Studies in Educational Evaluation 40 (2014) 43–49

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/stueduc

# An approach to estimate degree completion using drop-out rates

Stijn Luca [a,*], Marc Verdyck [b], Marc Coppens [c]

[a] Department of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
[b] Thomas More Kempen, Kleinhoefstraat 4, 2440 Geel, Belgium
[c] Department of Mathematics, KU Leuven, Celestijnenlaan 200B, B-3001 Leuven, Belgium

ARTICLE INFO

*Article history:*
Received 14 April 2013
Received in revised form 27 September 2013
Accepted 9 December 2013
Available online 6 January 2014

*Keywords:*
Completion rate
Graduation rate
Binomial distribution
School effectiveness
Student success

ABSTRACT

A completion rate of an institution $I$ of year $x$ is defined as the proportion of starting students during year $x$ that succeed in graduating at $I$ at some point in the future. In this paper, a new method is proposed to estimate such completion rate. This indicator is entirely based on the population of drop-out students during one academic year $x$ at the institution $I$. The proposed method is not based on a cohort of students so it allows an estimation of the current school effectiveness without substantial delay. Furthermore a statistical framework is presented in which completion rates can be studied. The proposed methodology results in a statistical estimator with a bias that stays small under appropriate assumptions.


## Introduction

Student success plays a crucial role in student careers, the accountability of educational institutions and the development of our society. Completion rates have a long history of being used as indicators of student success. In this paper a new measure for completion rate is studied. A completion rate of an institution $I$ of year $x$ is defined as the proportion of starting students during year $x$ that succeed in graduating at $I$ at some point in the future.

In post-secondary institutions, such a completion rate plays a more crucial role than an on-time graduation rate. Here it is more common that students choose to extend their study period by one or more semesters, e.g. in order to be able to combine studying with a part-time job.

Ideally an unbiased estimator of completion rate can be obtained in a cohort study where students are individually tracked through time. The duration of such cohort implies a substantial delay on the calculation of the completion rate. Moreover the obtained measure does not reflect the current state of school effectiveness. On the contrary, it is rather a result of the education process during the complete cohort period.

This paper aims for two main contributions. On one hand a new measure for completion rate of an institution $I$ is developed. In contrast to cohort studies, substantial delay is avoided and the most up-to-date information that is available is used. For this purpose a measure is developed that is completely based on the population of drop-out students during a specific academic year $x$, e.g. the current year.

On the other hand the performance of this measure is studied as a statistical estimator for completion rate that does not depend on study duration. The estimator is investigated in a statistical framework to study under which assumptions statistical inferences can be made.

This paper discusses completion rate in an educational context. However completion rates are also found in other disciplines. In computer science for instance a completion rate can be viewed as the number of complete files that are successfully transferred from one server to another. Therefore it should be noted that what follows can be adapted for its use in other disciplines.

## Completion rates in education, a review

Very related to completion rates are on-time graduation rates that depend on study-duration. In the United States all institutions of higher education are required by law to publish completion rates. For this reason there is a rich literature on an on-time graduation rate as defined in the No Child Left Behind Act (NLCB, 2002). However in using such graduation rates as a measure for school effectiveness caution is needed. As suggested in Astin (2005), on-time graduation rates and more generally completion rates cannot be adequately addressed without considering the kinds of students who initially enroll.

Nonetheless, differences among the measures that are published has lead to much debate in the United States over the correct

* Corresponding author. Tel.: +32 14562310.
*E-mail address:* stijn.luca@kuleuven.be (S. Luca).

0191-491X/$ – see front matter 
http://dx.doi.org/10.1016/j.stueduc.2013.12.001

rate to use to meet the demands of the NCLB act (Swanson & Chaplin, 2003). The two main difficulties in the calculation of completion rates are a lack of comprehensive sources for data on completion and a lack of consensus on the conceptual and technical definitions related to completion rates. An extensive overview of this matter is given in Hauser and Koenig (2011).

Ideally completion rates can be calculated in a cohort study. In such studies a set of starting students is followed during their study careers and completion rates are calculated after graduation (see e.g. Boden, 2011–2012; Ensminger & Slucarcick, 1992). In practice however this leads to several problems. A cohort study implies the need of longitudinal databases that track individual students through time. The availability or collection of such a dataset is far from obvious (Hauser & Koenig, 2011).

Moreover the duration of the cohort immediately implies a substantial delay on the calculation of a completion rate. This delay can be avoided by the use of an estimator that is based on available data through current and historical records. Such estimations do not require the tracking of individual student over time. However, they can only be viewed as proxies for a true cohort indicator.

In literature many estimator for on-time graduation rates are found. Comprehensive overviews can be found in Miao and Haney (2004), Seastrom et al. (2006), and Swanson and Chaplin (2003). They range from very simple estimators to more complicated estimators which also require more complex demands on data collection systems.

The simple on-time graduation rate (SGR) for instance is defined as the ratio of the number of students $G_x$ that graduate during year $x$ and the total number $n_{x-T}$ of enrollments at institution $I$ in year $x-T$ where $T$ denotes the duration of the study-program under consideration:

$$SGR = \frac{G_x}{n_{x-T}} \tag{1}$$

This simple estimator will likely differ from the true cohort rate because students move in and out of the institution $I$ during the period of $T$ years. Alternatives proposed by Greene and Winters (2002) and Haney (2001) try to reduce this bias by adjusting the number of enrollments in the denominator.

The 'cohort graduation rate' and 'exclusion-adjusted cohort graduation indicator' (Seastrom et al., 2006), the latter being proposed by the National Institute of Statistical Sciences, are examples of cohort graduation rates requiring detailed data on individual students over time.

Swanson and Chaplin (2003) developed the Cumulative Promotion Index for high school graduation rate. This indicator estimates on-time graduation rate as a probability that a starting student will graduate on time. The estimator is based on a so-called synthetic cohort consisting of shortened time periods and only requires data from two school years.

In this paper a new estimator for completion rate is proposed that is completely based on the current population of drop-out students. In contrast to estimators for an on-time graduation rate, commonly found in literature, the indicator does account for students that graduate in a period that is longer than the standard number of years. It is therefore more suitable for post-secondary institutions.

The 'NCES high school completion rate' is another indicator that is based on drop-out counts. However this rate requires drop-out counts in each level of the study-program over the last $T$-years. Such drop-out statistics are often not available (Seastrom et al., 2006).

## From drop-out rates to completion rates

In this section a method is developed to estimate the completion rate of an institution $I$. This rate is defined as the proportion of starting students, that are recruited by an institution $I$ during an academic year $x$ and will graduate at some point in the future.

In particular the goal in this section is to develop an estimation in a way that:

1. The most up-to-date information that is available is used. This is in contrast with existing estimator that are mainly based on historical records.
2. The method is not based on a cohort study. Therefore, information is obtained without substantial delay.
3. The estimation is independent of the study duration and therefore very useful in post-secondary institutions.

### *Methodology*

Up-to-date information can be found by retrieving information from the set of students that drop out during academic year $x$. The number of these drop-out students can be calculated after the end of the enrollment period of academic year $x+1$ by registering those students enrolled during year $x$ but did not reregister the subsequent year. In this way a delay of maximal one year is obtained in contrast to cohort studies where several years are needed to obtain an estimation.

This set can be subdivided into subsets according to the academic year $x-i$ (a previous academic year where $i$ denotes a natural number) during which a student started his study at the institution $I$. Fig. 1 illustrates such partitioning of the population of drop-out students.

When the school is performing well, the set of drop-out students that started during a year $x-i$ should be small. One also expects that the size of the subsets decrease with $i$ and will be negligible from some index $i = T$. Generally it will suffice to apply the partition up till 5–6 subsets (i.e. $T = 3$–$4$). Denote $s$ as the sum of the ratios obtained by dividing the sizes of these subsets by the number of starting students during the academic years $x-i$. The complement of this sum $1-s$ is the new proposed estimation of the completion rate of $I$.

### *Example*

For illustrative purposes an example of the use of the formula is presented based on data of Thomas More University College (UC). The drop-out ratios presented in Table 1 were calculated until the academic year 2010–2011. The table presents the drop-out ratios according to the starting year and the number of years from enrollment to drop-out. Most standard study-programmes at

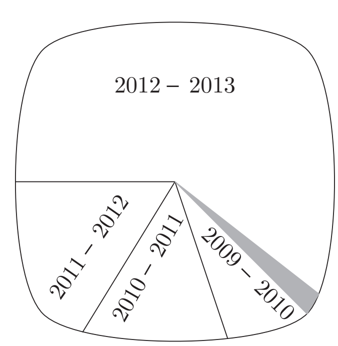


**Fig. 1.** Schematic illustration of the partition of the population of drop-out students. Academic year $x$ is taken as the academic year, $x$ = 2012–2013. For $T$ = 3 the population of drop out-students is partitioned into 5 subsets according to the starting years. Students that drop-out at year $x$ and have started more than 4 years ago correspond to the shaded area and are expected to be small.