Christophe Kolodziejczyk, Torben Pilegaard Jensen and Britt Østergaard Larsen

Measuring School Performance by Student Retention

Danish Institute
for Local and Regional
Government Research

The publication *Measuring School Performance by Student Retention* is available at www.kora.dk

Publisher: KORA
ISBN: 978-87-7488-779-9
Project 10267
August 2013

KORA
Danish Institute for Local and Regional Government Research

KORA is an independent governmental institution, the purpose of which is to improve administrative praxis in public bodies and provide them with the best possible decision-making tools.

Danish Institute for Local and Regional Government Research

Preface

In this paper we review the use of two different criteria to evaluate the performance of educational institutions in terms of student retention. The paper describes one application for each criterion. Each application is based on Danish data extracted from administrative registers.

The first criterion evaluates the performance of vocational training establishments and takes drop-out rates as an indicator of school performance. The second criterion is the average time elapsed before dropping out and is applied to evaluate professional bachelor courses. Based on each criterion we estimate the effectiveness of the institution, correcting for factors beyond its control.

Correcting for covariates linked to social background, previous educational record and local labour-market conditions, we find that such correction has a high impact on the ranking of vocational training schools, but a considerably lower one for professional bachelor courses.

Torben Pilegaaard Jensen
August 2013

Contents

1 Introduction

High dropout rates from the education system are of huge concern. A large fraction of young people in western countries starts on a course of study, but does not complete it and drop out of the education system without any qualifications. This represents a loss for society in terms of potential human capital and higher productivity. It also makes it more difficult for dropouts to find a job because of the negative signal this might send to future employers. According to Rumberger (1987), dropping out has negative individual and social consequences, like low-paying jobs, and difficulty finding a steady job. As a consequence it lowers tax revenue for governments. It can also have consequences for health in general and mental health in particular by lowering the self-esteem. Because dropping out can have such negative consequences it seems appropriate to take this criterion as a measure of school performance.

Because high dropout rates result in a waste of human capital and potential loss of productive labour, it is desirable to reduce the number of people who do not complete their studies. In order to increase the number of highly-qualified citizens, it is important to know what actions by educational institutions affect students' risk of dropping out. A first step in answering this question is to have an indicator of what a well-performing institution is.

Dropout rates can be affected at the personal level by individual characteristics, endowment effects like social and family background, local labour-market conditions and by school characteristics. Governments can try to reduce dropout rates either by trying to reduce social inequality or by devoting more resources to people with poorer endowments. Or they can try to invest more in the quality of their schools in order to lower dropout rates. In this case how can we measure school performance in terms of student retention? A natural solution would be to use raw dropout rates as an indicator. However, as we mentioned before, dropout rates at school level are influenced by individual characteristics such as social background. It is therefore not enough to look at raw dropout rates to gauge the performance of schools and compare them to each other. As a corollary we will measure dropout rates corrected for factors independent of the quality of the school, or at least factors beyond the control of the school. These factors include family background, previous academic achievement, student composition and local labour-market conditions.

The Danish government has as a goal that 95% of an age cohort should complete their higher secondary education (British: sixth form) and that 60% of an age cohort should complete a course of higher education. Approximately, 95% of a cohort starts on a secondary higher education, but only 80% completes it. According to OECD (2010) only 83% of the age cohort which started an upper secondary education in 2008 is expected to complete. This is substantially below that of other countries to which Denmark is often compared, such as Germany (97%), Finland (93%) or Norway (91%). With respect to the goal that 50% of an age cohort should have some form of higher education, the goal is considerably closer as 49.4% are expected to achieve this, see Undervisningsministeriet (2011). Substantial dropout rates are also observed from higher education institutes in general and from professional bachelor courses in particular. As a consequence, reducing dropout rates is seen as a desirable social goal. In this paper we propose two methods of measuring school performance in terms of student retention.

The objective of this paper is to discuss two methods allowing us to build a performance indicator of student retention at institutional level. Instead of looking at general measures of variation of the institution effects as Rumberger & Palardy (2004) does, we are interested in using our indicator to rank the institutions. This ranking can then, with the help of qualitative methods, be used to investigate specific institutions which have been shown to be good or bad performers. In the previous literature, the objects of investigation have generally been primary or secondary schools and the parameter considered grades or dropout rates (Rumberger & Palardy 2004 & 2005). In this paper we use either dropout rates or mean time before dropout as measures of a school's performance. We model the probability of dropping out at the individual level. From these estimates we build school performance indicators measured by student retention by comparing observed dropout rates at institutional level to the predictions of the model. The method can also be used to investigate the impact of external factors on dropout rates and the degree of correction shown by our indicators. Our definition of dropout is less restrictive than that used previously. Dropout rates generally only include people who disappear from the education system, whereas in this paper we consider people who take a break from their studies as dropouts.

The method presented in this paper is illustrated by two examples taken from the Danish educational system. Usually the literature has focused on upper secondary schools (sixth-form colleges). Here we propose two applications which look at vocational training courses and professional bachelor courses. These two types of study are important for the Danish economy, since students completing these courses will enter the labour market afterwards. Danish educational institutions have much discretion on how to organise their courses. For these two applications we use data from Danish administrative registers, where we look at (a) the propensity of students from a particular school continuing their education by enrolling in a vocational training course and (b) the average length of time elapsing before dropping out for students on a professional bachelor course. We find that external factors have an impact on students on vocational training courses and we obtain a substantial correction. For professional bachelor courses these factors have much less impact and the degree of correction is much less pronounced.

The paper is organised as follows. In the next section we discuss the rationale behind focusing on dropout rates as a measure of school performance. In section 3, we discuss how to measure performance and how to correct for factors related to external factors. In section 4, we present two types of application which use these methods. The two applications use data on Danish students on vocational training and professional bachelor courses, respectively, which results are presented in section 5. We provide some concluding comments in section 6.

2 Why focus on dropout rates

A primary goal for governments is to evaluate the performance of the educational system. The government can improve the educational system for the most part by specifically targeting the school level. It is difficult for a government to influence short-term socio-economic factors to improve the performance of the education system, however desirable it may be in a long-term perspective. A well-performing system at primary and lower secondary level is also a condition for good performance at higher levels of the system. If we want to understand which factors influence the education system at school level, a first step is to evaluate school performance. In doing so we have, however, to take into account factors which affect the probability of students' gaining a degree, but which have nothing to do with schools' specific efforts to help their students complete their courses. We will call them external factors, meaning that these factors are external to and beyond the schools' control. These factors can be individual characteristics, such as family background, composition of the school (e.g. the proportion of students from immigrant backgrounds) or variables in local labour-market conditions.

Many of the previous studies of school performance have focused on test scores (Rumberger & Palardy 2005; Rumberger & Thomas 2000). As argued by Rumberger & Thomas (2000), this type of performance indicator is problematic because it suffers from selection effects and can be manipulated by the schools themselves to show that their test scores are improving. Grades are typically observed only for those who pass an exam, and focusing on the population of people who have completed their studies will obviously result in a sampling problem. Since people who drop out will never be recorded, schools which are better at retaining their students will on average have lower grades, where highly selective schools will have higher average grades. These schools have, on average, better students due to selecting students with, probably, more favourable unobserved characteristics and not entirely due to a better quality of teaching (see Jensen, Larsen & Rangvid 2010). We assume here that on average academically poor students are more likely to drop out.

There are two reasons why we are interested in how good schools are at retaining their students instead of course completion rates. Firstly, the use of course completion rates as a success criterion rules out right-censored data. So if we consider only completion rates, we may be in the position where the sample period is too short to observe anyone completing the course. In the case where we are able to observe individuals who complete the course, we may still experience data limitations due to the right-censoring of the data. We will observe some students who drop out, but an important part of them can be observed still studying. In order to consider individuals who either drop out or complete the course, we will be forced to restrict the sample period. Therefore, these limitations can either result in a sample which is too small or too old data to give a ranking reflecting the present reality. Looking at the retention of the students and modelling the lengths of the course before dropout will solve these issues. Secondly, there is a positive correlation between the time a student has spent on his course and the probability he will complete it. If the course is split into two parts, as is the case with vocational training in Denmark, it might be useful to look at the first part of the course if we are worried about the current relevance of the data. In this case we can also use duration analysis to incorporate as many data as we can and deal with right-censoring. In order to investigate dropout rates we can think of two types of indicator. The first one is an indicator of whether students have completed their courses within a time considered normal for such a course. This will typically lead to a binary outcome, to which we can assign the value 1 if the student has dropped out and 0 otherwise (completed or still under study). But dropout rates can also be measured with the help of duration models. We can estimate the expected survival time at the individual level. Since we are interested in measuring the risk of dropping out, duration models can help us to see the flow of people who drop out from their studies at any given point in time. This is interesting because we can give a more detailed picture of when students are most likely to drop out. In section 3, we show that discrete outcome models and duration models are two tools which can be used in this context.

3 Method: How to measure schools' student retention ability

In this section we discuss how we can measure schools' success at retaining their students.

3.1 How to rank institutions in relation to dropouts

Our method consists of two steps. In the first step, we estimate the parameters of a statistical model describing the outcome at the individual level. In our case it will be a measure of dropping out at the individual level. Then in the second step we predict the outcome for each individual and take the average of these predicted outcomes at school level. These predictions are then compared to the observed outcomes at school level. The difference between the two constitutes our indicator of school performance. This residual method is not new. It has been used to evaluate the performance of hospitals, production sites within a firm, schools and integration of immigrants (Husted, Heinesen & Andersen 2009; Andersen & Heinesen 2009). Previous literature on the evaluation of schools has used indicators related to test scores, which is a continuous outcome. The interested reader is referred to Raudenbusch & Willms (1995) for a discussion of how to estimate school efficiency where the outcome is a continuous dependent variable. In our examples we have to deal with limited-dependent variables. In the first case the outcome is discrete and indicates whether the individual has dropped out after a predefined amount of time. A second outcome is the duration (time elapsed before dropping out) or survival time. We have to take into account that durations can be censored variables.

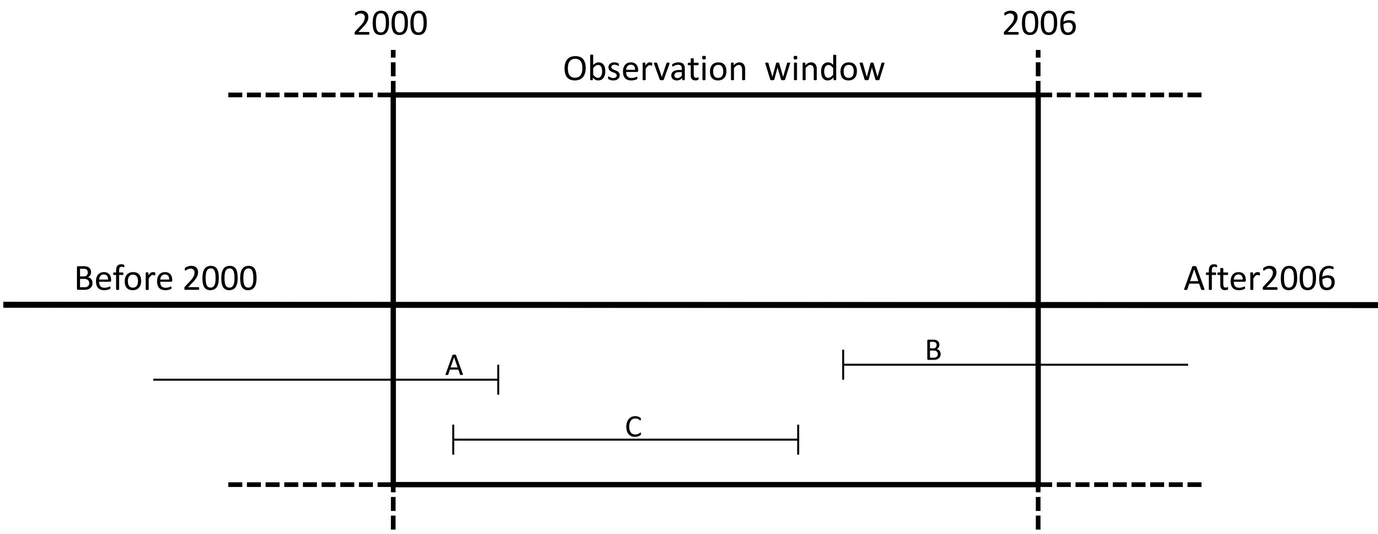
As emphasised in Andersen & Heinesen (2009), we are not trying to estimate causal effects but rather effects corrected for the individual school. Note that there will be a bias if there is a correlation between practice and local external factors (Andersen & Heinesen 2009). There may be a correlation between practice and external factors both at individual and institutional level. The latter is most likely to cause serious problems. Therefore, we need data at the individual level in order to correct for relevant external factors. In assessing the performance of educational institutions it is preferable to use a corrected measure of drop-out rates even if some explanatory factors are missing in the model. It will eventually reduce the potential bias.

In order to evaluate institution performance we have to take into account that personal factors and the composition of the student body affect the likelihood of dropping out and are beyond the control of the educational institutions. The factors included in our models have been used in previous research on school effectiveness. One can refer to Rumberger (1983 & 1987) and Rumberger & Palardy (2004) for a discussion of these factors. These factors include parental education, income, ethnicity and previous educational records such as grades in secondary or high school[1]. We also include information about the student composition of the institutions.

3.2 Dropout rates: methodological considerations in terms of sample selection

Studying is a dynamic process and obviously until students complete their course they are at risk of dropping out. These characteristics of the process have methodological implications for the selection of the sample. Figure 3.1 illustrates the specificity of the data related to the measurement of dropout rates and student retention. Suppose we have data about students' course enrolment and that the window covers the years 2000 to 2006. In case A, we have the example of a student who begins before the year 2000 and completes her course or drops out before 2006. This instance is left-truncated. In example B, the student begins after 2000 and ends after 2006. This instance is right-censored. Finally, in example C our hypothetical student starts after 2000 and ends before 2006. If example C is not problematic, examples A and B pose the problem that we do not observe the actual length of the study period. For example B the observed duration is less than the true duration. For example A the observed duration is also less than the actual duration, but we have no information about what happened when the individual started the course. A fourth case which is not illustrated in the figure is when subjects start and end their spell outside the observation window.

Figure 3.1 Data window and study duration



At the same time it is important when studying dropout rates to have a sufficiently long observation window and the latest data, because of possible changes in school structure. These two objectives are in conflict with each other. Figure 3.1 suggests applying survival analysis, but it is not always possible to apply such models in our context, simply because

[1] High school is the equivalent to grammar school in the UK.

data in the form of sufficiently precise durations are not available. If the interest lies in the final dropout rates, i.e. how many people drop out during the normal time for completing their courses, it is important to ensure that most of the students observed during the sample period had the possibility to either complete their course or drop out. In other words, we have to limit as much as possible the number of people whose duration is right-censored. This will limit the sample considerably since we will have to study individuals who started during a limited period of time depending on how long it takes to complete a course of study. If it takes for example two years to complete a given course, this will imply that we have to consider individuals who started no later than 2004.

If we have duration data and we are interested in the flow of people who drop out at each point, then the problem of right-censoring is less problematic and is well handled by duration models. Left-truncation is more problematic because we do not have any information of what happened before the entry in the study. Therefore, it has been decided in the empirical example to consider people who have started their studies during the period of observation. We have also restricted the starting date in order to reduce right-censoring.

3.3 Performance in terms of dropout rates

Here we discuss how to measure schools' performance, when we only observe whether an individual drops out. We observe for each individual *i*, the outcome y_{ik} which is a discrete variable and which takes value 1 if student *i* in institution *k* has dropped out and 0 otherwise. We assume that there exists a latent variable y_{ik}^* which measures the propensity for the individual to leave the course. We assume that y_{ik}^* is influenced by a set of observed characteristics x_{ik} consisting of variables not (directly) influenced by schools, such as social background. The parameter vector β measures the influence of x_{ik} on the latent variable y_{ik}^*. The outcome is also influenced by u_k which is an unobserved component at the school level. Finally, v_{ik} is an error-term which measures idiosyncratic variation at the individual level. This term is assumed to be of zero mean and uncorrelated with both x_i and u_k.

$$y_{ik}^* = x_{ik}\beta + u_k + v_{ik}$$

We do not observe y_{ik}^* but only y_{ik}. We assume that if y_{ik}^* is positive, then the subject drops out, that is

$$y_{ik} = 1(y_{ik}^* > 0).$$

By assuming a normal or a logistic distribution for v_{ik}, we obtain the probit and logit models, respectively. These models can be estimated by traditional maximum likelihood methods.

The computation method for the indicator consists of estimating β from the model presented previously and predicting the probability of the outcome at the individual level. Then the averages of these predictions are computed at school level and compared to the observed percentages of positive outcomes at school level. The difference between these two figures gives us a performance indicator cleansed of factors beyond the control of the institution. Note that the model is assumed additively separable in v_i and u_k and that the terms v_i and u_k have to be independent of x_i and have a mean of zero. Formally, the indicator is defined by

$$I_k = \bar{p}_k - \bar{\hat{p}}_k,$$

where $\bar{p}_k$ is the share of students in school k who have dropped out and $\bar{\hat{p}}_k$ is the average of individual predicted dropping out in school k. This will be interpreted as the difference between the observed dropout rates and the expected dropout rate given the characteristics of students in school k.

As discussed in the previous section, the use of this method has some implications in the design of the sample period. As we want to limit right-censoring as much as possible, we have to consider a sample period where all the students had a chance to complete their studies.

3.4 Performance in terms of expected mean duration before dropping out

In this section we discuss how to construct a performance indicator for duration data. Our focus will be on the average length of time before a student drops out of his course at a given institution. Obviously, this is influenced both by how many students drop out and when they drop out. It is our conviction that the later the students drop out the better. Rumberger (1987) and Rumberger & Palardy (2005) refer to the concept of "Holding power"[2], which has been used in the literature. Indeed, schools which are able to retain their students longer must also be the ones with lower dropout rates. It is also likely that dropping out later and the final dropout rate are (negatively) related.

Our goal is to compare expected survival time across the different institutions. As in the previous model, we want to correct for factors which influence expected survival time but are not controlled by the institution. In order to do so, we estimate a duration model. The choice of a duration model is motivated by the choice of the indicator, which is the mean duration before dropping out. Another argument for choosing a duration model resides in the nature of the course. This is a dynamic process where people are at risk of dropping out during the time they are at school/college and therefore fit in well in the duration models/survival analysis framework.

Once our duration model is estimated we can compute our indicator which consists in subtracting predicted mean durations at school level from observed durations. The problem in this case is that we do not observe true durations due to right-censoring. One advantage of using a duration model is that we can model right-censoring and in this way, contrary to the model in the previous section, extend our sample period and include subjects which are right-censored at the end of the period. This method has been applied when evaluating

[2] See also Rumberger & Thomas (2000).

Danish municipalities' performance in terms of integration of immigrants on the labour market (Husted, Heinesen & Andersen 2009). In this study the authors take as a criterion for successful integration on the labour market the time taken, before individuals have their first employment of at least six months. The rest of this section closely follows the methodology used in Husted, Heinesen & Andersen (2009).

Since expected survival time is intimately related to the form of the hazard rate (see below) our starting point will be the modelling of the hazard rate of dropping out. The hazard rate is defined as the probability of leaving the initial state in a short interval of time given survival up to time t. The hazard rate is defined by

$$\lambda(t) = \lim_{h \to 0} \frac{P(t \leq T < t + h | T \geq t)}{h}$$

The term T denotes duration time. In our case this represents the probability of an individual dropping out in a short interval of time given that the student has been studying until T. One can also define the hazard rate conditional on covariates, i.e.

$$\lambda(t|x;\beta) = \lambda_0(t) exp(x\beta).$$

Here we impose a proportionality assumption. It means that the effect of a covariate is proportional to $\lambda_0(t)$ which represents the effect of duration dependency. The term $\lambda_0(t)$ is called the baseline hazard and measures the hazard rate in relation to time.

With the help of the previous model we compute the expected mean duration before dropping out at the individual level. The mean duration μ is simply the integral of the survival function S and is given by

$$\mu = \int_0^\infty S(t|\mathrm{x};\beta)dt = \int_0^\infty \exp\left(-\int_0^t \lambda(\mathrm{s}|\mathrm{x};\beta)\mathrm{ds}\right) dt.$$

The indicator for school performance in relation to student retention is computed as follows. After estimating the model we obtain the expected mean duration for each individual. This expected mean duration is averaged out over the individuals belonging to the same school/institution. This is the expected mean duration given the composition of covariates in the school. Each average expected mean duration at school level is compared to the unconditional mean duration estimated non-parametrically. The unconditional mean duration is computed by integrating the survival curve. The unconditional survival function is estimated by the Kaplan-Meier estimator. The difference between these two quantities constitutes our indicator. In this way we are able to compare the expected mean duration conditional on factors independent of institutions to the mean duration observed. If a specific school has a higher unconditional mean duration than that predicted for it, then we might conclude that this school retains its students longer than expected given the observed characteristics of its students.

By integrating the survival curve up to infinity, our estimate of the mean duration will depend on points which are extrapolations of the survival curve for durations not existent for some institutions and are not precisely estimated. Therefore, we compute mean durations up to a predetermined point. Another argument in favour of using restricted means is that we simply want to compute a mean duration over a specific period which represents the normal time for completing the training. If we denote as t_{max} the maximum time duration one can observe, we can compute restricted mean durations. The expected duration between 0 and t_{max} is equal to

$$E_{t_{max}}(T|x) = \int_0^{t_{max}} exp\left[-\int_0^t \lambda(\mathrm{s}|\mathrm{x};\beta)\mathrm{ds}\right] dt \ .$$

The Kaplan-Meier estimator is equal to:

$$E_{KM,t_{max}}(T) = \sum_{j:t_j \leq t_{max}} \hat{S}_{KM}(t_j),$$

where $\hat{S}_{KM}(t_j)$ is the surface under the different parts of the Kaplan-Meier survival function, where the hazard rate is constant. The difference

$$R_{t_{max}} = E_{KM,t_{max}}(T) - E_{t_{max}}(T|x)$$

is our indicator or the “residual”, which gives the average duration in a specific institution over and above the expected duration estimated from the model. For the interested reader more details on this method are given in Husted, Heinesen & Andersen (2009), in particular on how to perform statistical inference.

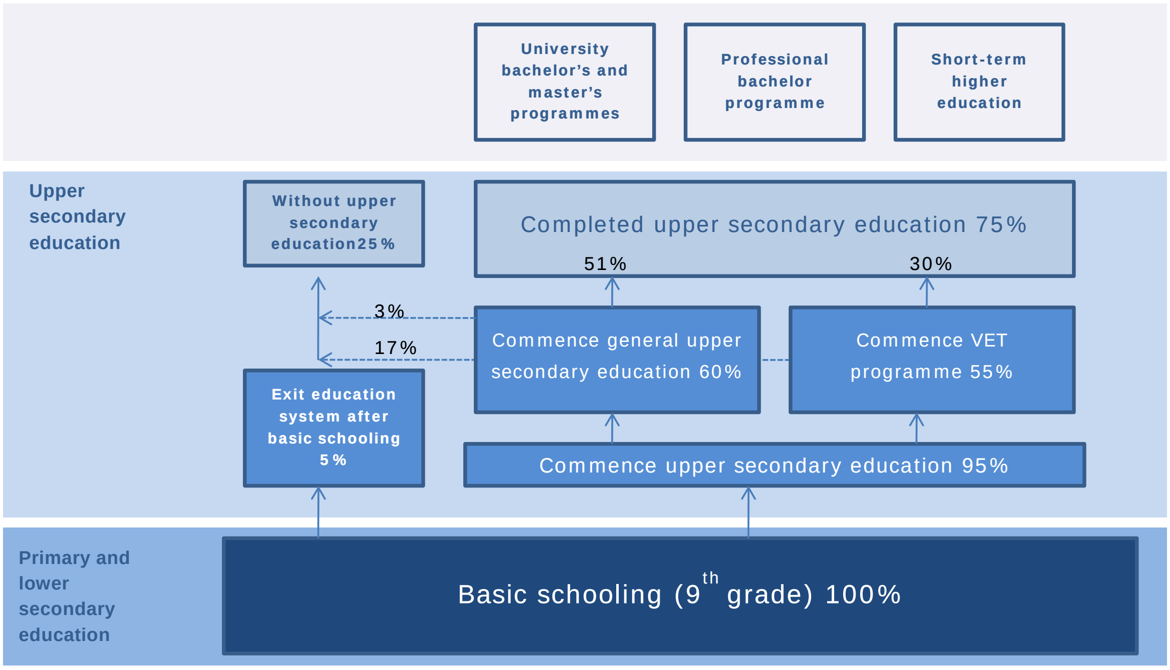
4 Data

In this section we present the results of two empirical applications to illustrate the methods described earlier. Each application is related to one of the models presented in section 3. The first application is related to vocational training in Denmark, whereas the second application is related to professional bachelor courses. First we discuss the Danish educational system. Then we discuss some issues about the source of the data used for our applications. Finally, we present the results of our analyses in section 5.

4.1 The Danish educational system

Figure 4.1 gives an overview of the Danish educational system and the different flows of students at the upper secondary stage. After completing secondary education (9^{th} grade), students can either choose an upper secondary education as an academic (corresponding to a British sixth-form college) or a Vocational Education and Training (VET). Around 95% of those finishing lower secondary school will start on an upper secondary education, but only 75% will complete their studies.

Figure 4.1 Flows in the Danish education system from basic schooling to youth education



Vocational Education and Training in Denmark is a dual training system based on interaction between school-based study and work-based training. The work-based training takes place in companies that have entered into a contract to train an apprentice. VET qualifies students for a skilled job when they enter the labour market. As shown in figure 4.1, around 30% of people who have finished lower secondary school complete Vocational Education Training. The dropout rate is around 54%, since around 55% of people who finished secondary school start Vocational Education and Training. Note that 17% of students who have started a Vocational Education and Training will drop out of school without having completed their course. A description of the organisation of the Danish educational system in general and the Vocational Education and Training system in particular can be found in Colding (2006a; 2006b).

Professional bachelor courses represent approximately one third of the students in higher education. Professional bachelor degrees cover a large number of professions. The most common professions are engineering, nursing and teaching. All these professions require a professional bachelor degree. Professional bachelor degrees are offered by a large number of Danish higher education institutions as a career-oriented professional educational programme.

4.2 Data

The data used in the applications come from Danish administrative registers. In Denmark it is possible to link an individual's data across the different administrative registers by using a Civil Registration Number. A unique number is given to every person in Denmark and is used by administrative authorities to register information about citizens' contacts with the State. It includes for example the population registry, the education system, the health-care system, the tax authorities or income benefits. Information on school records can therefore be linked to information from other registers. With this number it is also possible to link the data of children to information about their parents.

In the following analysis it has been possible to control for students' individual characteristics, such as ethnicity, age and gender, for social background: parents' education and income, parents' labour-market status, type of family etc. We have also been able to include information on past educational records. In particular, we can observe grades obtained at the previous stage of the education ladder before the student enrolled in a specific course of study. We have been able to control for past attempts to obtain vocational qualifications and whether people have attempted higher education before.

4.3 Application 1: Dropout rate for Vocational Education and Training in Denmark

4.3.1 Facts about the Danish vocational training system and their consequences on the design

In this application, the focus is on two types of Vocational Education and Training (VET). We look at technical and commercial training. Vocational Education and Training in Denmark is very close to the apprenticeship system available in German-speaking countries (Germany, Austria and Switzerland)[3], where it is a mix of learning at school and in a firm.

[3] A description of these vocational education and training systems can be found in Bosch & Charest (2010). Another description of vocational education and training in Denmark can be found in Cort & Wiborg (2009).

Technical and commercial courses consist of two steps. During the first step (the basic programme) students learn general and mainly theoretical knowledge, since they will not be at work. The purpose of this period is to lay the grounds for more specialised training afterwards. The first part takes 20-60 weeks for the technical programme according to the subject, where this part of the commercial programme is normally completed in two years. Within these two types of VET, students will have to choose a specialisation[4]. In order to continue to the main programme after the basic programme students will have to have a signed contract for an apprenticeship with a firm. The apprenticeship has to be closely related to the field of specialisation which the student has chosen. A considerable number of people actually drop out because they are unable to find an apprenticeship place. The second part will involve the learning of skills both at school and at the workplace. Figure 4.2 and 4.3 illustrate the timing of these two steps.

The Danish VET Programmes

Figure 4.2 Illustration of the VET programme – commercial courses

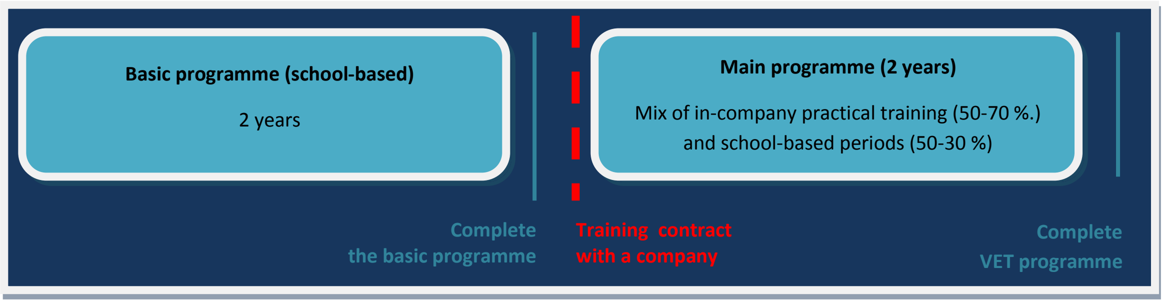


Figure 4.3 Illustration of the VET programme – technical courses

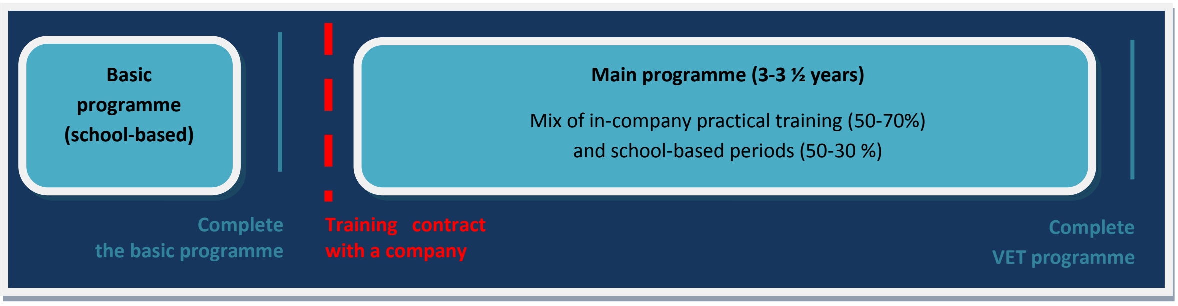


Table 4.1 illustrates dropout rates at different stages for students taking the commercial and technical courses. Figures are for students starting in 2003 who should have had completed their courses by 2007. The percentages of those who complete their course are similar for the two types of courses (40-50%) and few students drop out during the second part. Nevertheless, table 4.1 shows different patterns of timing when dropping out. Most students who take a technical course leave during the first part, whereas a lot of commer-

[4] Overall there are 123 different VET programmes – divided into seven basic programmes: one commercial course and six technical courses which consist of Service Industries, Building and Construction, Technology and Communication, Mechanical Engineering, Transport and Logistics, Food Production and Catering, and Crafts and Engineering Trades.

cial students complete the first part, but a high fraction of them does not start the second part.

Table 4.1 Drop-out rates commercial/technical courses

	Commercial	Technical
Completed	42	48
Dropout during first part	14	32
Do not start second part	39	12
Dropout during second part	5	8

These features have implications as to how to analyse dropout rates and on the choice of the success criterion, which means that it is necessary to look at specific phases of the process. For commercial students we decided to model the probability of completing the basic programme. This choice is motivated by the fact that a lot of students on commercial courses choose a different subject after having completed the basic programme. Generally, dropouts occur either during the basic programme or between the end of the basic programme and the start of the main programme. Moreover, as the problem of changing subjects is not so widespread for students who choose technical studies, we decided to consider the probability of starting the main programme as a success criterion.

4.3.2 Design of the study and data

In this section we describe the dependent variables chosen for students on commercial and technical courses. A lot of students change their subject during the first step of their training. Conditional on the type of programme most of the students change subjects once or twice. This cannot necessarily be considered as dropping out, since in most cases these students continue their education in another subject. Education can be considered an experience good, as it takes time for students to find out which type of basic programme is best suited to their taste and abilities. Therefore, it has been decided to focus on the last spell of vocational training.

Given the length of the studies and the availability of the data we have had to delimit the sample period. The important decision is to choose a starting date. Since the latest data available are for the year 2007, and given the length of the studies, which is about two years, it has been chosen to focus on people starting in 2004 and 2005.

The original data were extracted from Danish administrative registers and cover everybody who enrolled for a vocational training course during the period 2000–2007. In this application we have chosen to select students who started the last part of their commercial or technical studies in 2004–2005. Table 4.2 reports the number of people who started in 2004 or 2005 and who dropped out up to 2006 or 2007, respectively.

Table 4.2 Dropouts commercial and technical courses – students starting in 2004/2005

	N	%
Commercial courses		
Number of students starting in 2004/2005	16,041	
Dropout rates during first step	3,207	20
Remaining in 2006/2007	12,834	80
Technical courses		
Number of students starting in 2004/2005	43,588	
Dropout rates during first step	15,054	34
Dropout rates between first and second steps	4,706	11
Remaining in 2006/2007	23,828	55

4.4 Application 2: Duration analysis of dropouts on professional higher education courses in Denmark

4.4.1 Design of the study and data

In this section we analyse dropouts among students on professional bachelor courses for nurses and teachers. A professional bachelor course takes approximately 3.5 years.[5] Our sample consists of individuals who started between 2000 and 2005. We observe dropouts up to October 2006. The data consist of two variables indicating respectively, the start and end dates of the period and a variable indicating whether the student has dropped out, completed his course or is still studying. In the last case the observation is considered right-censored. Therefore, we can observe study duration until drop out (survival time) and estimate a duration model.

We use information about the social background of the parents measured at the time when the student was eighteen. We also collected information about high-school grades and whether these students have been previously enrolled in a different professional course. Other demographics such as age, gender and ethnicity are incorporated in the model. A short description of the data is given in table A3 in the appendix.

5 Some students might take a longer time than normal. These students will tend to make an artificially positive contribution to our indicator. Therefore, the study time for these students has been set equal to the time normally expected to complete the course.

5 Results

5.1 Application 1: Results and effect of corrections

In this section we give a brief summary of the results of the binary-response model for dropping out. The estimation results are given in the appendix (see table A2). The model specifies that the probability of dropping out depends on age, gender, ethnicity and type of family (i.e. whether the biological parents live together). Parental income and education have been included to take into account the fact that parental background has some impact on the probability of dropping out. Records from secondary school, i.e. grades, are also included in the model. One issue with Vocational Education and Training, and which is related to the fact that students have to obtain a training contract with a firm, is the availability of training places. We have to find an indicator of the potential chances for the students to find such a contract, especially for the technical courses, since it will affect the likelihood of their dropping out of the basic programme or continuing their training after the basic programme, and this factor is beyond the control of the schools. We have incorporated as proxies the unemployment rate in the municipality where the school is situated, the proportion of people in the municipality with no education higher than lower secondary school and the proportion of immigrants in the municipality. All these variables were not statistically significant and were dropped from the models.

Here we report some results on the effect on the rankings of the different schools of the correction due to our model. This will allow us to judge whether correcting for the socio-economic composition of students has an impact on the ranking of the institutions. Note that 46 institutions were analysed for commercial studies and 42 for technical studies.

Table 5.1 shows some descriptive statistics for our indicator. The institutions are ranked according to our indicator and grouped according to which quintile of the indicator's distribution they belong. The last three columns indicate for each quintile the number of institutions which indicator is statistically different from zero (where the significance level of the test is 1, 5 and 10 per cent). For each quintile we also report the number of observations per quintile, the minimum, the maximum, the mean and the standard deviation of the indicator. Columns 4 to 7 give the minimum, the maximum, the mean and the standard deviation for each quintile, respectively. For example, for the first quintile for commercial studies, dropout rates are on average 6 percentage points lower than expected given the explanatory variables of the model, where the minimum and the maximum are 9 and 3 percentage points lower, respectively. We can see quite large differences in the correction. Still for commercial studies, the last quintile contains an institution for which the dropout rate is 11 percentage points higher than expected by the model. If we look at the indicator computed for the technical training, the computations for the first and the last quintiles are bigger with a minimum of -19.5 percentage points and a maximum of -12 percentage points compared to what would be expected by the model. We can see that for a number of institutions their indicator is statistically different from zero at the 5% level. In the first quintile 6 out of 9 institutions offering commercial studies have an indicator statistically different from zero, whereas, in the last quintile, the indicator is different from zero for 7 out of the 10 institutions. Looking at the middle of the distribution, the correction at the

10% level is statistically different only for a few institutions. The results for the technical courses are very similar.

Table 5.1 Descriptive statistics for indicator

Quintile	q	N	Min	Max	Mean	sd	1%	5%	10%
Commercial									
1	-0.031	9	-0.086	-0.033	-0.058	0.019	3	3	0
2	-0.013	9	-0.031	-0.013	-0.022	0.006	0	1	2
3	0.015	9	-0.013	0.012	0.002	0.01	0	0	0
4	0.03	9	0.015	0.027	0.021	0.004	0	0	0
5	.	10	0.03	0.119	0.076	0.037	2	5	1
Technical									
1	-0.036	7	-0.121	-0.04	-0.068	0.029	5	1	0
2	-0.012	8	-0.036	-0.017	-0.023	0.008	0	0	1
3	0.013	7	-0.012	-0.001	-0.007	0.004	0	0	0
4	0.042	8	0.013	0.039	0.022	0.009	0	1	0
5	.	8	0.042	0.195	0.092	0.051	2	4	2

Figure 5.1 is a scatter plot of the corrected and the uncorrected indicator for each type of course. The red line represents the regression line between the corrected and the uncorrected indicator. One can see that the variables included in the model have an impact on the dropout rates at school level. Looking at the dispersion of the scatter plot of the corrected vs. the uncorrected indicators in figure 5.1, we observe that the correction has an impact on the rankings and that some institutions swap places.

Figure 5.1 Uncorrected indicator vs correction

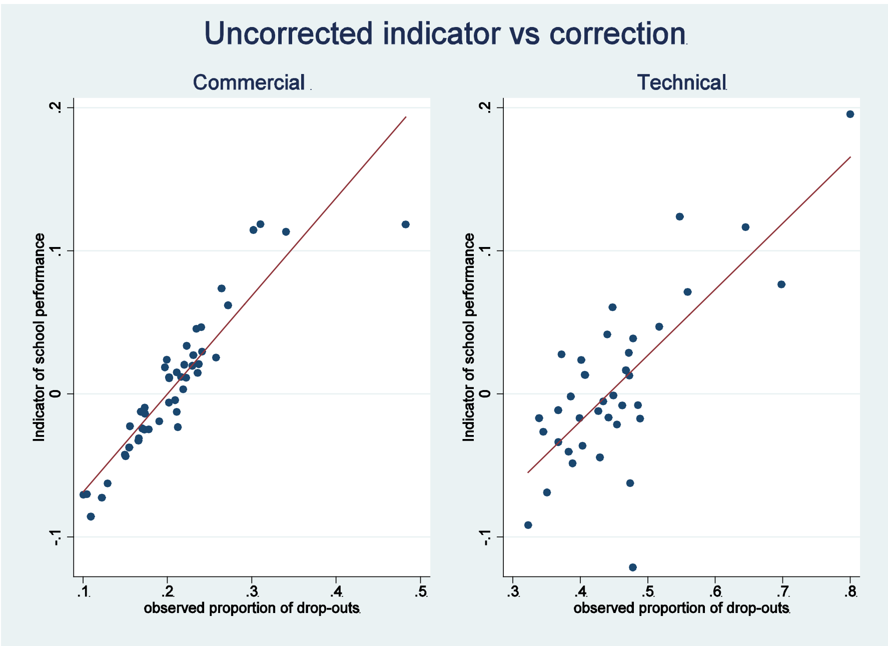
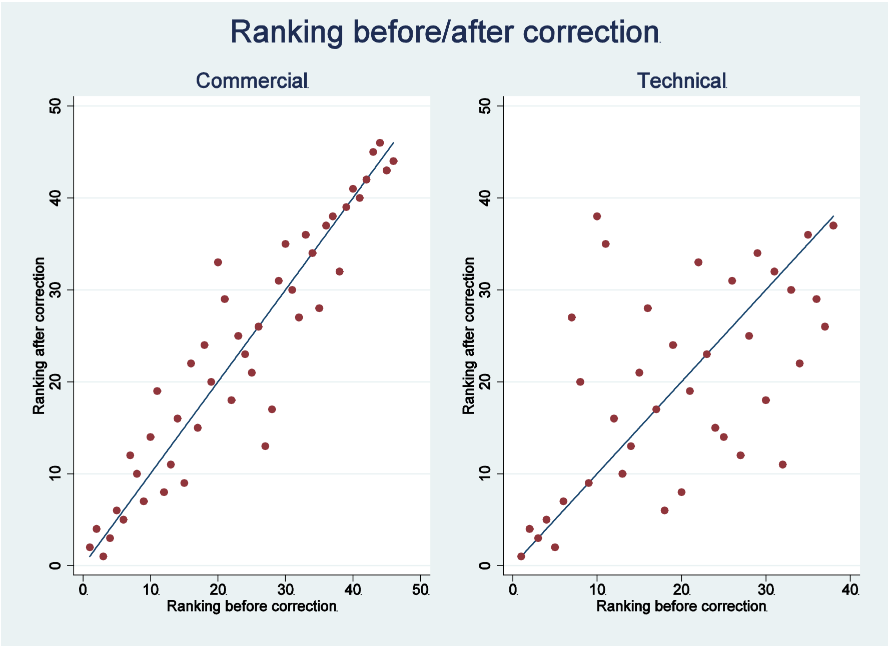


Figure 5.2 is a scatter plot of the rankings before and after correction. On the x-axis we have plotted the ranking of the institutions before correction and on the y-axis the ranking after correction. The red line is the 45-degrees line. If an institution does not change place in the ranking, then its point will lie on the 45-degrees line. One can see that there are changes in ranking after correction and these are more pronounced for the technical institutions.

Figure 5.2 Ranking before/after correction



In table 5.2 we report the transition matrix for the rankings before and after correction. We look at the changes by quintile in order to see if correction has caused more dramatic changes. As we can expect from figure 5.2, the institutions offering commercial courses change place mostly to the next quintile, whereas for institutions offering technical courses we observe more dramatic changes. Indeed, we can observe in table 5.2 moves beyond to one quintile. It appears that the correction has more impact on the technical training than on the commercial one.

Table 5.2 Transition matrix after correlation quintile

	Quintile				
Commercial					
Before\After	1	2	3	4	5
1	8	2	0	0	0
2	2	4	3	0	0
3	0	3	4	2	0
4	0	0	2	6	1
5	0	0	0	1	8
Technical					
Before\After	1	2	3	4	5
1	6	0	1	1	0
2	0	4	1	1	2
3	2	1	4	0	1
4	0	2	1	2	2
5	0	1	1	3	2

Finally, in table 5.3 we report some descriptive statistics for the indicators and the prediction of the model at school level. These statistics include the mean, the minimum, the maximum and the variance for respectively the observed (uncorrected), predicted and corrected dropout rates. The last column reports the reduction in variance in percentage due to our correction. The correction implies a noticeable reduction in variance at school level. For commercial institutions the reduction is about 44% and for technical institutions the reduction is up to 62% of the variation in dropout rates. The correction shows that a lot of the variation in dropout rates at school level is due to the variation of the explanatory variables of the model, which are essentially individual characteristics having nothing to do with the school. It shows that it is important to perform this correction for school composition with regard to students' background and ability.

Table 5.3 Descriptive statistics indicator at school level

	Mean	Min	Max	Variance	Δ V in %
Commercial					
Uncorrected	0.208	0.100	0.482	0.004	
Prediction	0.203	0.171	0.364	0.001	
Corrected	0.005	-0.086	0.119	0.002	44.193
Technical					
Uncorrected	0.453	0.323	0.8	0.009	
Prediction	0.448	0.345	0.622	0.004	
Corrected	0.005	-0.121	0.195	0.004	62.002

Note: Δ V indicates the reduction in the variance of the indicator at the institutional level.

5.2 Application 2: Results of the ranking due to the correction

In this section we give a brief summary of the results of the duration model. The estimation results behind the results of our correction are reported in table A4 in the appendix. The effects of age at the start of the course are not equal across the different types of bachelor studies. The probability of dropping out increases for male students. Note that these two courses have an overrepresentation of women. The indicator variable for Danish ethnicity is insignificant. Parental education and income have not had a strong impact so these two factors were left out of the model. On the other hand, the type of education the students have had before starting their course seems to play a more important role. Grades have a significant impact on the hazard rate.

Table 5.4 shows some descriptive statistics about our performance indicator for nurses and teachers. As we have only 22 institutions for nurses and 18 for teachers we have chosen to rank the different institutions into 3 equal sized groups according to their rank in the distribution of our indicator. Here the indicator computed for each institution measures the average number of months in difference of the mean survival at the national level. For nurses the average value of the indicator for the first quintile indicates a mean duration approximately two months lower compared to what would be expected given the explanatory variables of the model, where the last quintile has an average value of 1 month in excess of the national average. Again, the results are similar for teachers. Although the overall variation in observed means that duration is low, we still see some important differences. For example, the school with the lowest indicator has a difference of five months compared to the value expected from the model. The school with the highest indicator has a difference of three months. These figures are not that low if we remember that we are considering expected values at school level.

Table 5.4 Descriptive statistics for indicator by quintile

	q	N	Min	Max	Mean	sd	1%	5%	10%
Nurses									
1	-0.131	7	-5.35	-0.3	-1.93	1.971	2	0	0
2	0.155	7	-0.13	0.105	-0.03	0.081	0	0	0
3	.	8	0.155	2.994	1.027	0.866	0	2	1
Teachers									
1	-0.987	6	-1.59	-1.04	-1.29	0.193	0	3	1
2	0.7	6	-0.94	0.557	-0.18	0.628	0	0	0
3	.	6	0.7	1.999	1.376	0.526	0	4	0

Figure 5.3, like figure 5.1, is a scatter plot of the corrected and the uncorrected indicator for each type of course. As one can see, individual characteristics have some impact on dropout rates at school level since observations are spread around the regression line.

Figure 5.3 Uncorrected indicator vs correction

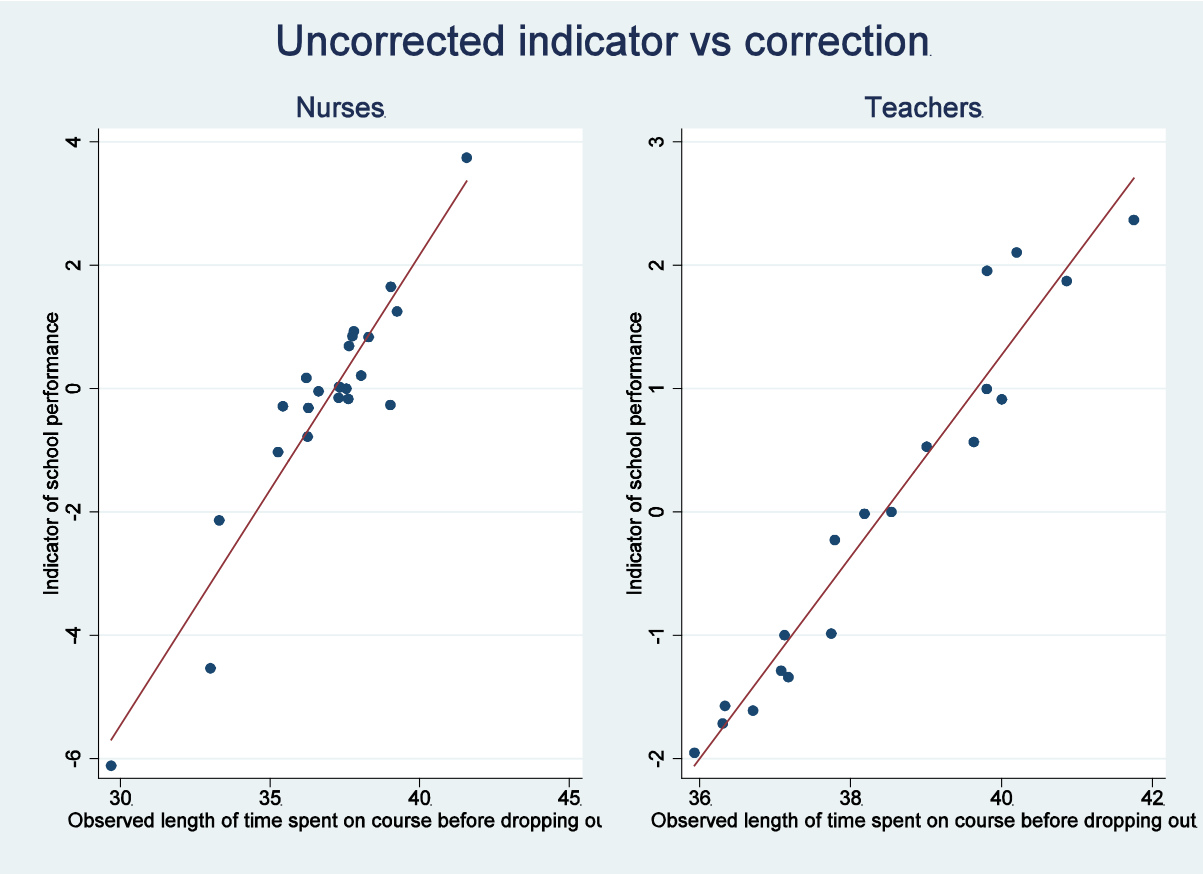


Figure 5.4 shows some changes in the rankings before and after correction for both nurses and teachers, but the change seems more pronounced for nurses. This observation is confirmed by table 5.5, which reports for each type of course the transition matrices issued from the movement from the rankings before and after the correction and the movements at the top and bottom of the rankings. We observe some changes to the rankings, but the impact of correction is rather limited. Some institutions swap places, where others stay in the same place. But most of the changes occur within quintiles and not across.

Figure 5.4 Ranking before/after correction

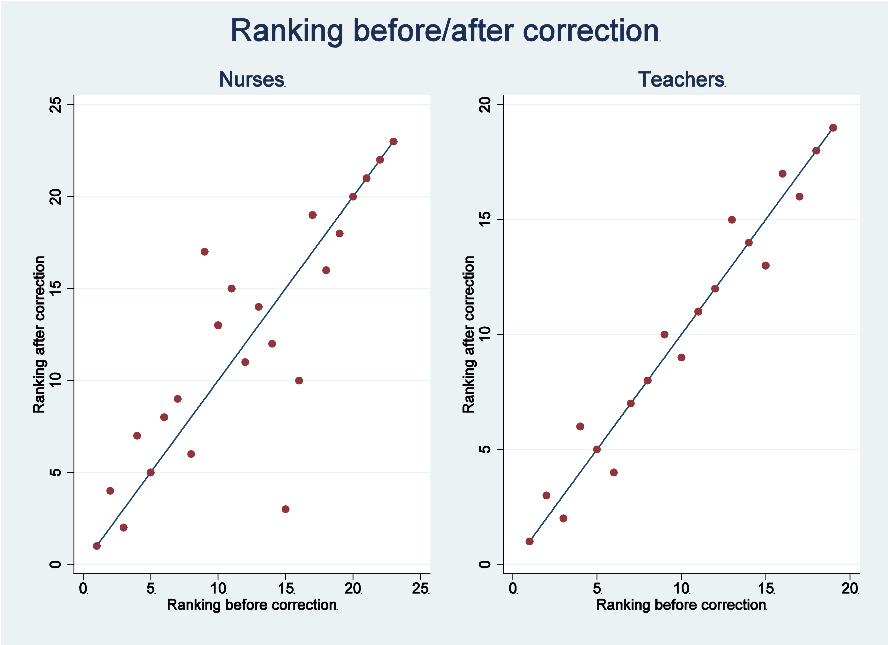


Table 5.5 Transition matrix for the ranking before and after correction

Nurses			
Before\after	1	2	3
1	7	1	0
2	1	6	1
3	0	1	6
Teachers			
Before/after	1	2	3
1	7	0	0
2	0	5	1
3	0	1	5

Table 5.6 reports the reduction in variance due to correction. For each type of course we have a noticeable reduction in the variance of about 30% for both nurses and teachers.

Table 5.6 Descriptive statistics indicator at the school level

	Mean	Min	Max	Variance	Δ V in %
Nurses					
Uncorrected	36.82	29.69	41.575	6.139	
Prediction	37.067	35.439	39.287	0.78	
Corrected	-0.247	-6.114	3.744	4.013	33.286
Teachers					
Uncorrected	38.414	35.93	41.753	3.077	
Prediction	38.436	37.853	39.386	0.225	
Corrected	-0.022	-1.953	2.367	2.18	28.765

Note: Δ V indicates the reduction in the variance of the indicator at the institutional level.

6 Conclusion

We have presented two methods for evaluating the performance of schools in terms of their success in preventing students dropping out during their studies. We have also illustrated these two methods by applying them to two types of tertiary education in Denmark, vocational courses and professional bachelor courses. In order to obtain meaningful indicators we have to divest the raw data of factors independent of the schools' methods of retaining their students. The first challenge lies in choosing the model. In some cases it is possible to estimate duration models and use mean survival time at school level as a measure of the retention rate. In other words, with duration data it is possible to characterise the risk of dropping out (i.e. the hazard rate) at any given point in time. In other cases, due to data limitations and course design, it is better to analyse the probability of dropping out after some time has elapsed, which definition depends on the individual characteristics of the course. We have presented a series of tools to gauge the effect of the correction on the ranking of the institution according to the success criterion. The discrete-response model has been applied to evaluate the performance of institutions offering vocational courses in Denmark, whereas the duration model has been applied to evaluate the retention rate of Danish institutions offering professional bachelor degrees.

The second challenge lies in the design of the study. Since educational programmes are dynamic processes, the outcome which will measure the dropout rates needs to be properly defined. This will have implications on how we delimit the sample. The definition for the vocational application has relied on the different institutional rules linked to the two types of courses (technical and commercial). This is particularly relevant because these courses are composed of two phases and the second phase is conditional upon finding an apprenticeship. It should therefore be done on a case-by-case basis. For the professional bachelor course application we were able to observe survival time which allowed us to apply a survival analysis.

For these two applications we have used data from administrative registers covering the entire population. Therefore, we have large samples which allow us to perform meaningful statistical inferences. The first application uses data from students enrolled in vocational training courses. We have constructed success criteria measuring dropout rates corrected for factors beyond the control of the educational institutions, such as family background and past educational records. We have shown that correction has an impact on the ranking of these institutions and that this impact is more pronounced for students doing a technical course. The second application uses data from students following a professional bachelor course. Our focus has been on student nurses and student teachers. Although individual characteristics have an impact on the hazard rate the magnitude of the correction in terms of rankings of the institutions was not great. Comparing the two applications we see that external factors have a higher impact on the ranking for vocational training courses and they explain a higher fraction of the between-school variation of the success criterion. Therefore, the conclusion is that what schools do – planned or not planned – to retain students, means most at institutions offering vocational courses compared to institutions offering professional bachelor degrees. But in both types of education, institutions play a significant role, where the good institutions invoke our interest of what they do. In this way our analyses are also very suitable for doing qualitative studies at institutional level.

7 References

Andersen, S.H. & E. Heinesen (2009): Estimating the Relative Success of Local Authorities at Labour-Market Integration of Immigrants. *European Journal of Population* 24:59-86.

Bosch, G. & J. Charest (2010): *Vocational Training, International Perspectives*. New York: Routledge.

Colding B. (2006a): Ethnicity, gender and vocational education in Denmark. International *Journal of Manpower* 27(4):342-357.

Colding, B. (2006b): A dynamic analysis of educational progression of children of immigrants. *Labour Economics* 13:479-492.

Colding, B.; L. Husted & H. Hummelgaard (2009): Educational progression of second-generation immigrants and immigrant children. *Economics of Education Review* 28: 434-443.

Cort, P. & S. Wiborg (2009): The vocational education and training system in Denmark: continuity and change. In B. Gerhard: *Vocational Training, International Perspectives*, 84-109. New York: Routledge.

Husted, L.; E. Heinesen & S.H. Andersen (2009): Labour market integration of immigrants: Estimating Local Authority Effects. *Journal of Population Economics* 22(4):909-939.

Jensen, K.B.; C. Kolodziejczyk & T.P. Jensen (2010): *Frafald på professionsbacheloruddannelserne. Hvordan klarer uddannelsesinstitutionerne sig?* mimeo.

Jensen, T.P.; B.Ø. Larsen & B.S. Rangvid (2010): *Institutionernes resultater: En registerbaseret analyse af ungdomsuddannelsesinstitutionerne og betydningen af institutionstype, skolestørrelse og geografisk placering*. AKF Working Paper.

Larsen, B.Ø. & T.P. Jensen (2010): *Fastholdelse af elever på de danske erhvervsskoler*. AKF Rapport.

OECD (2010): *Education at a Glance 2010*. OECD INDICATORS, ISBN 978-92-64-05598-8 (print), ISBN 978-92-64-07566-5 (PDF).

Raudenbusch, S.W. & J.D. Willms (1995): The Estimation of School Effects. *Journal of Educational and Behavioral Statistics* 20(4):307-335.

Rumberger, R.W. (1983): Dropping out of high school: The influence of race, sex and family background. *American Educational Research Journal* 20:199-220.

Rumberger, R.W. (1987): High School Dropouts: A Review of Issues and Evidence. *Review of Educational Research* 57(2):101-121.

Rumberger, R.W. & G.J. Palardy (2004): Multilevel Models for School Effectiveness Research. In D. Kaplan (Ed.): *Handbook of Quantitative Methodology for the Social Sciences*, 235-258. Thousand Oaks, CA: Sage Publications.

Rumberger, R.W. & G.J. Palardy (2005): Test Scores, Dropout Rates, and Transfer Rates as Alternative Indicators of High School Performance. *American Educational Research Journal* 42(1):3-42.

Rumberger, R.W. & S.L. Thomas (2000): The distribution of dropout and turnover rates among urban and suburban high schools. *Sociology of Education* 73(1):39-67.

Undervisningsministeriet [Danish Ministry of Education] (2011): *Profilmodel 2009 – fremskrivning af en ungdomsårgangs uddannelsesniveau*.

8 Appendix

Data

Vocational Training in Denmark

Table A1 Commercial and technical vocational training – descriptive statistics

	(1)		(2)	
	Commercial mean	sd	Technical mean	sd
Mother higher education/high school	0.156	0.362	0.174	0.379
Mother primary school	0.450	0.498	0.440	0.496
Father higher education/high school	0.096	0.294	0.106	0.307
Father primary school	0.450	0.498	0.444	0.497
Mother 2nd-4th income quartile	0.712	0.453	0.714	0.452
Mother income missing	0.050	0.218	0.048	0.213
Father 2nd-4th income quartile	0.641	0.480	0.644	0.479
Father's income missing	0.145	0.352	0.141	0.348
Parents not living together	0.395	0.489	0.416	0.493
Type of family missing	0.024	0.152	0.023	0.151
Mother non-participation	0.154	0.361	0.140	0.347
Mother unemployed	0.084	0.277	0.077	0.267
Father non-participation	0.089	0.285	0.087	0.282
Father unemployed	0.052	0.222	0.050	0.219
Age 17-21	0.756	0.429	0.718	0.450
Age 22-25	0.110	0.313	0.148	0.355
Woman	0.622	0.485	0.304	0.460
Non-western 1st generation immigrant	0.084	0.278	0.059	0.236
Non-western 2nd generation immigrant	0.055	0.227	0.025	0.158
Children	0.022	0.148	0.017	0.130
Apprenticeship	0.019	0.135	0.151	0.358
9th grade, boarding school	0.039	0.195	0.058	0.233
10th grade	0.541	0.498	0.458	0.498
High-school education	0.113	0.317	0.116	0.320
Education missing	0.012	0.108	0.019	0.136
Lower secondary school score 6.5-7	0.139	0.346	0.114	0.318
Lower secondary school score 7 and above	0.518	0.500	0.402	0.490
Lower secondary school score missing	0.230	0.421	0.353	0.478
Technology & communication			0.190	0.392
Craft & engineering trades			0.084	0.277
Food production & catering			0.199	0.400
Mechanic, transport & logistic			0.155	0.362
Service industries			0.127	0.333
Observations	16,041		43,131	

Table A2 Logit regression model: Drop-out rates Commercial and Technical vocational training

	(1)		(2)	
	Commercial parameters	Marginal effect	Technical parameters	Marginal effect
Mother higher education/high school	0.127	0.020	0.204***	0.050***
Mother primary school	0.108*	0.016*	0.225***	0.055***
Father higher education/high school	0.149*	0.023*	0.367***	0.091***
Father primary school	0.188***	0.028***	0.257***	0.063***
Mother 2nd-4th income quartile	0.008	0.001	-0.071	-0.017
Mother income missing	-0.153	-0.022	-0.121*	-0.029*
Father 2nd-4th income quartile	-0.026	-0.004	-0.197***	-0.048***
Father income missing	0.085	0.013	-0.082	-0.020
Parents not living together	0.449***	0.069***	0.419***	0.102***
Type of family missing	0.592***	0.104***	0.047	0.012
Mother non-participation	0.186**	0.029**	0.142**	0.035**
Mother unemployed	0.148	0.023	0.250***	0.062***
Father no-participation	0.248*	0.040*	0.124***	0.030***
Father unemployed	0.189	0.030	0.079	0.019
Age 17-21	0.370***	0.052***	0.285***	0.068***
Age 22-25	0.181	0.028	0.078	0.019
Woman	-0.272***	-0.042***	0.538***	0.132***
Non-western 1st generation immigrant	0.345***	0.056***	0.125	0.031
Non-western 2nd generation immigrant	0.294***	0.048***	0.061	0.015
Children	0.188	0.030	0.150	0.037
Apprenticeship	-0.611**	-0.075**	-2.528***	-0.436***
9th grade, boarding school	0.195*	0.031*	0.023	0.006
10th grade	-0.350***	-0.053***	-0.374***	-0.091***
High-school education	-1.198***	-0.132***	-0.670***	-0.153***
Schooling missing	0.643	0.115	0.247*	0.061*
Lower secondary schoolscore 6.5-7	-0.329***	-0.045***	-0.475***	-0.111***
Lower secondary schoolscore 7 and above	-0.514***	-0.077***	-0.735***	-0.175***
Lower secondary schoolscore missing	-0.106	-0.015	0.121**	0.030**
Technology & communication			1.068***	0.261***
Craft & engineering trades			0.379**	0.094**
Food production & catering			0.442***	0.109***
Mechanic, transport & logistic			0.510***	0.126***
Service industries			1.003***	0.246***
Constant	-1.444***	***	-0.638***	***
Observations	16,041		43,131	
N. groups	46		42	
pseudo R2	0.0607		0.179	
Log-likelihood	-7538.0		-24400.5	

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Professional Bachelor Education in Denmark

Table A3 Professional bachelor education – descriptive statistics

	(1)		(2)	
	Nurses mean	sd	Teachers mean	sd
Age	22.718	3.901	23.191	4.096
Female	0.952	0.214	0.651	0.477
Married/living together	0.303	0.460	0.265	0.441
Danish	0.942	0.234	0.953	0.211
Education missing	0.011	0.103	0.006	0.076
Primary school	0.138	0.345	0.133	0.340
Vocational training	0.146	0.353	0.117	0.321
Higher education	0.020	0.140	0.027	0.163
Tried a higher education before	0.165	0.371	0.251	0.433
log high-school grades	1.720	0.778	1.784	0.725
High-school grades	6.615	3.092	6.884	2.884
High-school grades squared	53.311	28.080	55.700	25.927
High-school grades missing	0.167	0.373	0.140	0.346
Parents: primary school	0.134	0.341	0.120	0.325
Parents: high school	0.015	0.121	0.017	0.129
Parents higher education	0.313	0.464	0.395	0.489
Parents: highest income	2.987	2.051	3.035	2.188
Parents: income missing	0.099	0.298	0.101	0.301
Mother single	0.276	0.447	0.256	0.436
Father single	0.279	0.449	0.259	0.438
Parents divorced	0.321	0.467	0.322	0.467
School size	0.201	0.137	0.266	0.087
Population 16-64 in municipality	10.821	12.212	8.333	9.912
Share of jobs with medium-high qualifications	0.159	0.032	0.166	0.037
Number of students in municipality	3.912	6.058	4.252	6.197
Observations	14,862		21,691	

Table A4 Regression model – hazard rate for dropout for nurses and teachers

	(1) Nurses Marginal effect		(2) Teachers Marginal effect	
0-3 months	0.002		-0.007	**
4-6 months	0.009		-0.002	
7-30 months	0.002		-0.007	*
31-39 months	-0.005	*	-0.009	***
40 months or more	-0.005		-0.008	***
Female	-0.007	***	-0.003	***
Married/living together	-0.001	***	-0.001	*
Danish	-0.002		-0.000	
High-school education	ref.		ref.	
Education missing	0.003	*	0.004	**
Primary school	0.004	***	0.005	***
Vocational training	-0.001	*	0.000	
Higher education	-0.004	***	0.000	
Tried course of higher education before	0.008	***	0.003	***
log high-school grades			-0.013	***
High-school grades	-0.007	***		
High-school grades squared	0.000	***		
High-school grades missing	-0.015	***	-0.012	***
School size	-0.003		-0.003	
Share of jobs with medium-high qualifications	-0.016		0.004	
Number of students in municipality	-0.000	***		
Parents higher education	-0.000			
N	56,573		81,926	
N clusters	22		18	
N subjects	14,771		21,550	

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Danish Institute
for Local and Regional
Government Research