# Which students apply for and succeed in STEM higher education? An analysis using Dutch registry data

Melvin Vooren[1]

28th September 2018

Lisbon Economics and Statistics of Education 2019

**Abstract**

In this paper we investigate which factors are important predictors for success in Science, Technology, Engineering, and Mathematics (STEM) higher education. We fit a sequential logit model in which students can either drop out from STEM higher education, or continue studying until they graduate in a STEM field. We utilise rich Dutch registry data on high school exam grades to explain the differences in success and dropout rates. Higher mathematics exam grades seem to increase study success in STEM-related higher education, especially for students that took advanced mathematics in high school. Students with higher grades for Dutch are more likely to graduate in STEM than students with higher grades for English.

**Keywords**: STEM, higher education, study success

**This is a preliminary draft. Please do not cite.**

## 1 Introduction

Improving study success in higher education is high on the agenda in many European countries (European Commission, 2015). A frequently used indicator for study success in higher education is the time-to-degree, which can be measured by the percentage of students that has graduated within the nominal duration of the programme plus at most one additional grace year. In the Netherlands, 59 per cent of the bachelor graduates from universities of applied sciences (UAS) and 72 per cent of the bachelor graduates

[1]Corresponding author. Amsterdam School of Economics, University of Amsterdam. P.O. Box 15867, 1001 NJ Amsterdam, The Netherlands. M.Vooren@uva.nl

from research universities of the cohort that graduated in 2017 meet this definition (Inspectie van het Onderwijs, 2018, p. 174). Apparently there is room to improve on study success, which can also help the shortage on the labour market for STEM graduates. For this, insights in the determinants of study success are needed.

In the United States, high drop out rates in STEM fields are a problem as well. Meyer and Marx (2014) describe the experiences of students who dropped out from engineering in the United States. The most common reasons that students drop out from engineering include performance-related issues, which are most likely the result of difficulties with fitting into the engineering field and the requirements of the programme. In terms of self-confidence and motivation, Litzler et al. (2014) break down the self-reported levels of STEM confidence between gender and ethnicity groups, and find that women on average report lower confidence than otherwise comparable men.

Various interventions aimed to improve student success have been evaluated over the past years (see Brock (2010) for a review). Career-based student involvement programmes are found to be effective in improving study success (Nitecki, 2011). These programmes primarily work through creating a unique institutional subculture. Another way to improve study success is by improving the fit between secondary education and higher education. Using structural equation modeling, Torenbeek et al. (2010) find that a closer resemblance between the higher education programme and the courses students take in secondary education, improve first year study success in the Netherlands significantly.

This paper explores the factors underlying the decision of students to pursue an educational career in higher education in a STEM-related field. We utilise register data from Statistics Netherlands to predict the probability of enrolment in STEM higher education. This data contains detailed information on students careers and grades, both in secondary education and in higher education. After estimating these study choice factors, we estimate a dynamic discrete choice model to model the probabilities of dropping out. We use the results of these models to unfurl the factors

We find that conditional on enrolment, women are more likely to graduate than men in STEM-related fields, especially in university bachelor's programmes. Higher grades for both mathematics and Dutch are associated with higher success rates, but higher grades for English correlates with higher first-year dropout rates and lower graduation rates. Students with a non-Western migration background do not seem to perform worse in terms of first-year dropout rates, but do tend to drop out more often during the year after the final year of the programme.

The remainder of this paper continues as follows. In the next section, we briefly describe the literature on major choice and study success. We give a description of our data in section 3. In section 4 we introduce and explain our model. We describe our results in section 5. Then, we come to a conclusion and a discussion in section 6.

## 2 Literature

### 2.1 Major choice

The traditional schooling decision of whether to obtain a degree or not is a well-researched topic. Closely related to this schooling decision are the returns to schooling, as introduced in Becker's (1962) human capital theory. The returns to education are frequently estimated using the model estimated by Mincer (1974), known as the Mincer equation. More recently, the returns to schooling are estimated using instrumental variable (IV) techniques (Ashenfelter et al., 1999). However, the schooling decision nowadays does not primary entail the decision whether to invest in schooling. In a society where people are becoming more and more educated, the schooling decision also applies to which type of education to attain.

To begin with, each different study programme is associated with a different pay-off in terms of future wages. Kirkeboen et al. (2016) exploit college admission standards in Norway to identify the different payoffs of several fields of study at the admission cutoff to account for self-selection. They find sizable differences in earnings between different studies. In terms of payoff the differences between field of study are also found to be bigger than the effects of attending a more selective institution. Given the differences in payoffs of attending different study programmes, and the self-selection into different programmes, it is important to incorporate wages when modeling the choice of major.

Altonji et al. (2012) estimate a dynamic discrete choice model and review the literature. They argue that the difficulties in finding a causal effect of field of study on wages is closely related to the irreversibility of study decisions. Heterogeneous human capital stocks in different students may lead them to self-select into the field of study based on skills and experiences from earlier life. Also ability, preferences, and uncertainty about students own abilities play a role. Finally, changing in wages during their educational career plays a role. Altonji et al. (2016) develop a dynamic model of educational decision-making. Their results show an effect of major choice on later earnings. However, the skills that lead to success in STEM related fields might also explain differences in earnings. It could very well be that these skills are the same skills that are rewarded in the labour market.

With the use of administrative data from the economics and business programme of the University of Ljubljana in Slovenia, Bartolj and Polanec (2012) investigate the effect of cognitive ability on major choice. The data is modelled in a choice model where students maximise expected utility. They use high school GPA as a measure for general ability, and university grades in major-specific courses as a measure for major-specific ability. The authors find that not only general ability, but also major-specific ability plays a role in determining major choice.

# 3 Data

## 3.1 Sample

We utilise data from Statistics Netherlands. This facility provides cross-sectional registry microdata about every Dutch inhabitant over the past years, split up into numerous different tables from different sources. The main source of the data we used is the Dutch *Dienst Uitvoering Onderwijs*, the division of the Department of Education which administers the educational data of Dutch citizens. Their rich data sets contains information about enrolments, degrees, and secondary education exam courses and grades. We transferred most of the variables directly from the respective source tables to our sample. Appendix X provides details on the sources of the variables we used in our analysis.

To infer whether a student has graduated or dropped out from the STEM programme, we took the following approach. We create two dummy variables *dropout* and *graduated*. The dropout variable takes value one when we observe a change in the main programme that the student is enroled into with respect to the preceding year, without observing a change in their level of education status, and zero otherwise. In this case the student has either switched to a different major, switched to a lower level of post secondary education, or dropped out from the educational system as a whole. When both the student's level of education status and the main programme variable changes, the student has graduated, so in that case the graduated variable takes value one.

## 3.2 Descriptive statistics

Table 1 gives descriptive statistics on the secondary school cohort that immediately enrolled into a STEM higher education programme, right after graduating from secondary school in 2007. In the Dutch system, students are free to elect the courses for their secondary school exam. However, they do have to follow a set of 'tracks', which are the same in all schools. There are four main tracks that students can elect, but there is also room for free electives, which gives students the possibility to combine two tracks.

Table 1: Descriptive statistics school cohort 2006-2007, subject on 2007 STEM enrolment

| | University of Applied Sciences | Reseach University | Total |
|---|---|---|---|
| | (1) | (2) | (3) |
| *(a) exam track* | | | |
| Culture | 100 | 8 | 108 |
| Economics | 336 | 34 | 370 |
| Health | 1,317 | 603 | 1,920 |
| Science | 1,571 | 371 | 1,942 |
| Science + health | 461 | 232 | 693 |
| Other combination | 14 | 0 | 14 |
| *(b) exam level* | | | |
| Havo | 3,363 | 0 | 3,363 |
| Vwo | 436 | 1,248 | 1,684 |
| *(c) migration history* | | | |
| Native | 3,276 | 1,051 | 4,327 |
| Immigrant | 523 | 197 | 720 |
| *(d) gender* | | | |
| Male | 3,122 | 835 | 3,957 |
| Female | 677 | 413 | 1,090 |
| *(e) mathematics level* | | | |
| Regular math (A) | 422 | 30 | 452 |
| Advanced math (B) | 3,377 | 1,218 | 4,595 |
| Total | 3,799 | 1,218 | 5,047 |

*Notes*:

Table 2: Average exam grades school cohort 2006-2007, subject on STEM enrolment

| | HBO | WO | Average |
|---|---|---|---|
| | (1) | (2) | (3) |
| Math (A or B) | 6.5 | 6.9 | 6.6 |
| Dutch | 6.1 | 6.3 | 6.2 |
| English | 6.4 | 6.6 | 6.5 |

*Notes*: the grades are the average grades on the centralised government exam.

The main difference between the four tracks is the level of mathematics. Students from the cultural track are not always obliged to take mathematics, while students in the economics and health are obliged to include mathematics A in their exam. Students that want to follow the science track are obliged to follow mathematics B instead of mathematics A, which is more challenging and has a deeper focus on calculus. Students from other tracks than the science track can also opt to substitute mathematics A for the more advanced mathematics B. Also, the admittance requirements for most STEM fields include mathematics. The final exam grade does not necessarily have to be a pass grade, however.

Nevertheless, the mathematics requirement for STEM education provides us with a high school mathematics grade for all students in our sample. All students are obliged to include both Dutch and English in their exam. The average grades for mathematics, Dutch, and English are displayed in table 2. We focus on the final grades on the centralised government exams. All students take this written exam designed by the designated government body in charge of designing these exams. The exam is the same for each student in a cohort, and is marked independently by both the student's own teacher and a randomly drawn teacher from a different school in the Netherlands. The grades range from scale 1 to 10, where 6 is the passing grade. Grade 10 represents a 100% score, and is therefore seldom awarded.

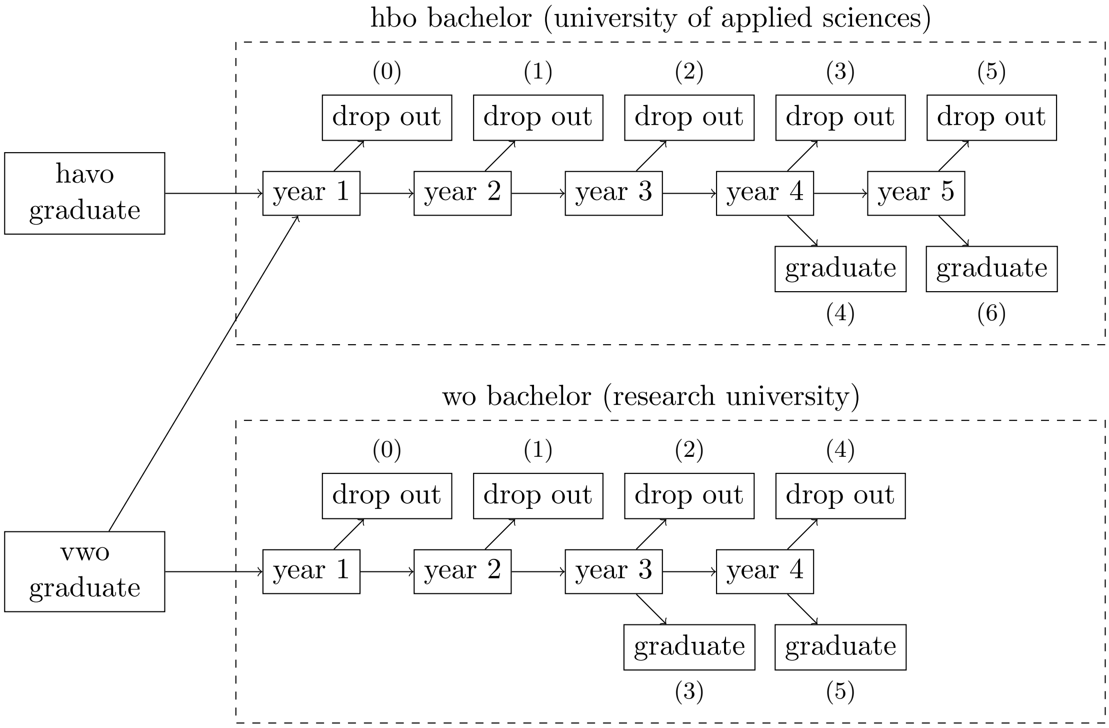
# 4 Methodology

## 4.1 Model

We estimate a sequential logit model (McFadden and Domencich, 1975) of the Dutch higher education system. To do so, we assume that each year, students can either decide to continue studying for another year, or drop out. After having studied for number of years, students can also graduate. In figure 1 we present a stylised version of the

Dutch higher education system. In the Dutch system, only vwo students can enrol into university (wo) bachelor programmes directly from secondary education. They can also choose to enrol into a hbo bachelor programme.

Figure 1: Simplified version of the Dutch higher education system




*Notes:*

All ending states are captured in the *tree* variable, ranging from 0 to 6. This variable is the outcome variable. The values of the tree variable correspond with the labels in figure 1. For instance, when a student dropped out from hbo in the fourth year, the variable takes value 3. Notice that for a wo student, value 3 corresponds with graduating from wo after 3 years, as depicted in figure 1. We estimate two sequential logit models, one for the hbo track, and one for the wo track. Descriptive statistics on the outcome variable are shown in table 3.
The numbers in table 3 show that the majority of students that drop out, do so during the first year. This pattern is the same in both hbo and wo. In hbo, about an equal share of students graduate after four years. The majority of hbo students graduate after five years. In wo, the number of students that graduate after four years is almost twice as high as the number of students that graduate after 3 years.

Table 3: Distribution of STEM study outcomes

| University of Applied Sciences (hbo) | | Research University (wo) | |
|---|---|---|---|
| (1) | | (2) | |
| drop out 1st year (tree=0) | 1,008 | drop out 1st year (tree=0) | 569 |
| drop out 2nd year (tree=1) | 312 | drop out 2nd year (tree=1) | 161 |
| drop out 3rd year (tree=2) | 123 | drop out 3rd year (tree=2) | 101 |
| — | — | graduate in 3 years (tree=3) | 120 |
| drop out 4th year (tree=3) | 99 | drop out in 4th year(tree=4) | 73 |
| graduate in 4 years (tree=4) | 989 | graduate in 4 years (tree=5) | 224 |
| drop out 5th year (tree=5) | 161 | — | |
| graduate in 5 years (tree=6) | 1,085 | — | |
| Total | 3,777 | Total | 1,248 |

*Notes*:

## 5 Results

The results of the dynamic discrete choice model are presented in table 4. Given the fact that most students drop out during the first year of study, we focus on this transition in our analysis, in addition to the transitions to the graduated states. Column 1 shows the odds ratios for the first transition, the decision to continue studying after the first year. Hbo students from the academic high school track are more likely to continue studying than hbo students from the general high school track. Also, a higher grade on the high school mathematics exam seems to correlate with improved succes rates. This relation is emphasised for students that took advanced mathematics in secondary education.

Students with a higher exam grade for Dutch are more likely to graduate in the nominal duration of a hbo programme, being four years. This is not true for the exam grade for English. Students with a higher grade for English are more likely to drop out during the first year, and less likely to graduate within the nominal duration of four years.

Students with a non-Western migration background perform worse in STEM-related higher education than native Dutch students or students with a Western migration background. This difference only takes effect at the last transition, where students can either graduate or decide to drop out. Students with a non-Western background are not less likely to drop out during the first year, but they are less likely to graduate after the nominal duration of the programme plus one additional year. This is an interesting result because most other bad covariates get sorted out during the first year of study.

Female students are equally likely to drop out during the first year of study than men, but are more likely to graduate eventually. This difference in graduation performance is the biggest for female students in a STEM-related bachelor's programme at a research university. Also, in research university STEM bachelor's programmes, this difference starts showing already after the nominal programme duration, while in STEM bachelor's programmes at universities of applied sciences, this difference is only appearing at after the nominal duration plus one additional year.

## 6 Conclusion

In this paper we model the Dutch educational system from the point where students start their higher educational career, up until one year after the nominal duration of the programme. In different phases of the model students can either drop out, continue studying, or graduate. We utilise rich Dutch registry data including grades on centralised government-designed high school exam. This allows for a robust comparison between students from different schools.

We find that female students perform better in STEM-related fields than male students. Female students are more likely to graduate, especially in university bachelor's programmes. This is an interesting finding because the recent literature primarily focuses on the low STEM enrolment rates among females (Arcidiacono et al. (2016) Hunt (2015) Reuben et al. (2014) Venkatesh et al. (2003) Volman and Van Eck (2001)). From our results it seems that, when controlling for ability, measured by high school mathematics and language grades, it seems that female students are more likely to graduate, conditional on enrolment.

A striking result is the low success rate for students with a non-Western migration background. Students with a non-Western migration background do not seem to be more susceptible to drop out during the first year of study, but they massively drop out during the year after the final year of the programme. Policy should be geared to prevent this phenomenon, and more data is needed to explain why this happens, for instance through credit accumulation in the higher education programme.

Table 4: Results sequential logit model

| | University of Appl. Sciences | | Research University | |
|---|---|---|---|---|
| | Odds ratio | St. Error | Odds ratio | St. Error |
| | (1) | (2) | (3) | (4) |
| *(i) continue versus drop out in first year* | | | | |
| Academic track (vwo) | 2.186 | 0.307*** | — | — |
| Math | 1.219 | 0.043*** | 1.025 | 0.073 |
| Math*Advanced | 1.182 | 0.021*** | 1.036 | 0.059 |
| Dutch | 1.029 | 0.043 | 1.000 | 0.071 |
| English | 0.923 | 0.034** | 1.056 | 0.051 |
| Female | 1.188 | 0.128 | 1.180 | 0.147 |
| Immigrant | 1.212 | 0.137* | 0.814 | 0.127 |
| Constant | 0.374 | 0.128*** | 0.549 | 0.273 |
| *(ii) graduate in nominal duration versus drop out* | | | | |
| Academic track (vwo) | 5.509 | 2.679*** | — | — |
| Math | 1.309 | 0.148*** | 1.542 | 0.296** |
| Math*Advanced | 1.035 | 0.067 | 0.891 | 0.143 |
| Dutch | 1.186 | 0.144** | 1.360 | 0.228* |
| English | 0.869 | 0.079** | 0.940 | 0.105 |
| Female | 1.000 | 0.557** | 3.506 | 1.056*** |
| Immigrant | 0.855 | 0.187 | 0.641 | 0.250 |
| Constant | 1.458 | 1.121 | 0.018 | 0.022*** |
| *(iii) graduate versus drop out in nominal duration plus one year* | | | | |
| Academic track (vwo) | 1.709 | 0.575 | — | — |
| Math | 1.020 | 0.087 | 1.229 | 0.280 |
| Math*Advanced | 1.044 | 0.054 | 1.126 | 0.220 |
| Dutch | 1.005 | 0.094 | 1.642 | 0.278*** |
| English | 0.914 | 0.075 | 0.807 | 0.096* |
| Female | 1.967 | 0.587** | 2.806 | 0.981*** |
| Immigrant | 0.472 | 0.100*** | 0.356 | 0.143** |
| Constant | 7.782 | 5.924*** | 0.061 | 0.076** |
| $N$ | 3,715 | | 1,248 | |
| $\chi^2$ | 453.53 | | 137.38 | |

*Notes*: ***, **, * denote 1%, 5%, and 10% significance levels, respectively.
Standard errors in parentheses.

## References


Altonji, J., P. Arcidiacono, and A. Maurel (2016). The analysis of field choice in college and graduate school: Determinants and wage effects. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume V, Chapter 7, pp. 305–396. Amsterdam: Elsevier.

Altonji, J. G., E. Blom, and C. Meghir (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics 4*(1), 185–223.

Arcidiacono, P., E. M. Aucejo, and V. J. Hotz (2016). University differences in the graduation of minorities in stem fields: Evidence from california. *American Economic Review 106*(3), 525–562.

Ashenfelter, O., C. Harmon, and H. Oosterbeek (1999). A review of estimates of the schoolingrearnings relationship, with tests for publication bias. *Labour Economics 6*(4), 453–470.

Bartolj, T. and S. Polanec (2012). College major choice and ability: Why is general ability not enough? *Economics of Education Review 31*(6), 996–1016.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy 70*(5), 9–49.

Brock, T. (2010). Young adults and higher education: Barriers and breakthroughs to success. *The Future of Children 20*(1), 109–132.

European Commission (2015). *Dropout and Completion in Higher Education in Europe, Main Report.* European Union.

Hunt, J. (2015). Why do women leave science and engineering? *ILR Review 69*(1), 199–226.

Inspectie van het Onderwijs (2018). *De Staat van het Onderwijs, Onderwijsverslag 2016-2017.* The Hague: Netherlands Inspectorate of Education.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics 131*(3), 1057–1111.

Litzler, E., C. C. Samuelson, and J. A. Lorah (2014). Breaking it down: Engineering student STEM confidence at the intersection of race/ethnicity and gender. *Research in Higher Education 55*(8), 810–832.

McFadden, D. and T. Domencich (1975). *URBAN TRAVEL DEMAND - A BEHAVIORAL ANALYSIS.* Oxford, England: North-Holland Publishing Company Limited.

Meyer, M. and S. Marx (2014). Engineering dropouts: A qualitative examination of why undergraduates leave engineering. *Journal of Engineering Education 103*(4), 525–548.

Mincer, J. (1974). *Schooling, Experience, and Earnings.* National Bureau of Economic Research, Inc.

Nitecki, E. M. (2011). The power of the program: How the academic program can improve community college student success. *Community College Review 39*(2), 98–120.

Reuben, E., P. Sapienza, and L. Zingales (2014). How stereotypes impair women's careers in science. *PNAS 111*(12), 4403–4408.

Torenbeek, M., E. Jansen, and A. Hofman (2010). The effect of the fit between secondary and university education on firstyear student achievement. *Studies in Higher Education 35*(6), 659–675.

Venkatesh, V., M. G. Morris, G. B. Davis, and F. D. Davis (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly 27*(3), 425–478.

Volman, M. and E. Van Eck (2001). Gender equity and information technology in education: The second decade. *Review of Educational Research 71*(4), 613–634.