Mind the Gap: Measuring Academic Underachievement Using Stochastic Frontier Analysis

By DENI MAZREKAJ, KRISTOF DE WITTE AND THOMAS P. TRIEBS*

Please do not circulate without the permission of the authors.

We propose using Stochastic Frontier Analysis to estimate pupils' academic underachievement. We model underachievement as the gap between expected achievement and actual achievement, not due to a learning disability. Our data are a panel for 2,228 Belgian pupils observed over six years of primary education. We find that the average underachievement gap is 23.5%. That is, the average pupil does not exploit about one fourth of her potential. Gifted pupils appear to underachieve as much as non-gifted pupils. We also find that class size is a determinant of underachievement. The association between class size and underachievement is non-monotonic; with an underachievement minimum at a class size of about 20 pupils.

* Deni Mazrekaj (*corresponding author*): University of Oxford, Department of Sociology, 42-43 Park End Street, OX1 1JD Oxford, United Kingdom (e-mail: deni.mazrekaj@sociology.ox.ac.uk); Leuven Economics of Education Research (LEER), KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium (e-mail: deni.mazrekaj@kuleuven.be); Kristof De Witte: Leuven Economics of Education Research (LEER), KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium (e-mail: kristof.dewitte@kuleuven.be); United Nations University – Maastricht Economic and Social Research Institute on Innovation and Technology (UNU-MERIT), Boschstraat 24, 6211 AX Maastricht, the Netherlands (e-mail: k.dewitte@maastrichtuniversity.nl); Thomas P. Triebs: School of Business and Economics, Loughborough University, LE11 3TU Loughborough, United Kingdom (e-mail: t.triebs@lboro.ac.uk). The data for this study are protected by a confidentiality agreement and we are precluded from sharing the data with others. Interested readers can contact the corresponding author for information on how to access the data. The Stata code to replicate our results is available upon request. For their helpful comments and suggestions, we would like to thank Hung-Jen Wang, Subal Kumbhakar, Eric Hanushek, John Ruggiero, Lori Taylor, Frank Verboven, Bieke De Fraine, Jan Van Damme, Ides Nicaise, the participants of the Conference of the International Association for Applied Econometrics, as well as the seminar and workshop participants at Carleton University, Loughborough University, Politecnico di Milano, University of Lisbon, Maastricht University, University of Extremadura, and KU Leuven. Deni Mazrekaj acknowledges funding by the Research Foundation Flanders (FWO) as Aspirant [grant number 1172519N] and by the and by the European Research Council [grant number 681546]. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

Keywords: Academic Underachievement; Gifted Pupils, Class Size; Stochastic Frontier Analysis.

1. Introduction

Underachievement in education is a waste of talent and resources. It is important for both the individual and society that education enables individuals to fully exploit their potential, i.e. transform ability into maximum possible outcome. Despite potential performance gains from reducing underachievement, most research has focused on the measurement of low achievement, not *under*achievement. Broadly defined, low achievers have poor outcomes relative to their peers but do not necessarily underperform given their potential. By contrast, underachievers exhibit a severe discrepancy between expected achievement and actual achievement.[1] Whereas low achievers can easily be identified by observing outcomes, e.g. test scores, the identification of underachievers is difficult as potential outcomes are unobservable. It is particularly challenging to identify gifted underachievers who typically have average or even high outcomes, but still perform below their true ability. Besides wasting resources and talent, not identifying underachievers risks boredom and demotivation (Acee, et al., 2010), potentially leading to a downward spiral in performance. Also, underachievement increases the risk of dropout (Peterson, 2000), which has severe consequences in terms of earnings, health, life expectancy, and overall happiness (Oreopoulos & Salvanes, 2011).

Our contribution is twofold. First, we propose to model underachievement using regression-based Stochastic Frontier Analysis (SFA) in the context of an Education Production Function (Hanushek, 1986). SFA is commonly applied to estimate unobserved managerial inefficiency in firm production where inefficiency is the gap between actual and potential output given inputs (Anaya & Pollitt, 2017; Badunenko & Kumbhakar, 2017; Ferrantino & Ferrier, 1995). Intuitively,

[1] As common in the underachievement literature, we assume this discrepancy is not due to learning disabilities (McCoach & Siegle, 2003).

SFA estimates a frontier (or benchmark) to obtain pupils' potential test scores given observed characteristics such as ability and socioeconomic status, and compares this frontier with actual test scores. SFA achieves this by decomposing the error term in a regression model into a symmetric normal random variable that represents measurement error, and an asymmetric negative half-normal random variable that represents underachievement. Thus, we define underachievement as the difference between an estimated best practice frontier and an individual's actual performance.

Our second contribution is to apply the model to the analysis of the influence of class size on underachievement. Class size is important because it is often at the discretion of the school manager and has budget implications (Denny & Oppedisano, 2013; Hoxby, 2000). Whereas there is a large literature on the influence of class size on achievement, to our best knowledge, we are the first to analyse the influence on underachievement. Prior results for the influence on achievement are mixed. Whereas some studies found that larger classes reduce pupils' achievement (Bressoux, 2009; Krueger, 1999), some studies found no effect (Dieterle, 2015; Hoxby, 2000), and some found that larger classes improve pupils' achievement (Denny & Oppedisano, 2013). We are able to potentially reconcile these diverse results, by applying a specific SFA model that allows for a non-monotonic influence of class size on academic underachievement (Wang, 2002).

We apply the model to unique longitudinal survey data for Belgian pupils. The data include 2,228 pupils from the Flemish community of Belgium observed over six years of primary education. Primary education in Flanders is a good setting to study underachievement, because there is no ability grouping or tracking in Flemish primary education and no standardized exams. Teachers are likely to teach for the average pupil (Van Klaveren & De Witte, 2014), which may result in underachievement for the entire distribution and demonstrate the usefulness of the SFA model.

We find that average underachievement is 23.5%. That is, the average pupil only uses about three quarters of his/her potential. There is no evidence that underachievement systematically varies with gender, ethnicity, or ability. Gifted pupils appear to underachieve as much as non-gifted pupils. We also observe that underachievement peaks in the third grade. Finally, we find that the association between class size and underachievement is non-monotonic; underachievement is lowest in classes of about 20 pupils. Below this threshold, larger classes decrease underachievement, and above this threshold, smaller classes decrease underachievement.

The SFA method for measuring underachievement departs from the three methods widely used in the previous literature. Irrespectively of the method used, previous estimates of underachievement vary from as low as 9% (Schick & Phillipson, 2009) to as high as 49% (Reis, Colbert, & Hébert, 2004); see White, Graham, and Blaas (2018) for a review. The first method, nomination, uses teachers', parents', or peer assessment to identify underachievement (Abelman, 2006; Lau & Chan, 2001). Nomination is widely used by practitioners such as student counsellors. Although easy to apply, this measure suffers from subjectivity bias and often fails to identify gifted underachievers. The share of underachievers may depend on who does the nominating. For instance, Lau and Chan (2001) found that out of 15 potential underachievers, only three were nominated by both teachers and pupils' peers. One potential solution may be to ask pupils themselves whether they are underachieving (Gohm, Humphreys, & Yao, 1998). However, pupils are unlikely to be aware of their true potential, especially at a young age. Ziegler and Stoeger (2003) found that most pupils assessed themselves to be of average intelligence, regardless of their IQ score.

The second method identifies underachievers through a comparison of aptitude test scores (e.g. IQ test) with achievement test scores (e.g. mathematics or reading). There are four varieties: the absolute split method, the simple difference score method, the regression method, and the optimal achievement model. First, the absolute split method defines underachievers as pupils who score higher than a certain threshold (e.g. top 5%) on the aptitude test but score lower than a certain threshold (e.g. bottom 5%) on the achievement test. This is the most popular method in academic research and is typically employed to identify gifted underachievers (Baker, Bridger, & Evans, 1998; Matthews & McBee, 2007; McCoach & Siegle, 2003; Schick & Phillipson, 2009; Figg, Rogers, McCormick, & Low, 2012; Reis, Colbert, & Hébert, 2004; Ritchotte, Matthews, & Flowers, 2014). Second, the simple difference score method calculates a discrepancy score by subtracting the standardized achievement test score from the standardized aptitude test score. If the discrepancy score is higher than a specified threshold, usually one standard deviation, a student is identified as an underachiever (Obergriesser & Stoeger, 2015; Stoeger & Ziegler, 2013; Ziegler & Stoeger, 2003). A third variety regresses achievement test scores on aptitude test scores and defines underachievers as observations that lie a certain distance below the regression line, i.e. a have a sufficiently large positive error term (Dixon, Craven, & Martin, 2006; Preckel & Brunner, 2015). Although these methods are less subjective than nomination, they require an arbitrary threshold, the choice of which influences the amount of underachievement (White, Graham, & Blaas, 2018). As a fourth variety, the optimal achievement model aims to correct for the arbitrariness of the threshold by converting both aptitude test scores and achievement test scores to logits and using a 95% confidence interval as a threshold for the discrepancy (Phillipson, 2008; Phillipson & Ka-on Tse, 2007). Although this method provides a less arbitrary threshold, it is highly sensitive to outliers and it does not account for any control variables. These regression-based methods are conceptually similar to SFA but there are important differences. In contrast to SFA, standard regression errors are deviations from the average performance not best practice performance. Also, as it is likely that achievement scores have a random error component, the SFA method explicitly separates this random component from underachievement.

The third strand of literature measures underachievement using Data Envelopment Analysis (DEA) (Silva Portela, 2001; Thanassoulis, 1999). This method is conceptually similar to SFA, and like SFA, it is widely used to estimate managerial inefficiency in production. DEA estimates a production frontier indicating the potential scores pupils could achieve and compares it with the scores pupils actually achieve. Consequently, the DEA solves most problems identified above: it is not based on subjective judgment, it does not include an arbitrary threshold, it compares the performance of a student to the best performers in the sample, and it allows for environmental factors. A drawback of DEA models is that they do not allow for a stochastic error, i.e. they assume that the outcome variable is measured without error. Given that both aptitude and achievement test scores are an imperfect proxy of aptitude and achievement respectively, this assumption is unlikely to hold. Therefore, DEA may yield biased underachievement estimates (Ehrgott, Holder, & Nohadani, 2018; Ruggiero, 2004; Schiltz, De Witte, & Mazrekaj, 2019). Moreover, even in a conditional DEA model, it is difficult to control for a wide variety of control variables as the underlying kernel function suffers from dimensionality issues (De Witte & Kortelainen, 2013).

The remainder of the paper is structured as follows. Section 2 provides information on the Flemish education system. Section 3 describes the data and sample restrictions. Section 4 formulates the Stochastic Frontier Analysis model. Section 5 estimates underachievement and examines the influence of class size on underachievement. The paper ends with a discussion of the results and several limitations of the analysis in Section 6.

2. The Flemish Education System

The Flemish education system provides compulsory education between the ages of six and 18 or until a younger age if a student has already obtained a high school diploma. Before children enter compulsory education, they can enrol into kindergarten from the age of 2.5. With a participation rate of about 98.8% (Eurydice, 2018), almost all children attend kindergarten. Although most children enter primary education at age 6, parents may decide to enrol their child into primary education already at the age of five.

Primary education lasts for six years until the age of 12. A pupil may spend at most eight years in primary education. The class committee (mostly consisting of the school principal and the teachers) decides whether a pupil may continue to the next school year or has to repeat the grade. In school year 2017-2018, grade retention was 1.94% in Flemish primary education (Flemish Parliament, 2018). If pupils complete all six years of primary education, they receive a certificate of primary education. Parents may choose any elementary school for their child, there are no catchment areas or standardized admission tests. Places are allocated on a first-come first-serve basis until the capacity of the school is reached. Also, there is no ability grouping. The school board decides how pupils are distributed among classes and the number of pupils per class. In general, one teacher teaches all the subjects, although specialists might teach in some schools. Each school year, a new teacher is assigned to the class. Thus, teachers do not follow their classes over the years. Upon successful completion of primary education, pupils enter a tracking system in secondary education at age 12 that includes four main tracks: the general track (ASO), the technical track (TSO), the vocational track (BSO), and the arts track (KSO).

3. Empirical Methodology

Our empirical model of underachievement starts with a standard Education Production Function (Hanushek, 1986) that relates an output to inputs as well as a number of control variables. In our study, outputs and inputs are at the level of the individual *i* and control variables are at the level of the teacher *t* or school s:

(1) $$y_i = f(x_i) + g(c_{ts}) + \epsilon_i$$

In **Equation 1**, y_i is an individual's output, x_i denotes the inputs, c_{ts} corresponds to a set of control variables at the level of the teacher t and school s and ϵ_i captures unobserved underachievement as well as random noise. Our measure of output is a mathematics test score and our input variables are ability (measured by IQ score), gender, ethnicity and socioeconomic status. Finally, our control variables are school and school year indicators as well as teacher's gender, experience, effort and motivation. Ignoring these school and teacher characteristics would overestimate pupils' underachievement (Goldhaber, 2016). We describe these variables in detail in the next section.

In this model, the error term captures unobserved underachievement and random noise. In a second step, we decompose the error term ϵ_i into two components: an i.i.d. random error v_i on the one hand, and underachievement u_i on the other. The complete model is:

(2a) $$y_i = f(x_i) + g(c_{ts}) + \epsilon_i$$

(2b) $$\epsilon_i = v_i + u_i$$

We estimate underachievement applying the Stochastic Frontier model (Aigner, Lovell, & Schmidt, 1977; Meeusen & van Den Broeck, 1977).

To identify the two error components, we require specific distributional assumptions. Following Wang (2002), we assume a normal distribution for the stochastic error v_i with zero mean and variance σ_v^2, and a truncated normal distribution at zero from above for underachievement u_i, with mean μ_i and variance σ_i^2. Thus, the distributional assumptions are as follows:

(3a) $v_i \sim N(0, \sigma_v^2)$

(3b) $u_i \sim N^+(\mu_i, \sigma_i^2)$

By imposing these two distributional assumptions, it is possible to separate underachievement from random noise using the calculation outlined in Jondrow, et al. (1982). Moreover, by construction, underachievement will be equal to or greater than zero (there are no overachievers).

Intuitively, the SFA model estimates an achievement frontier indicating the potential test scores pupils could achieve given observable characteristics and compares it with the test scores pupils actually achieved. **Figure 1** illustrates this method for a single input: ability measured by IQ. The potential test score is marked by point B and the actual test score by point A. The gap between the two is represented by the line AB which in turn is decomposed into underachievement AC, and measurement error BC. Thus, given her IQ score, the pupil obtains a test score marked by point A. The counterfactual is that with more effort (less underachievement) she could have obtained a higher test score marked by point C.

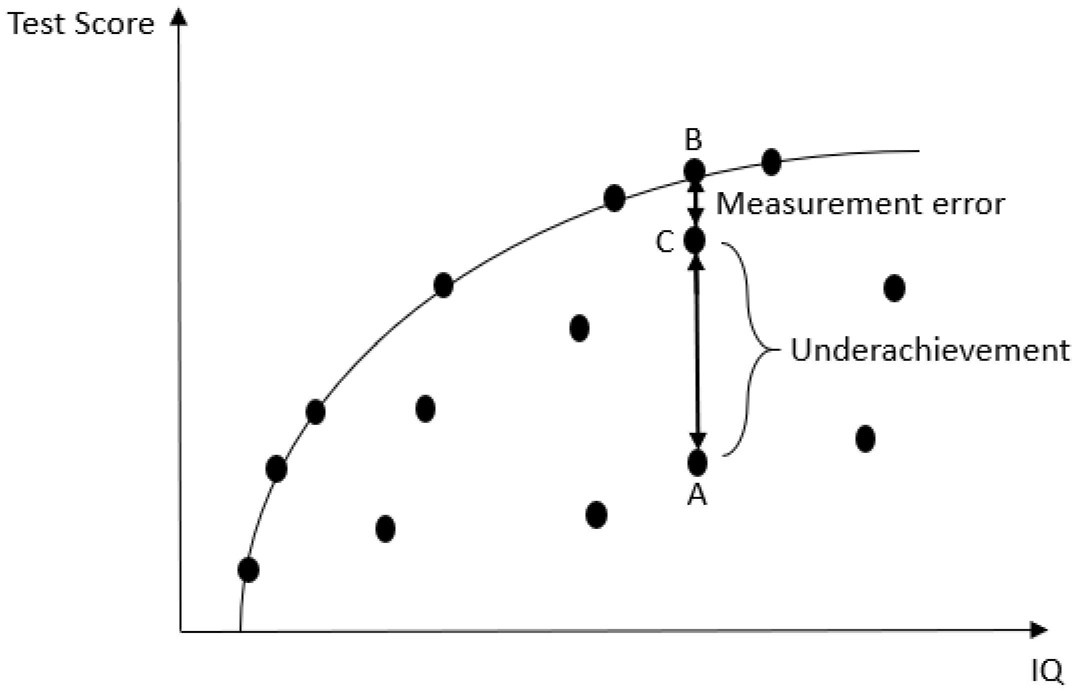

FIGURE 1: TWO-DIMENSIONAL REPRESENTATION OF UNDERACHIEVEMENT USING STOCHASTIC FRONTIER ANALYSIS

Notes. The discrepancy between the potential test score B and the actual test score A is interpreted as underachievement (AC) and measurement error (BC).

Whereas underachievement determines math scores, it would be reasonable to think that underachievement itself has its determinants. We augment the model above by adding class size as a determinant of underachievement. Instead of influencing achievement directly (as part of the frontier) it influences achievement indirectly through underachievement. Following the approach by Wang (2002), we let the mean and the variance of the underachievement distribution be a function of class size z.[2] We add to the model above:

(4a) $\mu_i = z_i\delta$

(4b) $\sigma_i^2 = e^{z_i\gamma}$

This parametrization for the determinants of underachievement is attractive for two reasons. First, it allows the relationship between the determinants of underachievement z and underachievement itself to be non-monotonic, i.e. the marginal influence of z on u_i can change signs. Second, underachievement and the influence of its determinants are estimated together. Estimating underachievement first and then estimating the influence of its determinants in a second stage would lead to biased estimates for underachievement (Wang & Schmidt, 2002). The reason is that the estimation of underachievement would exclude the determinants of underachievement z for the construction of the achievement frontier, introducing selection bias. The direction of this bias depends on the correlation between the inputs x and the determinants of underachievement z,

[2] It should be noted that alternative SFA model specifications can also be modelled in that way (e.g. the four-component model that estimates also persistent and time-varying inefficiency, see for instance Kumbhakar, Lien, and Hardaker (2014) and Lien, Kumbhakar, and Alem (2018)). However, as these more recent SFA specifications make strong assumptions on the decomposition of the error term, and consequently on our estimate of underachievement, we consider this as scope for future research.

but the bias exists even if the correlation is zero. In addition, the bias from the first stage would spill over to the second stage, leading to biased marginal coefficients of the determinants of underachievement. Using a one stage approach, we include the determinants of underachievement z directly in the estimation of underachievement, avoiding these endogeneity issues.

4. Data

We apply the proposed Stochastic Frontier Analysis model to data from the Flemish community of Belgium. The dataset, SiBO (*Schoolloopbanen in het Basisonderwijs*), includes a random sample of 6,138 pupils, nested in 196 schools, which were followed from the last year of kindergarten (2002-2003) until the first year of secondary education (2010-2011). Thus, most pupils were born in 1997. The data oversample pupils from a lower socioeconomic background as the goal of the survey was to study school outcomes of disadvantaged pupils. This is beneficial for our investigation of underachievement as prior evidence shows that pupils from a lower socioeconomic background are more likely to underachieve. For instance, Wyner, Bridgeland and Dilulio (2007) found that 44% of low-income pupils in the U.S. who scored in the top ten per cent in the first grade, did not score in the top ten per cent in the fifth grade. Moreover, gifted low-income pupils progressed at half the rate of their gifted high-income peers.

Our measure of output for the education production function is the mathematics test score from a test taken at the end of each school year.[3] To enable comparison of test scores across the years, we standardize the mathematics test score variable by year to have zero mean and unit variance.

[3] These tests were specifically designed each year to fit the needs of pupils. For instance, kindergarten focused on counting by means of picture associations, while in the third year of primary education, the emphasis lay on multiplication and division.

Although the data also include language (reading and writing) tests, we only use the mathematics score because language tests are subdivided into five different tests, each consisting of two different versions, making it difficult to interpret and compare these tests.

In the selection of the inputs, we follow the previous literature on efficiency in education (see De Witte and López-Torres (2017) for an extensive overview). The main input of interest is an IQ test score.[4] The test is a combined version of the CIT-3-4 verbal cognitive test (Stinissen, Smolders, & Coppens-Declerck, 1975) and the non-verbal Raven's Standard Progressive Matrices Test (Raven, 2000). Whereas the first test is particularly suited for the Flemish primary education pupils, the latter is a test widely used in the underachievement literature (Lau & Chan, 2001; Phillipson, 2008; Phillipson & Ka-on Tse, 2007; Obergriesser & Stoeger, 2015) as well as the psychological literature on gifted pupils overall (see Worrell, et al. (2019) for a review). Combined, these two cognitive tests assess pupils' verbal and non-verbal abilities. The detailed procedure of the construction of the combined Flemish version of the CIT-3-4 and the Raven's cognitive tests can be found in Hendrikx, et al. (2008). The cognitive test was administered in the school year 2005-2006 when pupils were in third grade (usually at age 9). As such, previous mathematics tests might influence the IQ score (Ritchie & Tucker-Drob, 2018), potentially introducing a source of endogeneity (Elwert & Winship, 2014). Nonetheless, if we take mathematics test scores at the beginning of kindergarten, or at the beginning of the first grade as an alternative proxy for ability, our average estimate of underachievement is robust. It only slightly decreases by 0.7 or 1.1 percentage points depending on the control variables included (not reported).

[4] Although our IQ measure closely follows the current practice in the underachievement literature, ideally, we would also include measures of non-cognitive skills. Unfortunately, our dataset does not include such measures.

We include the following additional inputs. Gender is an indicator with value 1 for males and 0 for females. Ethnicity is an indicator given a value of 1 if either the pupil or one of the parents was born abroad, and a value of 0 otherwise. The proxy for Socio-Economic Status (SES) is the first principle component for the following seven variables: highest diploma father, highest diploma mother, employment status father, employment status mother, occupational level father, occupational level mother, and family income. The data were collected through a questionnaire filled in by the parents when their children were in the first year of primary education. A higher value for the SES variable indicates a higher socioeconomic status. Further, we also include an indicator for grade retention given a value of 1 if the pupil repeats the current grade, and a value of 0 if the pupil progressed from the last grade. We use school and school year fixed effects to compare pupils only within the same school and year. This captures unobserved heterogeneity due to, e.g., differences in school policy. Finally, we add four variables to control for teacher characteristics. These include, gender (1 = male, 0 = female), experience in years, effort (hours at home spent on work per week in addition to regular hours), and (intrinsic) motivation. The latter is measured as an indicator given a value of 1 if the teacher agreed with the statement *"for me, there is no better job than being a teacher"*. Given that, within each grade, teachers generally teach only one class, teacher fixed effects are highly collinear with class size such that we cannot include teacher fixed effects.

We restrict the sample in three ways. First, we only include observations for primary education and drop observations from kindergarten and secondary education because following individuals across these types of education is difficult in the dataset. Second, we restrict the sample to observations that include information on the family's socioeconomic status. Although achievement and cognitive ability data were collected for the full sample of pupils, the parental survey was administered to a subsample of 3,534 pupils only. The exact subsampling procedure is reported in

Reynders, Nicaise, and Van Damme (2005). Third, we drop observations for 303 pupils with reported learning disabilities. As mentioned before, these pupils are generally not included in the definition of underachievement in the literature (McCoach & Siegle, 2003). Fourth, as we also analyse the development of underachievement over time, we remove 900 pupils for whom we do not observe all six grades of primary education. Finally, we drop 103 pupils with missing values for at least one of the variables.

Our final dataset is a balanced panel including 2,228 pupils in 168 schools over 6 years of primary education. Descriptive statistics for the pooled sample are provided in **Table 1**. Our sample includes slightly more girls than boys and about 19% of the pupils have a foreign background. Further, almost 4% of the pupils have repeated a grade at least once. This is slightly higher than the official statistics, and can be explained by the oversampling of pupils from a low socioeconomic status. Teachers appear to be mostly male with about 25 years of experience. Moreover, and in line with other information about the work-life balance of Flemish teachers,[5] they spend almost 20 additional hours outside regular 29 working hours and only about 21% do not feel motivated for their job. Finally, the average class size is 19. Maximum size is 32 and the minimum size is 1. The smallest classes are in rural areas, where schools are significantly smaller.[6] The class size distribution is roughly normal as shown by **Figure A1** in the Appendix.

[5] https://onderwijs.vlaanderen.be/nl/onderzoek-tijdsbesteding-leraren-basis-en-secundair-onderwijs

[6] Removing these very small classes leaves the results virtually unchanged (not reported).

TABLE 1 – DESCRIPTIVE STATISTICS

	Mean	Std. Dev.	Minimum	Maximum
Output				
Math test score	93.050	14.929	48	130
Math test score (z-score)	0.000	1	-4.238	2.774
Inputs				
IQ	13.608	2.171	4	18
IQ (z-score)	0.000	1	-4.455	2.284
Gender (1 = Male)	0.473	0.499	0	1
Ethnicity (1 = Foreign)	0.193	0.395	0	1
Socioeconomic status (index)	0.130	0.861	-2.088	2.073
Grade retention (1 = repeated grade)	0.037	0.188	0	1
Teacher gender (1 = Male)	0.662	0.473	0	1
Teacher experience (years)	25.178	7.506	2	43
Teacher additional hours	19.645	7.083	8	50
Teacher motivation (1=motivated)	0.791	0.221	0	1
Determinant of achievement				
Class size	18.915	4.995	1	32
Number of pupils		2,228		
Number of schools		168		
Number of observations[a]		13,368		

5. Results

In this section we give the estimation results for underachievement and the influence of class size. First, we show results for the full sample. Then, we report results by gender, ethnicity, and whether the pupil is gifted.

5.1. Overall Underachievement

Table 2 gives the estimation results for the full sample. Whereas column (4) includes all input and control variables, columns (1) to (3) report nested models. The nested models help us better understand the interactions between the independent variables and whether their inclusion has any influence on underachievement. All models control for school and school year fixed effects assuring that we compare pupils only within the same school and school year. Put differently, we control for all unobserved school and school year specific influences. Moreover, all models include underachievement and class size as a determinant. Column (1) only includes pupils' standardized IQ scores as an input which has a positive and sizable association with pupils' mathematics test scores. If the IQ score increases by one standard deviation, the mathematics test score increases by 0.569 standard deviations. But our main interest (and the advantage of the SFA model) is the estimate for underachievement, which amounts to about 22%, an economically and statistically significant estimate. That is, the average pupil, given her IQ score, could increase her mathematics score by 22%. We find no evidence that class size is an economically or statistically significant determinant of underachievement.

In column (2), we add three additional inputs to the education production function: gender, ethnicity, and socioeconomic status. Note that the influence of IQ (ability) hardly decreases. This confirms the intuition that ability is independent of these other background variables. However, even when controlling for ability, there is an economically and statistically significant influence for gender, ethnicity and socio-economic background. Boys outperform girls on the mathematics test by about 0.43 standard deviations. Having a foreign background (at least one parent born outside Belgium) reduces the score by 0.075 standard deviations. Increasing socioeconomic status by one standard deviation increases the score by about 0.012 standard deviations (0.861*0.141), a relatively small influence. The influence of socio-economic background is only about a quarter of the influence of ability.

In column (3), we add an indicator for grade retention. Unsurprisingly, pupils who repeat the grade score significantly lower on the mathematics test by about 0.3 standard deviations. This suggests that the objective of retention – to give pupils the ability to catch-up in terms of knowledge – is not achieved.

In the last column, our preferred model, we also add teacher characteristics. The coefficients for the pupil inputs remain virtually unchanged; as expected, teacher and pupil characteristics are independent. The results for the teacher characteristics show that achievement is higher when the teacher is male and motivated. However, the influence of the teacher's gender is much smaller than the pupil's gender. Surprisingly, contrary to earlier literature (Compen, De Witte, & Schelfhout, 2019), teacher's experience and effort (approximated by teachers' overtime) do not affect pupils' achievement. For our preferred model, average underachievement is 23.5%. This is slightly higher than for the other models, but the estimate varies little across columns. Once we control for ability, adding additional control variables does not change the estimate for underachievement. The same is true for the influence of class size.

TABLE 2 – ESTIMATING OVERALL UNDERACHIEVEMENT

	(1)	(2)	(3)	(4)
Inputs and Controls				
IQ (std.)	0.569***	0.539***	0.533***	0.533***
	(0.008)	(0.008)	(0.008)	(0.008)
Gender (1 = Male)		0.428***	0.427***	0.428***
		(0.013)	(0.013)	(0.013)
Ethnicity (1 = Foreign)		-0.075***	-0.069***	-0.068***
		(0.021)	(0.021)	(0.021)
Socioeconomic status (index)		0.141***	0.134***	0.134***
		(0.009)	(0.009)	(0.009)
Grade retention (1 = repeated grade)			-0.291***	-0.294***
			(0.037)	(0.037)
Teacher gender (1 = Male)				0.088***
				(0.023)
Teacher experience (years)				0.001
				(0.002)
Teacher additional hours				-0.002
				(0.002)
Teacher motivation (1=motivated)				0.548***
				(0.216)
Fixed effects:				
School year	Yes	Yes	Yes	Yes
School	Yes	Yes	Yes	Yes
Determinant of achievement				
Class size	0.001	0.002	0.000	0.000
	(0.003)	(0.002)	(0.001)	(0.001)
Underachievement				
Overall underachievement	0.219	0.230	0.230	0.235
	[0.066]	[0.077]	[0.075]	[0.077]
Number of pupils	2,228	2,228	2,228	2,228
Number of observations[a]	13,368	13,368	13,368	13,368

Notes. Standard errors are in parentheses. Standard deviations are in squared brackets. Outcome in all models is the mathematics test score standardized by school year.

[a] Pupils are observed in all six grades of primary education. Nonetheless, some pupils have repeated a grade.

*** Significance at the 1% level.

The estimate of average underachievement of 23.5% hides considerable heterogeneity across pupils. In **Figure 2,** we plot the distribution of underachievement. We see that underachievement is skewed to the right and ranges from as low as 9% to as high as 81%. That is, there is a long tail of pupils with considerable underachievement. We also consider underachievement per grade in **Figure 3**. It appears that underachievement peaks in the third grade at almost 31% and then gradually decreases to about 23% in the sixth grade. The lowest underachievement is observed in the second grade at only about 7%. This pattern can be potentially explained by concepts taught in each grade as part of the primary education curriculum in Flanders. The first grade of primary education focuses on learning how to read, whereas the second grade focuses on calculus. Both reading and calculus are relatively novel concepts for pupils, making underachievement unlikely (Acee, et al., 2010). Nonetheless, some pupils are already familiar with reading before entering primary education, whereas this is uncommon for the calculus taught in the second grade (e.g. time tables). This likely explains why underachievement in the second grade is even lower than in the first grade. The third and fourth grade foresee quite a bit of repetition of the earlier taught concepts, triggering underachievement due to boredom (Acee, et al., 2010). In the final two grades, foreign languages and algebra are taught. These new concepts help reduce underachievement, albeit gradually as underachievement is difficult to tackle in full once it has occurred (Dixon, Craven, & Martin, 2006).

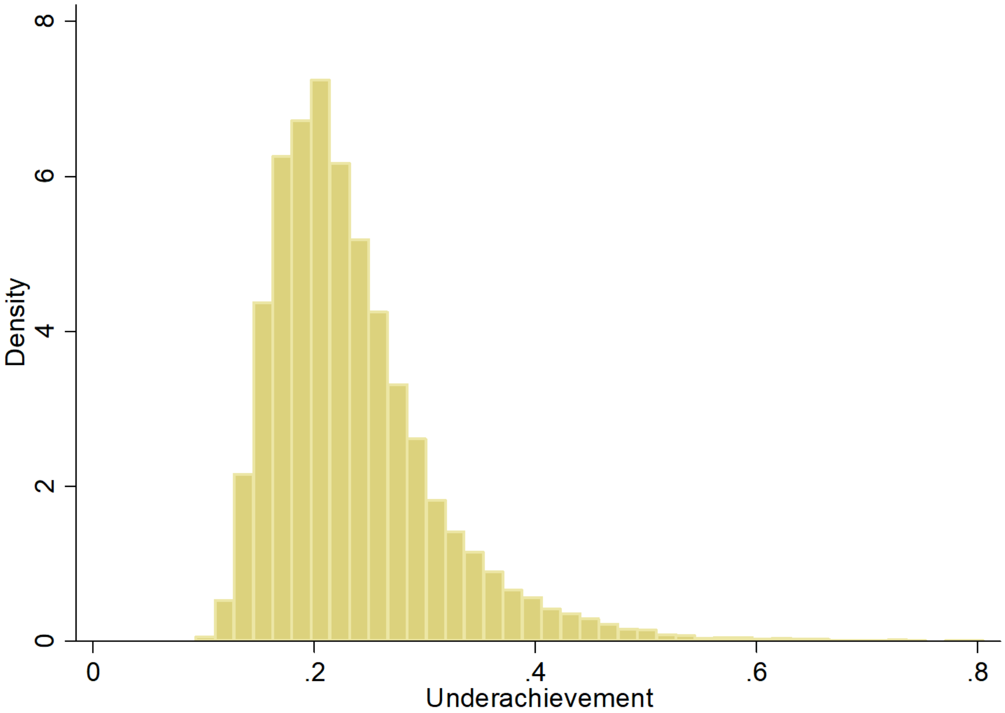

FIGURE 2: DISTRIBUTION OF UNDERACHIEVEMENT

Notes. This suggests that underachievement is skewed to the right and ranges from as low as 9% to as high as 81%.

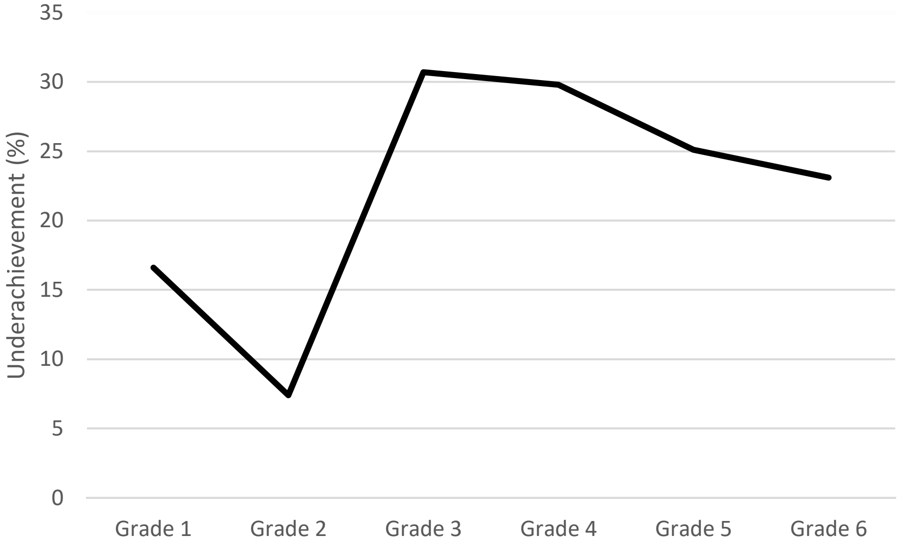

FIGURE 3: UNDERACHIEVEMENT BY GRADE

Notes. In each grade, underachievement has been estimated using the full set of inputs, controls, school and school year fixed effects, as well as class size.

5.2. Underachievement by Gender, Ethnicity, and Ability

In education policy there is a considerable interest in performance variation by gender, ethnicity, and ability. We re-estimate our preferred model for sub-samples by gender, ethnicity, and ability. **Table 3** gives the results. First, we split the sample by gender because the literature on underachievement has produced mixed results. Specifically, some studies found that boys' underachievement is two to three times larger than girls' underachievement (McCoach & Siegle, 2003; Peterson & Colangelo, 1996), whereas others have found that males underachieve just as much as girls (Preckel & Brunner, 2015). We find that boys' average underachievement is about one third larger than that of girls. However, a test of the equality of the coefficients shows that this difference is not statistically significant at the 10% level (p = 0.149). Second, we analyse underachievement by ethnicity. The prior literature suggests that pupils with foreign ethnicity are particularly prone to underachievement (Siegle, 2013; Thanassoulis, 1999). We find that pupils from a foreign origin underachieve 5.7 percentage points more than pupils from a Belgian origin. Here too, however, this difference is not significant (p = 0.121). Third, we divide the sample into gifted and non-gifted pupils because the prior literature on underachievement has focused almost exclusively on gifted underachievers. Following the definition of the National Association for Gifted Children (2019) gifted pupils are defined as pupils in the top ten percent of the IQ score distribution), while non-gifted pupils are in the lower 90% of the distribution. The results indicate that gifted pupils' underachievement is higher (27.4%) than that of non-gifted pupils (22.3%) However, this difference is not statistically significant (p = 0.481).

TABLE 3 – UNDERACHIEVEMENT BY GENDER, ETHNICITY, AND ABILITY

	Gender		Ethnicity		Ability	
	Boys	Girls	Foreign	Belgian	Gifted	Not gifted
Inputs and Controls						
IQ (std.)	0.512***	0.557***	0.418***	0.555***	0.772***	0.498***
	(0.011)	(0.011)	(0.018)	(0.009)	(0.087)	(0.009)
Gender (1 = Male)			0.420***	0.428***	0.364***	0.445***
			(0.033)	(0.014)	(0.044)	(0.014)
Origin (1 = Not Belgian)	-0.126***	-0.070***			0.216**	-0.089***
	(0.031)	(0.029)			(0.090)	(0.022)
Socioeconomic status	0.174***	0.098***	0.010	0.162***	0.150***	0.134***
	(0.014)	(0.013)	(0.026)	(0.010)	(0.034)	(0.010)
Grade retention (1 = repeated grade)	-0.364***	-0.205***	-0.478***	-0.202***	-0.811***	-0.290***
	(0.055)	(0.051)	(0.059)	(0.050)	(0.283)	(0.037)
Teacher gender (1 = Male)	0.063*	0.092***	0.083	0.061**	0.088	0.072***
	(0.034)	(0.031)	(0.059)	(0.026)	(0.066)	(0.025)
Teacher experience	-0.001	0.003	-0.008**	0.004	0.008	0.001
	(0.002)	(0.002)	(0.003)	(0.002)	(0.006)	(0.002)
Teacher additional hours	0.005**	-0.011***	0.001	-0.002	-0.010**	-0.000
	(0.002)	(0.003)	(0.004)	(0.002)	(0.005)	(0.002)
Teacher motivation (1=motivated)	0.510*	0.248	0.001	0.537***	0.001	0.564***
	(0.305)	(0.304)	(0.001)	(0.213)	(0.001)	(0.220)
Fixed effects:						
School year	Yes	Yes	Yes	Yes	Yes	Yes
School	Yes	Yes	Yes	Yes	Yes	Yes
Determinant of achievement						
Class size	-0.001	0.003	0.006	-0.002	0.002	0.001
	(0.002)	(0.003)	(0.005)	(0.005)	(0.002)	(0.001)
Underachievement						
Overall underachievement	0.255	0.178	0.271	0.214	0.274	0.223
	[0.097]	[0.050]	[0.010]	[0.068]	[0.134]	[0.069]
Number of pupils	1,053	1,175	430	1,798	224	2,004
Number of observations[a]	6,318	7,050	2,580	10,788	1,344	12,024

Notes. Standard errors are in parentheses. Standard deviations are in squared brackets. Outcome in all models is the mathematics test score standardized by school year.

[a] Pupils are observed in all six grades of primary education. Nonetheless, some pupils have repeated a grade.

*** Significance at the 1% level; ** Significance at the 5% level; * Significance at the 10% level.

5.3. Differential Influence of Class Size on Underachievement

Above we found that on average class size has no influence on underachievement. An important advantage of the SFA model applied here is that it allows the influence of class size on underachievement to be non-monotonic. In other words, the marginal influence of class size can vary with class size. We plot the marginal influence of class size on underachievement for different class sizes in **Figure 4**. As shown above, around the average class size, the influence is indistinguishable from zero. However, we see that this is not true away from the mean. For class sizes larger than 20 pupils, underachievement is reduced by making classes smaller. The marginal influence is negative below a class size of 20 pupils, indicating that in smaller classes an increase in class size decreases underachievement. As underlying mechanism for this finding, Jepsen and Rivkin (2009) showed that a larger number of smaller classes leads to an increase in lower quality teachers. Alternatively, Sims (2008) points to the higher frequency of small combination classes (classes that combine pupils from different grades into one small class), in which teachers have to split their attention over different groups.

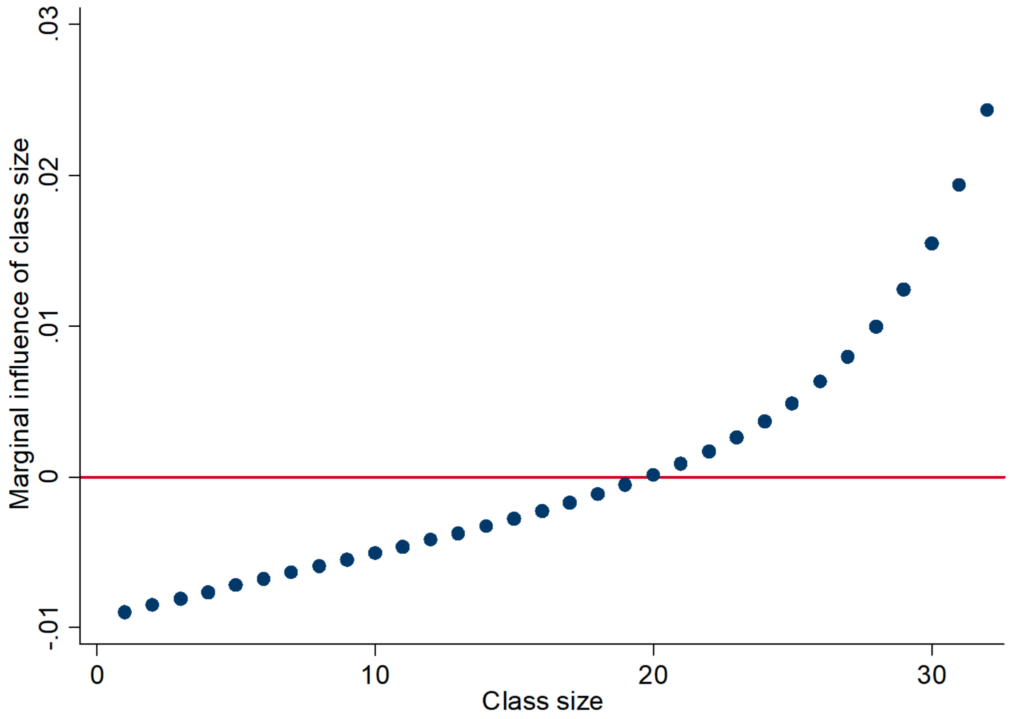

FIGURE 4: MARGINAL INFLUENCE OF CLASS SIZE ON UNDERACHIEVEMENT BY CLASS SIZE

Notes. This figure suggests a non-monotonic relationship between underachievement and class size. The marginal influence of class size appears to vary with class size. Above the class size of 20 pupils, larger classes seem to increase underachievement. Below this threshold, however, larger classes may actually reduce underachievement.

6. Conclusion

In this paper, we proposed to use regression-based Stochastic Frontier Analysis (SFA) to measure underachievement and its determinants in education. The key insight is that underachievement is unobservable – we never observe the counterfactual of maximum achievement – and needs to be modelled explicitly. SFA provides such a model, and compared to conceptually similar models, it has the advantages of allowing random noise, being robust to small samples, allowing for the consistent inclusion of determinants of underachievement, and straightforward statistical testing.

The results suggest that in Flemish elementary schools, pupils' average underachievement is 23.5%. This estimate falls somewhere in the middle of the estimates of underachievement in the prior literature (White, Graham, & Blaas, 2018). However, contrary to the suggestions in the policy debate, we found no evidence that underachievement systematically varies with gender, ethnicity, or ability. This also questions the prior literature's focus on gifted pupils. Finally, our evidence suggests that, in terms of underachievement, the optimal class size is 20. Both smaller and larger classes increase underachievement. One possible mechanism for this result is that with 20 pupils the teacher can optimally trade-off lecturing and one-on-one supervision in class (Bosworth & Caliendo, 2007). In larger classes, teachers are unable to provide individualised instruction to pupils, which is why they resort to lecturing. Moreover, larger classes might have a larger variance in abilities (especially in primary education where pupils are not tracked yet), making it more difficult for teachers to adopt the teaching style to the different ability levels of the pupils (Van Klaveren & De Witte, 2014). On the other hand, in classes with less than 20 pupils, larger classes may actually reduce underachievement. Underlying mechanisms for this finding might be the lower quality of teachers as a result of a larger number of smaller classes (Jepsen & Rivkin, 2009) and the occurrence of combination classes: pupils from different grades combined in one small class (Sims, 2008).

Although we introduced a new method to measure underachievement, this study is not without limitations. First, we do not claim to present causal evidence. It is possible that unobserved factors change the estimate of underachievement. Moreover, class size as well as teacher factors may be endogenous as schools may allocate pupils to particular classes with particular teachers. Future research may expand the range of inputs or environmental variables, or combine exogenous shocks with SFA to increase the causal interpretation of the findings. It is also useful to estimate underachievement beyond primary education and to investigate the influence of underachievement on potential high school dropout or later life outcomes. As a final line of future research, qualitative data should accompany these quantitative findings to explore why pupils underachieve in detail.

References

Abelman, R. (2006). Fighting the war on indecency: Mediating TV, internet, and videogame usage among achieving and underachieving gifted children. *Roeper Review, 29*(2).

Acee, T. W., Kim, H., Kim, H. J., Kim, J.-I., Chu, H.-n., Kim, M., . . . Wicker, F. W. (2010). Academic boredom in under- and over-challenging situations. *Contemporary Educational Psychology, 35*, 17-27.

Aigner, D., Lovell, K. C., & Schmidt, P. (1977). Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics, 6*, 21-37.

Anaya, K. L., & Pollitt, M. G. (2017). Using stochastic frontier analysis to measure the impact of weather on the efficiency of electricity distribution businesses in developing economies. *European Journal of Operational Research, 263*, 1078-1094.

Badunenko, O., & Kumbhakar, S. C. (2017). Economies of scale, technical change and persistent and time-varying cost efficiency in Indian banking: Do ownership, regulation and heterogeneity matter? *European Journal of Operational Research, 260*, 789-803.

Baker, J. A., Bridger, R., & Evans, K. (1998). Models of Underachievement Among Gifted Preadolescents: The Role of Personal, Family, and School Factors. *Gifted Child Quarterly, 42*(1), 5-15.

Bosworth, R., & Caliendo, F. (2007). Educational production and teacher preferences. *Economics of Education Review, 26*(4), 487-500.

Bressoux, P. (2009). Teachers' Training, Class Size and Students' Outcomes: Learning from Administrative Forecasting Mistakes. *Economic Journal, 119*(536), 540-561.

Compen, B., De Witte, K., & Schelfhout, W. (2019). The role of teacher professional development in financial literacy education: A systematic literature review. *Educational Research Review, 26*(1), 16-31.

De Witte, K., & Kortelainen, M. (2013). What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Applied Economics,* 2401-2412.

De Witte, K., & López-Torres, L. (2017). Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society, 68*(4), 339-363.

Denny, K., & Oppedisano, V. (2013). The surprising effect of larger class sizes: Evidence using two identification strategies. *Labour Economics, 23*, 57-65.

Dieterle, S. G. (2015). Class-size reduction policies and the quality of entering teachers. *Labour Economics, 36*, 35-47.

Dixon, R. M., Craven, R., & Martin, A. (2006). Underachievement in a whole city cohort of academically gifted children: what does it look like? *Australasian Journal of Gifted Education, 15*(2), 9-15.

Ehrgott, M., Holder, A., & Nohadani, O. (2018). Uncertain Data Envelopment Analysis. *European Journal of Operational Research, 268*(1), 231-242.

Elwert, F., & Winship, C. (2014). Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable. *Annual Review of Sociology, 40*, 31-53.

Eurydice. (2018, February 23). *Belgium - Flemish Community: Primary Education*. Retrieved September 15, 2019, from https://eacea.ec.europa.eu/national-policies/eurydice/content/primary-education-3_en

Ferrantino, M. J., & Ferrier, G. D. (1995). The technical efficiency of vacuum-pan sugar industry of India: An application of a stochastic frontier production function using panel data. *European Journal of Operational Research, 1995*, 639-653.

Figg, S., Rogers, K. B., McCormick, J., & Low, R. (2012). Differentiating Low Performance of the Gifted Learner: Achieving, Underachieving, and Selective Consuming Students. *Journal of Advanced Academics, 23*(1), 53-71.

Flemish Parliament. (2018, May 30). *Plenaire vergadering: Actuele vraag over het nut van zittenblijven in het basisonderwijs [Plenary meeting: Timely question about the usefulness of grade retention in primary education]*. Retrieved September 15, 2019, from https://www.vlaamsparlement.be/plenaire-vergaderingen/1258203/verslag/1259287

Gohm, C. L., Humphreys, L. G., & Yao, G. (1998). Underachievement Among Spatially Gifted Students. *American Educational Research Journal, 35*(3), 515-531.

Goldhaber, D. (2016). In schools, teacher quality matters most: today's research reinforces Coleman's findings. *Education Next*, 56-63.

Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature, 24*(3), 1141-1177.

Hendrikx, K., Maes, F., Magez, W., Ghesquière, P., & Van Damme, J. (2008). *Longitudinaal onderzoek in het basisonderwijs: Intelligentiemeting (schooljaar 2005-2006 [Longitudinal research in primary education: Intelligence assessment (school year 2005-2006)]*. Leuven, Belgium: Steunpunt SSL.

Hoxby, C. M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics, 115*(4), 1239-1285.

Jepsen, C., & Rivkin, S. (2009). Class size reduction and student achievement the potential tradeoff between teacher quality and class size. *Journal of human resources, 44*(1), 223-250.

Jondrow, J., Lovell, K. C., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics, 19*(2-3), 233-238.

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics, 114*(2), 497-532.

Kumbhakar, S. C., Lien, G., & Hardaker, B. J. (2014). Technical efficiency in competing panel data models: a study of Norwegian grain farming. *Journal of Productivity Analysis, 41*, 321-337.

Lau, K.-L., & Chan, D. W. (2001). Identification of Underachievers in Hong Kong: do different methods select different underachievers? *Educational Studies, 27*(2), 187-200.

Lien, G., Kumbhakar, S. C., & Alem, H. (2018). Endogeneity, heterogeneity, and determinants of inefficiency in Norwegian crop-producing farms. *International Journal of Production Economics, 201*, 53-61.

Matthews, M. S., & McBee, M. T. (2007). School Factors and the Underachievement of Gifted Students in a Talent Search Summer Program. *Gifted Child Quarterly, 51*(2), 167-181.

McCoach, B. D., & Siegle, D. (2003). The school attitude assessment survey-revised: A new instrument to identify academically able students who underachieve. *Educational and Psychological Measurement, 63*(3), 414-429.

Meeusen, W., & van Den Broeck, J. (1977). Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review, 18*(2), 435-444.

National Association for Gifted Children. (2019, September 23). *What is Giftedness?* Retrieved September 23, 2019, from http://www.nagc.org/resources-publications/resources/what-giftedness

Obergriesser, S., & Stoeger, H. (2015). The role of emotions, motivation, and learning behavior in underachievement and results of an intervention. *High Ability Studies, 26*(1), 167-190.

Oreopoulos, P., & Salvanes, K. G. (2011). Priceless: The Nonpecuniary Benefits of Schooling. *Journal of Economic Perspectives, 37*(2), 159-184.

Peterson, J. S. (2000). A follow-up study of one group of achievers and underachievers four years after high school graduation. *Roeper Review, 22*(4), 217-224.

Peterson, J. S., & Colangelo, N. (1996). Gifted Achievers and Underachievers: A Comparison of Patterns Found in School Files. *Journal of Counseling and Development, 74*, 399-406.

Phillipson, S. N. (2008). The optimal achievement model and underachievement in Hong Kong: an application of the Rasch model. *Psychology Science Quarterly, 50*(2), 147-172.

Phillipson, S. N., & Ka-on Tse, A. (2007). Discovering patterns of achievement in Hong Kong students: An application of the Rasch measurement model. *High Ability Studies, 18*(2), 173-190.

Preckel, F., & Brunner, M. (2015). Academic self-concept, achievement goals, and achievement: Is their relation the same for academic achievers and underachievers? *Gifted and Talented International, 30*(1-2), 68-84.

Raven, J. (2000). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology, 41*(1), 1-48.

Reis, S. M., Colbert, R. D., & Hébert, T. P. (2004). Understanding resilience in diverse, talented students in an urban high school. *Roeper Review, 27*(2), 110-120.

Reynders, T., Nicaise, I., & Van Damme, J. (2005). *Longitudinaal onderzoek in het basisonderwijs: De constructie van een SES-variabele voor het SiBO-onderzoek [Longitudinal research in primary education: The construction of a SES-variable for the SiBO-study]*. Leuven, Belgium: Loopbanen doorheen Onderwijs naar Arbeidsmarkt.

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How Much Does Education Improve Intelligence? A Meta-Analysis. *Psychological Science*, *29*(8), 1358-1369.

Ritchotte, J. A., Matthews, M. S., & Flowers, C. P. (2014). The Validity of the Achievement-Orientation Model for Gifted Middle School Students: An Exploratory Study. *Gifted Child Quarterly, 58*(3), 183-198.

Ruggiero, J. (2004). Data envelopment analysis with stochastic data. *Journal of the Operational Research Society, 55*, 1008-1012.

Schick, H., & Phillipson, S. N. (2009). Learning motivation and performance excellence in adolescents with high intellectual potential: what really matters? *High Ability Studies, 20*(1), 15-37.

Schiltz, F., De Witte, K., & Mazrekaj, D. (2019). Managerial efficiency and efficiency differentials in adult education: a conditional and bias-corrected efficiency analysis. *Annals of Operations Research*.

Siegle, D. (2013). *The Underachieving Gifted Child: Recognizing, Understanding, & Reversing Underachievement.* Waco, Texas, United States: Prufrock Press.

Silva Portela, M. C. (2001). Decomposing school and school-type efficiency. *European Journal of Operational Research, 132*, 357-373.

Sims, D. (2008). A strategic response to class size reduction: Combination classes and student achievement in California. *Journal of Policy Analysis and Management, 27*(3), 457-478.

Stinissen, J., Smolders, M., & Coppens-Declerck, L. (1975). *Handleiding bij de Collectieve Verbale Intelligentietest voor derde en vierde leerjaar (CIT-3-4) [Manual for the Collective Verbal Intelligence Test for Grades 3 and 4 (CIT-3-4].* Brussels, Belgium: Centrum voor Studie en Beroepsoriëntering.

Stoeger, H., & Ziegler, A. (2013). Deficits in fine motor skills and their influence on persistence among gifted elementary school pupils. *Gifted Education International*, *29*(1), 28-42.

Thanassoulis, E. (1999). Setting Achievement Targets for School Children. *Education Economics, 7*(2), 101-119.

Van Klaveren, C., & De Witte, K. (2014). How are teachers teaching? A nonparametric approach. *Education Economics*, *22*(1), 3-23.

Wang, H.-J. (2002). Heteroscedasticity and Non-Monotonic Efficiency Effects of a Stochastic Frontier Model. *Journal of Productivity Analysis, 18*, 241-253.

Wang, H.-J., & Schmidt, P. (2002). One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels. *Journal of Productivity Analysis, 18*, 129-144.

White, S. L., Graham, L. J., & Blaas, S. (2018). Why do we know so little about the factors associated with gifted underachievement? A systematic literature review. *Educational Research Review, 24*, 55-66.

Worrell, F. C., Subotnik, R. F., Olszewski-Kubilius, P., & Dixson, D. D. (2019). Gifted Students. *Annual Review of Psychology, 70*, 551-576.

Wyner, J. S., Bridgeland, J. M., & Dilulio, J. J. (2007). *Achievementrap: How America is Failing Millions of High-Achieving Students from Lower-Income Families.* Washington, D.C., United States: Civic Enterprises.

Ziegler, A., & Stoeger, H. (2003). Identification of Underachievement: An Empirical Study on the Agreement Among Various Diagnostic Sources. *Gifted and Talented international, 18*(2), 87-94.

Appendix

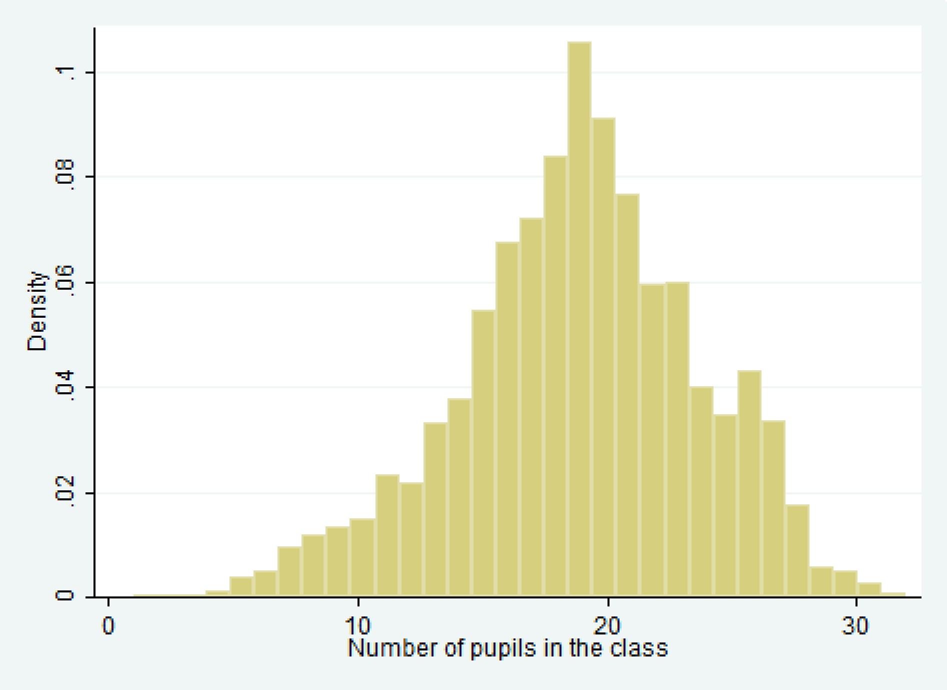

FIGURE A1: DISTRIBUTION OF CLASS SIZE

Notes. The average class consists of about 19 pupils. The largest class counts 32 pupils, while the smallest class comprises of only 1 pupil. The smallest classes are located in rural areas, where also schools are significantly smaller.