Using Data to Advance Educational Research, Policy, and Practice: Design, Content, and Research Potential of the Netherlands Cohort Study on Education

Carla Haelermans 1,2,*, Timo Huijgen1, Madelon Jacobs 1, Mark Levels 1, Rolf van der Velden 1,2, Lynn van Vugt 1 and Sanne van Wetten 1

1Research Centre for Education and the Labour Market (ROA), School of Business and Economics, Maastricht University, PO Box 616, 6200 MD Maastricht, the Netherlands and 2Netherlands Initiative for Education Research (NRO), the Netherlands

*Corresponding author. Email: carla.haelermans@maastrichtuniversity.nl

Submitted March 2020; accepted May 2020

Abstract

In many countries, the quality of (large-scale quantitative) educational research is threatened by data challenges. In this article, we present an innovative data research project from the Netherlands in which many of the challenges that come forward in previous literature are addressed. The Netherlands Cohort Study on Education [in Dutch abbreviated as NCO (Nationaal Cohortonderzoek Onderwijs)] uses longitudinal register data on track placement of cohorts of pupils in primary and secondary education from Statistics Netherlands, which will be combined with school admin data on pupils’ performance and data from the major (inter)national surveys. NCO maps pupils’ pathways and performances through primary and secondary education and their trajectory into tertiary education. NCO so far covers more than 10 complete cohorts and is complemented with the newest data every year. These registers are made available at Statistics Netherlands. A major contribution of NCO is that data from school administrative systems and additional research projects are linked with administrative register data, thereby creating a unique data set that enriches not only research, but also policy and practice. This data brief elaborates on the possibilities of this database by exploring and following one of the cohorts over time.

Introduction

Improving our understanding of how education systems work is of vital importance to get better insights in the explanations for inequalities in educational pathways and performance. However, in many countries, the quality of (large-scale quantitative) educational research, which are often survey-based cohort studies, is threatened by huge non-response rates (Van der Velden and Van der Maas, 2015). At the same time developments in data science offer interesting new possibilities...
for research (Hey, Tansley and Tolle, 2009; Bongers, Jager and te Velde, 2015). To exploit these possibilities for research, we need reliable and representative data of high quality. This entails a large data set, representative for different cohorts, longitudinal data with good information on high stakes tests, and good predictors regarding children, parental background, schools, and neighbourhoods.

A lack of reliable and representative high quality data is problematic for good research, but also for policy and practice. The problems that researchers commonly encounter include (i) the problem of insufficient access to data, (ii) the problem that large, representative, high quality (longitudinal) data sets on education are rare, and (iii) the problem that schools are increasingly unwilling to participate in (additional) research collections, which results in selective non-response on surveys and low response rates. These low response rates have several adverse effects. First, they pose problems for researchers, as non-response undercuts the quality of data. Second, they impose additional burdens on the remaining schools that are willing to participate. These problems may lead to unrepresentative samples and/or potential power issues in studies (Murnane and Willet, 2010). Third, they also pose a problem for policymakers who aim to base educational policies on sound analyses of high-quality evidence.

In the past few decades, large-scale educational research in the Netherlands, has increasingly been plagued by these problems. Most notably, selective non-response among schools is undermining data quality. It is not uncommon that only 25% of the schools or even less participate in large-scale research projects. The non-response in large-scale research projects has several reasons. First, participating in studies takes a lot of time and puts a burden on teachers and pupils. Second, it is often argued that the low response is partly due to stakeholders at school (principals, administrators, staff) not seeing an immediate benefit of participating in a certain research project. Although the stakeholders acknowledge a responsibility to participate in educational research, they assess the associated costs in terms of workload as too high and the immediate benefits as too low (Van der Velden and Van der Maas, 2015). A third reason is the lack of coordination between different national and international data collection projects that approach schools for participation in their project (examples of these are the PISA and TIMMS data collections, and large national research projects coordinated by the Netherlands Organisation for Scientific Research). This results in some schools being sampled multiple times in the same year while others get no request. The relevance of the specific research projects to the school or whether the school belongs to the network of the researcher, may directly influence the decision to participate, leading to selection bias in the sample (Van der Velden and Van der Maas, 2015).

NCOs Approach to Overcome These Problems

Administrative registers are becoming increasingly popular as data sources for research and policy (Reed and Raaum, 2003; Card et al., 2010; Einav and Levin, 2013) and are also seen as a real promise in education research (González-Sancho and Vincent-Lancrin, 2016; Figlio, Karbowik and Salvanes, 2017). However, using register data in isolation is not going to solve the problems described above, as much valuable information is not contained in registers. Consequently, many research questions need other, additional data such as perceptions or motivations, to be answered properly.

The key behind a sustainable model for producing high-quality data for educational research is to alleviate the burden of participation in surveys for schools, while at the same time maximizing the returns to research participation in such a way that schools can benefit from high quality research and data. One way of reducing the workload for schools is by making better use of data that are already available. In many countries, schools already provide a lot of information about their pupils, teachers and school practices to national organizations and governments. In the Netherlands e.g., comparable high stakes aptitude tests take place multiple times during the school year in primary education. These data are stored at schools, but usually not disclosed to researchers. By making data available to researchers and schools, the additional time pressure schools experience to cooperate in research is avoided, leading to more efficient use of research resources and better quality data.

In this vein, in 2016, the Netherlands Initiative for Education Research (NRO) initiated the Netherlands Cohort Study on Education [in Dutch abbreviated as NCO (Nationaal Cohortonderzoek Onderwijs)]. NCO systematically combines existing register data from the Netherlands and integrates them with relevant other data sources. The core of NCO is formed by register data from Statistics Netherlands on the full population of pupils in the Netherlands, which allows us to address some of the problems of survey data, such as response problems and data quality, but also improves statistical power and increases possibilities to link data. By hosting these data sets at Statistics Netherlands, an entrusted third party, the data can be linked to all other data, and
the data are available to all researchers, while at the same time ensuring security and confidentiality issues.

The NCO currently consists of four pillars. The first pillar maps pupils’ pathways through education and their trajectory into tertiary education and combines this with information on their background using register data from Statistics Netherlands. The second pillar consists of additional information at the school level. This information is available at the Dutch Ministry of Education and the Dutch Inspectorate of Education. It consists of data on e.g., school size, urbanization, and denomination, and in the future will be complemented with more detailed information on the educational process and school quality. A third pillar consists of microdata on pupil performance from school administrative systems. Primary schools in the Netherlands are required to monitor their pupils’ progress in domains like language and math. Most schools use some national test for this, which makes it possible to have information on the development of pupil’s performance between the age of 8 and 12. We are currently conducting a pilot to make these data available for NCO. The fourth pillar consists of linking the major national and international surveys to the NCO data set. The work on this pillar will start from 2020 onwards. It offers the possibility to enrich surveys with longitudinal data, e.g., linking PISA data at age 15 with previous pupil performance during primary education or later outcomes such as success in higher education.

In this data brief, we present the features and possibilities of the NCO data set. In Section 2, we shortly explain the Dutch education system. Next, in Section 3, we delineate the composition and structure of the NCO data set. Section 4 provides more information on the contents of the data set and Section 5 discusses data access procedures, future plans, and the school reports. Section 6 concludes.

**The Dutch Education System**

**General Properties**

To understand the structure and potential of the NCO, it is essential to have a good understanding of the Dutch education system. Figure 1 presents an overview of the structure of education in the Netherlands. Most schools in primary and secondary education in the Netherlands are publicly funded. Privately funded schools do exist but are very rare. Schooling in the Netherlands is compulsory from the age of 5 to 18 years. Most pupils however, start primary school at the age of 5 to 18 years. Privately funded schools exist but are very rare. Schooling in the Netherlands is compulsory from the age of 5 to 18 years. Most pupils however, start primary school at the age of 6.

4. They generally transition to secondary education around the age of 12. At this age pupils are tracked into different types of secondary education, which in turn enables them to enrol at different levels of tertiary education. Pupils can attend preparatory vocational secondary education—in Dutch called *voorbereidend beroeps onderwijs* (*vmbo*). Within preparatory vocational education, four tracks are distinguished depending on the amount of practical education versus theoretical education. The lowest track in preparatory vocational education is called *vmbo praktijkonderwijs* (*vmbo pro*), followed by *vmbo basisberoepsgerichte leerweg* (*vmbo bb*), *vmbo kaderberoepsgerichte leerweg* (*vmbo kb*), and *vmbo gemengde en theoretische leerweg* (*vmbo gltl*). These tracks all take 4 years to complete and enable access to upper secondary vocational education and training or *middenbaar beroeps onderwijs* (*mbo*) in Dutch. The structure of upper secondary vocational education in the Netherlands corresponds with the different tracks in preparatory vocational education. *Vmbo pro* grants access to *mbo level 1* which last 1 year. *Vmbo bb* grants access to *mbo level 2* which lasts 2 years, *vmbo kb* and *vmbo gltl* grants access to *mbo level 3* and *mbo level 4*. *Mbo level 3* education lasts 3 years and *mbo level 4* lasts 3 to 4 years.

After primary school, pupils can also attend senior general secondary education—in Dutch called *hoger algemeen vormend onderwijs* (*havo*). This track takes 5 years to complete and allows access to higher professional education or *hoger beroeps onderwijs* (*hbo*) offered at universities of applied sciences (UAS). The highest track of secondary education is university preparatory education—in Dutch called *voorbereidend wetenschappelijk onderwijs* (*vwo*). This track takes 6 years to complete and allows access to research-oriented higher education or *wetenschappelijk onderwijs* (*wo*) offered by research universities. Educational programmes at UAS and research universities lead to either bachelor’s or master’s diplomas, albeit UAS’ bachelor programmes have a longer duration (4 years) than research universities’ bachelor programmes (3 years). Another characteristic of the education system in the Netherlands is that mobility between and stacking of educational levels is allowed. For instance, after completing upper secondary vocational education at level 4, it is possible to continue studying at a UAS level programme.

**Grade System in the Netherlands**

In the Netherlands, pupils enrol at the age of 4 or 5 in kindergarten (KG) 1 and 2. After this, they follow...
6 years of education in their primary school (grade 1 to grade 6). After grade 6, they transfer to secondary school, which is always a different school compared to their primary school. In this school, they follow depending on their track, 4, 5, or 6 years of education. Pupils following the preparatory vocational secondary education (vmbo) tracks are enrolled in grades 7, 8, 9, and 10. Pupils in senior general secondary education (havo) follow 5 years of education in grades 7, 8, 9, 10, and 11. Lastly, pupils in university preparatory education (vwo) follow grades 7, 8, 9, 10, 11, and 12. In the Netherlands, it is not uncommon for pupils to repeat a certain grade. For example in primary education, pupils often repeat KG 2 when the pupils are not ready yet to start writing and reading. In secondary education, pupils often repeat classes as well. In university preparatory education (vwo), pupils are most likely to repeat the 9th grade. In primary education specifically, pupils sometimes skip grades when their reading and writing skills are of a higher level.

Testing and Standardized Tests in the Netherlands

At the end of primary education, pupils take part in standardized national tests to determine the right track of secondary education. This test results in a score which is accompanied by a tracking advice. Besides a test score and test advice, pupils receive a teacher advice based on the pupil's performance over the last years. It is important to note that in the Netherlands the law regarding standardized tests changed in 2014/2015 allowing only a limited number of tests to be offered and obliged all primary schools to implement one of these standardized tests. Therefore, after 2014/2015 one can see a change in the number of test suppliers. Each supplier has established a unique standardized test indicating the suitable track in
secondary education. With this new law, the timing of the teacher advice also changed. Before 2014/2015 the teacher gave an advice after the pupils took the standardized tests, as of 2014/2015 teachers give their advice before pupils take the standardized tests.

In secondary education, the performance of pupils is measured by their grades, on a 0–10 scale. In general, a minimum score of 5.5 is sufficient. In the final year of secondary education pupils take a centralized exam, existing of several school exams and also national exams on the specific courses they follow. Again, the scores on these exams can range from 0 to 10 where a 5.5 is seen as a pass. Furthermore, the pupils in the last year of secondary education receive a diploma if the combination of the school exams and the central exams are above 5.5.\(^7\)

Adult Education in the Netherlands

In the Netherlands, it is possible for adults (18+, and in some cases even at the age of 16 or 17) to get a complete degree or a specific subject certificate for three major tracks in secondary education. This is possible for the highest level of the pre-vocational secondary education track, for the senior general secondary education and for the pre-university track. The qualifications in adult education are equal to the ones in regular secondary education.

Composition and Structure of the NCO Data Set

In this section, we describe the first two pillars of the NCO data set, namely (i) pupils’ pathways through education and their trajectory into tertiary education, combined with information on their background using register data from Statistics Netherlands and (ii) additional information at the school level. These pillars are based on three different sources. The first source is register data concerning pupils’ educational careers. All schools in the Dutch education system provide yearly data about their pupils to the Dutch government’s Executive Agency for Education (DUO; Dutch abbreviation for Dienst Uitvoering Onderwijs), the data department of the Ministry of Education. The data that the Ministry of Education collects consist of annual data for all pupils with respect to track, grade, exam results, and some other school career aspects. This allows for the construction of a population-wide longitudinal data set on track placement and performance. The second source comes from Statistics Netherlands who obtains administrative register data on personal characteristics of these pupils, such as gender, age, country of origin, and on their parental background such as their income, marital status, work status, and education. These data provide a number of relevant background characteristics and are made available at Statistics Netherlands [in the so-called system of social statistical data sets (SSD) which we further describe in the next section]. Lastly, the Ministry of Education and the Inspectorate of Education have additional information available about schools, such as denomination and school structure. All pupils and schools have a unique personal identifier that can be used to link the different data sources. In the section below, we elaborate on the various sources of the NCO data set and how these are combined. The combination of data sets from these various sources for NCO is done by following a ‘cohort structure’ which is discussed in the last part of this section.

The System of Social Statistical Data Sets

The NCO data set is built within the context of the data infrastructure of the system of Social Statistical Data sets (SSD) at Statistics Netherlands as explained above. In the Netherlands, the SSD is the most important source for official social statistics from registry data (Bakker, van Rooijen and van Toor, 2014). SSD is the backbone of all the social data that is available at Statistics Netherlands.

In short, the SSD is a system of interlinked and standardized registers and surveys. It contains information on persons, households, jobs, benefits, pensions, education, health, crime reports, and more. The basis for the SSD are a large number of registers (more than 50) which sole purpose is to administrate records. In order to create one system of data sets out of these different registers three core elements are crucial: (i) data are centrally stored in a standardized way, (ii) different unit types like persons, companies, schools can be uniquely linked by assigned linkage keys, and (iii) the coordination of organizational, technical and content related aspects is of utmost importance. Some examples of national administrative registers being used in the SSD are the tax registers, the Dutch national education registers, pension funds registers, and the Dutch population register. The latter is also the central register for the process of assigning linkage keys to individual persons in the SSD. Together with households, companies, non-profit organizations, and buildings, individuals are central statistical unit types within SSD. These linkage keys are assigned by Statistics Netherlands for linkage purposes only and are anonymous and preclude direct identification of unit types.

Statistics Netherlands has access to all these registers. The access is formalized on the basis of the Statistics
the NCO population in the data set, we use the following three cohort types: primary education entry cohorts, primary education exit cohorts, and secondary education entry cohorts.

Firstly, this has practical reasons. One of the aims of NCO is to construct a national data infrastructure for (longitudinal) education research in the Netherlands. Simply using all available information on all individuals from BRON would not only lead to unworkable large data sets but would also make the NCO data set unfit for direct use for longitudinal studies. A preselection of variables as well as a clear definition of the relevant population is necessary to construct an organized, and maintainable data infrastructure. A central starting point for this infrastructure is therefore the definition of the different cohorts. These cohorts are a selection of all individuals in the national education registers.

Secondly, it has technical reasons. The information in the registers on entrants in primary education starts in 2010/2011. However, even though information is available on a yearly basis from 2010/2011 onwards, research possibilities for these entry cohorts are for the time being rather limited, as we are mostly interested in the final year of primary school, the transition to secondary school, and performance in secondary school and tertiary education. However, as primary school takes 8 years, it takes quite a number of years before we observe the primary education entry cohorts in secondary education, and currently we only have one cohort for which this is the case. This cohort is the primary education entry cohort of 2010/2011 which has only started their secondary education trajectory in 2018/2019 (see also Table 1). Consequently, successes in secondary and tertiary education cannot be mapped yet for this entry cohort, nor for later primary education entry cohorts at this point in time. In a few years’ time, after the addition of new yearly data sets, these longitudinal data become available for research. To still be able to analyse the transition between primary and secondary education the NCO data set currently focuses not only on these primary education entry cohorts but also on cohorts leaving primary education (primary education exit cohorts) and cohorts entering secondary education (secondary education entry cohorts). These three cohort types form the population of the NCO data set.

Primary education entry cohort
The first type of cohort is the primary education entry cohort; pupils who are in KG 1 of primary education in year \( t \) are part of entry cohort \( t \). The data on these...
primary education entry cohorts are available from school year 2010/2011 onwards, where each school year a new cohort is added to the data. Pupils entering primary education at a later stage than KG 1 are assigned to their relevant entry cohort. For example, a pupil enrolling in primary education in KG 2 in school year 2013/2014 is assigned to the 2012/2013 entry cohort. Limiting these entry cohorts to pupils who actually started in the first year of kindergarten, is achieved by selecting all pupils whose cohort year is the same as their starting year.

Primary education exit cohort

The second type of cohort is the primary education exit cohort, which we use to map the transition between primary and secondary education as long as we do not have enough primary education entry cohorts that have already made this transition. The primary education exit cohorts consist of pupils who are in the final year (grade 6) of primary education in year $t$. As mentioned earlier, this type of cohort has been established due to the relative short history of the NCO data set and the available data on primary education entry cohorts cover a too short time span for elaborate analyses on the transition from primary to secondary education. The exit cohorts also start from school year 2010/2011. For the earlier exit cohorts, information on their previous educational trajectory is (partly) missing since these pupils started before 2010/2011. As the NCO data set ages with the addition of new years and new cohorts, these primary education exit cohorts will at some point no longer be needed when a full transition from primary to secondary education becomes possible to study. For example, from the 2010/2011 primary education entry cohort the first pupils will graduate from secondary education in 2021/2022. For now, the exit cohorts allow analyses on the transition from primary to secondary education, since the transition decisions take place in grade 6. Therefore, all necessary information for the transition is present for these exit cohorts. Limiting these exit cohorts to pupils who actually left primary education, can be done by looking into the registration of these pupils in year $t+1$. These children should then not be observed in primary education anymore, but are most likely in secondary education.

Secondary education entry cohort

The third type of cohort is the secondary education entry cohort. Pupils who are in grade 7, the first grade of secondary education in year $t$, are part of the secondary education entry cohort $t$. Pupils entering secondary education at a later stage than grade 7 are assigned to their relevant secondary education entry cohort. Records from secondary education are available from school year 2007/2008 onwards. For the earlier secondary education entry cohorts (until school year 2010/2011) information on their previous educational trajectory (primary education) is partly missing as it is not possible to track their history in primary education before 2010/2011. Limiting these entry cohorts to pupils who actually started in the first year of secondary education, is done by selecting all pupils whose secondary education

### Table 1. Number of pupils per school year per cohort

| School year | Primary education entry cohort* | Primary education exit cohort | Secondary education entry cohort |
|-------------|---------------------------------|------------------------------|----------------------------------|
| 2007/2008   | —                               | —                            | 189,862                          |
| 2008/2009   | —                               | —                            | 193,323                          |
| 2009/2010   | 174,962                         | 194,477                      | 195,327                          |
| 2010/2011   | 175,140                         | 198,489                      | 200,980                          |
| 2011/2012   | 177,739                         | 199,900                      | 203,628                          |
| 2012/2013   | 175,883                         | 196,829                      | 205,298                          |
| 2013/2014   | 175,380                         | 194,980                      | 204,220                          |
| 2014/2015   | 171,994                         | 189,966                      | 205,826                          |
| 2015/2016   | 172,207                         | 185,038                      | 201,542                          |
| 2016/2017   | 165,577                         | 180,400                      | 194,278                          |
| 2017/2018   | —                               | —                            | 189,598                          |
| 2018/2019   | —                               | —                            | 189,598                          |

*The numbers for the primary education entry cohort are based on registrations in KG 2. This is done because in KG 1 children enter into primary education throughout the entire year, which is mainly based on their month of birth. However, the reference date on which the educational data is gathered is the first of October of each school year and does therefore not reflect the total number of children entering primary education in each year. Therefore the registration in KG 2 in $t+1$ is more accurate for entry cohort $t$.

Source: Netherlands Cohort Study on Education (NCO).
entry cohort year is the same as their start year in secondary education.

Implications cohort structure
The NCO cohort structure has implications for how the data should be used, and can be used. For example, it is relatively easy to follow and analyse the ‘class of 2010/2011’ yet it is difficult to obtain cross sectional data on an entire school in 2010/2011. It is important to notice that the notion of cohorts, i.e., the pupils in certain cohorts, form the ‘backbone’ of the NCO data infrastructure. Based on the secured, anonymized personal identification codes from these individuals it becomes possible, and relatively easy, to match other information from available register data at Statistics Netherlands.

Table 1 shows the (rounded) number of children into these different cohorts. As the bold numbers demonstrate, children from the primary education entry cohort 2010/2011 are the primary education exit cohort of 2017/2018. The large majority of this exit cohort forms the secondary education entry cohort of 2018/2019.

The table shows that there are differences between the entry cohorts numbers and the corresponding exit cohort numbers in primary education. We see that in 2010/2011 174,900 pupils enrol in primary education while 180,400 pupils leave primary education in 2017/2018. This difference can be explained by (i) pupils who leave primary education and enrol in special primary education, (ii) pupils who have to repeat classes in primary education and therefore do not leave primary education 8 years after their enrolment, (iii) pupils who skip a certain grade and have already left primary education in 2017/2018—therefore belonging to the primary education exit cohort 2016/2017, and (iv) migrant pupils entering primary education throughout these 8 years.

Furthermore, the table shows a discrepancy in the number of pupils leaving primary education and the number of pupils entering secondary education the following year. This difference can be explained by (i) pupils enrolling into regular secondary education after following special primary education, (ii) entering of first generation migrant pupils into secondary education—about 30–50% is a first generation migrant pupil, and (iii) pupils who have not followed (state-funded) primary education in the Netherlands.

Variables of the NCO Data Set
As mentioned in the previous section, the NCO data set consists of three different types of cohorts. Based on the personal identifier, several demographic, socio-economic, and educational variables are matched to the NCO data set at Statistics Netherlands. All of these variables stem from registers that are part of SSD and are available on a yearly basis. Within each type of cohort, the availability and completeness of these variables over the different entry and exit years is comparable.

Position in the education system
Figure 2 shows the NCO data set structure. The backbone of the NCO data set are the variables on individual’s educational careers (blocks 1 and 2 in Figure 2) derived from the national education registers. These registers hold a large number of variables on the pathways and performances of pupils in state-funded primary, secondary, and tertiary education.

Variables on primary education included in the NCO data set consist of, amongst others, information on enrolments, standardized test scores in the final year of primary school (grade 6) and their accompanying math and reading scores, primary school track placement recommendations regarding the follow up track in secondary education from teachers, switching to special education tracks, and lastly a school identifier variable. Variables concerning secondary education include information on enrolments, obtained qualifications, study fields, and exam results at detailed levels per course and a school identifier. This school identifier allows the matching of pupils within schools. For tertiary education, information is available on enrolments, gained qualifications, and study programmes. All variables regarding the level of education and study programmes can be linked to the ISCED-classification. Unfortunately, the national education registers, and therefore the NCO data set, do not include information on classes and teachers of the pupils.

Socio-economic and demographic information
Several socio-economic and demographic variables are available in the NCO data set (block 3, Figure 2). For each individual, information on age, gender, country of origin, marital status of the parents, household information, socio-economic status (SES) of both individual and his/her parents, and regional variables are available. Parental variables include variables on their highest obtained educational level, SES, working status, income, and wealth. The funding of Dutch primary schools is partly based on the socio-economic background of the school’s population. The NCO data set therefore holds information on whether or not and to what extent a school population consists of pupils with disadvantageous socio-economic statuses.
School characteristics

For all pupils, each enrolment and qualification are linked to a specific school identifier. Some schools are an own entity; for these schools the school identifier is unique. Some schools are either operating within a larger educational institute (e.g. schools can offer different tracks within the same institute) or are located in different locations. For these schools, there is an overlapping code within the school identifier and a unique location code to specify the specific school. The school identifier makes it possible to add several school characteristics to the NCO data infrastructure (block 4, Figure 2) at the most detailed, individual school level. Information at school level consists of variables on the denomination as well as regional variables like municipality and urbanisation level. Variables on the educational structure of the school and several indicators on the pupil population are also available, e.g., the size of the school and information about the pupil population enrolled in the school in year $t$.

Example Cohort

In this section, we present descriptives of one of the primary education exit cohorts to elaborate on the variables in the NCO data set. The primary education exit cohort that we use for this example finished primary education in 2010/2011. The majority of pupils in this primary education cohort enrols in secondary education in 2011/2012. Depending on their track, students from this cohort enter upper secondary vocational education or tertiary education from 2015/2016 onwards. As such, this specific primary education exit cohort allows us to present descriptives on primary education, secondary (vocational) education, as well as tertiary education.

With the exception of Table 5, the descriptives shown in the tables are yearly descriptives of the cohort. Although information on almost all variables is available for every year, we chose to present only one. Register data, like survey data, has outliers and rare cases. In the tables below, sometimes unlikely values are shown. While the NCO data have been cleaned, it is up to researchers to decide whether specific cases are included in analyses or not. As this is register data stored at Statistics Netherlands, this cannot be corrected at the original administrative source. Also, in the registers missing values are sometimes assigned to a value code without defining this specific code as a missing value. At first glance, it appears that the variable has valid information for all cases but in practice information is missing for some. We leave it up to researchers to decide what to do with these cases.

Background characteristics

Table 2 shows basic background characteristics of the pupils in the selected cohort. As can be seen, 50.7% of the pupils in this cohort are male and 49.3% of the pupils are female. The Netherlands is the country of origin for the majority of pupils in this primary education exit cohort. However, quite a large number of pupils are
from non-western countries (15.7%), while 6.0% of pupils come from other western countries. Pupils with an origin outside of the Netherlands are often second-generation migrants (18.8%) whereas only 2.9% is a first generation migrant.

Household and parental characteristics

Table 3 presents the variables related to pupils’ socio-economic status and background. Several indicators are created based on the socio-economic status of the parents as well. Besides the household composition and the number of children within the household, the data set also includes the cohabitation status which indicates whether parents are cohabiting at the same address in the Netherlands. There are several reasons why parents are not cohabiting anymore. The main reason is that one of the parents moved to a different address or both parents moved to a different but separate address (17.2%). ‘Not living together; unknown’ means that parents are not cohabiting anymore but the reason is unknown. ‘No registry of living together’ indicates that it has never been registered whether or not parents are or have been cohabiting. This last category can be seen as a missing value.

The NCO data set holds elaborate information on the socio-economic status (SES) of both parents. The values shown in Table 3 are combinations of underlying values, which are combined into employed, not employed and receiving benefit. For privacy reasons, the wealth and income statistics are transformed into percentile scores. For instance, a pupil from whom the father scores a 2 on income, is in the bottom second percentile with respect to income, meaning that 98% of the other pupils have fathers with higher incomes. In principle, the wealth descriptives of the father and mother range from the 1st percentile to the 100th percentile.

The highest obtained educational status of the parents is only available for 36.1% of the fathers and 41.2% of the mothers. This is due to the fact that educational diploma registers have only been established fairly recently. Therefore, educational attainment of parents from older cohorts is difficult to track. Additionally, the presented descriptives are shown for year 2010/2011. Note that it is possible that parents have obtained a higher degree in the period since 2011/2012 up to now.

Besides information about the household, every pupil has a neighbourhood/zip code with which information about the living area of the pupil can be matched. However, as the zip code is 8-digit detailed code, this results in quite an extensive list of frequencies and therefore this is not shown in the tables presented in this article. An important indicator that can be matched based on zip code is an identifier for pupils living in deprived areas, the so-called ‘APC-areas’.

Primary education characteristics

Table 4 shows the variables in the NCO-data set related to primary education of the pupils. In the Netherlands, financing of schools is based on the number of pupils within a school and on whether or not a pupil is viewed as being ‘disadvantaged’ and in need of extra care and attention of teachers. Pupils are seen as ‘disadvantaged’ when their parents’ educational attainment is quite low. On average, 81.6% of the pupils are not regarded as ‘disadvantaged’, whereas 13% of the pupils are seen as disadvantaged.

Furthermore, the data set includes information on the standardized national test pupils took during their last year in primary education. In our example cohort of 2010/2011, we see that more than six different tests were used at the time (note that this is before these tests were made mandatory in 2014/2015). The Cito13 test has always had a dominant position but in recent years the role of other test suppliers has increased. The test scores are often measured with a score based on the number of correct answers. In Table 4, the standardized score is shown. It is important to note that tests have their own range of scores, some ranging from 0 to 100,
others from 500 to 550. The test score variable is therefore only relevant in combination with the variable ‘standardized test’ as this variable indicates which test has been taken.

Table 3. Descriptive measurements of household and parental characteristics of the pupils from primary education exit cohort 2010/2011 in 2010/2011

| Household composition | N    | Mean | Percentage |
|------------------------|------|------|------------|
| Living with both parents | 152,641 | 76.9 |
| Living with mother     | 27,812 | 14.0 |
| Living with father     | 3,352  | 1.7  |
| Living with mother and partner | 11,380 | 5.7  |
| Other                  | 2,688  | 1.4  |
| Number of children in the household | 198,579 | 2.4  |
| Cohabitation status parents | 198,579 | 77.0 |
| Cohabiting together    | 152,813 | 77.0 |
| Not cohabiting; separate address | 34,195 | 17.2 |
| Not cohabiting; death  | 2,302  | 1.2  |
| Other/no registry      | 9,269  | 4.7  |
| Socio-economic status father\(^a\) | 186,994 | 85.4 |
| Employed               | 169,501 | 85.4 |
| Not-employed           | 3,351  | 1.7  |
| Receiving benefit      | 14,142 | 7.1  |
| Socio-economic status mother\(^a\) | 196,160 | 75.1 |
| Employed               | 149,221 | 75.1 |
| Not-employed           | 26,988 | 13.6 |
| Receiving benefit      | 19,951 | 10.0 |
| Income father (percentile score) | 186,338 | 69.7 |
| Income mother (percentile score) | 195,529 | 67.4 |
| Wealth father (percentile score) | 186,999 | 56.3 |
| Wealth mother (percentile score) | 196,162 | 55.2 |
| Highest educational attainment father | 71,602 | 36.1 |
| Primary education      | 5,833  | 8.2  |
| Vmbo bb, vmbo kb, mbo level 1 | 7,559 | 10.6 |
| Vmbo gl/tl or havo and vwo grade 7, 8 and 9 | 1,706 | 2.4 |
| Mbo level 2 and mbo level 3 | 8,178 | 11.4 |
| Mbo level 4             | 11,035 | 15.4 |
| Havo and vwo            | 8,836  | 12.3 |
| Hbo and wo bachelor     | 16,776 | 23.4 |
| Hbo and wo master       | 11,629 | 16.2 |
| Highest educational attainment mother | 81,723 | 41.2 |
| Primary education      | 8,791  | 10.8 |
| Vmbo bb, vmbo kb, mbo level 1 | 9,325 | 11.4 |
| Vmbo gl/tl or havo and vwo grade 7, 8 and 9 | 3,339 | 4.1 |
| Mbo level 2 and mbo level 3 | 10,180 | 12.5 |
| Mbo level 4             | 13,528 | 16.6 |
| Havo and vwo            | 9,765  | 11.9 |
| Hbo and wo bachelor     | 16,920 | 20.7 |
| Hbo and wo master       | 9,875  | 12.1 |

Source: Netherlands Cohort Study on Education (NCO).

\(^a\)The categories for SES of the father and mother are based on several underlying categories combined into one, which are: employed, director, self-employed, other self-employed, working household member, receiver unemployment benefit, receiver social assistance, receiver social services, receiver sickness benefit, receiver pension benefit, other without income, household without observed income.

The percentile scores math and reading which are presented below the standardized test score, are percentile scores based on the subset of questions related to either math or reading only for pupils who have taken the
Pupils with disadvantaged backgrounds for which the school receives financial support

| Description                                                                 | N    | Mean | Percentage |
|-----------------------------------------------------------------------------|------|------|------------|
| Missing                                                                     | 10,691 | 5.4  |            |
| 0.00 pupil is not disadvantaged                                             | 162,089 | 81.6 |            |
| 0.30 both parents of child have no higher educational attainment than vmbo pro | 15,395 | 7.8  |            |
| 1.20 one parent has no higher educational attainment than primary education, the other not higher than pro/vbo | 10,404 | 5.2  |            |

Standardized test

| Description                      | N    | Mean | Percentage |
|----------------------------------|------|------|------------|
| Pupil has not done standardized national test                               | 25,263 | 12.7 |            |
| Cito test                        | 143,838 | 73.4 |            |
| GPC intelligence test            | 4,414  | 2.2  |            |
| GPC school improvement test      | 2,406  | 1.2  |            |
| Drempelonderzoek                 | 5,947  | 3.0  |            |
| Other                            | 14,711 | 7.4  |            |

Standardized test score

| Description                      | N    | Mean | Percentage |
|----------------------------------|------|------|------------|
| Percentile score math            | 139,041 | 50.62 |            |
| Percentile score reading         | 139,041 | 50.56 |            |

Teacher advice (see Figure 3)

| Description                      | N    | Mean | Percentage |
|----------------------------------|------|------|------------|
| Number of schools within board   | 198,579 | 17.8 |            |

Denomination of the school

| Description                        | N    | Mean | Percentage |
|------------------------------------|------|------|------------|
| Public schools                     | 60,566 | 30.5 |            |
| Schools based on (educational, pedagogical or societal) philosophies | 11,255 | 5.7  |            |
| Schools based on religious beliefs | 126,637 | 63.8 |            |

Size of the school (number of pupils)

| Description                      | N    | Mean | Percentage |
|----------------------------------|------|------|------------|
| Percentage disadvantaged pupils within schoola | 198,579 | 290  |            |

| Source: Netherlands Cohort Study on Education (NCO). |
| aThis is the total percentage of pupils per school for which the school receives additional financial aid. |

Cito-test. A pupil with a math percentile score of 80 means that the pupil has done a better test than 80% of the other pupils.

The last few variables in Table 4 are characteristics of the school the pupils go to. The denomination of the schools is shown in four categories resulting from a combination of underlying categories based on either ideological/pedagogical or religious foundation of the school. Around 31% of the pupils go to a public school, whereas more than 64% of the pupils go to schools with a religious foundation. Furthermore, pupils go on average to schools existing within a board of 17.8 schools and have on average 290 pupils within the school.

Figure 3 shows the distribution of the teacher’s track advice for the primary education exit cohort 2010/2011. In addition to the standardized test score, the teacher also gives an advice which is not only based on this test score but also takes into account other factors based on his or her own experience with the pupil during primary education. This advice therefore reflects which track level is, according to the teacher, the most suitable after primary school. In the Netherlands, it is quite common to give pupils a ‘double’ advice. This for example entails that the advice of the pupils’ most suitable match is ‘higher general secondary education/pre-university education’, which is often given by teachers who are a bit conservative and/or want to give a child the chance to prove herself at a higher track. ‘Single’-advices (69%) are given more frequent than ‘double’-advices (29%). The categories presented in Figure 3 are combined in accordance with common practice of the Dutch Inspectorate of Education’s guidelines for combining advices. These guidelines are broadly used within educational research in the Netherlands and are also part of the Inspectorate’s methodology for evaluating individual schools. Almost 10% of the pupils have not received a teacher advice, half of them also do not have a standardized test score.

Secondary education characteristics

In Table 5, the variables related to secondary education of the pupils are shown for 1, 4, and 6 years after leaving primary education (2011/2012, 2014/2015, and 2016/2017, respectively). This format is chosen to highlight the longitudinal aspect of the NCO data set and to show
the implications this has for interpreting results and for the descriptive statistics. In the first year after leaving primary education (2011/2012), the exit cohort enters secondary education. Depending on their chosen track it takes them 4–6 years to obtain their diploma without delay. This means that students enrolled in pre-vocational secondary education (duration of 4 years) can graduate no earlier than 2014/2015, students enrolled in senior general secondary education (duration 5 years) can graduate no earlier than 2015/2016, and students enrolled in pre-university education (duration 6 years) can graduate no earlier than 2016/2017.15

The specific track in secondary education the pupil is attending is captured by a code in the data set. As this list of codes is very elaborate it is not shown here. These codes are based on a national register classification, and it is possible to match additional information within the SSD to the NCO data sets by using this track code as a linkage key. More information that is not shown in the table below concerns subjects pupils attended and the grades pupils obtained for these subjects. National final exam and school exam grades are available for a list of subjects which includes, but is not restricted to, Dutch, English, German and French, but also math, history, physics, economics, chemistry, and others.

Furthermore, the average national exam grade and school exam grade is provided. In the year 2011/2012, no national exam grades or school exam grades are available yet for the primary education exit cohort of 2010/2011 as this is their first year in secondary education. The average national exam grade for the year 2014/2015 is on average 6.49 on a scale of 1 to 10, the average school exam grade has an average of 6.48.16 For the year 2016/2017, the average national exam grade is on average 6.48, and the average school exam grade has an average of 6.57.17

The NCO data set also includes information about ‘adult education’ (not shown in Table 5); when students fail the national exams at the regular secondary school, they often enrol into ‘adult education’. The information in the data set about adult education includes the registration, the level of adult education and whether or not the pupil graduated from adult education.

Secondary education schools’ characteristics

Table 6 presents the school characteristics of pupils’ secondary school. These descriptives are based on their first year in secondary education in 2011/2012. The variables in this table are related to the denomination of the school the pupil is enrolled in, the number of schools within the board, the size of the school the pupil is enrolled in and the size of the different educational tracks. On average pupils go to schools who exist within a board of 13.9 schools and have on average 1086.5 pupils within their schools. Furthermore, most of the pupils go to schools with religious denominations (57.5%), followed by public schools (25.5%), schools based on educational, pedagogical, or societal philosophies (13.1%) and lastly schools who are multi-denominational (3.9%).

Furthermore, we have information about the structure of the secondary school. In the Netherlands, schools often provide education in several tracks. The variable
regarding the number of pupils within a track shows what type of tracks the school offers. Based on this, one can identify community schools and single-track schools. The pupils that belong to the track ‘other’ are pupils who for example attend a double track, such as a combined higher general secondary education/pre-university education track. All pupils have information about the size of the tracks, however, when the school does not offer this track, pupils are given a value of 0.

### Tertiary education characteristics

Tables 7, 8, and 9 show the flow of students into their further educational pathway such as upper secondary vocational education (Table 7), universities of applied sciences (Table 8), and research universities (Table 9). In Table 7, the descriptives of the students flowing into upper secondary vocational education are shown for the year 2015/2016. As mentioned earlier, this year is chosen specifically since the nominal duration of the pre-vocational tracks ‘vmbo’ last 4 years. So, 2015/2016 is theoretically the first year in which students of the primary education exit cohort 2010/2011 (that we used as an example throughout the article) can study in upper secondary vocational education.

#### Table 7. Descriptive measurements of secondary education characteristics after 1, 4, and 6 years of leaving primary education of the pupils from the primary education exit cohort 2010/2011

| School year | 2011/2012 | 2014/2015 | 2016/2017 |
|-------------|-----------|-----------|-----------|
| N/percentage| N/percentage| N/percentage|

| Grade | 192,092 | 191,429 | 63,238 |
|-------|---------|---------|--------|
| 7th grade | 99% | 91% | 1% |
| 8th grade | 99% | 91% | 1% |
| 9th grade | 1% | 12.2% | 1% |
| 10th grade | 1% | 87.3% | 6% |
| 11th grade | 1% | 44.8% | 1% |
| 12th grade | 1% | 48.7% | 1% |

| Sector (only vmbo pupils grades 9 and 10) | 192,092 | 191,429 | 63,238 |
|-------------------------------------------|---------|---------|--------|
| Technical sector | 6.6% | 1% | 1% |
| Agricultural sector | 4.0% | 1% | 1% |
| Economic sector | 6.0% | 1% | 1% |
| Healthcare sector | 7.5% | 1% | 1% |
| Combination | 6.8% | 1% | 1% |
| Not applicable | 69.1% | 99% | 99% |

| Specialization (only havo and vwo education grades 10th to 12th) | 192,092 | 191,429 | 63,238 |
|---------------------------------------------------------------|---------|---------|--------|
| NT (Science and technique) | 7.2% | 14.9% | 1% |
| NG (Science and health) | 9.9% | 21.5% | 1% |
| CM (Culture and society) | 3.6% | 9.3% | 1% |
| EM (Economics and society) | 13.0% | 30.5% | 1% |
| Combination NT and NG | 4.0% | 13.4% | 1% |
| Combination EM abd CM | 1.7% | 5.2% | 1% |
| Not applicable | 60.5% | 4.9% | 1% |

| Graduation result | 192,092 | 191,429 | 63,238 |
|-------------------|---------|---------|--------|
| Passed | 41.4% | 21.3% | 1% |
| Cum-laude | 0.0% | 0.0% | 1.1% |
| Failed | 2.0% | 1% | 1% |
| Not applicable | 56.6% | 75.1% | 1% |

| Exam grades | — | 86,279 | 49,533 |
| National exam grade | — | 6.49 | 6.48 |
| School exam grade | — | 6.48 | 6.57 |

Source: Netherlands Cohort Study on Education (NCO).

*For this variable, the NCO data set also includes the sector in which the pupil receives his/her diploma.

For this variable, the NCO data set also includes the profile in which the pupil receives his/her diploma.
about the specific study the student attends with a so-called study code (CREBO code). Based on this study code, ISCED-classifications and other study code related information can be matched. In the Netherlands, upper secondary vocational studies can offer two versions, either the ‘school’ track where students attend classes for 70% of their time and apply in practice what they have learned for the remainder 30%, or an apprenticeship track that offers ‘dual’ education. In this ‘dual’ track students work in an apprenticeship for about 80% of their time and attend classes for the remaining 20%. However, the majority (90.8%) attends the ‘school’ track, and 8% attend the ‘dual’ track. Furthermore, a small fraction (1.2%) consists of extraneous students, who pay an ‘exam fee’ but no tuition fee. They take the exams related to a study but do not attend classes.

Table 7 presents registration and enrolment into research universities. The table shows that in school year 2017/2018, 21,316 students enrolled into a university after graduating from secondary education in 2016/2017. Furthermore, the table shows the
sector of the study, of which there are slightly more categories compared to the other tertiary tracks: physics and law are added to the sector in which the student studies.

### Functionalities and Future Plans of NCO

#### Data Access Procedure

NRO and Statistics Netherlands carry out a specific data access procedure. Access to NCO data is only granted to researchers for scientific purposes; using NCO data for a commercial purpose is not permitted. Researchers are granted access to the NCO data set under several strict conditions that ensure compliance with national and international data protection regulations. The data access procedure consists of two parts.

First, the content of the research projects is assessed. To this end, researchers request permission to access the NCO data from NRO via an online application form. The application form requires researchers to provide information on the research project they plan to carry out with the NCO data. Applicants are asked to provide information on all researchers who are part of the research project. In addition, applicants are asked to provide a project title, goal, abstract, and future publications that are expected from the research project. Moreover, applicants should provide information regarding financing of the research project, possible conflicts of interests, and state whether an ethics committee has approved the research proposal. This information is used to determine whether the intended research project is in line with the objectives set by NRO with regard to NCO and whether the proposed study’s research questions can be answered with the NCO data.

Second, after a positive assessment of this application by NRO, researchers apply for access to microdata at Statistics Netherlands. In short, an application for microdata access at Statistics Netherlands is possible for researchers from organisations that are authorised to work with microdata. These authorised organisations include Dutch universities, institutes for scientific research, organizations for policy advice or policy analysis, statistical authorities in EU countries. If an organization is not authorised to work with microdata, a request for authorization can be submitted at Statistics Netherlands under several conditions. Such organisations should have a good name and reputation and have the primary mission to conduct statistical or scientific research. In addition, it is required to publish the results of the research. Foreign institutions applying for microdata access should preferably have working relations with a Dutch authorized institution. After receiving permission from Statistics Netherlands, researchers can access the NCO data through a secure remote access environment that has a double identification procedure with both a token and a password. The remote access environment is a secured working environment in which researchers can store draft versions, syntaxes, and output.

#### Data Tool

The NCO data set comprises of a vast amount of separate data files, with content as described above. All data files are separated by year and register source. Therefore, the NCO data infrastructure is quite fragmented. When data access is obtained, researchers gain access to the NCO data set through a data tool that has been developed to help researchers compose a workable data set from all data files underlying the NCO data set. This tool provides a user-friendly interface through

### Table 8. Descriptive measurements of higher education (universities of applied sciences) characteristics in 2016/2017 for the primary education exit cohort 2010/2011

| Specialization of study | N    | Percentage |
|-------------------------|------|------------|
| Education               | 2,136| 10.2       |
| Agriculture             | 730  | 3.5        |
| Technique               | 5,698| 27.3       |
| Health care             | 2,440| 11.7       |
| Economics               | 7,403| 35.4       |
| Social sciences         | 2,295| 11.0       |
| Other                   | 200  | 1.0        |

Source: Netherlands Cohort Study on Education (NCO).

### Table 9. Descriptive measurements of higher education (universities) characteristics in 2017/2018 for the primary education exit cohort 2010/2011

| Specialization of study | N    | Percentage |
|-------------------------|------|------------|
| Combination of sectors  | 1,040| 4.9        |
| Agriculture             | 942  | 4.4        |
| Physics                 | 2,531| 11.9       |
| Technique               | 3,912| 18.4       |
| Health care             | 3,040| 14.3       |
| Economics               | 3,442| 16.1       |
| Law                     | 1,926| 9.0        |
| Social sciences         | 2,985| 14.0       |
| Language and culture studies | 1,498| 7.0     |

Source: Netherlands Cohort Study on Education (NCO).
which it is possible to select cohorts and variables that are relevant for specific research purposes. After this selection, the tool creates a customised data set according to the researcher’s preferences, which is stored in Statistics Netherlands’ secured remote access environment.

**Linking Other Data to NCO**

There are several possibilities for linking other data to the NCO core data set. For instance, at additional costs, it is possible to link register data that is available through Statistics Netherlands to the NCO core data set. In addition, under several conditions it is possible to upload your own data sets and link these with NCO data through the Statistics Netherlands secured remote access environment. For this, it is necessary that the acquisition and the delivery of the data sets researchers wish to link to NCO are legitimate. Statistics Netherlands carries out anonymization of uploaded data sets and creates the link with the NCO data.

**Modules and Future Data Collection**

In addition to the possibility to link other data sets to the NCO core data set, additional data is collected based on so-called thematic ‘NCO modules’. This data collection only concerns a (representative) part of the student population. In these modules, the focus is on collecting information on social relevant issues that cannot be collected via existing data sources, thereby enriching the core NCO data set. The data generated from these modules is made available to the scientific community soon after the module project has ended. NCO’s first module study started in 2019 and is called PRIMS. In this module researchers investigate how relationships with peers in primary education influence school choice, school performance and social integration in secondary school, and what this means for segregation and (in)equality in education.

Furthermore, more variables will be added to the NCO data set in the coming years. These variables are collected from existing registers but also from school administrative systems and surveys. For instance, a current pilot aims to investigate how information from primary schools’ administrative systems on standardized tests scores can be used to provide insights into schools’ registration practices of early childhood education and the learning curve that pupils experience in reading and math during primary school. When successful, NCO aims to gather information from school administrative systems of all primary schools in the Netherlands to complement the NCO core data set.

**Concluding Remarks**

In 2016, NCO was initiated. NCO systematically combines existing register data from the Netherlands and integrates them with relevant other data sources. The basis of NCO is formed by register data from Statistics Netherlands, which overcomes some of the problems that large-scale educational research have faced recently, such as low response rates, low data quality, and lack of statistical power. By hosting these data sets at Statistics Netherlands, an entrusted third party, the data can be linked to all other data, and the data are available to all researchers, while at the same time ensuring security and confidentiality.

The NCO data set is set up to be valuable for both researchers and schools. The major added value of the NCO data set for researchers is longitudinal individual pupil data from school administrative systems which can be added to other data sets such as (international) surveys or other register data. This allows us to overcome a major challenge of survey data, which does not contain interesting information such as longitudinal performance data. In addition, by linking survey data from large national and international survey data collections (e.g. PIRLS, TIMMS, PISA), new promising possibilities for data analysis arise. However, the NCO data does currently not (yet) contain information on for example social-emotional skills or on pupils that require specific (health and pysosocial) support, which is information that was available in the previous cohort data sets that were based on survey data.

Furthermore, the advantages are large for researchers, but not directly for schools. To create a win–win situation where schools also profit, NCO provides annual reports to all primary and secondary schools in the Netherlands. These reports provide information on their students’ performance and educational careers, both during school as well as in further education. A major asset of these reports is that they provide information on how comparable schools (in terms of student population) perform (the so-called reference value), which is an important reference point for schools. Schools use the reports to improve their educational practices, to base their policy decisions on and to check whether they have reached their goals. This is specifically important since the intuition and assumptions of school leaders and teachers about the consequences of educational policies and practices do not always appear to be correct (Baltussen et al., 2019; Vanlommel et al., 2017). Schools also wish to use data, for instance on tests or student background, to improve their education (e.g. Schildkamp and Kuiper, 2010). By providing annual
reports to all schools, we hope to establish a mutual commitment so that schools will participate in research in the future, which should boost response rates.

In addition to providing annual reports to schools, NCO aims to co-ordinate (inter-)national surveys taking place in primary and secondary education in order to decrease the time pressure schools experience for taking part in research. Given the current and future flow of national and international surveys, we estimate that primary schools need to participate in a research project once every 8 years and secondary education schools once every 3 years. Once the coordination is in place, schools experience less pressure and data collection is dispersed more evenly over all schools in the Netherlands. Additionally, since all NCO data can be accessed by researchers, it becomes redundant for researchers to ask already available questions if these can be retrieved from the existing NCO data set. This makes surveys conducted in the future more efficient and the collection of new data at schools less necessary.

To conclude, NCO so far covers more than 10 complete cohorts and is complemented with the newest data every year, and maps pupils pathways and performances throughout the entire Dutch education system. The use of the Dutch register data on education is not new, yet it has never been structurally matched to student's background information, or their parents. A specific added value of NCO is to enrich the core data set with modules and other (inter)national surveys. Finally, a major contribution of the NCO is that data from school administrative systems on the development of students' performance in key domains will be linked to the core data set, thereby creating a unique data set that enriches not only research, but also policy and practice.

Funding
This work was supported by the Netherlands Initiative for Education Research (NRO).

Notes
1 There are some exceptions, such as the German NEPS data (https://www.neps-data.de/).
2 NRO oversees most of the budget for educational research in the Netherlands, and is part of the Netherlands Organisation for Scientific Research.
3 Link to website (in Dutch only): www.nationaalcohortonderzoek.nl
4 For a more elaborate explanation on the Dutch education system, see for instance https://www.nufic.nl/en/subjects/education-in-the-netherlands/
5 About 1% of the primary schools are private schools and about 2% in secondary education are private schools (Dutch Inspectorate of Education, 2019).
6 New test suppliers are: Route 8, IEP, and a few years later AMN, CESAN, DIA.
7 For more information about specific conditions that need to be met in order to receive a secondary education diploma, for example see: https://www.government.nl/topics/secondary-education/secondary-school-leaving-examination, [last accessed on May 29, 2020].
8 The registers also include information on adult education and special education, but these are only limitedly included in NCO.
9 Due to privacy regulations, school identifiers are recoded within SSD. This makes it possible to control for school differences but impossible to identify specific schools directly.
10 Except information on country of birth and gender which are not yearly updated.
11 Presenting descriptive statistics for all possible years would not only lead to very large tables and too much detailed information but it would also have little added value. Especially as the statistics for the chosen year are similar to and representative for all other years. This also holds true for the chosen primary education exit cohort 2010/2011 as a whole: the statistics for this specific cohort are comparable to those of all other primary education exit cohorts.
12 An example of an outlier in the data set is a child who is 5 years old in the last year of primary education.
13 Dutch abbreviation for ‘Centraal Instituut voor Toets Ontwikkeling’. In 2014/2015, Cito changed the name of the test to ‘Centrale Eindtoets’ abbreviated with CET.
14 For pupils of the primary education exit cohort (2010/2011) it was not mandatory to take tests. After 2014/2015, a new law passed in the Netherlands that made it mandatory to take standardized tests at the end of primary education for all pupils.
15 Whereas repeating classes in secondary education is quite common, skipping classes is very uncommon.
16 These averages most probably reflect pupils in vmbo-tracks, education tracks that last 4 years to complete.
17 These averages mostly reflect pupils in the vwo-track, and a small portion of retained havo.
pupils as this is 6 years after leaving primary education.

18 This online form is available at the NCO website, see: https://nationaalcohortonderzoek.nl/toegang-aanvragen-voor-onderzoek-met-nco-data/

19 NRO and Statistics Netherlands are proponents of open access publication, and publications based on data access via Statistics Netherlands are expected to be made fully, directly and in principle free of charge available to interested parties.

20 Microdata is the name for the service Statistics Netherlands offers to researchers for working with their register data.

21 More information regarding microdata access procedures and forms, is available at the website of Statistics Netherlands: https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research/applying-for-access-to-microdata, [last accessed on May 29, 2020].

22 Procedures and forms to apply for authorisation of organisations are available at the website of Statistics Netherlands: https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research/ importing-your-own-datasets, [last accessed on May 29, 2020].

23 Information on the procedures for adding your own data to NCO via the Statistics Netherlands secured remote access environment is available at the website of Statistics Netherlands, see: https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research/importing-your-own-datasets, [last accessed on May 29, 2020].

References

Baltussen, M. et al. (2019). De Staat Van Het Onderwijs 2019. Hoofdlijnen. Den Haag: Inspectie van het Onderwijs.

Bongers, F., Jager, C. and te Velde, R. (2015). Big Data in Education and Science: Inventarisation and Essays (Report No. 2014.086-1307) [Big data in onderwijs en wetenschap: Inventarisatie en essays]. Retrieved from https://www.researchgate.net/publication/301524152_Big_Data_in_onderwijs_en_ wetenschap_Inventarisatie_en_essays [last accessed 29 May 2020].

Card, D., Chetty, R., Feldstein, M. and Saez, E. (2010). Expanding access to administrative data for research in the United States. American Economic Association, Ten Years and Beyond: Economists Answer NSF’s Call for Long-Term Research Agendas. Retrieved from https://eml.berkeley.edu/ –saez/card-chetty-feldstein-saezNSF10dataaccess.pdf [Last retrieved 29 May 2020].

Dutch Inspectorate of Education (2019). Rapporten particuliere scholen (B3) basis onderwijs en voortgezet onderwijs, available from the Website of the Dutch Inspectorate of Education: <https://www.onderwijsinspectie.nl/onderwijsectoren/particulier-onderwijs/rapporten-particuliere-scholen-vo> and <https://www.onderwijsinspectie.nl/onderwijsectoren/particulier-onderwijs/rapporten-particuliere-scholen-po>, [last retrieved 29 May 2020].

Einav, L. and Levin, J. D. (2013). The Data Revolution and Economic Analysis. NBER Working Paper No. 19035.

Figlio, D., Karbownik, K. and Salvanes, K. (2017). The promise of administrative data in education research. Education Finance and Policy, 12, 129–136.

González-Sancho, C. and Vincent-Lancrin, S. (2016). Transforming education by using a new generation of information systems. Policy Futures in Education, 14, 741–758.

Hey, T., Tansley, S. and Tolle, K. (2009). The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond: Microsoft Research.

Murnane, R. J. and Willett, J. B. (2010). Methods Matter: Improving Causal Inference in Educational and Social Science Research. New York: Oxford University Press.

Røed, K. and Raaum, O. (2003). Administrative registers—unexplored reservoirs of scientific knowledge? The Economic Journal, 113, 258–281.

Schildkamp, K. and Kuiper, W. (2010). Data-informed curriculum reform: which data, what purposes, and promoting and hindering factors. Teaching and Teacher Education, 26, 482–496.

Van der Velden, R. K. W. and Van der Maas, H. (2015). Naar een toekomstbestendig cohortonderzoek. (ROA Reports; No. 482–496. Maastricht: Research Centre for Education and the Labour Market.

Vanlommel, K. et al. (2017). Teachers’ decision-making: data based or intuition driven? International Journal of Educational Research, 83, 75–83.

Dr. Carla Haelermans is Associate Professor in Education economics at the Research Centre for Education and the Labor Market (ROA), School of Business and Economics, Maastricht University, and one of the national coordinators of the Netherlands Cohort Study on Education for the Netherlands Initiative for Education Research (NRO; part of the Netherlands Organization for Scientific Research NWO). Her main research interests are in education economics and labour economics and inequality. She has done quite some work on the effectiveness of ICT-tools in education, parental involvement in education, efficiency of education, and the effectiveness of education and labour market policies.
Timo Huijgen is a statistical analyst at the Research Centre for Education and the Labor Market (ROA), Maastricht University. His activities mainly focus on research projects based on large national and international (register-) data sets.

Madelon Jacobs is a researcher at the Research Centre for Education and the Labor Market (ROA), Maastricht University. She is interested in research related to work-life balance, gender inequalities, social inequalities in education.

Prof dr. Mark Levels is professor at Maastricht University, research program director at Maastricht University’s Research Centre for Education and the Labor Market (ROA), and fellow of the WZB. He studies the extent to which laws, institutions, and policies affect individual decision-making. He is also a program coordinator of the Technequality project.

Prof. Dr. Rolf van der Velden (1955) is professor at Maastricht University and director of the Research Centre for Education and the Labour Market (ROA). He supervised several (inter)national studies on the transition from school to work. Currently he is coordinator of the Netherlands Cohort Study on Education and one of the coordinators in the PIAAC project. In 1983 he finished his study sociology at the University of Groningen. Current research focuses on education and skills development, transition from education to work, skills mismatches and the acquisition and decline of skills over the life course.

Lynn van Vugt is a PhD candidate at the Research Centre for Education and the Labor Market (ROA), Maastricht University. Her research interests include the school-to-work transition, social inequalities in education and on the labor market, and cross-national research.

Dr. Sanne van Wetten is a postdoctoral researcher at the Research Centre for Education and the Labor Market (ROA), Maastricht University. She is a member of the Netherlands Cohort Study on Education (NCO) project, for which she mainly supports the coordination of thematic research modules. Her research interests include social inequalities in education and school to work transitions, effectiveness of education and skill development.