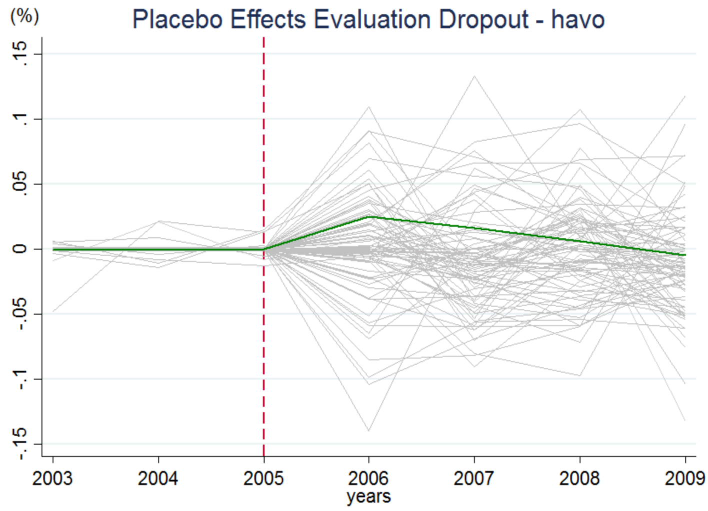
Notes: Each of the (lines within) the subfigures are based on the most complete specifications of the synthetic control procedure inherent to the particular outcome variable. The mean squared prediction error is minimized over the period 2003-2005. The year 2002 is used in the matching procedure as covariate intrinsic to each study. Labels on the vertical axis represent percentage points within a 0-1 range (e.g 0.1 reflects 10 %). The enrollment plots are based on 87 studies, including the pabo.

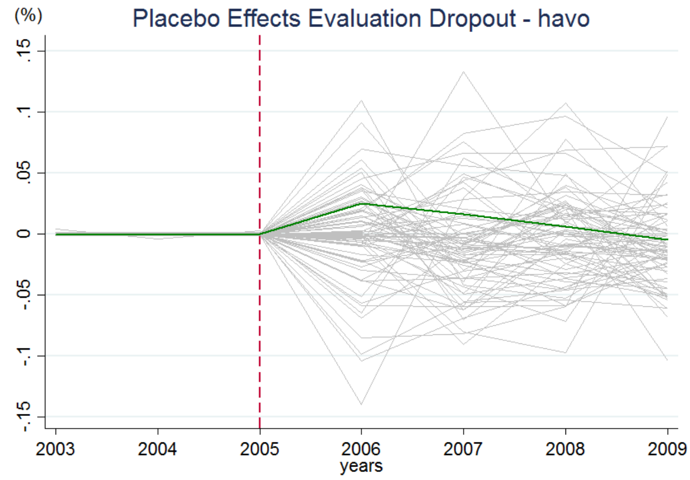
Appendix Figure 10 –The Placebo Effects of Entry-Tests per Highest Previously Attained Education
Dropout rate

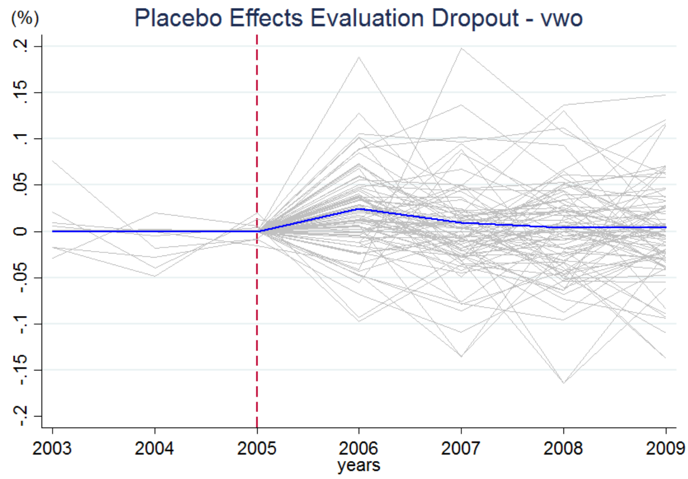
Exclusion 10% worst MSPE

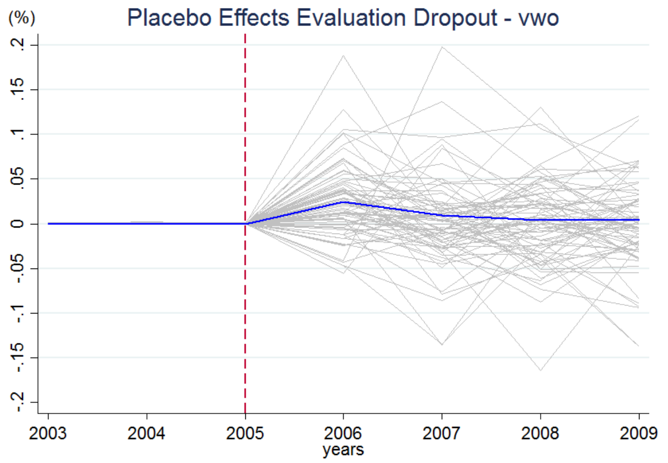








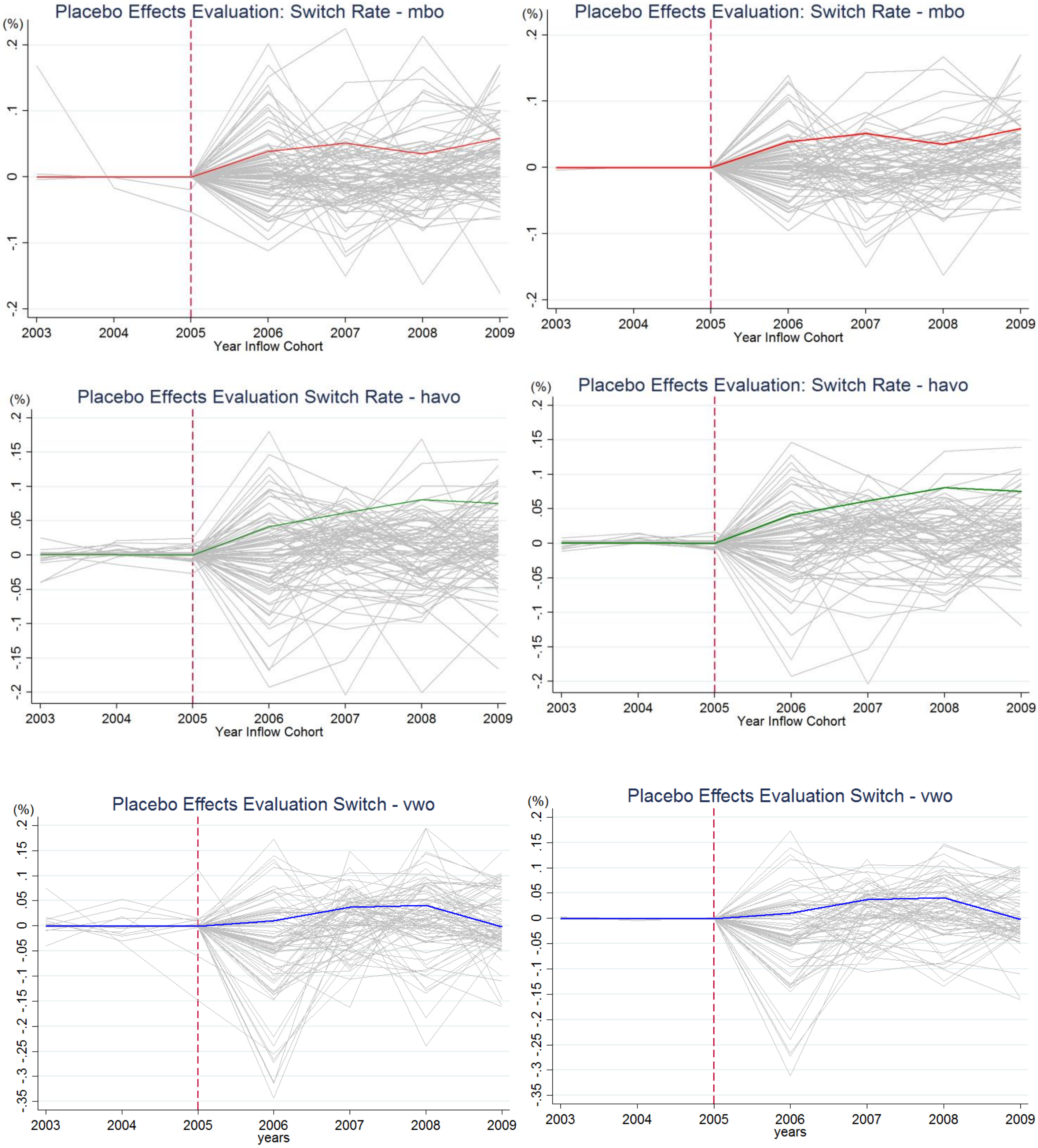




Notes: Each of the (lines within) the subfigures are based on the most complete specifications of the synthetic control procedure inherent to the particular outcome variable. The mean squared prediction error is minimized over the period 2003-2005. The year 2002 is used in the matching procedure as covariate intrinsic to each study. Labels on the vertical axis represent percentage points within a 0-1 range (e.g 0.1 reflects 10 %). The placebo runs on each of the outcome variables entail the complete donor pool of studies, i.e. 110 with the pabo included.

Appendix Figure 11–The Placebo Effects of Entry-Tests per Highest Previously Attained Education
Switch rate

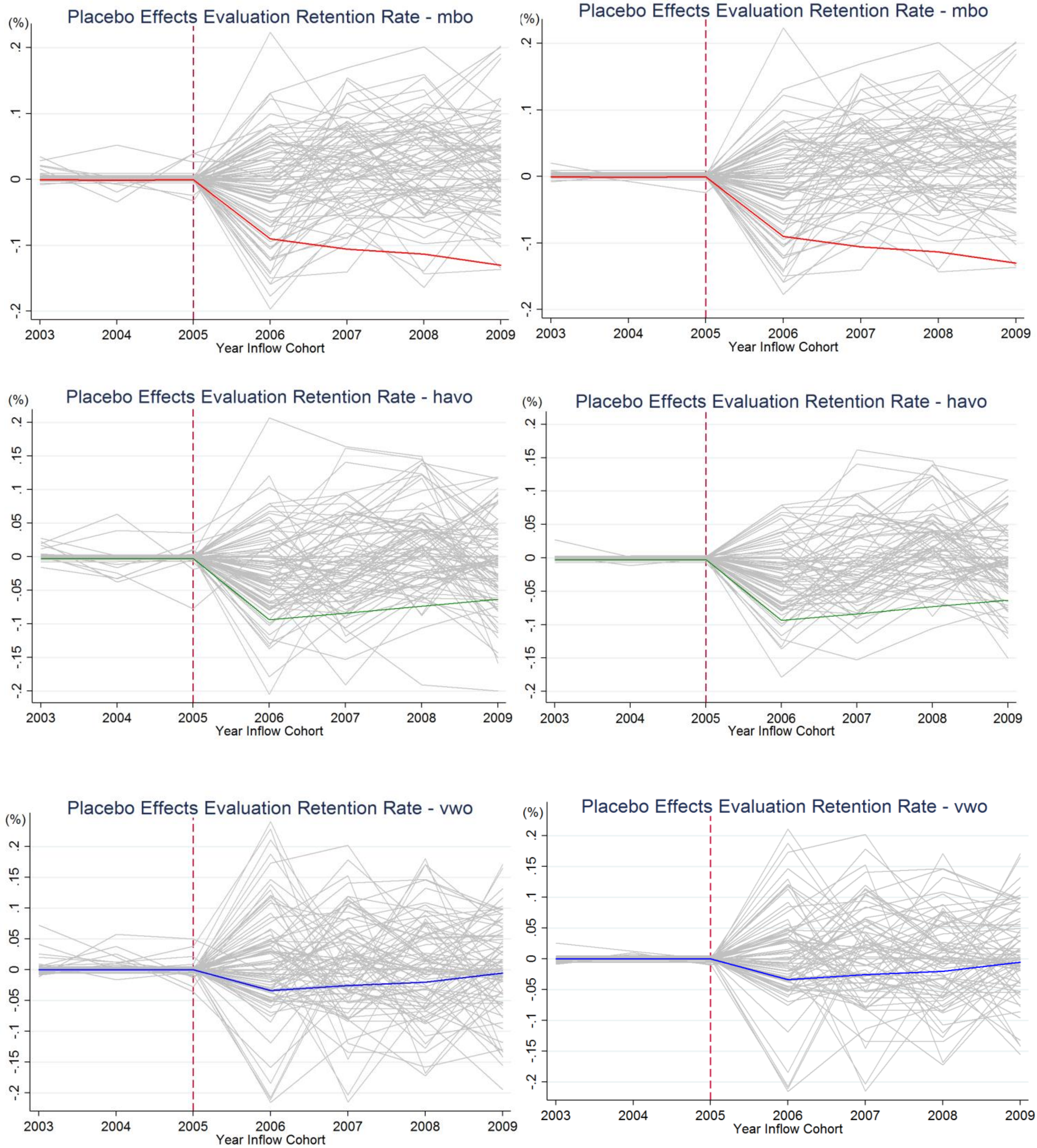
Exclusion 10% worst MSPE



Notes: Each of the (lines within) the subfigures are based on the most complete specifications of the synthetic control procedure inherent to the particular outcome variable. The mean squared prediction error is minimized over the period 2003-2005. The year 2002 is used in the matching procedure as covariate intrinsic to each study. Labels on the vertical axis represent percentage points within a 0-1 range (e.g 0.1 reflects 10 %). The placebo runs on each of the outcome variables entail the complete donor pool of studies, i.e. 110 with the pabo included.

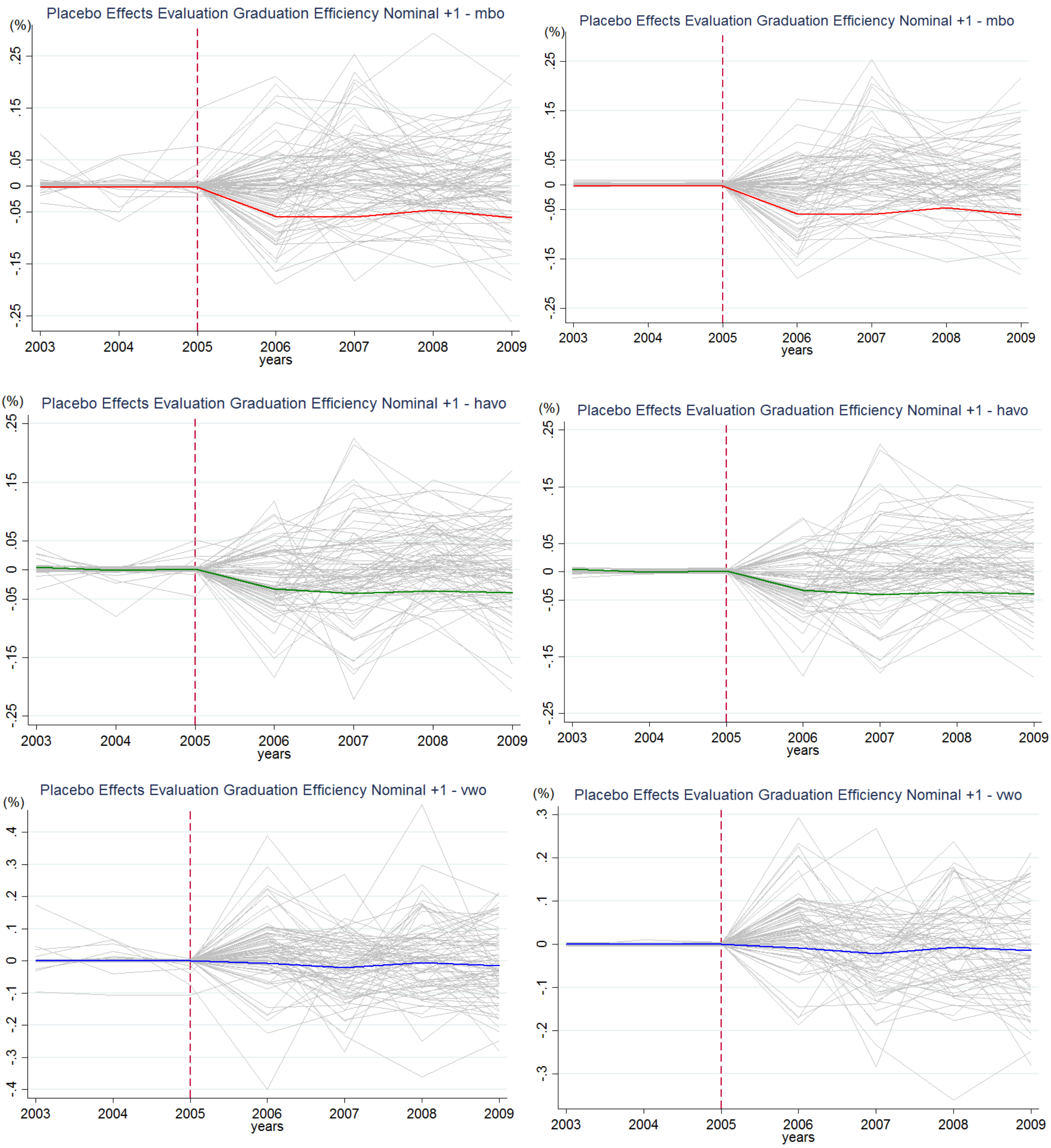
Appendix Figure 12 –The Placebo Effects of Entry-Tests per Highest Previously Attained Education Retention rate

Exclusion 10% worst MSPE



Notes: Each of the (lines within) the subfigures are based on the most complete specifications of the synthetic control procedure inherent to the particular outcome variable. The mean squared prediction error is minimized over the period 2003-2005. The year 2002 is used in the matching procedure as covariate intrinsic to each study. Labels on the vertical axis represent percentage points within a 0-1 range (e.g 0.1 reflects 10 %). The placebo runs on each of the outcome variables entail the complete donor pool of studies, i.e. 110 with the pabo included.

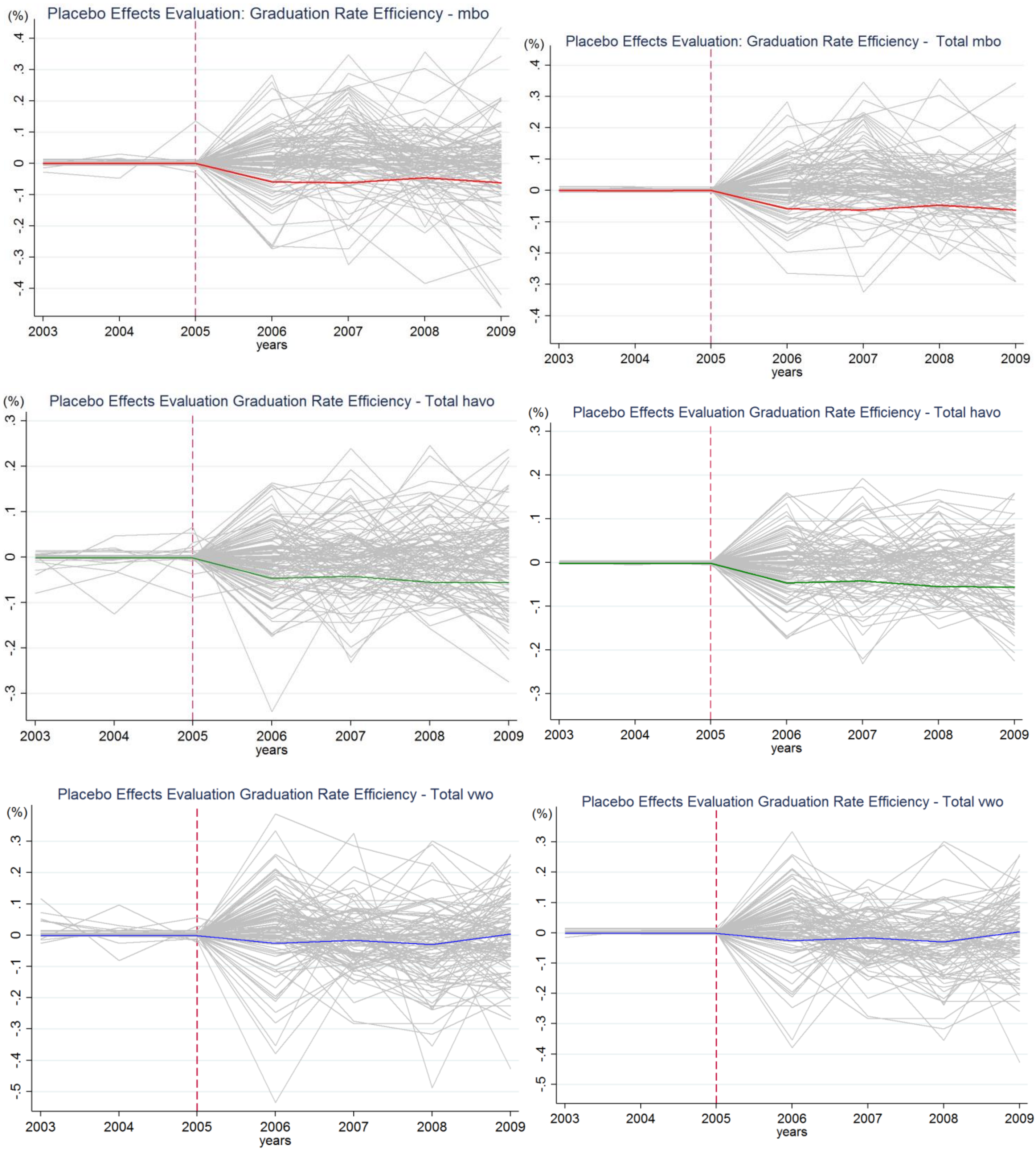
Appendix Figure 13 –The Placebo Effects of Entry-Tests per Highest Previously Attained Education
Graduation rate Efficiency – Nominal +1
Exclusion 10% worst MSPE



Notes: Each of the (lines within) the subfigures are based on the most complete specifications of the synthetic control procedure inherent to the particular outcome variable. The mean squared prediction error is minimized over the period 2003-2005. The year 2002 is used in the matching procedure as covariate intrinsic to each study. Labels on the vertical axis represent percentage points within a 0-1 range (e.g 0.1 reflects 10 %). The placebo runs on each of the outcome variables entail the complete donor pool of studies, i.e. 110 with the pabo included.

Appendix Figure 14 –The Placebo Effects of Entry-Tests per Highest Previously Attained Education Graduation rate Efficiency - Total

Exclusion 10% worst MSPE



Notes: Each of the (lines within) the subfigures are based on the most complete specifications of the synthetic control procedure inherent to the particular outcome variable. The mean squared prediction error is minimized over the period 2003-2005. The year 2002 is used in the matching procedure as covariate intrinsic to each study. Labels on the vertical axis represent percentage points within a 0-1 range (e.g 0.1 reflects 10 %). The placebo runs on each of the outcome variables entail the complete donor pool of studies, i.e. 110 with the pabo included.

Appendix C: Tables

Appendix Table 1: The Decreasing Test Scores of Dutch Pupils

Year	*Mathematics score (rank)*	*Reading score (rank)*	*Natural sciences score (rank)*
TIMSS 2003/ PIRLS 2001	540 (5th of the 24)	554 (2nd of the 28)	525 (11th of the 24)
TIMSS 2007/ PIRLS 2006	535 (8th of the 24)	547 (9th of the 28)	523 (13th of the 24)

Notes: PIRLS 2001 and 2006 consider reading skills, whereas TIMMS 2003 and 2007 consider mathematics and natural sciences
Source: Van der Steeg, Vermeer and Lanser (2011)

Appendix Table 2 - Did the Cognitive Abilities of pabo students Improve over time?

	Final Average			
	havo		vwo	
	Entrance	*Retention*	*Entrance*	*Retention*
Average Treatment Effect	-0.00312	0.00701	-0.0204	-0.106
	(-0.31)	(0.66)	(-1.04)	(-0.52)
Observations	213787	136562	51353	40754
R^2	0.301	0.309	0.351	0.350
Treatment (2006)	-0.00974	-0.00832	-0.000986	-0.0172
	(-0.76)	(-0.64)	(-0.40)	(-0.68)
Treatment (2007)	0.0106	0.0235	-0.00121	0.0155
	(0.73)	(1.50)	(-0.52)	(0.56)
Treatment (2008)	0.00364	0.0413	-0.00445	-0.0214
	(0.13)	(1.21)	(-0.99)	(-0.59)
Treatment (2009)	-0.0136	0.0138	-0.00520	-0.0397
	(-0.51)	(0.51)	(-1.15)	(-1.01)
Observations	213787	136562	51353	40754
R^2	0.301	0.309	0.351	0.350

Notes: Dependent variables in italics. All coefficients reflect the average marginal change of the variable of interest with respect to the treatment for all/unique enrollment cohorts, based on individually estimated OLS models. The t-statistics of standard errors clustered at the institutional-study cohort level are presented between parenthesis. Significance at the 10, 5 and 1 percent levels are respectively denoted by * / ** / ***.

Appendix Table 3: The Effects of Entry Tests on Enrollment Cohorts: Enrollment and Retention Estimates per Highest Previously Attained Education

	ln (Enrollment mbo)				*ln (Enrollment havo)*				*ln (Enrollment vwo)*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Treatment (2006)	-0.0791	-0.122	-0.108	-0.102	-0.0565	-0.0757	-0.0483	-0.0606	-0.0809	-0.129	-0.0961	-0.121
	(-0.94)	(-1.35)	(-1.17)	(-1.10)	(-0.51)	(-0.40)	(-0.43)	(-0.53)	(-0.49)	(-0.97)	(-0.75)	(-0.92)
Treatment (2007)	-0.205**	-0.255**	-0.249**	-0.231***	-0.166	-0.172	-0.175	-0.168	-0.291	-0.281	-0.317*	-0.322*
	(-2.86)	(-2.91)	(-3.27)	(-3.40)	(-1.64)	(-1.41)	(-1.48)	(-1.40)	(-1.36)	(-1.76)	(-2.36)	(-2.25)
Treatment (2008)	-0.351**	-0.344***	-0.341***	-0.380***	-0.223*	-0.328**	-0.311*	-0.336**	-0.351	-0.453***	-0.437**	-0.388*
	(-2.94)	(-4.22)	(-3.57)	(-3.51)	(-2.31)	(-2.48)	(-2.29)	(-2.49)	(-1.69)	(-3.97)	(-2.65)	(-2.43)
Treatment (2009)	-0.393***	-0.357***	-0.387***	-0.428***	-0.284**	-0.363*	-0.352*	-0.394**	-0.327	-0.381**	-0.368*	-0.363*
	(-3.58)	(-3.67)	(-3.71)	(-3.61)	(-2.82)	(-2.66)	(-2.44)	(-2.79)	(-1.41)	(-2.84)	(-2.30)	(-2.28)
Observations	3927	3982	3982	3982	3944	1938	3944	3944	3828	3742	3828	3800
(Pseudo) R^2	0.677	0.800	0.803	0.804	0.708	0.837	0.781	0.786	0.729	0.798	0.804	0.805

	Retention mbo				*Retention havo*				*Retention vwo*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Treatment (2006)	-0.0828***	-0.0759***	-0.0777***	-0.0812***	-0.0926***	-0.0910***	-0.0957***	-0.0938***	-0.0437**	-0.0464***	-0.0446**	-0.0450**
	(-4.97)	(-4.71)	(-4.78)	(-4.62)	(-3.31)	(-3.17)	(-3.41)	(-3.37)	(-2.16)	(-2.59)	(-2.49)	(-2.39)
Treatment (2007)	-0.0831**	-0.0964**	-0.107***	-0.102***	-0.0867**	-0.0888*	-0.0838**	-0.0812**	-0.0293*	-0.0240**	-0.0224**	-0.0216**
	(-2.03)	(-2.26)	(-2.93)	(-2.81)	(-2.38)	(-2.83)	(-2.84)	(-2.70)	(-1.92)	(-2.17)	(-2.05)	(-2.38)
Treatment (2008)	-0.0949***	-0.0942***	-0.0938**	-0.0989***	-0.0661***	-0.0749***	-0.0747***	-0.0755***	-0.0284**	-0.0311**	-0.0163	-0.0339
	(-3.54)	(-3.70)	(-2.57)	(-2.97)	(-2.93)	(-2.92)	(-3.26)	(-3.37)	(-1.98)	(-2.26)	(-0.46)	(-1.07)
Treatment (2009)	-0.125***	-0.123***	-0.116***	-0.109***	-0.0535**	-0.0572**	-0.0620***	-0.0597***	-0.0183	-0.0282	-0.00886	-0.0152
	(-3.14)	(-3.07)	(-3.41)	(-3.56)	(-2.53)	(-2.56)	(-3.39)	(3.14)	(-0.78)	(-1.26)	(-0.31)	(-0.54)
Observations	248074	248074	248074	248074	249664	249664	249664	249664	64728	64728	64728	64728
Pseudo R^2	0.023	0.040	0.042	0.044	0.081	0.123	0.094	0.123	0.014	0.037	0.037	0.037
Year and Study Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Individual Controls	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Study Controls	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
Institution and City Controls	N	N	N	Y	N	N	N	Y	N	N	N	Y

Notes: Dependent variables in italics. All coefficients reflect the average marginal change of the variable of interest with respect to the treatment for unique enrollment cohorts, based on individually estimated probit models. Four different specifications are estimated for each of the outcome variables, the build-up being presented at the lower region of the table. The t-statistics of standard errors clustered at the institutional-study cohort level are presented between parenthesis. The only exception are the coefficients of the change in enrollment outcome variable, being estimated using OLS regressions at the institutional-study level. Here standard errors are clustered at the institutional level. Significance at the 10, 5 and 1 percent levels are respectively denoted by * / ** / ***.

Appendix Table 4: The Effects of Entry Tests on Enrollment Cohorts: Graduation Rate Efficiency Estimates per Highest Previously Attained Education

	Graduation Rate Efficiency - Nominal +1 mbo				*Graduation Rate Efficiency - Nominal +1 havo*				*Graduation Rate Efficiency - Nominal +1 vwo*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Treatment (2006)	-0.0587***	-0.0639***	-0.0606***	-0.0619***	-0.0319**	-0.0422**	-0.0374**	-0.0371**	-0.00476	-0.0134	-0.0143	-0.0160
	(-3.03)	(-3.12)	(-3.62)	(-3.97)	(-2.24)	(-1.96)	(-2.42)	(-2.07)	(-0.11)	(-0.35)	(-0.38)	(-0.49)
Treatment (2007)	-0.0439*	-0.0546**	-0.0554**	-0.0626***	-0.0374	-0.0354	-0.0351*	-0.0289	-0.0302	-0.0340	-0.0271	-0.0273
	(-1.73)	(-2.16)	(-2.29)	(-3.16)	(-1.44)	(-1.24)	(-1.78)	(-1.23)	(-1.17)	(-1.31)	(-0.98)	(-1.28)
Treatment (2008)	-0.0196	-0.0295	-0.0274	-0.0307	-0.0397**	-0.0381***	-0.0421**	-0.0414**	0.00103	-0.000656	0.00557	0.00375
	(-0.86)	(-1.42)	(-1.06)	(-1.47)	(-2.34)	(-3.01)	(-1.84)	(-2.16)	(0.04)	(-0.02)	(0.20)	(0.13)
Treatment (2009)	-0.0314	-0.0394*	-0.0397***	-0.0381**	-0.0602***	-0.0435*	-0.0410***	-0.0384***	-0.00749	-0.0246	0.0192	0.00279
	(-1.26)	(-1.82)	(-2.73)	(-2.25)	(-3.41)	(-1.87)	(-2.59)	(-2.87)	(-0.30)	(-1.15)	(0.37)	(0.06)
Observations	248074	248074	248074	248074	248784	248784	248784	248784	64728	64728	64728	64728
(Pseudo) R^2	0.026	0.055	0.056	0.056	0.053	0.117	0.121	0.124	0.025	0.073	0.074	0.074
	Graduation Rate Efficiency – Total mbo				*Graduation Rate Efficiency - Total havo*				*Graduation Rate Efficiency - Total vwo*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Treatment (2006)	-0.0558***	-0.0536***	-0.0659***	-0.0542***	-0.0355***	-0.0433***	-0.0437**	-0.0457**	-0.0215	-0.0236	-0.0277*	-0.0275
	(-3.10)	(-2.97)	(-3.80)	(-4.15)	(-2.79)	(-2.72)	(-2.13)	(-2.17)	(-0.98)	(-1.59)	(-1.69)	(-1.54)
Treatment (2007)	-0.0496**	-0.0628***	-0.0575**	-0.0592***	-0.0346	-0.0376	-0.0374	-0.0319	-0.0143	-0.0216	-0.0249	-0.0221
	(-2.53)	(-2.88)	(-2.25)	(-3.45)	(-1.24)	(-1.12)	(-1.31)	(-1.12)	(-0.53)	(-1.17)	(-1.07)	(-1.23)
Treatment (2008)	-0.0229	-0.0455*	-0.0386*	-0.0378*	-0.0580**	-0.0579***	-0.0551**	-0.0534**	-0.0228	-0.0239	-0.0279	-0.0223
	(-1.15)	(-1.87)	(-2.00)	(-1.72)	(-2.10)	(-2.99)	(-1.96)	(-2.30)	(-1.01)	(-1.62)	(-0.75)	(-1.13)
Treatment (2009)	-0.0591**	-0.0588**	-0.0550**	-0.0621***	-0.0567*	-0.0565**	-0.0548**	-0.0519**	0.0111	-0.0109	-0.00539	0.00478
	(-2.46)	(-2.55)	(-2.44)	(-2.90)	(-1.67)	(-2.37)	(-2.59)	(-2.38)	(0.52)	(-0.60)	(-0.22)	(0.15)
Observations	248074	248074	248074	248074	249598	249598	249598	249598	64728	64728	64728	64728
Pseudo R^2	0.021	0.044	0.049	0.049	0.060	0.112	0.113	0.115	0.020	0.060	0.060	0.061
Year and Study Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Individual Controls	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Study Controls	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
Institution and City Controls	N	N	N	Y	N	N	N	Y	N	N	N	Y

Notes: Dependent variables in italics. All coefficients reflect the average marginal change of the variable of interest with respect to the treatment for unique enrollment cohorts, based on individually estimated probit models. Four different specifications are estimated for each of the outcome variables, the build-up being presented at the lower region of the table. The t-statistics of standard errors clustered at the institutional-study cohort level are presented between parenthesis. The only exception are the coefficients of the change in enrollment outcome variable, being estimated using OLS regressions at the institutional-study level. Here standard errors are clustered at the institutional level. Significance at the 10, 5 and 1 percent levels are respectively denoted by * / ** / ***.