The interaction between dropout, graduation rates and quality ratings in universities

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Download date: 25 Apr. 2024

The interaction between dropout, graduation rates and quality ratings in universities

Eline Sneyers[1]* and Kristof De Witte[1,2]

[1]*Maastricht University, Maastricht, the Netherlands; and* [2]*Katholieke Universiteit Leuven (KU Leuven), Leuven, Belgium*

This paper investigates in a non-parametric framework whether academic programmes maximize their student graduation rates and programme quality ratings given the first-year student dropout rates. In addition, it explores what institutional and programme characteristics explain this interaction. The results show a large variation in how academic programmes are able to deal with the selective nature of first-year dropout. Nevertheless, we can accurately explain the variation among programmes by programme and institutional characteristics. It seems that universities can maximize the relation between first-year dropout, graduation rates and quality ratings in several ways: (1) by improving student programme satisfaction, (2) by better preparing certain groups of students for higher education, (3) by supporting male students, (4) by supporting ethnic minority students, (5) by attracting older staff, and (6) by strengthening the selective nature of the first year (ie, increasing the academic dismissal policy threshold).

1. Introduction

Higher education institutions are increasingly held accountable for the quality and efficiency of education they deliver (Archibald and Feldman, 2008). In order to increase the accountability of higher education institutions, many governments have introduced some form of performance-based funding (PBF). PBF links funding to prescribed performance measures agreed upon by the government and the institution(s). Most PBF mechanisms use a number of different indicators (eg, the number of degrees awarded or the number of research publications) (Jongbloed and Vossensteyn, 2001).

This paper will focus on three important indicators in many PBF-mechanisms: student dropout, programme quality ratings and graduation rates. These indicators are seen as indicators of institutional excellence and performance (eg, Cave *et al*, 1991). Academic programmes (ie, programmes delivered at universities) with high graduation rates and low dropout rates are often considered to be more efficient. They also tend to receive higher quality ratings than academic programmes with high dropout and low graduation rates (Kokkelenberg *et al*, 2008; OCW, 2011). Academic quality is often assessed by accreditations in which higher education institutions obtain quality ratings. The procedure involves a process of self-evaluation, a peer review

by outside experts and discussion of statistical information and performance indicators (Van Vught and Westerheijden, 1994; Van Damme, 2000). The accreditation is normally done by a national accreditation agency. Consequently, the quality ratings are measured in a nationwide standardized way and allow for benchmarking between institutions and even academic programmes. It should be noted that a high dropout rate should not necessarily lead to low quality ratings as it can be the result of selection of the best and most motivated first-year students. By introducing high grading standards, a high difficulty level, heavy workloads and time investments in homework, academic programmes try to select the students who have the highest chance of graduating within the nominal study time.

This paper explores the interaction between student dropout, student graduation and programme quality ratings. This is important for two reasons.

First, PBF systems include agreements between the government and the institutions on a (simultaneous) decrease of student dropout, an increase of graduation rates and high programme quality ratings (OCW, 2011). Although the literature lacks empirical evidence on whether it is possible to observe these three directions simultaneously, the PBF system pays significant attention to them (eg, in the Netherlands 7% of the total higher education budget is reserved for performance agreements which have dropout, completion rates and quality as main indicators). Therefore, it is important to provide empirical evidence regarding the relationship between dropout, graduation rates and quality ratings.

Correspondence: Eline Sneyers, Faculty of Humanities and Sciences, Maastricht University; TIER, PO Box 616, Maastricht 6200MD, the Netherlands.
E-mail: eline.sneyers@maastrichtuniversity.nl

Second, we observe in the underlying data that some academic programmes combine low first-year dropout rates with high graduation rates (measured as the percentage of students who successfully passed the first year and graduated with maximum one year of delay). Other programmes exploit the selective nature of the first year of higher education more (ie, high grading standards, high workload, etc) and combine high dropout rates with high graduation rates. However, some academic programmes fail in obtaining high graduation rates independently of their dropout rates. This is an alarming finding given the increased attention from governments, higher education institutes, students and parents towards student success. It can be expected that the observed variation between dropout and graduation rates is correlated to programme and institutional characteristics. Consequently, it can be interesting to identify programme and institutional characteristics that, for given levels of dropout rates, lead to high graduation rates.

Using a non-parametric model, we explore the relationship between student dropout, student graduation and programme quality ratings. Further, we explain what makes some academic programmes more efficient than others. In the context of this study, a more efficient academic programme indicates that, in comparison to academic programmes with less or equal dropout, this programme is able to achieve a higher graduation rate and quality rating. *In particular, we answer the following research questions: (1) Do programmes maximize their student graduation rates and programme quality ratings given the first-year student dropout rates? And (2) what institutional and programme characteristics explain this relative efficiency?*

Besides its research questions, this paper is innovative in more technical ways. To conceptualize the interaction, we rely on recent innovations in efficiency models. While this is not the first article dealing with efficiency in higher education, it is the first focusing on efficiency with academic programmes as a unit of analysis. Earlier work used the institution, the academic department, non-academic or auxiliary units within institutions as identification level (eg, Jongbloed *et al*, 2003). Higher order units of analysis might result in aggregation bias as academic programmes differ. This paper avoids similar biases. Next, the paper applies a recently developed fully non-parametric conditional efficiency method with continuous and discrete environmental variables. This method allows us to account for heterogeneity in performance assessment by evaluating the impact of environmental variables. Further, this method enables us to test the significance of the control variables on academic programmes' performance. Earlier literature studying higher education efficiency mainly applied standard regression analysis (eg, Archibald and Feldman, 2008), Stochastic Frontier Analysis (SFA) (eg, Robst, 2001), traditional Data Envelopment Analysis (DEA) (eg, Archibald and Feldman, 2008) or the Free Disposal Hull (eg Agasisti, 2011). Each of those methodologies faces one or more of the following disadvantages: (1) specification biases due to their (semi-)parametric nature; (2) outlying observations influence the results of deterministic

models; (3) the influence of the control variables is required to be monotone; (4) whether the control variables serve as an input or output has to be chosen *a priori*; (5) it is often not possible to include multiple control variables; and (6) a separability condition needs to be imposed if one wants to take the operational environment into account (De Witte and Kortelainen, 2013). The non-parametric conditional efficiency model as described by Cazals *et al* (2002) and Daraio and Simar (2005) does not suffer from these drawbacks and allows us to test for the direction of the influence of continuous institutional and programme characteristics. Although Bonaccorsi and Daraio (2008) were the first to apply a conditional efficiency model to higher education institutions, their approach was still unable to estimate direction of the influence of discrete control variables and to test the significance of the control variables on study programmes' performance. Using recent extensions by De Witte and Kortelainen (2013), this paper will do so.

This paper focuses on Dutch universities for which we obtained a rich data set on the universe of study programmes. It includes various programme and institutional characteristics such as student satisfaction and student-teacher ratios. The data on dropout, graduation, student satisfaction and quality rating are measured in a nation-wide standardized way. The latter avoids endogeneity issues arising from measurement errors. Despite its focus on the Netherlands, our study is of interest to an international audience. PBF mechanisms are also implemented in many other European and non-European countries. The mechanisms have in common that they aim for high graduation rates and low dropout rates (Benneworth *et al*, 2011).

The remainder of the paper unfolds as follows. The next section presents a literature review on graduation rates and quality ratings and on the applied methodologies for measuring efficiency in higher education. Section 3 outlines the empirical model. Section 4 presents the Dutch setting and discusses the data. Section 5 presents the results, while Section 6 points to endogeneity issues and presents a robustness test. Section 7 includes a conclusion and policy implications.

2. Literature review

Despite the multiproduct nature of higher education institutions, that is, they have to do both research, teaching and provide public services, we focus on the educational outcomes of higher education institutions (see De Witte *et al*, 2013a,b for more discussion concerning the teaching-research nexus of higher education institutions). Particularly our focus on first-year dropout, student graduation rates and programme quality ratings assure that concentrating on a single-product (ie, education) is still valid, as the former three variables are less (or not) influenced by the research capacities of institutions.

There is an increasing attention towards the concepts of first-year dropout, student graduation and programme quality ratings. although higher education institutions are able to tightly
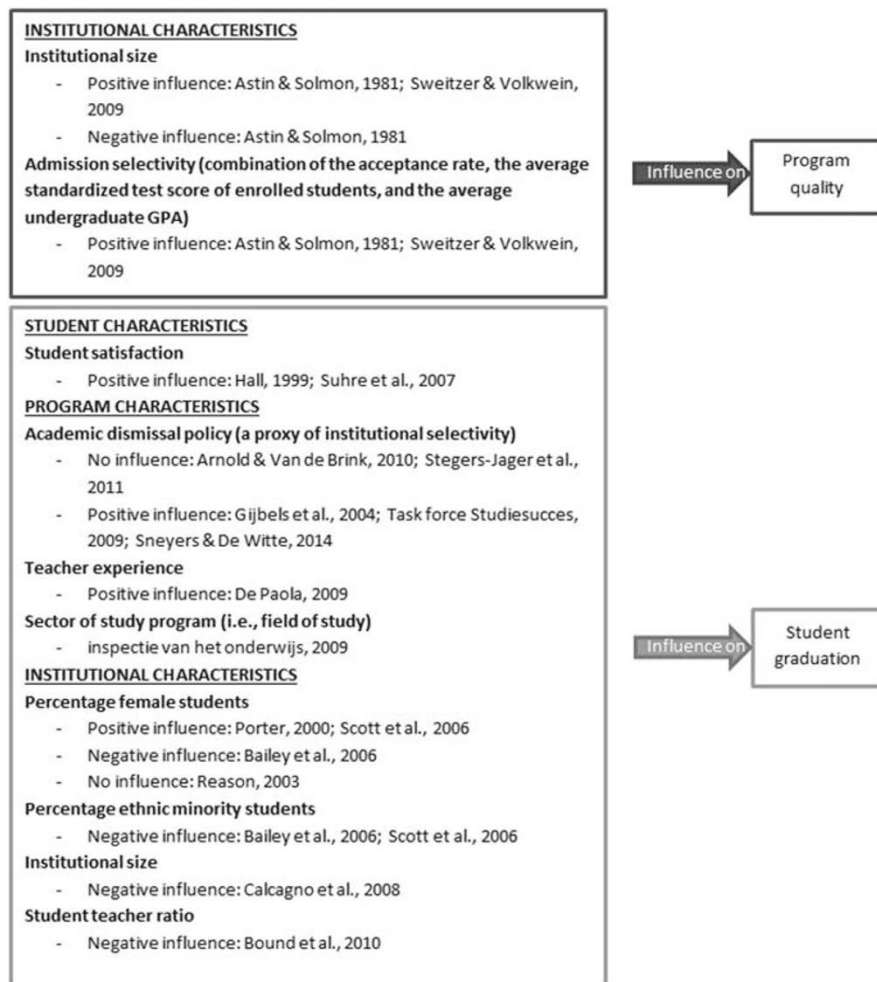
```
INSTITUTIONAL CHARACTERISTICS
Institutional size
    -   Positive influence: Astin & Solmon, 1981; Sweitzer & Volkwein,
        2009
    -   Negative influence: Astin & Solmon, 1981
Admission selectivity (combination of the acceptance rate, the average
standardized test score of enrolled students, and the average
undergraduate GPA)
    -   Positive influence: Astin & Solmon, 1981; Sweitzer & Volkwein,
        2009
```

```
STUDENT CHARACTERISTICS
Student satisfaction
    -   Positive influence: Hall, 1999; Suhre et al., 2007
PROGRAM CHARACTERISTICS
Academic dismissal policy (a proxy of institutional selectivity)
    -   No influence: Arnold & Van de Brink, 2010; Stegers-Jager et al.,
        2011
    -   Positive influence: Gijbels et al., 2004; Task force Studiesucces,
        2009; Sneyers & De Witte, 2014
Teacher experience
    -   Positive influence: De Paola, 2009
Sector of study program (i.e., field of study)
    -   inspectie van het onderwijs, 2009
INSTITUTIONAL CHARACTERISTICS
Percentage female students
    -   Positive influence: Porter, 2000; Scott et al., 2006
    -   Negative influence: Bailey et al., 2006
    -   No influence: Reason, 2003
Percentage ethnic minority students
    -   Negative influence: Bailey et al., 2006; Scott et al., 2006
Institutional size
    -   Negative influence: Calcagno et al., 2008
Student teacher ratio
    -   Negative influence: Bound et al., 2010
```

Program quality

Student graduation

Figure 1   Influence of programme and institutional characteristics on student graduation rates and programme quality ratings observed in previous literature.

manage dropout rates by organizational factors, graduation rates and quality ratings are less controllable. Therefore, this paper provides a framework to examine how student graduation and quality ratings can be increased for given levels of student dropout. We start by summarizing earlier literature that investigates the influence of programme and institutional characteristics on graduation rates and quality ratings. Given that we rely on frameworks from the efficiency literature, we discuss in section 2.2 earlier applied efficiency models in higher education.

2.1. Graduation rates and quality ratings

By following the incentives provided in the allocation of resources, universities are increasingly focusing on student graduation rates (Huisman and Currie, 2004). Indeed, in the academic year 1998–1999 countries such as Australia, Germany, Japan, the Netherlands and the US allocated part of their PBF budget based on graduation rates (Jongbloed, 2001). In order to receive funding, universities were required to focus more on obtaining high graduation rates. Simultaneously, increased attention towards quality emerged in the European higher education sector due to the Bologna Process. This process stimulated international benchmarking to enable comparison between study programmes in Europe. Hence, many European countries implemented an accreditation system because it leads to independent quality ratings of academic programmes or institutions (De Corte, 2014; Teichler, 2007). While in 1998 only six out of 20 countries participating in the Bologna Scheme had an accreditation scheme implemented, 18 out of 20 countries had an accreditation mechanism in place in 2003. Note that accreditation systems had already been in place in the US since the 1980s (Schwartz and Westerheijden, 2004).

Given those trends, it is important for higher education institutions to know what programme and institutional characteristics influence graduation and quality ratings. The earlier literature is summarized in Figure 1, which makes a distinction between the influence on programme quality and student graduation.

First consider the characteristics that influence student graduation.[1] A first variable is student satisfaction. Hall (1999) observed that students who took more than four years to graduate indicated that this was due to dissatisfaction with the teacher and decided to drop the course. Studies often point towards a positive relationship between graduation and satisfaction (eg, Suhre *et al.*, 2007) because of a sense of belonging at and loyalty to the institution derives from student satisfaction (Tinto, 1993).[2]

Next, there is no consensus regarding the influence of an academic dismissal (AD) policy (see section 4.2 for more information regarding AD polices). Some studies (eg, Gijbels *et al*, 2004; Task force Studiesucces, 2009; Sneyers and De Witte, 2015) point towards a positive influence of an AD policy on graduation rates. This may indicate that an academic dismissal policy is an effective tool for the selection of well-performing students. Other studies (eg, Arnold and Van de Brink, 2010; Stegers-Jager *et al*, 2011) found no significant correlation of academic dismissal policies with first-year completion rates.

Previous studies that investigated teacher experience found a positive influence on course graduation rates (eg, De Paola, 2009).

Next, a study by the Education Inspectorate (2009) found that student graduation rates differ significantly across faculties and subjects (eg, ICT and Law). These differences may be due to variations in teaching across various academic disciplines (eg, Neumann, 2001).

The literature is inconclusive concerning the influence of the percentage of first-year female students. Porter (2000) and Scott *et al* (2006) showed that institutions with a higher ratio of female students have higher six-year graduation rates. These gender differences in degree performance may, for example, be due to differences (i) in characteristics that are correlated with attainment (eg, family background), and (ii) in psychological and/or biological factors (Mellanby *et al*, 2000; McNabb *et al*, 2002). Reason (2003), however, found no significant relationship of the percentage of female students on student performance. Bailey *et al* (2006) observed a negative relationship between the percentage of female students and the graduation rate.

Further, a high share of ethnic minority students has a negative influence on student graduation (eg Bailey *et al*, 2006). This can be explained by the fact that various racial groups have different educational experiences (Reason, 2009).

Also, institutional size has a negative relationship with earning a degree (eg, Calcagno *et al*, 2008). If the institution increases in size, the amount of students increases faster than the number of facilities. Consequently, the ratio of people to facilities increases and, in turn, academic and social support suffers (Chickering and Reisser, 1993).

Finally, a low student-teacher ratio correlates to high graduation rates (Bound *et al*, 2010). Owing to a low student-teacher ratio, there will be frequent interaction between students and teachers which, in turn, will act as a key factor in promoting student academic integration and persistence (Tinto, 2002).

Second, consider the characteristics that influence programme quality. Sweitzer and Volkwein (2009) showed that quality ratings, measured by reputational scores, are positively influenced by a combination of institutional size and student admission selectivity (see Figure 1 for a definition of student admission selectivity). In the same line, Astin and Solmon (1981) concluded that large, highly selective institutions have substantive correlations with overall quality of undergraduate education. Note, however, that while selectivity in itself has a positive relationship to quality ratings, institutional size in itself has a negative relationship to quality ratings. We also observe that the student-staff ratio is negatively associated with quality measures (Drennan and Beck, 2001).

2.2. Applied methodologies for measuring efficiency in higher education

One possible way to focus on the relationship between student dropout, student graduation rates and quality ratings is by using an efficiency framework (see next section). Although this is the first paper to connect the three variables, it is not the first to apply efficiency models to higher education settings. There exist various techniques, which lead to different insights.

Regression analysis was the first methodology applied to study university efficiency. Johnes and Taylor (1990) showed that variations in a small number of input variables determine large inter-university variations in output measures (eg, student completion rates and labour market success of graduates). Furthermore, it became clear that rankings of universities based on raw output measures differ from those based on input variables. Johnes (2006a) concluded that although some information on performance can be derived from applying regression analysis, the construction of performance indicators by regression analysis is not recommended. Indeed, those performance indicators are derived from a production function which is an average line through the data rather than a frontier around the data.

Frontier methods such as SFA and DEA address the above problem and are both widely used in the context of higher education (see Salerno, 2003 for an excellent summary of efficiency studies in six countries using SFA and DEA). However, the production process of universities is largely unknown. Nevertheless, SFA requires a parametric specification and assumptions on how efficiency is distributed. Incorrect assumptions concerning the functional form will invariably lead to biased and inconsistent estimators (Salerno, 2003).

The non-parametric DEA effectively deals with this limitation. As efficiency estimates in DEA are based on the behaviour of the other observations, no assumptions about efficiency have

---

[1] Note that graduation rates and quality ratings are often not measured in a consistent way. We incorporated the used definitions/measurements in Tables A1 and A2 in Appendix A.

[2] However, few studies have investigated the direct relationship between satisfaction and graduation rates.

to be made *a priori*. Furthermore, due to the inclusion of distance functions in DEA it is also possible to examine multiple inputs and multiple outputs (Athanassopoulos and Shale, 1997; Salerno, 2003; Johnes, 2006b). Not surprisingly, DEA is widely applied to universities. Examples exist for the United States (eg, Kokkelenberg *et al*, 2008), the United Kingdom (eg, Stevens, 2001), Australia (eg, Abbott and Doucouliagos, 2003), Canada (McMillan and Datta, 1998), and the Netherlands (eg, Jongbloed *et al*, 2003).

Efficiency estimations that do not account for the operational environment only have a limited value. Consequently, traditional nonparametric methods to estimate efficiency, such as DEA, are being directed towards the inclusion of exogenous covariates. Indeed, it is possible that some of the observations in the same sample are operating in different environments. For example, in measuring the grade averages of higher education institutions it is needed to consider the type of higher education institution. Failure to account for this external factor may well discriminate the efficiency results in favour of the higher education institutions which are private. Private institutions select their students better and students in private institutions are often from more educated families. Hence, we can expect that these institutions reach a higher average grade. Failure to account for environmental factors is bound to confound results and may lead to unreliable economic decisions (Avkiran and Rowlands, 2008). The literature discusses various approaches (eg, one-stage approach, frontier separation approach, two stage approach) to take external control variables into account in nonparametric efficiency analysis. The conditional efficiency approach seems to be the most promising method to include exogenous control variables into nonparametric frontier models (see for a discussion De Witte and Kortelainen, 2013). Not surprisingly, the conditional efficiency approach is increasingly used in research questions, including studies concerning the productivity of universities (eg, Bonaccorsi *et al*, 2007; Bonaccorsi and Daraio, 2008; De Witte and Hudrlikova, 2013).

3. A non-parametric technique to estimate the interaction

University programmes can choose to exploit the selective nature of the first year in order to increase completion rates and quality ratings. Some university programmes allow most students to enrol in the second year. Other programmes select during the first year only the best students, who can continue to the second year. Consequently, this choice will influence the first-to-second-year dropout rate. This in turn will influence the quality and graduation rate of the programmes.

From a conceptual point of view, this interaction resembles a production process where university programmes are characterized by a set of inputs ($x \in \mathbb{R}^p_+$; in our case dropout rates) and outputs y ($y \in \mathbb{R}^q_+$; in our case completion rates and quality ratings). All feasible combinations of the relationship are characterized by $\Psi = \{(x, y) \in \mathbb{R}^{p+q}_+ \mid x \, can \, produce \, y\}$.[3] Within this interaction, we can identify programmes which obtain for

a given dropout rate (x) the highest completion rates and quality ratings (y). Vice versa, there are programmes that do obtain the highest possible graduation rate and quality rating given their dropout level. By combining these 'best practice' observations we obtain a 'trade-off line', which represents the maximum trade-off between dropout, completion and quality.[4] Indeed, observations on this trade-off line are non-dominated observations in the sense that they are the academic programmes which reach the highest completion rates and quality ratings for a given dropout rate. Vice versa, observations below the trade-off line could increase their completion and quality rates for a given level of dropout. The radial increase in completion rates and quality ratings can be estimated by λ:

$$\lambda(x, y) = \sup \{\lambda \mid (x, \lambda y) \in \Psi\}, \qquad (1)$$

where $\lambda(x,y) \geqslant 1$ is the proportionate increase of outputs, which the academic programme operating at level (x,y) should attain to operate at the efficient combination of dropout, completion and quality rate (ie, $\lambda(x,y) = 1$).

We can operationalize the estimation of λ by imposing a free disposability assumption. It is defined as: $\forall (x, y) \in \Psi$, *if* $\tilde{x} \geqslant x$ *and* $0 \leqslant \tilde{y} \leqslant y$ *then* $(\tilde{x}, \tilde{y}) \in \Psi$. The production possibility set is then estimated as:

$$\hat{\Psi}_{\text{FDH}} = \{(x, y) \in \mathbb{R}^{p+q}_+ \mid y \leqslant y_i, x \geqslant x_i, (x_i, y_i) \in x_n\}, \qquad (2)$$

where x_n represents the sample set. Equation (2) resembles the Free Disposal Hull model by Deprins *et al* (1984). The estimated shortfall in completion rates and quality ratings, for a given dropout level, is then measured by:

$$\lambda(x, y) = \sup \left\{\lambda \mid (x, \lambda y) \in \hat{\Psi}\right\},$$

This non-parametric model has two major drawbacks: (1) it is deterministic, and (2) it does not take heterogeneity among programmes into account. We start by discussing the first issue. The deterministic nature of the estimator arises from the fact that all n programmes in the sample x_n are considered to be potential best practices: $Prob((x,y) \in \Psi) = 1$. Consequently, outliers and atypical observations heavily affect the trade-off line and thus potential increase in outputs for given inputs. We follow Cazals *et al* (2002) in estimating Equation (1) relatively to a partial frontier. This partial frontier depends on a random set of $m < n$ programmes. These observations are repeatedly drawn (B times) with replacement among the population of academic programmes which use less inputs than the level of x. By considering the expected value of this less extreme benchmark, the robust efficiency estimate $\lambda_m(x,y)$ is obtained. A robust estimate $\lambda_m(x,y)$ can acquire a 'super-efficiency' score (ie, an output-efficiency score of $\lambda_m(x,y) < 1$) when it is on average performing superior than its m randomly

---

[3]Note that academic programme, given their dropout, can choose not to obtain the highest graduation rate and quality rating as possible (ie, the Free Disposability Assumption). Further in the section, we discuss this more deeply.

[4]Note that in the DEA literature, this 'trade-off line' is known as the 'efficient frontier'.

drawn (with $x_i \leqslant x$) reference units. These 'super-efficient' observations indicate that the programme produces for a given input vector x a higher output vector y than the average m evaluated programmes in the reference set. In other words, the super-efficient programme is thus doing better than expected (De Witte and Kortelainen, 2013).

Next, we tackle the issue of heterogeneity. Traditional non-parametric models assume a separability condition such that the operational environment does not influence the level of inputs and outputs. However, in real life this is not realistic as, for example, the student characteristics influence both the dropout and completion rates. The conditional efficiency approach, formulated by Cazals *et al* (2002) and Daraio and Simar (2005, 2007), allows us to take the operational environment into account. While using the robust efficiency estimate, the idea is to draw the subsample of size m so that similar programmes have a higher chance of being drawn. The weights are obtained by estimating a kernel density function around the control variables:

$$K\left(\frac{(z-z_i)}{h}\right),$$

where z denotes a vector of continuous and discrete control variables, $K(.)$ is an appropriate Kernel and h is the band-width (see De Witte and Kortelainen, 2013 for a discussion on the appropriate Kernel and bandwidth). More precisely, the procedure works as follows. For a given x, draw a subsample of size m with replacement and with a probability $K((z-z_i)/h)/\sum_{j=1}^{n} K((z-z_i)/h)$ among those y_i such that $x_i \leqslant x$. Compute for this subsample with only comparable observations the FDH output-oriented efficiency score. Finally, repeat this procedure at least 2000 times and take the average of the 2000 efficiency scores. It should be noted that the integral formulation as discussed in Daraio and Simar (2007) is faster to compute. The *conditional* efficiency estimates $\lambda_m(x,y|z)$ compare like with likes (comparable to a matching approach).

De Witte and Kortelainen (2013) present a procedure for statistical inference regarding the influence of the control variables. This procedure is implemented by non-parametrically regressing the control variables on the ratio of the conditional $\lambda_m(x,y|z)$ to the unconditional $\lambda_m(x,y)$ model. An increasing regression line indicates a positive influence, a decreasing line indicates an average negative influence and a horizontal line indicates a neutral influence. A positive relationship implies that the control variable Z acts as an additional free disposal input, while a negative relationship means that the control variable Z uses more inputs than expected and thus constraints the production like an undesired output (Daraio and Simar, 2007). Finally, using a non-parametric bootstrap on this non-parametric regression permits us to obtain statistically inference on the correlation between the efficiency score and the control variables.

4. The Dutch setting and variables

4.1. Dutch setting

Universities prime students for independent scientific work in an academic or professional setting. Approximately one-third of the Dutch higher education students enrol in universities (Huisman, 2008). Academic education comprises of 17 governmental funded universities. Since secondary education, students are tracked for continuing education and prepared for specific disciplines (De Koning *et al*, 2014). In the Netherlands, there is an open higher education system in place. Students who hold a certificate of pre-university education or a first-year certificate of vocational education can enter universities (Huisman, 2008). There are no other entry requirements in place.[5]

In line with the European Credit Transfer System, Dutch universities offer programmes of 60 study points per year since 2002. Credit points represent the workload or 'study time' necessary to complete a course or programme. The workload consists of the actual hours spend in classes and the hours needed to prepare for classes, exams and other assessments. In the Netherlands, as in most European countries, a credit point represents a workload of 28h, meaning that a programme year comprises of 1680h.

4.2. The data

The data are provided by the Dutch Ministry of Education ('Dienst Uitvoering Onderwijs', DUO). The data consist of information at programme-level regarding the number of first-year students, the dropout rate, the switch rate (ie, students changing from programme within an institution), the number of students who re-enrol after the first year, and student success for Dutch universities for the academic years 2010–11 and 2011–12. All indicators concern bachelor students.

Student dropout and student graduation rates, two programme-level variables of interest, are defined as follows. *Student dropout* is defined as the percentage of full-time bachelor students that cease their education at the institution during the first year of enrolment. Note that in order to be included in this variable, students need to dropout from the institution. Student dropout can be a voluntary action (ie, dropout due to illness of mismatch with the programme/institution), or an involuntary action (ie, because one did not acquire sufficient credits). *Student graduation rates* are defined as the share of re-enrolled full-time bachelor students that complete their degree at the institution one year after the nominal study time.[6] Consequently, in order to be included in the latter indicator, students

---

[5]The only exceptions are the numerus fixus programmes in medicine, dentistry and other areas that have a limited number of student places. Numerus fixus means that programmes can choose which students they will accept. Note that only 0.30% of the available places can be filled in by this practice (Jongbloed, 2003).

[6]Thus, the student graduation rate is measured conditional on first-year dropout.

Table 1 Overview of the definition of the variables

Variables	Definition
Student dropout	The percentage of full-time bachelor students that cease their education at the institution during the first year of enrolment. These students will not enrol in the second year.
Graduation rate	The share of re-enrolled full-time bachelor students that complete their degree at the institution one year after the nominal study time. The student graduation rate is measured conditional on first-year dropout.
Quality ratings	A composite indicator of three underlying scores: *the level of the academic programme, the orientation of the programme* and *domain-specific requirements.* It is measured by an independent organization NVAO.
Student programme satisfaction	Measured by the following question 'give your overall judgment about the programme'.
First-year female students (%)	The percentage of first-year female students enrolled in the academic programme.
First-year ethnic minority students (%)	The percentage of first-year foreign students enrolled in the academic programme.
AD-policy threshold (credit points)	The threshold of the academic dismissal policy in place at the academic programme. This variable serves as a proxy of selectivity.
Academic programme	Denotes the subject of the academic programme (eg, math, ICT or law).
Number of students	The number of students enrolled at the institution.
Student-teacher ratio	The number of students per teacher employed at the institution.
Staff > 50 years (%)	The percentage of staff older than 50 years employed at the institution.

enrolled in universities need to graduate in 4 years (see Table 1 for the definition of the variables).

Next, we add a quality rating to the data. This latter variable is based on accreditations of the Dutch-Flemish Accreditation Organization (NVAO). Accreditation by the NVAO is a precondition to receive funding from the government. Furthermore, a positive accreditation gives academic programmes the right to award diplomas and to grant financial assistance to students. Accreditation thus entails that an academic programme satisfies certain quality standards. The report of an external peer-reviewed panel leads to the decision to accredit. This report, in turn, is based on the NVAO framework which consists of 30 criteria (Stensaker and Harvey, 2006). We focus on one of these criteria, that is, the programme-level variable *programme quality ratings*. This latter variable ranges between 0 (insufficient) and 3 (excellent). It is a composite indicator of three underlying scores: *the level of the academic programme, the orientation of the programme* and *domain-specific requirements* (see Table 1 for an overview of this variable and its definition). The *level of the academic programme* represents the degree to which the achieved qualifications correspond to general, internationally accepted descriptions of the qualifications of a Bachelor. Next, the accreditation *domain-specific requirements* is linked to the extent to which the programme fulfils the requirement for similar (foreign) programmes. The accreditation *orientation* reflects (i) whether holders of university Bachelor's degrees have obtained the qualifications to allow admission to at least one subsequent university course at the Master's level and/or to allow access to the labour market, and (ii) whether the achieved qualifications are derived from the requirements of scientific disciplines, international scientific practice and relevant practice in the occupational field.

The data are further enriched by information at the programme-level obtained from the annual national questionnaire on student satisfaction ('nationale studentenenquête', NSE), which is annually carried out by the organization Studiekeuze123. The survey includes data on the opinion of students about the higher educational programme they are taking. The data also contain programme features and institutional information (see Table 1 for an overview of the variables and their definition). In particular, we are interested in the following programme characteristics. *Programme satisfaction* of students (measured by the following question 'give your overall judgment about the programme') that gives qualitative information at programme-level about student satisfaction and which is proxied by a variable between 1 (very dissatisfied) and 5 (very satisfied). Further, we include the variables *percentage of first-year female students* (ie, the percentage of first-year female students enrolled in the academic programme), *percentage of first-year ethnic minority students* (ie, the percentage of first-year foreign students) and *AD policy threshold* (ie, the threshold of the academic dismissal policy). The latter variable serves as a proxy of selectivity. The academic dismissal policy is grounded on the number of credits. Students who do not earn sufficient credits after the first year are dismissed from the programme and from the institution. The higher the AD policy threshold of an academic programme, the higher the degree of selectivity since students have to earn more credits to enrol in the second year. The variable *Academic programme* denotes the subject of the academic programme (eg, math, ICT or law). Including it in the analysis allows us to capture the unobserved heterogeneity at subject level which could arise, for example, from the different nature of the academic curriculum.

As institutional characteristics we include: *the number of students* as a proxy of institutional size, the *student-teacher ratio*, and the *percentage of staff older than 50 years*. This latter variable measures the percentage of staff older than 50 years employed at the institution. The former variable measures the number of students per teacher employed at the institution.

Although the dropout rate deals with the students that started in 2010–2011 or 2011–2012, student graduation concerns students that graduated in 2010–2011 and 2011–2012. The underlying assumption is that the trend of the dropout

Table 2   Descriptive statistics at programme-level

		Minimum	Median	Mean	Maximum	St. Dev.
Input	Student dropout (%)	2.00	19.05	19.96	52.20	8.94
Output	Student graduation rates (%)	8.30	62.50	61.42	93.80	16.53
	Programme quality ratings (0 = insufficient; 1 = sufficient; 2 = good; 3 = excellent)	1.00	1.00	1.23	3.00	0.45
Control variables	Student programme satisfaction	3.42	4.01	4.01	4.67	0.19
	First-year female students (%)	2.04	52.38	49.29	100.00	24.68
	First-year ethnic minority students (%)	2.12	19.05	23.19	96.43	14.25
	AD-policy threshold (credit points)	24.00	40.00	41.31	60.00	5.27
	Number of students	6456	20763	21590	32739	7277
	Student-teacher ratio	14.46	32.04	29.21	43.11	6.35
	Staff > 50 years (%)	13.85	29.34	29.16	66.01	8.01

rates between different years, and thus between different cohorts of students, is comparable. This seems very likely given observations by the Dutch association of universities (VSNU, 2012) that showed that the trend of the first-year dropout rate remained stable between the academic years 2005–2006 and 2010–2011.

Given the different nature of the Open University, which focuses on distance learning, and three religious universities, which prepare students for religious duties (eg, priest), we leave them aside in the analysis even though they are state-funded institutions. The sample is further restricted to bachelor programmes since the data only contain information on bachelor programmes. Next, we eliminate programmes for which the student success or student dropout rate is zero or 100%. Such atypical rates can influence our analysis and are probably linked to programmes that have just started or that are going to be terminated. The final sample includes 420 higher educational programmes (ie, an average of 210 study programmes per year) at 13 universities. For the academic year 2010–2011 we have a sample size of 200, while we have a sample size of 220 for the academic year 2011–2012.[7]

4.3. Descriptive statistics of the interaction

Given the dropout level in the first year, academic programmes have to maximize completion rates and quality ratings. Dropout is an appropriate input variable given the Dutch open higher education system. In the Netherlands, students with the right qualifications can enter the study programmes of their choice. Therefore we expect that the student body of academic programmes is fairly homogenous and that differences in dropout level are due to choices of the programmes themselves

(ie, guidance of students, grading standards, workload). Note, that the combination of the output variables (ie, graduation rates and quality ratings) can be interpreted as a quality corrected graduation rate.

Although this paper is innovative in the way it deals with the relationship between student dropout, graduation rates and programme quality ratings, those variables have been used in earlier efficiency literature. Avkiran (2001) conducted an efficiency analysis of 36 Australian universities in which the student retention rate has been included in the analysis. Completion and graduation rates have been used as outputs before by, for example, Johnes (1996, 2006a), Johnes and Taylor (1990), and Kokkelenberg *et al* (2008). As argued by Johnes (2006a), the advantage of such degree measures is that they capture an element of both quantity and quality of teaching output.

Some descriptive statistics at programme-level are presented in Table 2. We observe that academic programmes have an average first-year student dropout of 20%. This suggests that, on average, 20% of the full-time bachelor students cease their education at the institution after the first year of enrolment. However, while some programmes have an extremely low dropout rate (ie, 2%), other programmes have to deal with an extremely high dropout rate (ie, 52.2%). Moreover, it appears that medical programmes (such as a dentistry programme) have on average the lowest dropout rate. Artistic programmes (such as music science) often have high dropout rates. Academic programmes have, on average, a student graduation rate of 61.4%. This means that, on average, 61.4% of the students who re-enrolled after the first year complete their degree in maximum 4 years. Again we observe that some programmes achieve a high student graduation rate (ie, 93.8%), while others achieve very low graduation rates (ie, 8.3%). We observe that study programmes of liberal arts colleges and Medical programmes have, on average, the highest graduation rate. More technical study programmes appear to have, on average, the lowest graduation rate. Next, we see that the average programme achieves a quality rating of 1.2 (ie, sufficient).

---

[7]Note that these sample sizes differ because we only included programmes which have information on all the control variables. Programmes with missing values are thus removed. There are more academic programmes with missing values in the academic year 2010–2011 compared with 2011–2012. This is not surprising since the data gathering of Studiekeuze123 is improving every year.

Table 3 Correlation of the input variable (dropout) on the output variables (graduation rates and quality ratings)

Independent variable	Graduation rates	Graduation rates	Quality ratings	Quality ratings
student dropout	− 0.369***	− 0.020	− 0.001	− 0.007*
	(0.132)	(0.117)	(0.003)	(0.004)
Year	2010–2011	2011–2012	2010–2011	2011–2012
R^2	0.038	0.0001	0.0003	0.02
N	200	220	200	220

Note: The regression includes a constant. Standard error between brackets. *** and * denote significance at 1% level and 10% level respectively.

A limited number of linguistic academic programmes achieved an excellent rating. Concerning programme satisfaction, the summary statistics show that students rate their academic programmes on average 4.1. This means that students are, on average, satisfied with their study programme. Furthermore, we observe that the first-year student population of academic programmes consists on average of 49.3% female students and 23.2% minority students. Academic programmes with an AD policy have on average a threshold of 41.3 credit points. Finally, the average programme is part of an institution with 21590 students enrolled, a student-staff ratio of 29.2 and a staff that consists of 29.2% of people older than 50 years.

The optimal size of the partial frontier m is selected following Daraio and Simar (2005). They suggested to use the value of m for which the percentage of super-efficient observations (ie, $\lambda_{FDH}^m < 1$) stabilizes. In the sample under study, m corresponds to 100. Finally, for statistical inference, we use 2000 bootstrap replications.[8]

5. Results

5.1. Relationship between the input and the output variables

We start by correlating the input and output variables in order to obtain some intuitive information on the data. The results are presented in Table 3. We observe a significant negative relationship between student dropout and student graduation for the academic year 2010–2011 and a negative insignificant correlation for the academic year 2011–2012. These findings suggest that a lower dropout rate often leads to a higher graduation rate. This is in line with, for example, Hosch (2008) and Lau (2003). However, the relatively weak negative correlation also indicates that programmes with a high dropout rate do not necessarily succeed in obtaining high graduation rates. This is surprising since a high dropout rate after the first year is not necessarily a bad thing in the open higher education system in place in the Netherlands. Moreover, a high dropout rate can be an indication of selection in the first year for retaining the best and most motivated students while rejecting the 'worst' students (eg, based on the accumulated study points). If this is the case, a high dropout rate should enable a

Table 4 Efficiency scores that exclude (unconditional) and include (conditional) the programme and institutional characteristics

	Unconditional (robust FHD)	Conditional	
		Model 1	Model 2
Average	1.371	1.002	1.004
St. Dev.	0.354	0.024	0.040
Min	0.821	1.000	1.000
Max	2.766	1.490	1.562

The efficiency scores denote the proportionate increase of completion rates and quality ratings, which an academic programme with a given dropout level could achieve if it would operate as efficient as the best practice observation.

high graduation rate. Since this is not observed for all programmes, it is interesting to identify the characteristics that enable programmes to obtain a higher graduation rate.

In a similar vein, we regress programme quality ratings (ie, the second output variable) on student dropout. The results in the third and fourth column of Table 2 indicate an insignificant negative relationship between quality ratings and student dropout for the academic year 2010–2011 and a significant negative relationship for the academic year 2011–2012. This suggests that programmes that obtain higher quality ratings suffer less from student dropout in comparison to programmes that achieved lower quality ratings. This finding is in line with studies that imply that retention is a relevant measure of academic quality improvement[9] (eg, Lee and Buckthorpe, 2008).

5.2. Student graduation rates and programme quality ratings for given dropout rates

We estimate the extent to which programmes are able to exploit student dropout in order to acquire higher quality ratings and graduation rates. The results of this interaction are presented in Table 4. In column 1, we summarize the results of the robust order-m estimations (ie, the unconditional efficiency). We obtain an average score of 1.37. This indicates that, on average and for a given first-year dropout level, Dutch academic programmes can increase their quality adjusted graduation rates

---

[8]The number of bootstrap replications only matters for the statistical inference. The conditional order-m model has been estimated in R by using the integral formulation, as this procedure is more time efficient and precise.

[9]Note that when we combine the two years in one data set and we do the above analyses again we find similar results. Moreover, we find a significant negative correlation between the graduation and dropout rate and an insignificant negative relation between graduation rate and the quality ratings.

by approximately 37% if they would operate as good as the best practices (ie, those academic programmes that are obtaining the highest student graduation rates and quality ratings for a given student dropout). However, there is a significant variation in the performance across academic programmes, as can be seen from the sizeable standard deviation around the mean. We observe that technical programmes are, on average, the least efficient. Furthermore, some academic programmes have an efficiency score below 1. These 'super-efficient' programmes (mostly medical study programmes) are performing better than the average m ($m = 100$) academic programmes in their reference sample. Concerning linguistic studies we find mixed results. Some linguistic programmes obtain a very low efficiency rating while others are 'super-efficient'.

To examine how this relationship is influenced by institutional and academic programme characteristics, we apply the conditional efficiency model with a mix of eight discrete and continuous control variables.[10] Our previous findings change in two important ways. First, we can accurately explain the variation among programmes. After including the control variables, the average efficiency score no longer significantly deviates from 1. Taking all exogenous variables into account, the conditional efficiency scores reduce to 1.002 (Model 1). By excluding the percentage of staff older than 50 years as exogenous variable, the conditional efficiency reduces to 1.004 (Model 2). Now, the least efficient academic programmes are the technical academic programmes. In both models there are also no 'super-efficient' units anymore. Second, the standard deviation around the average reduces significantly in both models.

These results show that a large part of the initial variation in programme efficiency can be explained by the control variables. Consequently, we are able to determine the variables which explain why some programmes do not succeed in obtaining high student graduation rates and programme quality ratings for a given dropout level.

We examine the direction of the influence (ie, favourable or unfavourable) of the programme and institutional characteristics by non-parametrically regressing the control variables on the ratio of the conditional to the unconditional efficiency scores. After examining the significance tests and the partial regression plots for the discrete and continuous variables, we observe that the programme characteristics have a significant influence on the efficiency scores. In other words, programme characteristics matter for the degree to which completion rates and quality ratings can improve, for given dropout rates. For institutional characteristics we only find a significant influence for the variable *percentage of staff older than 50 years* (see Table 5).

---

[10]In line with Jongbloed *et al* (1994, 2003), we divided the academic programmes into an arts, a sciences and a medical cluster and reran the analyses. The results of the arts cluster are robust with the current findings. The results of the sciences cluster showed differences concerning the influence of the control variables. Owing to insignificant power we did not find results for the medical cluster.

Table 5 Influence and significance of programme and institutional characteristics that influence the relationship between dropout, graduation rates and quality ratings

	Model 1	*Model 2*
Student programme satisfaction	Favourable (<2E-16)***	Favourable (<2E-16)***
First-year female students (%)	Favourable (0.004)***	Favourable (0.022)**
First-year ethnic minority students (%)	Unfavourable (<2E-16)***	Unfavourable (<2E-16)***
AD-policy threshold	Favourable (<2E-16)***	Favourable (<2E-16)***
Academic programme	Favourable (<2E-16)***	Favourable (0.022)**
Number of students	Unfavourable (1.000)	Unfavourable (1.000)
Student-staff ratio	Unfavourable (1.000)	Unfavourable (1.000)
Staff > 50 years (%)	Favourable (<2E-16)***	

Note: Bootstrapped *p*-values between brackets; ***, **, * denote significance at 1, 5 and 10% level.
Where 'favourable' ('unfavourable') denotes that the variable has a positive (negative) influence on the proportionate increase of completion rates and quality ratings, which an academic programme with a given dropout level could achieve if it would operate as efficient as the best practice observation.

We start by discussing the relationship of the programme characteristics. The favourable influence of programme satisfaction indicates that, for a given dropout level, programmes with a high level of student programme satisfaction succeed in obtaining higher quality adjusted graduation rates in comparison to programmes with a low level of student programme satisfaction. Next, we observe a favourable significant influence of the percentage of first-year female students. This indicates that, for a given dropout rate, a high percentage of first-year female students is a characteristic of programmes with high student graduation rates and high quality ratings. For first-year ethnic minority students we found a significant unfavourable influence. Thus, for a given level of student dropout, the presence of a high share of ethnic minority students will result in low graduation rates and low quality ratings. For the variable academic dismissal policy threshold we observe a favourable significant influence on student graduation and programme quality ratings (for given dropout rates). We further observe that there are significant differences between subjects.

Although we did not find a significant relationship for all the different institutional characteristics, we discuss the results of all variables for completeness. First, we observe a significant positive influence of the percentage of staff older than 50 years. Thus, for a given dropout level, programmes with high-quality adjusted graduation rates are characterized by a high percentage of staff older than 50 years. The next two institutional variables, number of students enrolled and student-teacher ratio, have an unfavourable insignificant relationship on student graduation and programme quality ratings (for given dropout rates). The

former indicates that student graduation and quality ratings (for given dropout rates) is harmed when more students are enrolled. Finally, a low student-staff ratio is a significant characteristic of programmes with high graduation rates and high quality ratings (for a given dropout rate).

Although the non-parametric technique does not allow us to directly examine whether the control variables are more influential for one output or another, we run some correlations to investigate this. We observe that the satisfaction of the students has a larger favourable influence on quality ratings than on the graduation rate. The opposite is found for the variables *academic dismissal policy* and *percentage of first-year female students*. The level of the academic dismissal policy and the percentage of female students have a larger favourable influence on graduation rates compared with the influence on quality ratings. We also found that the percentage of ethnic minority students and the percentage of staff older than 50 years have a similar influence on both output variables.

6. Self-selection of students and a robustness test

The results might be prone to endogeneity, due to self-selection of students in study programmes. The presented results assume that students do not choose a study programme on the basis of the selection criteria. We expect that this assumption holds as all students, with the necessary certifications, can start in a study programme in the Netherlands (ie, there is little or no selection at the gate). Consequently, study programmes have a fairly homogenous student population.

However, recently, selection after the gate became possible. The AD policy selection tool only allows students to enrol in the second year when they have earned a sufficient number of credits. Study programmes decide on the credit norm that students have to pass. In this context, it is possible that students base their choice of study programme on the presence and size of the AD policy. Students who are talented and motivated will attend study programmes with a high AD policy threshold. Students with less motivation and talent will probably attend study programmes without or with a low AD policy threshold. As a result, the student population of study programmes becomes more heterogeneous and endogenous.

To account for this possibility, we make a robustness check with the AD policy threshold as input (ie, a proxy for selectivity). The size of the AD policy might be a good input since study programmes decide upon the level of the AD policy threshold. The output and control variables remain unchanged. This robustness check delivers similar outcomes as the previous models (see section 5.2) (detailed results presented in Appendix B). Moreover, we find that, for a given AD policy threshold, programmes with a high student satisfaction succeed in obtaining higher graduation rates and high quality ratings. The percentage of female first-year students and the percentage of staff older than 50 years also have a significant positive influence. In contrast to our previous results, we find that the percentage of ethnic minority students has a significant positive influence. This indicates that a high percentage of ethnic minority students is a characteristic of programmes with a high graduation rate and quality rating. We can conclude that our results are fairly robust.

In a next step, we examine which programme characteristic has the largest influence on the efficiency model. We do not focus on the institutional information since these variables were not significantly different from 0 (see section 5.2). The question is examined by rerunning model 1 and by removing the significant variables one by one. The results show that student satisfaction influences the conditional efficiency estimates the most. Removing this variable from the model results in a decrease of 5% of the R^2. Furthermore, it leads to an increase in variation in performance of the programmes (ie, the standard deviation increased to 0.0031). The average performance of the programmes, however, remains fairly stable (ie, 1.003) (detailed results presented in Appendix C).

We would like to remark that the influence of the variable *average programme satisfaction* on the conditional model is larger than the other variables, but not much. Removing the other variables from the model specification results in decreases of the R^2 as well. This indicates that each single variable has no major influence on the conditional model. This implies that institutions should focus on multiple characteristics in order to improve their performance. Efficiency cannot be improved by, for example, only improving student satisfaction or by becoming more selective.

7. Conclusions, policy implications and further research

This paper models the interaction between student dropout, programme quality ratings and graduation rates. Using data from Dutch universities, we examine the heterogeneity in the interaction between the three variables, and the programme and institutional characteristics that influence it.

The results suggest large variation in how academic programmes are able to use the selective nature of first-year dropout to increase student graduation and programme quality ratings. Although some programmes could increase their quality ratings and graduation rates if they would learn from best practice observations, other programmes have higher than expected quality ratings and success rates. Our findings indicate that programme characteristics can significantly influence the relationship between student graduation and programme quality ratings (for given dropout rates). Institutional characteristics, on the other hand, have a limited impact on the relationship between dropout, graduation and programme quality ratings. Moreover, we observe the following characteristics of programmes that succeed in obtaining high graduation rates and high quality ratings: (i) high student programme satisfaction, (ii) high percentage of first-year female students, (iii) low percentage of first-year ethnic minority students, (iv) the presence of an academic dismissal policy, and (v) high

percentage of staff older than 50 years. Finally, we found that the student success and programme quality (for given dropout rates) differ between subjects (eg, ICT, math, economics) of academic programmes.

This paper yields some insightful policy advice. The results indicate that not all academic programmes effectively exploit first-year dropout to increase programme quality ratings and graduation rates. By making this relation more explicit, policymakers can increase the efficiency of higher education. Moreover, the results indicate certain programme and institutional characteristics that explain the variation in outcomes among study programmes. In this respect, policymakers can introduce some measures to improve programme satisfaction. High satisfaction can lead to high rates of performance due to a sense of belonging and loyalty to the institution (eg, Tinto, 1993). Further, we show that a lower share of male students and minority students lead to higher student success. Therefore, a policy implication for society is to better prepare male students and ethnic minority students for higher education. This is in line with recommendations by Bradley *et al* (2010) in vocational and further education. On the other hand, policymakers of academic programmes can try to improve the support services for these groups of students at risk. Next, academic programmes should increase the academic dismissal policy threshold. Owing to the increase in the academic dismissal policy threshold, the selectivity of the first year will increase, such that there will arise a better match between students and the academic programme (Sneyers and De Witte, 2015). Lastly, one can opt to attract older staff. However, the mechanism on why older staff can influence the efficiency of programmes is not clear yet. Older teachers may have, for example, more experience which may lead to a higher student's performance. Finally, our results show an insignificant influence of the student-staff ratio on the efficiency of study programmes. This has an important policy implication since higher education institutions often try to improve graduation rates and quality ratings by focusing on the student-staff ratio.

This paper provides several lines for further research. First, due to data constraints, student graduation and student dropout does not deal with the same cohort of students. Consequently, our results may capture some changes related to different cohorts of students. Further research can replicate our method using data from similar cohorts of students. Second, the study used a sample of Dutch academic programmes. The extent to which the results are generalizable to academic programmes in other countries is unknown. Future research should explore the external validity of our findings. To facilitate further applications, the R-code underlying the empirical application is available upon request.

References

Abbott M and Doucouliagos C (2003). The efficiency of Australian universities: A data envelopment analysis. *Economics of Education Review* **22**(1): 89–97.

Agasisti T (2011). Performances and spending efficiency in higher education: A European comparison through non-parametric approaches. *Education Economics* **19**(2): 199–224.

Archibald RB and Feldman DH (2008). Graduation rates and accountability: Regressions versus production frontiers. *Research in Higher Education* **49**(1): 80–100.

Arnold I and Van den Brink W (2010). Naar een effectiever bindend studieadvies [Towards a more effective binding study advice]. *Tijdschrift voor Hoger Onderwijs en Management* **2010**(5): 10–13.

Astin AW and Solmon LC (1981). Are reputational ratings needed to measure quality? *Change: The Magazine of Higher Learning* **13**(7): 14–19.

Athanassopoulos AD and Shale E (1997). Assessing the comparative efficiency of higher education institutions in the UK by the means of data envelopment analysis. *Education Economics* **5**(2): 117–134.

Avkiran NK (2001). Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. *Socio-Economic Planning Sciences* **35**(1): 57–80.

Avkiran NK and Rowlands T (2008). How to better identify the true managerial performance: State of the art using DEA. *Omega* **36**(2): 317–324.

Bailey T, Calcagno JC, Jenkins D, Leinbach T and Kienzl G (2006). Is student-right-to-know all you should know? An analysis of community college graduation rates. *Research in Higher Education* **47**(5): 491–519.

Benneworth P *et al* (2011). *Quality-Related Funding, Performance Agreements and Profiling in Higher Education*. Center for Higher Education Policy Studies: Enschede.

Bonaccorsi A and Daraio C (2008). The differentiation of the strategic profile of higher education institutions. New positioning indicators based on microdata. *Scientometrics* **74**(1): 15–37.

Bonaccorsi A, Daraio C and Simar L (2007). Efficiency and productivity in European universities: Exploring trade-offs in the strategic profile. In: Bonaccorsi A and Daraio C (eds). *Universities and Strategic Knowledge Creation: Specialization and Performance in Europe*. Specialization and Performance in Europe. Edward Elgar PRIME collection: Cheltenham.

Bound J, Lovenheim MF and Turner S (2010). Why have college completion rates declined? An analysis of changing student preparation and collegiate resources. *American Economic Journal. Applied Economics* **2**(3): 129.

Bradley S, Johnes J and Little A (2010). The measurement and determinants of efficiency and productivity in the further education sector in England. *Bulletin of Economic Research* **62**(1): 1–30.

Calcagno JC, Bailey T, Jenkins D, Kienzl G and Leinbach T (2008). Community college student success: What institutional characteristics make a difference? *Economics of Education Review* **27**(6): 632–645.

Cave M, Hanney S, Kogan M and Trevett G (1991). *The Use of Performance Indicators in Higher Education: A Critical Analysis of Developing Practice*. Jessica Kingsley: London.

Cazals C, Florens J-P and Simar L (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics* **106**(1): 1–25.

Chickering AW and Reisser L (1993). *Education and Identity*. Jossey-Bass: San Francisco.

Daraio C and Simar L (2005). Introducing control variables in nonparametric frontier models: A probabilistic approach. *Journal of Productivity Analysis* **24**(1): 93–121.

Daraio C and Simar L (2007). *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*. Springer: New York.

De Corte E (2014). *Evaluation of Universities in Western Europe: From Quality Assessment to Accreditation*. Belgium, Leuven: KU Leuven, from: http://vo.hse.ru/data/2015/01/14/1106361466/2014-4_DeCorte_En.pdf, accessed 13 April 2015.

De Koning BB, Loyens SMM, Rikers RMJP., Smeets G and van der Molen HT (2014). Impact of binding study advice on study behavior and pre-university education qualification factors in a problem-based psychology bachelor program. *Studies in Higher Education* **39**(5): 835–847.

De Paola M (2009). Does teacher quality affect student performance? Evidence from an Italian university. *Bulletin of Economic Research* **61**(4): 353–377.

Deprins D, Simar L and Tulkens H (1984). Measuring labor-efficiency in post offices. In: Marchand M, Pestieau P and Tulkens H (eds). *The Performance of Public Enterprises*. Elsevier Science Publishers: Amsterdam, pp 243–267.

De Witte K and Hudrlikova L (2013). What about excellence in teaching? A benevolent ranking of universities. *Scientometrics* **96**(1): 337–364.

De Witte K and Kortelainen M (2013). What explains performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete control variables. *Applied Economics* **45**(17): 2401–2412.

De Witte K, Rogge N, Cherchye L and Van Puyenbroeck T (2013a). Accounting for economies of scope in performance evaluations of university professors. *Journal of the Operational Research Society* **64**(11): 1595–1606.

De Witte K, Rogge N, Cherchye L and Van Puyenbroeck T (2013b). Economies of scope in research and teaching: A non-parametric investigation. *Omega—International Journal of Operational Research* **41**(2): 305–314.

Drennan LT and Beck M (2001). Teaching quality performance indicators—key influences on the UK universities' scores. *Quality Assurance in Education* **9**(2): 92–102.

Gijbels D, Van der Rijt J and Van de Watering G (2004). Het bindend studieadvies in het hoger wetenschappelijk onderwijs: worden de juiste studenten geselecteerd? [The binding study advice in higher scientific education: Are the right students selected?]. *Tijdschrift voor Hoger Onderwijs* **22**(2): 62–72.

Hall M (1999). Why students take more than four years to graduate. Paper presented at the Association for Institutional Research Forum, Seattle, WA.

Hosch BJ (2008). Institutional and student characteristics that predict graduation and retention rates. Paper presented at the North East Association for Institutional Research Annual Meeting.

Huisman J (2008). Shifting boundaries in higher education: Dutch hogescholen on the move. In: Taylor J, Ferreira JB, De Lourdes Machado M and Santiago R (eds). *Non-University Higher Education in Europe*. Springer: Milton Keynes, pp 147–167.

Huisman J and Currie J (2004). Accountability in higher education: Bridge over troubled water? *Higher Education* **48**(4): 529–551.

Inspectie van het onderwijs (2009). Uitval en rendement in het hoger onderwijs: achtergrondrapport bij werken aan een beter rendement, from http://www.onderwijsinspectie.nl/binaries/content/assets/Actueel_publicaties/2009/Werken + aan + een + beter + rendement + - + achtergro ndrapport.pdf, accessed 24 October 2013.

Johnes J (1996). Performance assessment in higher education in Britain. *European Journal of Operational Research* **89**(1): 18–33.

Johnes J (2006a). Measuring teaching efficiency in higher education: An application of data envelopment analysis to economics graduates from UK universities 1993. *European Journal of Operational Research* **174**(1): 443–456.

Johnes J (2006b). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review* **25**(3): 273–288.

Johnes J and Taylor J (1990). *Performance Indicators in Higher Education: UK Universities*. Open University Press and the Society for Research into Higher Education: Milton Keynes.

Jongbloed B (2001). *Performance-based funding in higher education: An international survey*. Centre for the Economics of Education and Training, Monash University, Australia, Working Paper n.35.

Jongbloed B (2003). Marketisation in higher education, Clark's triangle and the essential ingredients of markets. *Higher Education Quarterly* **57**(2): 110–135.

Jongbloed BWA, Koelman JBJ, Goudriaan R, de Groot H, Haring HMM and van Ingen DC (1994). *Kosten en doelmatigheid van het hoger onderwijs in Nederland, Duitsland en Groot-Brittannië*. Beleidsgerichte studies Hoger Onderwijs en Wetenschappelijk onderzoek 57, Ministerie van Onderwijs, Cultuur en Wetenschappen. SDU, Den Haag.

Jongbloed B, Salerno C and Kaiser F (2003). *Kosten per Student,* Methodologie Schattingen en een Internationale Vergelijking. CHEPS, Enschede.

Jongbloed B and Vossensteyn H (2001). Keeping up performances: An international survey of performance-based funding in higher education. *Journal of Higher Education Policy and Management* **23**(2): 127–145.

Kokkelenberg EC, Sinha E, Porter JD and Blose GL (2008). *The efficiency of private universities as measured by graduation rates,* Cornell Higher Education Research Institute Working paper, 113.

Lau LK (2003). Institutional factors affecting student retention. *Education* **124**(1): 126–136.

Lee C and Buckthorpe S (2008). Robust performance indicators for non-completion in higher education. *Quality in Higher Education* **14**(1): 67–77.

McMillan ML and Datta D (1998). The relative efficiency of Canadian universities: A DEA perspective. *Canadian Public Policy* **24**(4): 485–511.

McNabb R, Pal S and Sloane P (2002). Gender differences in educational attainment: The case of university students in England and Wales. *Economica* **69**(275): 481–503.

Mellanby J, Martin M and O'Doherty J (2000). The 'gender gap' in final examination results at Oxford University. *British Journal of Psychology* **91**(3): 377–390.

Neumann R (2001). Disciplinary differences and university teaching. *Studies in Higher Education* **26**(2): 135–146.

OCW (2011). Hoofdlijnenakkoord OCW-VSNU, from http://www.vsnu.nl/files/documenten/Domeinen/Accountability/HLA/Hoofdlijnenakkoord_universiteiten_DEF_20111208.pdf, accessed 24 October 2013.

Porter SR (2000). The robustness of the 'graduation rate performance' indicator used in the U.S. news and world report college ranking. *International Journal of Educational Advancement* **1**(2): 10–30.

Reason RD (2003). Using an ACT-based merit-index to predict between-year retention. *Journal of College Student Retention: Research, Theory and Practice* **5**(1): 71–87.

Reason RD (2009). Student variables that predict retention: Recent research and new developments. *Naspa Journal* **40**(4): 172–191.

Robst J (2001). Cost efficiency in public higher education institutions. *Journal of Higher Education* **72**(6): 730–750.

Salerno C (2003). *What We Know About the Efficiency of Higher Education Institutions: The Best Evidence*. CHEPS, Universiteit Twente: the Netherlands.

Schwarz S and Westerheijden DF (2004). *Accreditation and Evaluation in the European Higher Education Area*. Springer Science & Business Media: Dordrecht, the Netherlands.

Suhre CJ, Jansen EP and Harskamp EG (2007). Impact of degree program satisfaction on the persistence of college students. *Higher Education* **54**(2): 207–226.

Scott M, Bailey T and Kienzl G (2006). Relative success? Determinants of college graduation rates in public and private colleges in the US. *Research in Higher Education* **47**(3): 249–279.

Sneyers E and De Witte K (2015). The effect of an academic dismissal policy on dropout, graduation rates and student satisfaction. Evidence from the Netherlands. *Studies in Higher Education*, *in press*.

Stegers-Jager KM, Cohen-Schotanus J, Splinter TAW and Themmen APN (2011). Academic dismissal policy for medical students: Effect on study progress and help-seeking behaviour. *Medical education* **45**(10): 987–994.

Stensaker B and Harvey L (2006). Old wine in new bottles? A comparison of public and private accreditation schemes in higher education. *Higher Education Policy* **19**(1): 65–85.

Stevens PA (2001). *The Determinants of Economic Efficiency in English and Welsh universities*. National Institute of Economic and Social Research: London.

Sweitzer K and Volkwein JF (2009). Prestige among graduate and professional schools: comparing the U.S. News' graduate School reputation ratings between disciplines. *Research in Higher Education* **50**(8): 129–148.

Task force studiesucces (2009). *Eindadvies studiesucces*. Universiteit Leiden: Leiden.

Teichler U (2007). Accreditation: The role of a new assessment approach in Europe and the overall map of evaluation in European higher education. In: Schwarz S and Westerheijden DF (eds.) *Evaluation and Accreditation in the European Higher Education Area*. Kluwer Academic Publishers: Dordrecht.

Tinto V (1993). *Leaving College: Rethinking the Causes and Cures of Student Attrition*. 2nd edn, The University of Chicago Press: Chicago.

Tinto V (2002). Promoting student retention: Lessons learned from the United States. Paper presented at the 11th Annual Conference of the European Access Network, Prato, Italy.

Van Damme D (2000). Internationalization and quality assurance: Towards worldwide accreditation? *European Journal for Education Law and Policy* **4**(1): 1–20.

Van Vught FA and Westerheijden DF (1994). Towards a general model of quality assessment in higher education. *Higher Education* **28**(3): 355–371.

VSNU (2012). Prestaties in perspectief: Trendrapportage universiteiten 2000–2020, from http://www.vsnu.nl/files/documenten/Publicaties/Trendrapportage_DEF.pdf, accessed 13 November 2013.

Appendix A

Definitions used in earlier literature

Table A1   Definitions in earlier literature of 'graduation rates'

Author(s)	Definition/measurement
Gijbels *et al* (2004), Stegers-Jager *et al* (2011)	Percentage of students who complete the first-year
Sneyers and De Witte (2015)	Percentage of students who graduate from the institutions one year after the nominal study time. In order to be included students at university programmes need to graduate within 4 years, students at higher vocational programmes need to graduate within 5 years
De Paola (2009)	The percentage of students who complete the course
Inspectie van het onderwijs (2009)	The percentage of students who graduate within the nominal study time
Porter (2000); Scott *et al* (2006)	The percentage of students who graduate within 6 years
Bailey *et al* (2006)	The percentage of students who graduate within 150% of the normal study time
Calcagno *et al* (2008)	Number of students who earn a degree (no time limit)
Bound *et al* (2010)	The proportion of students who attend college within 2 years of cohort high school graduation and obtain a BA within 8 years of cohort high school graduation

Table A2   Definitions in earlier literature of 'quality ratings'

Author(s)	Definition/measurement
Sweitzer and Volkwein (2009)	Measured by reputation surveys filled in by deans and senior faculty members in each discipline on each campus. Each survey respondent judges the overall 'academic quality of programmes' in their field on a scale from 1 ('marginal') to 5 ('outstanding').
Astin and Solmon (1981)	Measured as the mean rating on six quality criteria (ie 'scholarly and professional accomplishment of the institution', 'preparation of students for graduate or professional school', 'preparation of students for employment after college', 'overall quality of undergraduate education', 'faculty commitment to undergraduate teaching', and 'innovativeness of curriculum and pedagogy')
Drennan and Beck (2001)	Total Quality Assessments obtained from institutional or departmental self-assessment, the production of a self-assessment document and a visit by a team of academic peers. The ratings are graded on a 24-point scale

Appendix B

Robustness test: the AD policy as input variable

Table B1   Efficiency scores that exclude (unconditional) and include (conditional) the programme and institutional characteristics

Unconditional (robust FHD)		Conditional	
		Model 1	Model 2
Average	1.312	1.004	1.008
St. Dev.	0.331	0.038	0.048
Min	0.740	1.000	1.000
Max	2.631	1.741	1.793

The efficiency scores denote the proportionate increase of completion rates and quality ratings, which an academic programme with a given dropout level could achieve if it would operate as efficient as the best practice observation.

Table B2   Influence and significance of programme and institutional characteristics that influence the relationship between dropout, graduation rates and quality ratings

	Model 1	Model 2
Student programme satisfaction	Favourable (<2E-16)***	Favourable (<2E-16)***
First-year female students (%)	Favourable (0.004)***	Favourable (<2E-16)***
First-year ethnic minority students (%)	Favourable (<2E-16)***	Favourable (<2E-16)***
Academic programme	Unfavourable (1.000)	Favourable (0.994)
Number of students	Favourable (1.000)	Unfavourable (1.000)
Student-staff ratio	Unfavourable (0.986)	Favourable (1.000)
Staff > 50 years (%)	Favourable (<2E-16)***	

Note: Bootstrapped p-values between brackets; ***, **, * denote significance at 1, 5 and 10% level.

Appendix C

Which programme characteristics have the largest influence on the efficiency model?

Table C1   Average efficiency scores when some variables are ignored

	Model 1	Conditional Model 2 (Model 1 without variable staff > 50 years)	Model 3 (Model 1 without variable average satisfaction)	Model 4 (Model 1 without variable % of first year female students)	Model 5 (Model 1 without variable % of first year ethnic minority students)	Model 6 (Model 1 without variable AD policy)
Average	1.004	1.004	1.002	1.022	1.013	1.003
St. Dev.	0.038	0.040	0.031	0.091	0.066	0.0267
Min	1.000	1.000	1.000	1.000	1.000	1.000
Max	1.741	1.562	1.617	1.989	1.814	1.516
R^2	0.989	0.985	0.948	0.984	0.987	0.993

The efficiency scores denote the proportionate increase of completion rates and quality ratings, which an academic programme with a given dropout level could achieve if it would operate as efficient as the best practice observation.