The Effects of a Dutch High School Curriculum Reform on Performance in and After Higher Education

Ernest Berkhout · Peter Berkhout · Dinand Webbink

Published online: 11 January 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract This paper investigates the effect of a major reform in the final years of the two highest levels of Dutch secondary education. The reform focused on increasing active and independent learning and aimed to improve the match between secondary and higher education. We use data from six graduation cohorts from Dutch higher education and exploit the fact that two recent graduation cohorts contain both students that were treated with the new curriculum and students that were treated with the old curriculum. The effects of the curriculum change have been estimated by using a difference-in-differences approach for matched samples of treated and untreated graduates. We find that the reform had no effect on student performance and labour market outcomes.

Keywords Curriculum · Higher education · Labour market

JEL Classification I2

1 Introduction

The quality of teachers is one of the main determinants of the quality of education (Rivkin et al. 2005). However, little is known about the factors that determine the
quality of teaching and the most effective role of teachers in the learning process of students. Do students learn most when teachers play their traditional role as instructors in front of the classroom, or do students benefit more from active and independent learning where teachers act as coaches? A curriculum change in the last years of Dutch secondary education, introduced in the school year 1999–2000, provides the opportunity to investigate this question. This curriculum change, which was called ‘The Second Phase’ (TF), mainly focused on increasing active and independent learning of students. In addition, the coherence of the choice of subjects was strengthened, in order to improve the match between secondary and higher education, and to stimulate performance in higher education. This paper investigates the effect of this curriculum reform on student performance in higher professional education and on the position on the labour market in the first years after graduation. For the analysis we use unique data from six graduation cohorts from Dutch higher education. We investigate the effect of ‘the Second Phase reform’ on performance in and after higher education among graduates from higher professional education.

In the Netherlands educational reforms are generally introduced on a nationwide scale. This makes it difficult to identify the effects of the reforms because a simple before-after comparison might be confounded with differences between cohorts or time trends. In this paper we use a difference-in-differences approach that exploits variation in study durations. In Dutch higher education large proportions of graduates study longer than the nominal duration. Hence, a typical cohort of graduates contains students that study ‘on time’ and students that need more time than the nominal duration. Two recent graduation cohorts contain students that were treated with the new curriculum and students that were treated with the old curriculum. Our data allow us to match the treated and untreated students from these two graduation cohorts with students from four previous cohorts that all followed the old curriculum. We identify the effects of the curriculum change by estimating difference-in-differences models for these matched groups.

This approach basically compares the outcomes of graduates who were very similar on a large set of covariates when they enrolled in higher education and who decided to enrol in the same type of study in higher professional education. A concern with this approach is that the sample consists of graduates and we do not observe dropout from higher education. If the new curriculum had an effect on drop out from higher education our estimates might suffer from selection bias. We check the robustness of our findings by re-estimating the main models for samples of graduates for whom we expect that drop out is not very likely.

Our paper contributes to the economic literature on the relationship between curriculum and educational attainment and labour market. The empirical economic literature on this topic is quite small. Most studies focus on the effect of college curriculum on earnings (see for instance, Hamermesh and Donald 2008). These studies investigate the long term effects of the college major while taking account of the endogeneity of the college major. The effects of high school curriculum, which is more related to the topic of this paper, is only investigated in a few studies. Altonji (1995) is the first systematic attempt to measure the effects of specific high school courses on years of college and wage. Variation in US high schools is used to identify the effects on wages and educational attainment of specific courses of study. The main finding
The Effects of a Dutch High School Curriculum Reform

from this study is that the return to additional courses in high school is very small and there seems to be no effect on postsecondary educational attainment. A recent study estimates the effect that six types of high school math courses have on students’ earnings nearly a decade after graduation (Rose and Betts 2004). They find that more-advanced courses have larger effects than less-advanced ones. In addition, math courses can help close the earnings gap between students from low-income and middle-income families, and specific course combinations can explain the earnings premium related to an additional year of school. Our study adds to this literature by investigating the effect of a reform that included various elements such as, introducing new courses, restricting the choice of courses and putting more emphasis on active and independent learning (see also Sect. 2). In addition, educational reforms are often nationwide introduced which raises problems for estimating the effects. Our identification strategy may provide a solution for these problems. A further contribution is that most studies focus on the US and our study provides evidence for a European country.

Our main finding is that the reform had no effect on student performance and labour market outcomes. For five dependent variables about student performance and labour market outcomes we find that all estimates of the effect of the reform are statistically insignificant. These findings are robust for various sensitivity checks.

The remainder of this paper is organized as follows. Section 2 describes the reform in Dutch secondary education. Sections 3 and 4 explains the data and empirical strategy. The main estimation results are presented in Sect. 5. Section 6 presents some sensitivity tests and Sect. 7 concludes.

2 The Reform in Dutch Secondary Education

Dutch secondary education is a tracked system. After primary education, on average at the age of twelve, pupils enrol in tracks with different levels of education. The two highest levels of secondary education, which are called HAVO (higher general secondary education) and VWO (pre university education), have a duration of 5 and 6 years respectively. The reform was targeted at the last 2 years of HAVO and the last 3 years of VWO. The name of the reform, Second Phase, refers to the final stage of these two types of secondary education tracks. In this paper we focus on graduates from HAVO because of the timing of the reform and the sample (see also Sect. 3). Graduation from HAVO gives access to higher professional education (in Dutch Hoger BeroepsOnderwijs, abbreviated as HBO). HBO-education prepares students for specific (categories of) professions. It is taught at about 60 special institutes evenly spread over the Netherlands. On average, 50,000 students graduate each year from HBO.

The aim of the reform was to improve the match between Dutch secondary and Dutch higher education. The reform included two main elements. The first element of the reform focused on increasing the autonomy of schools. The detailed yearly tables of lessons for specific subjects were replaced with a new system that allowed schools to choose the time needed for specific subjects, and rotate between multiple lesson
tables per year.\textsuperscript{1} An important part of this first element was the introduction of the concept of independent learning. The idea was that schools would promote active and independent learning of students and increase the students responsibility for her/his own learning process. This also meant a new role for the teacher with more emphasis on supporting independent skills development of students and less on direct class teaching.

The second element consisted of a major change in the curriculum of the last years of secondary education. Before the reform students in the second stage of HAVO could choose four subjects for their final exam, next to two compulsory subjects (Dutch & English). The reform restricted this choice into four so-called profiles (Nature & Science, Nature & Health, Economics & Society, Culture & Society) that prepared for specific types of higher education.\textsuperscript{2} In addition, several new subjects were introduced (general science, culture & arts, management & organisation, computer science) and the content of some existing subjects changed. All students had to follow a much broader program of compulsory subjects. To accommodate this the number of hours per subject was lowered in some cases, or subjects were divided in multiple parts, although a compulsory minimum number of 1,000 teaching hours per year was established by law. For each subject a detailed list of requirements for the national exam was made. The main idea was to strengthen the coherence in the choice of subjects which would improve the match with higher education programs. In addition, cross-curriculum competencies and skills became more important next to reproduction of knowledge.

These changes seem to have had an effect on the skills of the graduates from secondary education. de Vries and van der Velden (2005) find that post-reform students indicate more often than pre-reform students that they have mastered the following skills at HAVO secondary education (differences being more than 10% points): collecting & processing information, cooperating, working precisely, computer skills, communicative skills, taking initiative. The opposite is true regarding language & mathematical skills, post-reform students indicate more often that these skills have been learned only in tertiary education.\textsuperscript{3} The authors conclude that “a possible reason is that since the Second Phase Reform the curriculum lays more emphasis on teaching skills and less on classical teaching”. In 2007 some small revisions were made to the curriculum, regarding the subjects history and geography and mathematics. Also the separated partial subjects were merged together again (Tweede Fase Adviespunt 2007). These revisions did not effect our sample.

In the school year 1998/1999 approximately 25% of the schools started with reforming their second stage of secondary education. The other schools followed in the next school year. This introduction led to remarkable changes in the number of candidates

\textsuperscript{1} Tweede Fase Adviespunt (2005) states that in 2001 over 40% of the schools did not use yearly tables anymore.

\textsuperscript{2} Berkhout et al. (2002) show that 51% of the pre-reform HAVO-students had chosen subject combinations that were made impossible by this reform.

\textsuperscript{3} This finding matches the judgement of teachers in tertiary education. Tweede Fase Adviespunt (2005), page 89, shows that the teachers are very unsatisfied with the language & mathematical skills of the post-reform students.
The Effects of a Dutch High School Curriculum Reform

Fig. 1 Candidates for the final exam of HAVO 1998–2004. Source: Tweede Fase Adviespunt (2005)

for the final exam of HAVO. Figure 1 shows the population statistics of the candidates for the final exam for the years around the introduction of the reform.

In the school year 1999–2000 all schools participated in the reform and all students in grade 4 of HAVO were treated with the new curriculum. The HAVO-students could take their exam in 2001. In Fig. 1 we observe an strong decrease in the number of candidates of HAVO in 2001. Hence, in the first exam year that all students in HAVO had followed the new curriculum the number of candidates was much lower than in the previous years. The large drop in HAVO-candidates can be explained by three factors:

- Less enrolment of graduates from MAVO (general intermediate education);
- Less retention in grade 4 of HAVO (8,652 in 1998 and 5,089 in 1999);
- More graduation in 2000 (89%) than in 1999 (85%).

The fact that there was less retention in grade 4 and more graduation in 2000 suggest that schools and students tried to avoid that students had to change from the old curriculum, which they had followed in grade 4, to the new curriculum. Hence, the last cohorts of the old curriculum took on average less time for graduating from HAVO. This could have a negative impact on the preparation of these students for higher education. We will address this potential confounder in the empirical part of the paper.

The graduation rates from HAVO are slightly higher after the introduction of the reform. In 2000 89% of the HAVO candidates graduated against 90% in 2001. In the years after the introduction the graduation rates slowly moved upward.

Figures on enrolment in higher education show that the proportion of graduates of HAVO that enrol in higher education steadily increased for the period 1998–2004. The introduction of the reform did not create a discontinuity in this trend (Second Phase Advice Centre, 2005).

Although the reform consisted of two main elements we will, in the remainder of this paper, refer to the reform as the change in curriculum for reasons of convenience.
3 Data

The data were collected in surveys held among graduates from tertiary education, by SEO Economic Research institute (SEO) jointly with Elsevier magazine. SEO cooperates with the magazine to produce annual cover stories on the labour market position of graduates from higher education. The distinct advantage for our purpose is the large number of educations that are distinguished, and, unique in a large and broad sample, the availability of exam grades. A new cohort of graduates has been interviewed every year since 1997, especially about their labour market outcomes in the first 20 months. At HBO-level students can choose between 250 different courses of study. Most of them, however, produce only small numbers of graduates, making statistical analysis unreliable. Over 90% of the HBO-student population is concentrated in the 60 largest degree subjects. The survey is restricted to these 60 degree subjects (studies). This means that the survey is representative of more than 90% of the yearly outflow of graduates at HBO-level. Each year on average 4,150 graduates participate in the survey. The special feature of the survey is the large number of studies within tertiary education and the focus on the initial phase of the labour market career. The surveys collected a broad range of information on the performance in secondary and higher education and about early labour market experiences. In addition, information was collected on the personal history and the socioeconomic background of the graduates.

3.1 The Main Estimation Sample

In this paper we can use data from six cohorts of graduates of higher professional education (2000/2001–2005/2006). We selected graduates who completed the full-time study program, and took the main track from HAVO to HBO. We cannot include students who took the track from VWO to HBO because this track takes one additional year in secondary education. As a consequence, the number of treated students who took the VWO-track that can be observed in our data is quite small (62 observations). The structure of our sample is shown in Table 1. The rows in Table 1 show the year in which the student enrolled in higher education for the first time, the columns show the year of graduation from higher education. The left panel shows all students by year of enrolment and graduation, the right panel shows only the number of students that were treated by the reform.

The reform started in the school year 1998/1999 and was implemented in approximately 25% of the schools. All other schools followed in 1999/2000. As a result, the first students who were treated with the new curriculum could enrol in higher education in 2000/2001. These are students that followed higher secondary education (HAVO) in the schools that first implemented the reform, and followed the new curriculum for 2 years. After 4 years these students can graduate from higher vocational education, which is the graduation cohort 2003/2004. The question on treatment by the reform was first asked to graduates from the cohort 2004/2005. In our data we observe that 61 (28) students who graduated after 5 (6) years were treated with the new curriculum. In the next cohort, which started in 2001/2002, we observe a strong increase of the proportion of students who were treated with the new curriculum. This
Table 1: Graduates by enrolment (E) and graduation (G) cohort

| Enrolment cohort | Graduation cohort | Treated | 2004/2005 | 2005/2006 | Total |
|------------------|-------------------|---------|-----------|-----------|-------|
| 1997/1998        |                    | 522     | 396       | 191       | 1,261 |
| 1998/1999        |                    | 576     | 481       | 192       | 1,364 |
| 1999/2000        |                    | 23      | 197       | 600       | 1,083 |
| 2000/2001        |                    | 0       | 40        | 628       | 634   |
| 2001/2002        |                    | 0       | 0         | 132       | 186   |
| Total            |                    | 545     | 1,007     | 1,336     | 5,304 |
strong increase is caused by the nationwide implementation of the curriculum in the school year 1999/2000. This enrolment cohort (2001/2002) also includes substantial proportions of students who were not treated with the new curriculum, that is 132 (71) students who graduated after 4 (5) years followed the old curriculum in secondary education. These are students who lost one or more years between graduating from secondary education and graduating from higher education. There are various reasons for losing a year, for instance a year of working or travelling after secondary education or switching between studies in higher education. Unfortunately, we do not have complete information on the reasons for not taking the fastest route through higher education. To identify the effect of the curriculum reform we exploit the fact that a cohort contains substantial proportions of students who are ‘on time’ and students with some study delay in a difference-in-differences approach for matched groups (see next section).

The main groups of graduates for the analysis in this paper are graduates who enrolled in 2001/2002 and who graduated in 2004/2005 or 2005/2006. In the remainder of this paper we refer to these cells in Table 1 as E1G4 and E1G5. The enrolment cohort 2000/2001 also contains treated students. However, the question on treatment by the reform was not yet in the survey of the graduation cohort of 2003/2004. As a result, from the enrolment cohort 2000/2001 we only observe the treated students who have some delay. Therefore, for the treatment group we prefer to focus on the enrolment cohort 2001/2002 and use the other enrolment cohorts for our control groups.

Furthermore, we selected graduates who started higher education between the age of 16–21 and within the school years 1997/1998–2001/2002. This selection gives a homogenous group of students that did not previously enroll in higher education. The age restriction excludes older students that might have finished secondary education many years before. Graduates from medical studies were excluded from our sample because of their different curriculum which prevents the correct calculation of the ‘higher education entry date’. Moreover, we only selected graduates for whom the recorded study was actually their highest completed or last study. HBO graduates who also completed a university program (Master after Bachelor for example) or graduates who are still in education at the time of survey were excluded from our sample. This increases comparability with the most recent cohort because these groups of students can not yet be observed in the most recent cohort. In addition, 420 graduates were excluded from the main estimation sample because of checks on measurement error regarding the starting date of higher education.

3.2 Treatment by the Reform

Our main explanatory variable is a dummy variable that measures whether the graduate was treated by the new curriculum or by the old curriculum. In the survey graduates were asked to report whether they had participated in the new or old curriculum. An important difference between the two curricula is the number and type of exams. These variables were used to check the self reports of the main reform question. Table 1 shows the number of graduates that were treated with the new curriculum.
3.3 Dependent Variables

We investigate the effect of the reform on five dependent variables. Three variables indicate the performance in higher education. First, we investigate the effect on the duration of the last successfully completed study (study duration). The study duration is based on a comparison between the self reported dates of enrolment and graduation of the last study. The second dependent variable is the total duration in higher education. This duration includes time spent in all studies in higher education, completed or uncompleted. The third measures of performance in higher education is the average of the grades in higher education. The other two dependent variables indicate the labour market experiences. The fourth dependent variables is the job search duration which is the time between graduation and the start of the first job at the appropriate level. The fifth measures in the natural logarithm of the real net hourly wage, measured 20 months after graduation.

3.4 Covariates

The controls used in the estimation can be distinguished in three groups. First, we control for various personal characteristics: gender, age at the start of higher education (age 16–21), ethnicity (Dutch or foreign born), education of parents (9 levels), region (29 categories), labour market tightness in month of graduation (start of job search). The second group of controls includes indicators of ability: performance in secondary education (average grade, total number of subjects, number of language subjects, number of science subjects, secondary education track (higher general education ‘HA VO’ or MAVO-HA VO’). The third group includes controls for the type of study in higher education (69 categories) and a dummy for a shorter version of the curriculum in higher education.

Table 2 shows summary statistics for the dependent and explanatory variables for the treated and not treated groups of graduates in the main estimation sample. This sample consists of students who graduated after 4 or 5 years in higher professional education (the bold cells in the left panel of Table 1).

The standard nominal study duration for graduates of higher professional education in the Netherlands was 4 years. Because some students can get waivers for parts of the curriculum, the average is just below 4 years. Total time spend in higher education is more than 4 years because some students switch from one study to another. We observe only small differences between the treated and non treated students in the dependent variables. A comparison of the covariates for the two groups suggests that the treated students might have a somewhat higher ability. We observe that the treated students are nearly a year younger and less often followed the indirect track through secondary education (MAVO-HA VO). Moreover, we observe a lower proportion of ethnic minorities in the group of treated students and they included more subjects in their secondary education curriculum than the non-treated students.

A plausible explanation for these differences is that we focus on the first cohorts of graduates that were treated with the new curriculum. For instance, we do not observe treated students that repeated a year in the final stage of secondary education.
Table 2  Summary statistics of the main estimation sample

| Graduation cohort       | Old curriculum | New curriculum |
|-------------------------|----------------|----------------|
|                         | N              | N              |
| 2000/2001               | 522            | 0              |
| 2001/2002               | 972            | 0              |
| 2002/2003               | 1.199          | 0              |
| 2003/2004               | 1.028          | 0              |
| 2004/2005               | 463            | 387            |
| 2005/2006               | 71             | 158            |
| Total diagonal selection| 4.255          | 545            |

| Dependent variables       | Mean  | SD    | N    | Mean  | SD    | N    |
|---------------------------|-------|-------|------|-------|-------|------|
| Study duration            | 3.91  | 0.66  | 4255 | 3.89  | 0.46  | 545  |
| Total duration            | 4.24  | 0.42  | 4255 | 4.14  | 0.36  | 545  |
| Grade higher education    | 7.09  | 0.50  | 4131 | 7.09  | 0.49  | 534  |
| Job search duration       | 6.08  | 6.71  | 3850 | 5.86  | 6.31  | 493  |
| Net hourly wage           | 10.00 | 6.49  | 3752 | 9.99  | 7.53  | 504  |

*Covariates*

| Male (%)                  | 34    | 0.47  | 4,255 | 34    | 0.47  | 545  |
| Age                      | 18.4  | 1.05  | 4,255 | 17.7  | 0.69  | 545  |
| Ethnic minority (%)      | 13    | 0.33  | 4,255 | 11    | 0.31  | 545  |
| Labour market indicator  | 6.48  | 0.95  | 4,255 | 6.19  | 0.98  | 545  |
| Grade secondary education| 6.79  | 0.52  | 4,119 | 6.83  | 0.49  | 531  |
| Indirect secondary track (MAVO) (%) | 18 | 0.38 | 4,255 | 11 | 0.31 | 545 |
| # Subjects               | 5.9   | 0.79  | 4,119 | 6.4   | 1.31  | 531  |
| # Language subjects      | 2.6   | 0.68  | 4,119 | 2.6   | 0.61  | 531  |
| # Science subjects       | 2.0   | 1.18  | 4,119 | 2.0   | 1.37  | 531  |
| Short curriculum dummy (%) | 4 | 0.19 | 4,255 | 3 | 0.16 | 545 |
| Education parents (1–9)  | 4.8   | 1.75  | 4,071 | 4.7   | 1.71  | 527  |

4 Empirical Strategy

4.1 The Naïve OLS Approach

We investigate the effect of the curriculum reform (C) on student performance (P) in or after higher professional education. The reform was introduced in the school year 1998/1999 in approximately 25% of the schools. All other schools started in the school year 1999/2000. As shown in Table 1 we observe students who were treated with the new curriculum and students who followed the old curriculum within enrolment and graduation cohorts (within each cell in Table 1). A naïve approach would be to compare the performance of treated students with the performance of the non treated...
The Effects of a Dutch High School Curriculum Reform

students within enrolment cohorts. This can be done by estimating OLS-regressions like:

\[ P_{it} = \alpha_0 + \alpha_1 C + \alpha_2 E + \alpha_3 X_{it} + \epsilon_{it} \] (1)

with P is performance of student i in year t, C is a dummy which has value one if the graduate was treated with the new curriculum and value 0 for the old curriculum, E is a vector of dummies for enrolment cohorts, X is vector of covariates, \( \epsilon \) is error term and \( \alpha_0, \alpha_3 \) are parameters to be estimated. A major concern with this approach is that the treated students might not be a random sample. For instance, schools that already started with the reform in the school year 1998/1999 might be better or more ambitious schools with better teachers and higher quality education. The estimated effect of the reform \( \alpha_1 \) would then not only pick up the effect of the reform but also the effect of higher quality education of the schools that started first. In addition, the fact that we observe substantial proportions of treated and untreated graduation within enrolment-graduation cells (Table 1) is caused by the fact that untreated graduates lose a year on their way through higher education and treated graduates do not. Hence, the untreated graduates within the same cell do not seem an adequate control group.

4.2 A Matching Approach

A logical extension of the naïve OLS approach would be to include a comparison with previous untreated cohorts of students and to estimate difference-in-differences models. For this approach we need an identifier, such as the secondary school, to link treated graduates with untreated graduates from the previous cohort. Unfortunately, we do not have such an identifier in our dataset and cannot estimate standard difference-in-differences models. However, instead of using one identifier we can use a large set of covariates to construct a control group of untreated graduates from the previous cohorts that match the treatment group. In other words, we can construct a before group with propensity score matching. The control group consists of graduates with a similar propensity score (the conditional propensity of receiving treatment given the set of covariates x) as the graduates that received treatment. A comparison of the outcomes of the treatment and control group, constructed from previous cohorts, gives us a before-after estimator.

The general formula for a matching estimator of the average effect of the treatment on the treated is

\[ d = \frac{1}{N} \sum_{i \in (D = 1)} \left( y_{1,i} - \sum_j w(i, j) y_{0,j} \right) \] (2)

with \((D = 1)\) is the set of treated individuals, \(N\) is the number of cases in the treatment group, \(y\) is the outcome, \(j\) is an element of the set of matched comparisons, \(w(i, j)\) is the weight given to the jth case in making a comparison with the treated case (Cameron and
Different estimators can be generated by varying the choice of $w(i,j)$. In our estimates we will use kernel matching as our main approach but we will also present estimates using other matching approaches. The kernel matching estimator is generated by using the following weight factor:

$$w(i, j) = \frac{K \left( \frac{p_i - p_j}{n} \right)}{\sum_{j \in \{D=0\}} K \left( \frac{p_i - p_j}{n} \right)}$$

with $p$ is the propensity score and $K$ is a kernel.

We construct our treatment and comparison group from the enrolment-graduation cells in Table 1. The after group, which is our treatment group, consist of graduates from the enrolment cohort 2001/2002 who graduated after 4 or 5 years and were treated with the new curriculum. Table 1 shows that 519 and 229 persons from the cohort that started in 2001/2002 graduated after 4 and 5 years. From these groups 387 and 158 persons followed the new curriculum in secondary education (right panel Table 1). This is our main experimental group. We construct the before group by propensity score matching using all covariates used in the previous analysis. This control group is drawn from previous enrolment cohorts who also graduated after 4 or 5 years (and did not indicate that they were treated by the reform). In Table 1 this control group is drawn from the cells of the downward diagonal slope. These graduates followed the old curriculum.

It should be noted that we match treated students with untreated students from previous cohorts. Hence, we expect that our treated and untreated students on average have lost the same time between graduating from secondary education and enrolling in higher education because we also use age (at the start of tertiary education) as a matching variable. The untreated students are found in cohorts before the introduction of the new policies. It is important to note that the students in our control group did not select themselves in the situation of not being treated. These students did not get the treatment of the new curriculum because of the timing of the introduction of the new policies. As such, our approach is similar to the approach used recently for the evaluation of the Educational Maintenance Allowance (EMA) in the UK (Dearden et al. 2009). The EMA had been introduced in some regions in the UK. The authors constructed a control group from other regions by using propensity score matching. In our approach we do not use regional variation in the treatment but variation in treatment caused by the timing of the introduction of the policy. The use of variation in the treatment induced by the timing of the new policies (a before and after comparison) has the disadvantage that other factors might also have changed over time. We address this disadvantage by controlling for a second difference (the before-after estimator for the non treated graduates).

4.3 A Difference-in-Differences Approach for Matched Groups

A major concern with the before-after matching is that this estimator might be biased by other factors that have changed in time. To improve on this approach we use a
second difference to control for these unobserved time factors. The second difference is the before-after estimator for the non treated graduates. For this, we exploit the fact that we also have substantial proportions of non treated graduates in the same enrolment cohort. By matching these non treated students with students from previous cohorts we get an estimate of the effect of the time change on the outcome variables.

The sample of untreated persons from enrolment cohort 2001/2002 consists of $519 - 387 = 132$ persons who graduated after 4 years and $229 - 158 = 71$ students who graduated after 5 years. These $132 + 71$ students are matched with students from previous cohort who also graduated after 4 and 5 years. This gives an estimate of the effect of time changes. Hence, we assume that the differences in outcomes for these delayed students captures the effect of changes in time for our experimental group. In other words, we assume that the before-after difference for the untreated graduates would have been the before-after difference for the treated graduates in the absence of the reform.

4.4 Selection Bias From Enrolment and Drop Out

A concern with this difference-in-differences approach with matched groups comes from the fact that our estimation sample consists of graduates. We compare the outcomes of groups of graduates who looked very similar at the time of their enrolment in higher education. The only observable difference is that the groups of graduates come from different enrolment cohorts. However, by focusing on graduates we do not observe the effect that the reform might have had on drop out in higher education. If the reform lowered the probability of drop out this would bias our estimates for the sample of graduates. For instance, it is possible that more low ability students were able to graduate due to the reform. In that case, a comparison of the treated and untreated graduates on their outcomes would be biased because the group of treated graduates would include more low ability students. As a consequence, we might underestimate the effect of the reform. To avoid this bias by dropping out we re-estimate the main models on a sample of graduates for whom drop out seems not very likely. To this end we selected a group students with scores on the final exam of secondary education above a certain threshold.

5 Main Estimation Results

This section presents the main estimation results for the effect of the curriculum change on performance in and after higher education.

5.1 Regression Estimates of the Reform on Performance in and After Higher Education

As a first step we estimate linear regression models for the five dependent variables with different sets of controls using our main estimation sample (the bold cells in the left panel of Table 1).
Table 3  OLS-estimates of the effect of the reform on study performance and labor market position

|                  | (1)     | (2)     | (3)     | (4)     |
|------------------|---------|---------|---------|---------|
| **Study duration** |         |         |         |         |
| Reform           | −0.018  | 0.023   | 0.033   | 0.018   |
| SE               | 0.040   | 0.043   | 0.043   | 0.034   |
| t-value          | −0.44   | 0.54    | 0.75    | 0.53    |
| N                | 4,861   | 4,657   | 4,543   | 4,543   |
| **Time in HE**   |         |         |         |         |
| Reform           | 0.095   | 0.061   | 0.062   | 0.078   |
| SE               | 0.029   | 0.030   | 0.031   | 0.030   |
| t-value          | 3.31    | 2.01    | 1.98    | 2.59    |
| N                | 4,861   | 4,657   | 4,543   | 4,543   |
| **Test scores**  |         |         |         |         |
| Reform           | −0.006  | −0.034  | −0.027  | −0.033  |
| SE               | 0.035   | 0.038   | 0.038   | 0.036   |
| t-value          | −0.18   | −0.90   | −0.70   | −0.91   |
| N                | 4,723   | 4,530   | 4,430   | 4,430   |
| **Search duration** |        |         |         |         |
| Reform           | 1.368   | 1.926   | 1.863   | 2.037   |
| SE               | 0.482   | 0.512   | 0.519   | 0.508   |
| t-value          | 2.84    | 3.76    | 3.59    | 4.01    |
| N                | 4,397   | 4,240   | 4,137   | 4,137   |
| **Current wages** |         |         |         |         |
| Reform           | 0.018   | 0.049   | 0.058   | 0.069   |
| SE               | 0.023   | 0.025   | 0.025   | 0.024   |
| t-value          | 0.76    | 1.95    | 2.28    | 2.86    |
| N                | 4,305   | 4,179   | 4,078   | 4,078   |
| **Controls**     |         |         |         |         |
| Person           | No      | Yes     | Yes     | Yes     |
| School           | No      | No      | Yes     | Yes     |
| Study            | No      | No      | No      | Yes     |
| Year             | Yes     | Yes     | Yes     | Yes     |

Table 3 shows the estimated effect of the curriculum change on the five dependent variables. All models include dummies for enrolment cohort. Column (1) shows the estimation results without further controls, column (2) includes controls for personal characteristics, column (3) additionally controls for school factors and column (4) also controls for characteristics of the type of study in higher education.

The top panel of Table 3 shows that the curriculum change did not have an effect on study duration. The estimates of the effect of the curriculum change on the average scores in higher education are all negative but statistically insignificant. In addition, we find that the reform slightly increased the total time spent in higher education.
(with one additional month), increased the search duration needed for the first job with two months and increased earnings with nearly 7%. However, these regression results might be biased by selectivity or cohort effects. In the next section we try to take these biases into account by using a matching approach.

5.2 Estimates using Propensity Score Matching

In this section we match graduates from the main cells, E1G4 and E1G5, in Table 1 who were treated with the curriculum reform with graduates from previous enrolment and graduation cohorts. This means, we seek matching candidates in the cells from the downward sloping DIAGONals. For instance, candidates might be found in the cells E0G3, E0G4 or E9G2 and E9G3. Next, we apply the same procedures for the untreated graduates in the main cells E1G4 and E1G5. This procedure will generate four groups of graduates which we use for the difference-in-differences estimates. For the propensity score matching we use all controls from the previous regression analysis. As our main matching procedure we use kernel matching.

The results of the matching approach are shown in Table 4. For each dependent variable Table 4 shows the mean of the treated and the matched group and the difference between these two groups. In addition, Table 4 shows the mean of the untreated group and the matched group and the difference between these two groups. Finally, the difference-in-differences estimator is shown. For instance, the average study duration of the graduates who were treated with the curriculum change is 3.89 years. The average study duration for the matched group from previous cohorts is 3.93 years, which gives a difference of −0.04 years. The average study duration for graduates from the same enrolment and graduation cells as the treated students but who did not follow the new curriculum is 3.91 years. The average study duration of their matched counterparts is 3.93 years, and the differences between the untreated and the matched group is −0.03 years. The difference-in-differences estimator is −0.01 years which is statistically insignificant. Nearly all observations have common support for the propensity score, only 11 observations fail to have common support. Standard errors have been calculated with bootstrapping.

The estimates from the matching approach suggest that the curriculum change did not have an effect on the performance in or after higher education.4 We only find that the treated graduates spent less time in higher education than their matched counterparts from previous cohorts. However, the untreated graduates also spent less time in higher education. Therefore, the difference-in-differences estimator is statistically not significant.

In the public debates about the effects of the curriculum reform questions were raised about differential effects for boys and girls and for students with lower socioeconomic backgrounds.5 To investigate these issues we estimated the models of

---

4 The number of observations used in the matching analysis is somewhat smaller than the number of observations in the linear regressions in Table 3. Due to perfect collinearity in the probit regression of the matching procedure we lose observations.

5 See for instance the debate in NRC Handelsblad (Dutch newspaper) of June 13, 2009 and June 20, 2009.
Table 4 Matching estimates of the effect of the reform (kernel matching)

| New curriculum | Enrolment cohort 2001/2002 | Matched cohorts 1997/2001 | Average treatment effect |
|----------------|---------------------------|--------------------------|-------------------------|
|                | Mean | N    | Mean | N    | Coef. | S.E | t-value |
| Study duration | Yes  | 3.89 | 502  | 3.93 | 3,648 | −0.04 | 0.03 | −1.31  |
|                | No   | 3.91 | 182  | 3.93 | 3,467 | −0.03 | 0.05 | −0.53  |
|                | DiD  | 4.19 | 182  | 4.24 | 3,467 | −0.05 | 0.04 | −1.32  |
| Time in HE    | Yes  | 4.15 | 502  | 4.25 | 3,648 | −0.10 | 0.02 | −4.56  |
|                | No   | 4.19 | 182  | 4.24 | 3,467 | −0.05 | 0.04 | −1.32  |
|                | DiD  | 4.38 | 182  | 4.43 | 3,467 | −0.05 | 0.04 | −1.15  |
| Test scores HE| Yes  | 7.09 | 495  | 7.09 | 3,551 | 0.00  | 0.03 | 0.15   |
|                | No   | 7.08 | 175  | 7.13 | 3,371 | −0.05 | 0.04 | −1.07  |
|                | DiD  | 4.27 | 175  | 4.32 | 3,371 | 0.05  | 0.05 | 0.98   |
| Search duration| Yes  | 5.99 | 453  | 5.55 | 3,337 | 0.44  | 0.37 | 1.20   |
|                | No   | 5.48 | 156  | 6.12 | 3,084 | −0.64 | 0.60 | −1.08  |
|                | DiD  | 3.98 | 156  | 4.08 | 3,084 | 1.08  | 0.70 | 1.55   |
| Current wages | Yes  | 2.23 | 464  | 2.22 | 3,266 | 0.01  | 0.02 | 0.61   |
|                | No   | 2.21 | 165  | 2.25 | 3,088 | −0.04 | 0.02 | −1.56  |
|                | DiD  | 3.94 | 165  | 3.95 | 3,088 | 0.05  | 0.03 | 1.60   |

Table 4 for separate samples of boys and girls and samples of graduates with a lower socioeconomic and a higher socioeconomic background. However, we find that all difference-in-differences estimates are statistically not significant.6

Hence, the results in this section suggest that the curriculum changes did not result in better performance during or after higher education.

A cautionary note about our findings may be in order. Our sample consists of graduates from higher vocational education. With our data we can not observe the effect of the reform on drop out in higher vocational education. Hence, our analysis is not informative about the potential effect of the reform on dropout.

### 6 Robustness

6.1 Alternative Matching Procedures

To test the sensitivity of these results we used various alternative matching procedures. Table 5 shows the estimation results. The main finding is that the results are robust for alternative matching procedures. Using six alternative procedures we find no statistical significant estimate on any of the five outcomes.

---

6 Estimates can be obtained from the authors on request.
## Table 5  Difference-in-difference estimates using alternative matching procedures

|                          | TF-effect | SE  | T-stat | N    |
|--------------------------|-----------|-----|--------|------|
| **Duration of study**    |           |     |        |      |
| Kernel                   | −0.01     | 0.06| 0.20   | 4,389|
| Radius                   |           |     |        |      |
| Caliper (0.005)          | 0.02      | 0.06| 0.28   | 4,389|
| Caliper (0.001)          | 0.05      | 0.07| 0.78   | 4,389|
| Caliper (0.0005)         | 0.07      | 0.07| 0.97   | 4,389|
| n(1)                     | −0.03     | 0.08| 0.39   | 4,389|
| n(2)                     | 0.01      | 0.07| 0.13   | 4,389|
| n(5)                     | 0.00      | 0.06| 0.07   | 4,389|
| **Time in HE**           |           |     |        |      |
| Kernel                   | −0.05     | 0.04| 1.15   | 4,389|
| Radius                   |           |     |        |      |
| Caliper (0.005)          | −0.02     | 0.04| 0.53   | 4,389|
| Caliper (0.001)          | −0.02     | 0.05| 0.38   | 4,389|
| Caliper (0.0005)         | −0.03     | 0.05| 0.51   | 4,389|
| n(1)                     | −0.06     | 0.06| 1.00   | 4,389|
| n(2)                     | −0.01     | 0.05| 0.19   | 4,389|
| n(5)                     | −0.04     | 0.05| 0.92   | 4,389|
| **Test scores HE**       |           |     |        |      |
| Kernel                   | 0.05      | 0.05| 0.98   | 4,277|
| Radius                   |           |     |        |      |
| Caliper (0.005)          | 0.03      | 0.05| 0.62   | 4,277|
| Caliper (0.001)          | 0.03      | 0.05| 0.60   | 4,277|
| Caliper (0.0005)         | 0.02      | 0.06| 0.35   | 4,277|
| n(1)                     | 0.11      | 0.07| 1.63   | 4,277|
| n(2)                     | 0.10      | 0.06| 1.61   | 4,277|
| n(5)                     | 0.08      | 0.05| 1.42   | 4,277|
| **Search duration**      |           |     |        |      |
| Kernel                   | 1.08      | 0.70| 1.55   | 3,989|
| Radius                   |           |     |        |      |
| Caliper (0.005)          | 0.90      | 0.75| 1.19   | 3,989|
| Caliper (0.001)          | 0.43      | 0.79| 0.54   | 3,989|
| Caliper (0.0005)         | 0.66      | 0.83| 0.80   | 3,989|
| n(1)                     | 0.94      | 0.95| 0.99   | 3,989|
| n(2)                     | 0.97      | 0.85| 1.14   | 3,989|
| n(5)                     | 0.45      | 0.77| 0.59   | 3,989|
| **Current wages**        |           |     |        |      |
| Kernel                   | 0.05      | 0.03| 1.60   | 3,942|
| Radius                   |           |     |        |      |
| Caliper (0.005)          | 0.04      | 0.03| 1.35   | 3,942|
| Caliper (0.001)          | 0.06      | 0.03| 1.73   | 3,942|
| Caliper (0.0005)         | 0.06      | 0.04| 1.54   | 3,942|
| n(1)                     | 0.04      | 0.04| 1.00   | 3,942|
| n(2)                     | 0.04      | 0.03| 1.17   | 3,942|
| n(5)                     | 0.05      | 0.03| 1.58   | 3,942|
Table 6 Matching estimates of the reform using a subsample of students which scored above 6.5 on their final exam

| Dependent variable          | Curriculum change | Treated Mean | Treated N | Matched Mean | Matched N | ATET | SE  | t    |
|-----------------------------|-------------------|-------------|-----------|--------------|-----------|------|-----|------|
| Duration of study           | Yes               | 3.88        | 351       | 3.95         | 2,407     | −0.08| 0.04| −2.06|
|                             | No                | 3.92        | 122       | 3.90         | 2,129     | 0.02 | 0.06| 0.39 |
|                             | DiD               |             |           | 2.916        |           | −0.10| 0.07| −1.43|
| Time in HE                  | Yes               | 4.15        | 351       | 4.25         | 2,407     | −0.10| 0.03| −3.56|
|                             | No                | 4.20        | 122       | 4.23         | 2,129     | −0.03| 0.04| −0.68|
|                             | DiD               |             |           | 2.916        |           | −0.07| 0.05| −1.28|
| Test scores HE              | Yes               | 7.18        | 343       | 7.16         | 2,349     | 0.02 | 0.03| 0.65 |
|                             | No                | 7.12        | 117       | 7.16         | 2,074     | −0.04| 0.05| −0.70|
|                             | DiD               |             |           | 2.845        |           | 0.06 | 0.06| 0.94 |
| Search duration             | Yes               | 5.63        | 317       | 5.61         | 2.157     | 0.02 | 0.45| 0.05 |
|                             | No                | 5.23        | 109       | 5.86         | 1.873     | −0.62| 0.70| −0.90|
|                             | DiD               |             |           | 2.651        |           | 0.64 | 0.83| 0.78 |
| Current wages               | Yes               | 2.24        | 329       | 2.23         | 2.160     | 0.01 | 0.02| 0.39 |
|                             | No                | 2.23        | 109       | 2.30         | 1.873     | −0.07| 0.03| −2.34|
|                             | DiD               |             |           | 2.632        |           | 0.08 | 0.04| 2.14 |

6.2 Restricting the Sample to Students that are Less Likely to Dropout

A concern with the difference-in-difference approach is that the sample consists of graduates and we do not observe the effect of the reform on drop out. If the reform changed the probability of dropping out this might bias our estimates. In that case it is possible that the pool of graduates that were treated with the reform will include more low ability students than the pool of graduates that were not treated with the reform. To avoid this bias we re-estimated the main models for groups of graduates that we expect to be less likely to drop out. Therefore, we focused on the group of treated students with a score of over 6.5 for the final exam of secondary education. This cut off value for the final exam generates a sample of 373 treated graduates and 2,591 untreated graduates which we use for the re-estimation of the main models. Table 6 shows estimation results of the matching approach from the previous section.

In line with the findings in the previous section we find no effect of the reform on performance in higher education. All three difference-in-differences estimates for the indicators of performance in higher education are statistically insignificant. In addition, we find no effect of the reform on the search duration in the labour market. However, we find that the reform increased wages with approximately 8%. Moreover, we tested the robustness of these findings using alternative matching procedures (see “Appendix A”). These tests show that the wage effect is sensitive for the type of matching procedure.

In Sect. 2 we discussed the changes in the number of graduates during the introduction of the reform (see also Fig. 1). As a consequence, the last cohorts of students that
followed the old curriculum took less time for graduating from HAVO and might have been less well prepared for higher education. This might bias our estimates. As a third robustness test we investigated this bias by excluding the enrolment cohort 2000/2001 from our estimation sample. However, the estimates were very similar to our previous findings\(^7\).

### 7 Conclusions

This paper investigated the effect of a major reform in the final years of the two highest levels of Dutch secondary education. Our main finding is that the reform had no effect on student performance and labour market outcomes of graduates from higher professional education. For five dependent variables about student performance and labour market outcomes we find that all estimates of the effect of the reform are statistically insignificant. The findings are robust for various alternative matching procedures. In addition, restricting the sample to students who are less likely to dropout during higher education, based on their exam scores in secondary education, does yield similar results. For this restricting sample we find some evidence that the reform increased wages. However, this effect is not robust for alternative matching procedures. We conclude that the curriculum change did not improve performance in higher education or labour market outcomes of the first cohorts in higher vocational education after the reform.

In 2008 a special committee of members of Dutch parliament concluded that the nationwide introduction of several educational reforms in the nineties, including the reform studied in this paper, was not based on solid evidence and that the effects of the reforms on student performance were not clear. The findings in this paper show that one of these reforms did not have an effect on the performance in and after higher education. Hence, a cost-benefit analysis of this policy would yield a negative result. In general, a nationwide introduction of new policies seems a risky strategy. Even policies that have proven to be successful in other countries might be ineffective in a different context, such as the Dutch educational system. Policies that turn out to be ineffective will induce high social costs. An approach in which major new policies are only introduced after serious testing using small scale experiments can reduce the risk of wasting substantial amounts of public resources.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

**Appendix A**

See Table 7.

---

\(^7\) The estimation results can be obtained from the authors on request.
Table 7  DD-estimates, alternative matching procedures for students with scores above 6.5

| Duration of study | TF-effect | SE  | T-stat | N   |
|-------------------|-----------|-----|--------|-----|
| Kernel            | -0.10     | 0.07| 1.43   | 2,916|
| Radius Caliper (0.005) | -0.09     | 0.07| 1.24   | 2,916|
| Radius Caliper (0.001) | -0.09     | 0.08| 1.08   | 2,916|
| Radius Caliper (0.0005) | -0.03     | 0.09| 0.35   | 2,916|
| n(1)              | -0.08     | 0.10| 0.87   | 2,916|
| n(2)              | -0.10     | 0.08| 1.26   | 2,916|
| n(5)              | -0.15     | 0.07| 2.08   | 2,916|

| Time in HE        | TF-effect | SE  | T-stat | N   |
|-------------------|-----------|-----|--------|-----|
| Kernel            | -0.07     | 0.05| 1.28   | 2,916|
| Radius Caliper (0.005) | -0.06     | 0.05| 1.18   | 2,916|
| Radius Caliper (0.001) | -0.10     | 0.06| 1.70   | 2,916|
| Radius Caliper (0.0005) | -0.03     | 0.06| 0.43   | 2,916|
| n(1)              | -0.07     | 0.07| 1.01   | 2,916|
| n(2)              | -0.05     | 0.06| 0.79   | 2,916|
| n(5)              | -0.07     | 0.06| 1.32   | 2,916|

| Test scores HE    | TF-effect | SE  | T-stat | N   |
|-------------------|-----------|-----|--------|-----|
| Kernel            | 0.06      | 0.06| 0.94   | 2,845|
| Radius Caliper (0.005) | 0.00      | 0.06| 0.07   | 2,845|
| Radius Caliper (0.001) | -0.02     | 0.07| 0.25   | 2,845|
| Radius Caliper (0.0005) | -0.07     | 0.08| 0.88   | 2,845|
| n(1)              | -0.06     | 0.08| 0.72   | 2,845|
| n(2)              | 0.02      | 0.07| 0.27   | 2,845|
| n(5)              | 0.03      | 0.06| 0.47   | 2,845|

| Search duration   | TF-effect | SE  | T-stat | N   |
|-------------------|-----------|-----|--------|-----|
| Kernel            | 0.64      | 0.83| 0.78   | 2,651|
| Radius Caliper (0.005) | 0.40      | 0.91| 0.44   | 2,651|
| Radius Caliper (0.001) | 0.91      | 1.03| 0.89   | 2,651|
| Radius Caliper (0.0005) | -0.16     | 1.08| 0.15   | 2,651|
| n(1)              | -0.27     | 1.13| 0.24   | 2,651|
| n(2)              | 0.54      | 0.96| 0.56   | 2,651|
| n(5)              | 0.32      | 0.88| 0.37   | 2,651|

| Current wages     | TF-effect | SE  | T-stat | N   |
|-------------------|-----------|-----|--------|-----|
| Kernel            | 0.08      | 0.04| 2.14   | 2,632|
| Radius Caliper (0.005) | 0.11      | 0.04| 2.79   | 2,632|
| Radius Caliper (0.001) | 0.04      | 0.04| 0.92   | 2,632|
| Radius Caliper (0.0005) | 0.04      | 0.05| 0.93   | 2,632|
| n(1)              | 0.05      | 0.06| 0.77   | 2,632|
| n(2)              | 0.07      | 0.05| 1.46   | 2,632|
| n(5)              | 0.08      | 0.04| 1.95   | 2,632|
References

Altonji, J. (1995). The effects of high school curriculum on education and labor market outcomes. *Journal of Human Resources, 30*, 409–438.

Berkhout, E. E., Berkhout, P. H. G., & de Graaf, D. (2002). *Verplichte profielen in het voortgezet onderwijs: effect op studieduur en inkomen van hoger opgeleiden*. (Compulsory profiles in secondary education: effects on study spells and wages of higher educated), SEO-report nr. 639, SEO Economic Research, Amsterdam.

Cameron, C. A., & Trivedi, P. K. (2005). *Microeconometrics, methods and applications*. Cambridge: Cambridge University Press.

Dearden, L., Emmerson, C., Frayne, C., & Meghir, C. (2009). Conditional cash transfers and school dropout rates. *Journal of Human Resources, 44*(4), 827–857.

de Vries, R., & van der Velden, R. (2005). *Brug of Kloof?* (Bridge or gap?), ROA report 2005/8, Maastricht.

Hamermesh, D. S., & Donald, S. G. (2008). The effect of college curriculum on earnings: Accounting for non-ignorable non-response bias. *Journal of Econometrics, 144*, 479–491.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica, 73*(2), 417–458.

Rose, H., & Betts, J. (2004). The effect of high school courses on earnings. *The Review of Economics and Statistics, 86*(2), 497–513.

Tweede Fase Adviespunt. (2005). Zeven jaar Tweede Fase, een balans; Evaluatie Tweede Fase. Seven years of the Second Phase an evaluation, The Hague.

Tweede Fase Adviespunt. (2007). *Zakboek Tweede Fase; Tweede fase oude regeling en vernieuwde tweede fase*. Second Phase Handbook original and renewed Second Phase curriculum, The Hague.