

Mind the Gap: Measuring Academic Underachievement Using Stochastic Frontier Analysis

Exceptional Children
2022, Vol. 88(4) 442–459
© The Author(s) 2022



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00144029211073524
journals.sagepub.com/home/ecx



Deni Mazrekaj ^{1,2}, Kristof De Witte^{2,3},
and Thomas P. Triebbs⁴

Abstract

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 6 years of primary education. We found that the average underachievement gap is 23.5%. That is, the average pupil does not exploit about one fourth of their potential. Gifted pupils appear to underachieve as much as non-gifted pupils. We also found 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.

Keywords

academic underachievement, class size, gifted pupils, Stochastic Frontier Analysis

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, that is, transform ability into maximum possible outcome. Despite potential performance gains from reducing underachievement, most research has focused on the measurement of low achievement, not *underachievement*. 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, not due to learning disabilities (McCoach & Siegle, 2003). While low achievers can easily be identified by observing outcomes, for example, 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 estimated potential. 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 may be positively associated with the risk of dropout (Peterson, 2000), which has severe consequences in

¹Utrecht University, Department of Sociology, Utrecht, The Netherlands

²KU Leuven, Leuven Economics of Education Research (LEER), Leuven, Belgium

³United Nations University – Maastricht Economic and Social Research Institute on Innovation and Technology (UNU-MERIT), Maastricht, the Netherlands

⁴Loughborough University

Corresponding author:

Deni Mazrekaj, Utrecht University,
Department of Sociology, Padualaan 14, 3584 CH
Utrecht, The Netherlands; KU Leuven, Leuven
Economics of Education Research (LEER), Naamsestraat
69, 3000 Leuven, Belgium.
Email: d.mazrekaj@uu.nl

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 et al., 2021; Ferrantino & Ferrier, 1995). Intuitively, 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 investigate 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). While there is a large body of literature on the influence of class size on achievement, to our best knowledge we are the first to analyze the influence on underachievement. Prior results for the influence of class size on achievement are mixed. Whereas some studies found that pupils' achievement is reduced in larger classes (Bressoux, 2009; Krueger, 1999), other studies found no effect (Dieterle, 2015; Hoxby, 2000), and still others found that pupils' achievement improves in larger classes (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 applied the model to unique longitudinal survey data for Belgian pupils. The

data include 2,228 pupils from the Flemish community of Belgium, observed over 6 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. The absence of ability grouping may lead teachers to orientate their teaching towards the average student (Van Klaveren & De Witte, 2014), exacerbating underachievement at the tails. Moreover, without a system of external accountability, there is less incentive for teachers to improve the achievement of low-performing students.

The SFA method for measuring underachievement departs from the methods used in the previous literature. Irrespective of the method used, previous estimates of the share of underachievers vary from as low as 9% (Schick & Phillipson, 2009) to as high as 49% (Reis et al., 2004); see White et al. (2018) for a review. Although methods to measure underachievement vary by disciplinary origin, most of the literature on underachievement is in education and psychology disciplines. We identified two main methods: nomination, and a comparison of aptitude tests with achievement tests. The first method, nomination, uses teachers', parents', or peer assessment to identify underachievers (Abelman, 2006; Lau & Chan, 2001; Snyder et al., 2021). For instance, Snyder et al. (2021) asked teachers: "How well is this child performing in reading compared to how well you believe she could?" 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 et al., 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 on the aptitude test but score lower than a certain threshold on the achievement test. This method is typically employed to identify gifted underachievers (Matthews & McBee, 2007; Ritchotte et al., 2014; Schick & Phillipson, 2009). For instance, Schick and Phillipson (2009) define underachievers as students who score above 115 on an IQ test and below 4 on a 5-point scale grade-point average. 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 (Obergruesser & 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, that is, have a sufficiently large positive error term (Dixon et al., 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 et al., 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.

Outside the education and psychology literature, a third method identifies underachievement using Data Envelopment Analysis (DEA; Silva Portela, 2001; Thanassoulis, 1999). This method stems from economics and is widely used to estimate managerial inefficiency in production. DEA estimates a production frontier indicating the potential scores pupils could achieve and compares this with the scores pupils actually achieve. Consequently, DEA solves most problems identified in the previous methods: 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; 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 et al., 2018; Ruggiero, 2004; Schiltz et al., 2020). 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 method proposed in this paper, Stochastic Frontier Analysis, is conceptually similar to DEA and also stems from production economics. Unlike DEA, however, SFA can include a large number of control variables and it allows for measurement error.

The Flemish Education System

The Flemish education system provides compulsory education between the ages of 6–18, or until a younger age if a student has already

obtained a high school diploma. Before children enter compulsory education, they can enroll 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 enroll their child into primary education already at the age of 5.

Primary education lasts for 6 years, until the age of 12. A pupil may spend at most 8 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 must repeat a grade. In school year 2020–2021, grade retention was 2% in Flemish primary education (Flemish Ministry of Education and Training, 2021). If pupils complete all 6 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 served basis until the capacity of the school is reached. Pupils are not grouped according to ability. The school board decides how pupils are distributed among classes and the number of pupils per class. In school year 2019–2020, a total of 442,961 pupils were enrolled into primary education, of which about 10% were of non-Belgian origin and 22% did not speak Dutch at home (Flemish Ministry of Education and Training, 2021). The most represented groups of non-Belgian origin were children with Moroccan and Turkish parents. About 20% were from a disadvantaged socioeconomic background. The student base was equally distributed by gender. In general, one teacher teaches all the subjects, although specialists might teach in some schools. In general, a new teacher is assigned to the class each school year. Thus, teachers generally do not follow their classes over the years. Upon successful completion of primary education, pupils enter a tracking system in secondary education at the age of 12 that includes four main tracks: the general track (ASO), the technical track (TSO), the vocational track (BSO), and the arts track (KSO).

Empirical Method

Our empirical model of underachievement starts with a standard education production function (Hanushek, 1986) that relates an output to inputs as well as 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 :

$$y_i = f(x_i) + g(c_t, c_s) + \epsilon_i \quad (1)$$

In Equation 1, y_i is an individual's output, x_i denotes the inputs, c_t and c_s correspond to a set of control variables at the level of the teacher t and school s respectively, 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 (proxied by the IQ score), gender, origin, 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 independent and identically distributed (i.i.d.) random error v_i on the one hand, and underachievement u_i on the other. The complete model is:

$$y_i = f(x_i) + g(c_t, c_s) + \epsilon_i \quad (2a)$$

$$\epsilon_i = v_i - u_i \quad (2b)$$

We estimate underachievement by applying the Stochastic Frontier Model (Aigner et al., 1977; Meeusen & van Den Broeck, 1977).

To identify the two error components, we required specific distributional assumptions. Following Wang (2002), we assumed 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:

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

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

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 over-achievers). This is because the achievement of an actual highest-performing pupil in the data will be used to construct the frontier.

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 their IQ score, the pupil obtains a test score marked by point A. The

counterfactual is that, with lower underachievement, they could have obtained a higher test score marked by point C.

Further, it would be reasonable to think that underachievement itself has its determinants. We augment the model 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 . We add to the model:

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

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

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, that is, 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

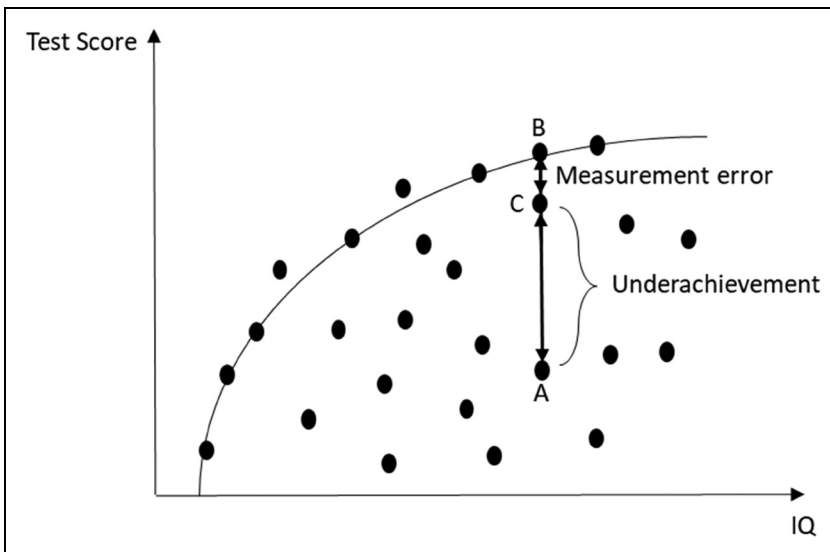


Figure 1. Two-Dimensional Representation of Underachievement Using Stochastic Frontier Analysis. *Note.* The discrepancy between the potential test score B and the actual test score A is interpreted as underachievement (AC) and measurement error (BC).

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 from 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 , 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.

Data

We applied 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 1st 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.

Our measure of output for the education production function was the mathematics test score from a test taken at the end of each school year. 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 3rd year of primary education, the emphasis lay on multiplication and division. 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. As an indication of reliability, the mathematics test has a Cronbach's alpha of 0.89, well above the conventional threshold of 0.7. Although the data also include language tests (reading and

writing), 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 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. We also used the mathematics score in the beginning of the 1st year of primary education instead of IQ in Table A1 in the online supplemental appendix. This is consistent with earlier literature that measures mathematics potential by early mathematics ability (Fong & Kremer, 2020). Given that the results are very similar, we opted for the use of IQ as underachievement is mostly defined in the literature as a discrepancy between a measure of ability and a measure of achievement. The test is inspired by the CIT-3–4 verbal cognitive test (Stinissen et al., 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; Obergrösser & Stoeger, 2015; Phillipson, 2008; Phillipson & Ka-on Tse, 2007) as well as the psychological literature on gifted pupils overall (see Worrell et al. [2019] for a review). The detailed procedure of the construction of the test can be found in Hendrikx et al. (2008) and the reliability, as measured by Cronbach's alpha, is 0.93. The cognitive test was administered in the school year 2005–2006 when pupils were in third grade (usually at age 9). As such, mathematics test scores obtained before the third grade might influence the IQ score, leading to reverse causality and biased estimates of underachievement. For instance, Ritchie and Tucker-Drob (2018) found that an additional year of education increases cognitive abilities by approximately 1–5 IQ points. Nonetheless, if we take mathematics test scores 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 1.1 percentage points (see Table A1 in the

online supplemental appendix). Moreover, we do not claim to present causal evidence.

We include the following additional inputs. Gender is an indicator with value 1 for males and 0 for females. Previous literature is mixed on the association between gender and underachievement, with many but not all studies (e.g., Preckel and Brunner [2015]) reporting a higher incidence of underachievers among boys (McCoach & Siegle, 2003; Peterson & Colangelo, 1996). Origin is an indicator that is given a value of 1 if either the pupil or one of the parents was born abroad, and a value of 0 otherwise. Although this variable is rather a crude proxy of pupils' immigration status, it has been widely used in previous educational studies in Flanders (Mazrekaj & De Witte, 2020; Van Houtte & Stevens, 2015), and is also reported by the Flemish Ministry of Education and Training. The prior literature suggests that pupils with a foreign origin are at a greater risk of underachievement (Siegle, 2013; Thanassoulis, 1999). The proxy for socioeconomic status is the first principal 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 1st year of primary education. A higher value for the socioeconomic status variable indicates a higher socioeconomic status. The correlation between origin and socioeconomic status amounts to 0.35. Prior evidence shows that pupils from a lower socioeconomic background are more likely to underachieve. For instance, Wyner et al. (2007) found that 44% of low-income pupils in the United States who scored in the top 10% in the first grade, did not score in the top 10% in the fifth grade. Moreover, gifted low-income pupils progressed at half the rate of their gifted high-income peers.

Furthermore, we also include an indicator for grade retention given a value of 1 if the pupil repeated the current grade, and a value of 0 if the pupil progressed from the last grade. Earlier studies have found that grade

retention can both increase and decrease pupils' mathematics test scores in primary education, depending on the country studied (see Valbuena et al. [2021] for an overview). We used school and school year fixed effects to compare pupils only within the same school and year. This captures unobserved heterogeneity due to, for example, differences in school policy. Finally, we added four variables to control for teacher characteristics. Previous literature has shown that teachers are instrumental to tackling underachievement. Specifically, in their meta-analysis, Steenbergen-Hu et al. (2020) found that (gifted) underachievers particularly benefit from positive teacher–student relationships and one-to-one mentoring. Moreover, Siegle et al. (2014) found that knowledgeable teachers build students' self-efficacy to learn, leading to lower underachievement. We included four collected characteristics: 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 ($\alpha = 0.96$). 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 restricted 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 restricted 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 et al. (2005). Third, we dropped 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,

Table 1. Descriptive Statistics.

	Mean	SD	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
Origin (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	13,368			

as we also analyze the development of underachievement over time, we removed 900 pupils for whom we do not observe all six grades of primary education. Finally, we dropped 103 pupils with missing values for at least one of the variables.

Our final sample 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. Almost 4% of the pupils 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. They spend almost 20 additional hours outside the regular 29 working hours and only about 21% do not feel motivated by their job. Finally, the average class size was 19. Maximum class size was 32 and the minimum class size was 1. The smallest classes are in rural areas, where schools are significantly smaller (removing these very small classes leaves the results virtually unchanged—see Figure A2 and Table A1 in the online supplemental appendix). The class size distribution is roughly normal as

shown by Figure A1 in the online supplemental appendix.

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, origin, and whether or not the pupil is gifted.

Overall Underachievement

Table 2 gives the estimation results for the full sample. Whereas Column 4 includes all input and control variables, columns 1–3 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

Table 2. Estimating Overall Underachievement.

	(1)	(2)	(3)	(4)
Inputs and Controls				
IQ (std.)	0.569*** (0.008)	0.539*** (0.008)	0.533*** (0.008)	0.533*** (0.008)
Gender (1 = male)		0.428*** (0.013)	0.427*** (0.013)	0.428*** (0.013)
Origin (1 = foreign)		-0.075*** (0.021)	-0.069*** (0.021)	-0.068*** (0.021)
Socioeconomic status (index)		0.141*** (0.009)	0.134*** (0.009)	0.134*** (0.009)
Grade retention (1 = repeated grade)			-0.291*** (0.037)	-0.294*** (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 (0.003)	0.002 (0.002)	0.000 (0.001)	0.000 (0.001)
Underachievement				
Overall underachievement	0.219 [0.066]	0.230 [0.077]	0.230 [0.075]	0.235 [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

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

^aPupils are observed in all six grades of primary education. Nonetheless, some pupils have repeated a grade.

***Significance at the 1% level.

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. Our main interest (and the advantage of the SFA model) is the estimate for underachievement, which amounts to about 22%, a statistically significant estimate. That is, the average pupil, given their IQ score, could increase their mathematics score by 22%. We found no evidence that class size is a statistically significant determinant of underachievement.

In Column 2, we add three additional inputs to the education production function: gender, origin, and socioeconomic status. Note that the association with IQ hardly decreases. This confirms the intuition that

ability is independent of these other background variables. However, even when controlling for ability, there is a statistically significant influence for gender, origin, and socioeconomic 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) is associated with a lower score by 0.075 standard deviations. Increasing socioeconomic status by one standard deviation increases the score by about 0.121 standard deviations (0.861×0.141). The influence of socioeconomic background is only about a quarter of the influence of ability.

In Column 3, we add an indicator for grade retention. Pupils who repeat the grade score

significantly lower on the mathematics test by about 0.3 standard deviations. In Table A1 in the online supplemental appendix, we also report the results solely for students who have not repeated a grade. These results are very similar to the results presented in Table 2. In the last column in Table 2, 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 less than the pupil's gender. Contrary to earlier literature (Compen et al., 2019), teachers' 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.

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 calculations. Both reading and calculations are relatively novel concepts for pupils, making underachievement unlikely (Acee, et al., 2010). However, some pupils are already familiar with reading before entering primary education, whereas this is uncommon for the calculations taught in the second grade (e.g., time tables). This likely explains why

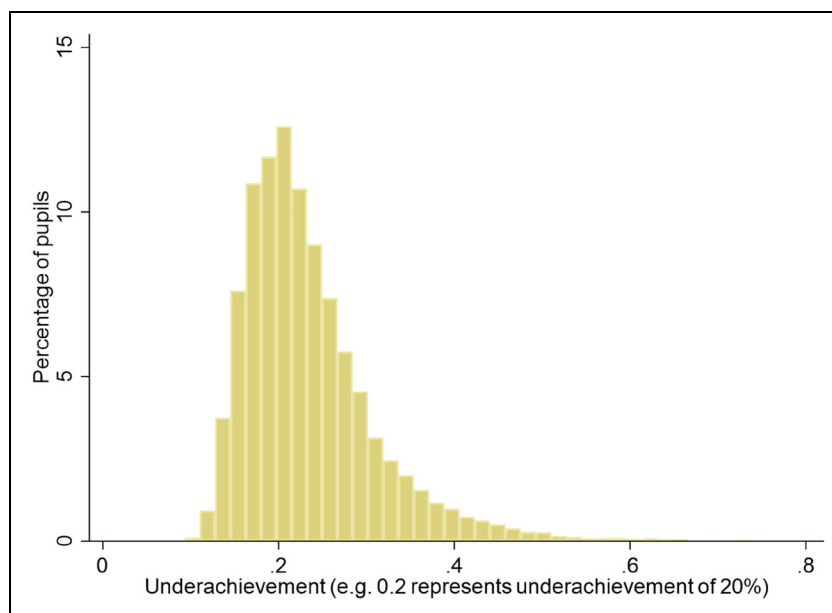


Figure 2. Distribution of Underachievement.

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

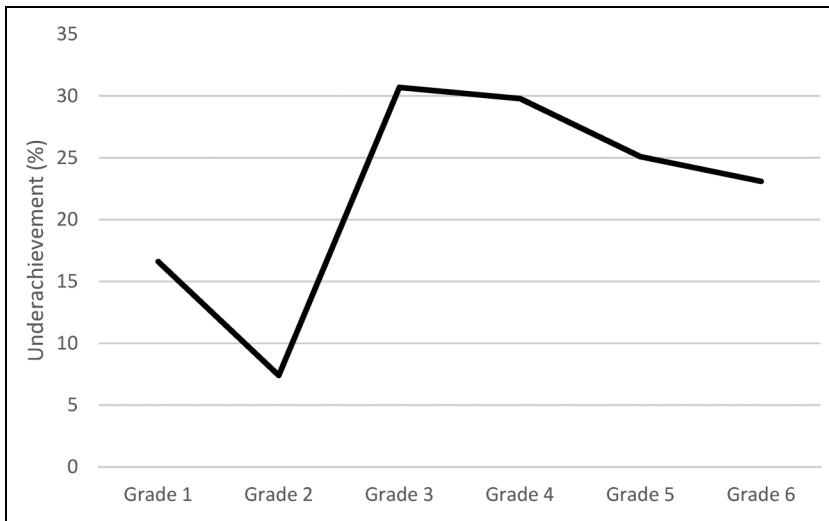


Figure 3. Underachievement by Grade.

Note. 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.

underachievement in the second grade is even lower than in the first grade. In the third and fourth grades there is 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 et al., 2006).

Underachievement by Gender, Origin, and Ability

In education policy there is a considerable interest in performance variation by gender, origin, and ability. We re-estimate our preferred model for sub-samples by gender, origin, 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 the incidence of male underachievers is 2–3 times larger than that of female underachievers (McCoach & Siegle, 2003; Peterson & Colangelo, 1996), whereas others have found that males underachieve just as much as females (Preckel & Brunner, 2015). We found that boys' average underachievement was

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 ($p = .149$). Second, we analyzed underachievement by origin. The prior literature suggests that pupils with a foreign origin are particularly prone to underachievement (Siegle, 2013; Thanassoulis, 1999). We found that pupils of foreign origin underachieved 5.7 percentage points more than pupils of Belgian origin. Again, this difference is not significant ($p = .121$). Third, we divided the sample into gifted and non-gifted pupils because the prior literature on underachievement has focused almost exclusively on gifted underachievers. To distinguish between gifted and non-gifted pupils, we chose an IQ threshold of top 10%. Thus, gifted pupils are defined as pupils in the top 10% of the IQ score distribution, whereas non-gifted pupils are in the lower 90% of the distribution. We opted for the 10% threshold given that “many consider children who are in the top 10% in relation to a national and/or local norm to be a good guide for identification and services” (National Association for Gifted Children, 2021). 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 ($p = .481$).

Table 3. Underachievement by Gender, Origin, and Ability.

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

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

^aPupils 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.

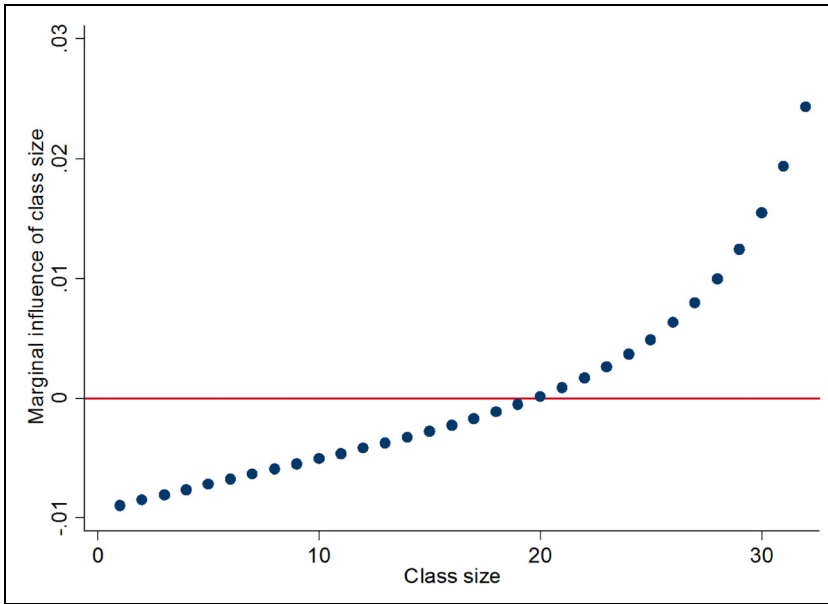


Figure 4. Marginal Influence of Class Size on Underachievement by Class Size.

Note. 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 a class size of 20 pupils, class size seems to be positively correlated with underachievement. Below this threshold, however, class size may actually be negatively correlated with underachievement.

Differential Influence of Class Size on Underachievement

In Table 2, 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 plotted the marginal influence of class size on underachievement for different class sizes (see Figure 4). 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 positively correlated with class size. The marginal influence is negative below a class size of 20 pupils, indicating that, in smaller classes, class size is negatively correlated with underachievement. One potential mechanism for this finding is that smaller classes might imply more classes and thus lower average teacher

quality (Jepsen & Rivkin, 2009). 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.

Conclusion

In this paper, we propose using 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—it needs to be modeled and estimated.

The results suggest that in Flemish elementary schools, pupils' average underachievement is 23.5%. This estimate falls somewhere between the estimates of underachievement in the prior literature (White et al., 2018). However, contrary to the suggestions in the policy debate, we found no evidence that underachievement systematically

varies with gender, origin, or ability. This also questions the prior literature's focus on gifted pupils. Finally, our evidence suggests that the optimal class size for underachievement is 20. 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 individualized 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 yet tracked), making it more difficult for teachers to adopt the teaching style to the different ability levels of the pupils (Van Klaveren & De Witte, 2014). For instance, Peters et al. (2017) found that within a classroom, 16%–37% of students scored a year or more above their current grade level, and this was 20%–49% for language in U.S. elementary and middle schools. On the other hand, in classes with less than 20 pupils, class size may actually be negatively associated with underachievement, meaning that larger classes may be beneficial. 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 introduce 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. Similarly, the data oversampled pupils from a lower socioeconomic background which may have affected the results. Future research may expand the range of inputs or control 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 (see for instance Almkhambetova and Hernández-Torrano [2020]) and to investigate

the influence of underachievement on potential high school dropout or later life outcomes. From a methodological perspective, future research could examine the existence of “overachievement.” Recent advancements in SFA models make use of bootstrapped samples in which the evaluated observation is excluded from the reference sample (Wanke et al., 2020). These technical innovations allow for the measurement of over-achievement. As a final line of future research, qualitative data should accompany these quantitative findings to explore in detail why pupils underachieve (see for instance Desmet et al. [2020]).

References

- Abelman, R. (2006). Fighting the war on indecency: Mediating TV, internet, and video game usage among achieving and underachieving gifted children. *Roepers Review*, 29(2). 10.1080/02783190709554393
- Ace, 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. 10.1016/j.cedpsych.2009.08.002
- Aigner, D., Lovell, K. C., & Schmidt, P. (1977). Formulation and estimation of Stochastic Frontier production function models. *Journal of Econometrics*, 6, 21–37. 10.1016/0304-4076(77)90052-5
- Almkhambetova, A., & Hernández-Torrano, D. (2020). Gifted students' adjustment and underachievement in university: An exploration from the self-determination theory perspective. *Gifted Child Quarterly*, 64(2), 117–131. 10.1177/0016986220905525
- 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. 10.1016/j.ejor.2017.05.054
- Badunenko, O., Mazrekaj, D., Kumbhakar, S. C., & De Witte, K. (2021). Persistent and transient inefficiency in adult education. *Empirical Economics*, 60, 2925–2942. 10.1007/s00181-020-01966-5
- Bosworth, R., & Caliendo, F. (2007). Educational production and teacher preferences. *Economics*

- of *Education Review*, 26(4), 487–500. 10.1016/j.econedurev.2005.04.004
- Bressoux, P. (2009). Teachers' training, class size and students'. Outcomes: Learning From Administrative Forecasting Mistakes. *Economic Journal*, 119(536), 540–561. <https://EconPapers.repec.org/RePEc:ecj:econjl:v:119:y:2009:i:536:p: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. 10.1016/j.edurev.2018.12.001
- 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. 10.1080/00036846.2012.665602
- 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. 10.1057/jors.2015.92
- Denny, K., & Oppedisano, V. (2013). The surprising effect of larger class sizes: Evidence using two identification strategies. *Labour Economics*, 23, 57–65. 10.1016/j.labeco.2013.04.004
- Desmet, O. A., Pereira, N., & Peterson, J. S. (2020). Telling a tale: How underachievement develops in gifted girls. *Gifted Child Quarterly*, 64(2), 85–99. 10.1177/0016986219888633
- Dieterle, S. G. (2015). Class-size reduction policies and the quality of entering teachers. *Labour Economics*, 36, 35–47. 10.1016/j.labeco.2015.07.005
- 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. 10.1016/j.ejor.2018.01.005
- 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*, 80(3), 639–653. 10.1016/0377-2217(94)00142-Y
- Flemish Ministry of Education and Training. (2021, January 12). Dataloop: Aan de slag met cijfers over onderwijs [Data magnifier: Getting started with education figures]. <https://onderwijs.vlaanderen.be/en/node/2664>
- Fong, C. J., & Kremer, K. P. (2020). An expectancy-value approach to math underachievement: Examining high school achievement, college attendance, and STEM interest. *Gifted Child Quarterly*, 64(2), 67–84. 10.1177/0016986219890599
- Gohm, C. L., Humphreys, L. G., & Yao, G. (1998). Underachievement among spatially gifted students. *American Educational Research Journal*, 35(3), 515–531. 10.3102/00028312035003515
- Goldhaber, D. (2016). In schools, teacher quality matters most: Today's research reinforces Coleman's findings. *Education Next*, 16(2), 56–63. <https://www.educationnext.org/in-schools-teacher-quality-matters-most-coleman/>
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3), 1141–1177. <http://hanushek.stanford.edu/publications/economics-schooling-production-and-efficiency-public-schools>
- 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)]*. 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. 10.1162/003355300555060
- 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. 10.1353/jhr.2009.0008
- 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. 10.1016/0304-4076(82)90004-5
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly*

- Journal of Economics*, 114(2), 497–532. 10.1162/003355399556052
- 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. 10.1080/03055690120050419
- 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. 10.1177/0016986207299473
- Mazrekaj, D., & De Witte, K. (2020). The effect of modular education on school dropout. *British Educational Research Journal*, 46(1), 92–121. 10.1002/berj.3569
- 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. 10.1177/0013164403063003005
- Meeusen, W., & van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18(2), 435–444. 10.2307/2525757
- National Association for Gifted Children. (2021, September 19). What is giftedness? Retrieved September 23, 2019, from <http://www.nage.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. 10.1080/13598139.2015.1043003
- Oreopoulos, P., & Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. *Journal of Economic Perspectives*, 37(2), 159–184. 10.1257/jep.25.1.159
- Peters, S. J., Rambo-Hernandez, K., Makel, M. C., Matthews, M. S., & Plucker, J. A. (2017). Should millions of students take a gap year? Large numbers of students start the school year above grade level. *Gifted Child Quarterly*, 61(3), 229–238. 10.1177/0016986217701834
- 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. 10.1080/02783190009554041
- 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. 10.1002/j.1556-6676.1996.tb01886.x
- 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. <https://psycnet.apa.org/record/2008-11745-005>
- 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. 10.1080/13598130701709640
- 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. 10.1080/15332276.2015.1137458
- Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1), 1–48. 10.1006/cogp.1999.0735
- 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. 10.1080/02783190509554299
- 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]*. 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. 10.1177/0956797618774253
- 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. 10.1177/0016986214534890
- Ruggiero, J. (2004). Data envelopment analysis with stochastic data. *Journal of the Operational Research Society*, 55, 1008–1012. 10.1057/palgrave.jors.2601779
- 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. 10.1080/13598130902879366
- Schiltz, F., De Witte, K., & Mazrekaj, D. (2020). Managerial efficiency and efficiency differentials in adult education: A conditional and bias-corrected efficiency analysis. *Annals of Operations Research*, 288(2), 529–546. 10.1007/s10479-019-03269-0
- Siegle, D. (2013). *The underachieving gifted child: Recognizing, understanding, & reversing underachievement*. Prufrock Press.
- Siegle, D., DaVia Rubenstein, L., & Mitchell, M. D. (2014). Honors students' perceptions of their high school experiences: The influence of teachers on student motivation. *Gifted Child Quarterly*, 58(1), 35–50. 10.1177/0016986213513496
- Portela, Silva, & C, M.. (2001). Decomposing school and school-type efficiency. *European Journal of Operational Research*, 132, 357–373. 10.1016/S0377-2217(00)00157-0
- 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. 10.1002/pam.20353
- Snyder, K. E., Carrig, M. M., & Linnenbrink-Garcia, L. (2021). Developmental pathways in underachievement. *Applied Developmental Science*, 25(2), 114–132. 10.1080/10888691.2018.1543028
- Steenbergen-Hu, S., Olszewski-Kubilius, P., & Calvert, E. (2020). The effectiveness of current interventions to reverse the underachievement of gifted students: Findings of a meta-analysis and systematic review. *Gifted Child Quarterly*, 64(2), 132–165. 10.1177/0016986220908601
- 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)]*. 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. 10.1177/0261429412440649
- Thanassoulis, E. (1999). Setting achievement targets for school children. *Education Economics*, 7(2), 101–119. 10.1080/09645299900000010
- Valbuena, J., Mediavilla, M., Choi, Á., & Gil, M. (2021). Effect of grade retention policies: A literature review of empirical studies applying causal inference. *Journal of Economic Surveys*, 35(2), 408–451. 10.1111/joes.12406
- Van Houtte, M., & Stevens, P. A. (2015). Tracking and sense of futility: The impact of between-school tracking versus within-school tracking in secondary education in Flanders (Belgium). *British Educational Research Journal*, 41(5), 782–800. 10.1002/berj.3172
- Van Klaveren, C., & De Witte, K. (2014). How are teachers teaching? A nonparametric approach. *Education Economics*, 22(1), 3–23. 10.1080/09645292.2011.560448
- Wang, H.-J. (2002). Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. *Journal of Productivity Analysis*, 18, 241–253. 10.1023/A:1020638827640
- 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. 10.1023/A:1016565719882
- Wanke, P., Tsionas, M. G., Chen, Z., & Moreira Antunes, J. J. (2020). Dynamic network DEA and SFA models for accounting and financial indicators with an analysis of super-efficiency in stochastic frontiers: An efficiency comparison in OECD banking. *International Review of Economics & Finance*, 69, 456–468. 10.1016/j.iref.2020.06.002
- 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. 10.1016/j.edurev.2018.03.001
- Worrell, F. C., Subotnik, R. F., Olszewski-Kubilius, P., & Dixon, D. D. (2019). Gifted students. *Annual Review of Psychology*, 70, 551–576. 10.1146/annurev-psych-010418-102846
- Wyner, J. S., Bridgeland, J. M., & Dilulio, J. J. (2007). *Achievementtrap: How aAmerica is failing millions of high-achieving students from lower-income families*. 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. 10.1080/15332276.2003.11673019

Acknowledgments

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.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Funding

Deni Mazrekaj acknowledges funding by the Research Foundation Flanders (FWO) as Aspirant [grant numbers 1172519N and V439919N].

Supplemental Material

Supplemental material for this article is available online.

ORCID iD

Deni Mazrekaj  <https://orcid.org/0000-0002-3311-6056>

Manuscript received May 2021; accepted December 2021.