Value-added measurement under high teacher turnover *

Pedro Freitas †1, Pedro Carneiro 2, Rodrigo Ferreira 1, Luís Catela Nunes1, Ana Balcão Reis1, and Carmo Seabra 1

1 *Nova School of Business and Economics, Universidade Nova de Lisboa*
2 *University College of London, IFS, CEMMAP*

June 8, 2018

Abstract

Teacher value-added (TVA) usually focuses on teacher quality during one school year, measuring how the teacher contributes to the evolution of the students scores between a baseline and a final test. However, in many of the available databases, it is not possible to measure the TVA yearly. That is the case of Portugal, where, for example, students are subject to standardized tests in the 4th grade and then in the 6th grade. Additionally, during this period, more than half of the students have a different teacher in the 5th and 6th grade. These data limitations bring additional challenges to the estimation of TVA. This works proposes a new identification strategy to recover TVA when standardized tests are not available on a yearly basis and when high level of teacher turnover is observed. We also estimate what is the technology behind the combination of teachers in the 5th and 6th grade. We conclude on how the elasticity between TVAs is different whether we consider the probability of the students to have an average or an high score. Additionally, we use the estimated TVAs to assess how the teacher turnover between the 5th and 6th grade affects the students performance, concluding on the low impact of changing teacher from one grade to the other.

JEL classification: I21,C61,C67.

Keywords: Teacher value-added, Teacher turnover, Test scores.

*Financial support from Fundação Belmiro de Azevedo through the EDULOG initiative is greatly acknowledged.We thank the statistics department of the Portuguese Ministry of Education for providing the data.

†Corresponding author: pedro.freitas@novasbe.pt

1 Introduction

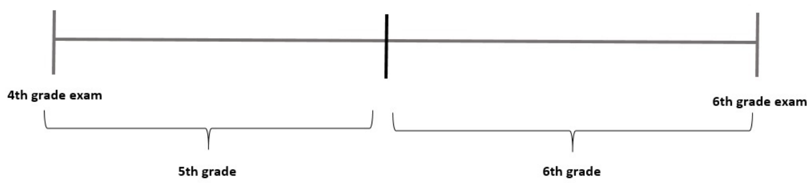
Teacher's contribution to students' results remains a key issue in education policy. It is recognized that teachers may vary in their quality both within and across schools.
The first attempts to estimate the teachers' contribution to students' outcomes focused on the observable characteristics of the teachers. However little evidence could be gathered from this analysis ([Hanushek and Rivkin, 2012] and [Rivkin et al., 2005]).
The most recent literature has focused on the measurement of teacher value-added (TVA). This measure assesses what is the contribution of each individual teacher for the progression of a student between a baseline and a final test. The use of teacher value-added as a measure of teacher quality has been subject to an intense debate, particularly about the potential bias of these estimations. [Cotfelter et al., 2006]; [Rothestein, 2010] and [Horváth, 2015] argue for the potential bias in this measure due to the non random allocation of students to classrooms and to teachers. On the other hand [Kane and Staiger, 2008], [Kane et al., 2013], [Chetty et al., 2014a] and [Chetty et al., 2014b] argue that after controlling for a baseline test, teacher value-added is an unbiased estimate of teacher's quality. Few works assessed teacher quality through the random assignment of teachers to classrooms, with the exceptions of the MET project ([Kane et al., 2013]) and [Araujo et al., 2016] for Equador.
The works which use value-added measures to assess teacher's quality normally dispose of a baseline test at the begging of the academic year and a final test at the end. Additionally the teacher during the academic year is expected to be the same. However, traditionally students are not subject to standardized tests in every academic year and between two standardized tests each class may have different teachers, particularly when the students move from one grade to the other.
The non existence of standardized tests in every grade and the teacher turnover between grades makes, that for many education systems, it is not possible to use the traditional strategies to measure teacher value-added. The goal of this paper is to propose an identification strategy that allows for TVA estimation in such cases. In order to do so, we use data from the portuguese education system. To the extend of our knowledge, this is the first TVA estimations are performed for an education system outside the US and a developing country. We particularly focus on the students in the 2nd cycle of studies, in the 5th and 6th grade. As a baseline test we use the national exam in reading and mathematics at the end of the 4th grade. As the outcome we use the national exams in the same subjects at the end of 6th grade.This database also allows to exploit TVA identification in the presence of high levels of teacher turnover, since in the portuguese system around 60% of the students change their teacher between the 5th and the 6th

grade. Given the high level of teacher turnover we aim to estimate its impact in students' performance. So far, the topic of teacher turnover was mainly addressed at school level by [Feng and Sass, 2011], [Ronfeldt et al., 2013] and [Hanushek et al., 2016]. This question has relevant policy implications in terms of teacher allocation, particularly for a system like the portuguese one which by law advises the school to keep the same teacher in each subject from one grade to the other.[1] This heterogeneity in terms of students who keep and do not keep the same teacher across the 5th and 6th grade creates an insightful source of variation to exploit what is the technology behind the combination of different teachers. This means, if the student has two teachers across the grades, which contributes more for the academic achievement? Another relevant question concerns the elasticity of substitution between the two teachers. We are able to identify this technology, since we observe more than 13 000 different teacher combinations in reading and mathematics.
Section 2 presents the identification strategy; section 3 describes the data used, section 4 presents the main results and uses the estimated teacher value-added to simulate what would be the impact of replacing the worst teachers by a median teacher; as well as the correlation between teacher value-added measure and observable teacher characteristics. Finally, section 5 concludes.

2 Identification strategy

Contrary to previous works on teacher value-added we do not observe a baseline test and a final test at the beginning and end of each school year.
As illustrated in Figure1, in our case we look at students in the 5th and 6th grade, considering the national exams in the 4th grade and in the 6th grade as the baseline and final outcome.

Figure 1



[1] Particularity in the law of 18th January 2001 it is written: "the efficient and rational allocation of human resources as well as the possibility of keeping the same teacher across different grades bring clear advantages for the educational system, increasing its quality"

During the 5th and 6th grade, the teacher assigned to each class may or may not be the same. The existence of teacher turnover between the 5th and 6th grade makes that we are not able to directly measure teacher value-added. What we are able to measure is a teacher combination vale added (VAC), which can be composed by the same or by different teachers in the 5th and 6th grade. A combination may correspond to one or more classes, and since teachers may have different classes across the years, they belong to different combinations.
The first step of the identification strategy is to estimate the combination value-added. In order to do so we run the following linear model, considering each student, i, who belongs to teacher combination, c, in school, s in cohort t:

$$Y_{i,c,s,t,6} = X_{i,c,s,t}\beta + Z_{s,t}\eta + Y_{i,c,s,t,4}\gamma + \mu_c + \epsilon_{i,c,s,t,6} \tag{1}$$

$Y_{i,c,s,t,6}$ stands for the score in the national exam in the 6th grade and $Y_{i,c,s,t,4}$ for the score in the national exam in the 4th grade. $X_{i,c,s,t}$ corresponds to a set of controls at the individual level and $Z_{s,t}$ to a set of controls at school level. μ_c stands to the fixed effect of each combination of teachers, and then it represents the teacher combination value-added. Both the scores in the 4th and 6th grade are reported in a discrete scale from 1 to 5. Thus, the specification above is treated as a linear probability model, considering the probability of the student to have a positive score (> 2), a good score (> 3) or a very good score (> 4). The specification is run for the three cases, in order to assess how sensitive is the TVA estimation to the score the student obtains. After collecting the fixed effect of each combination, we observe that each teacher belongs to one or several combinations. If we consider that we observe J teachers who belong to C combinations, we can summarize the relations between teachers and combinations as:

$$\begin{cases} f(TVA_A, TVA_A) = \mu_1 \\ f(TVA_A, TVA_B) = \mu_2 \\ f(TVA_B, TVA_B) = \mu_3 \\ f(TVA_B, TVA_C) = \mu_4 \\ f(TVA_C, TVA_B) = \mu_5 \\ ... \end{cases} \tag{2}$$

We define function f as a CES production function with an additive penalty in case of those combinations with two teachers, meaning that the students changed their teacher in reading or mathematics between the 5th and 6th grade:

$$\mu c = [a(TVA_j)^{\rho} + (1-a) * (TVA_j)^{\rho}]^{\frac{1}{\rho}} + penalty * I(Teacher_5 \neq Teacher_6) \tag{3}$$

a is the weight of the teacher in the 5th grade for the combination value-added and symmetrically $(1-a)$ is the weight of the teacher value added in the $6th$ grade. $\rho = \frac{s-1}{s}$, where s stands for an elasticity of substitution between the teachers.[2] *penalty* is the parameter in case there is teacher turnover between the two grades.
In order for us to have identification of the parameters in (3) we need the following sources of variation: 1.To identify the TVA_j we need the system in (2) to be at least just identified: then number of combinations must be higher than the number of teachers, $N_{teacher} \leq \mathrm{N}_{combinations}$; 2.To estimate the weight and elasticity parameters, a and ρ, we need that teachers with different TVAs to teach the $5th$ grade students in some combinations and the $6th$ grade students in other combinations - around 60% of the teachers fulfill this requirement; 3. To identify the *penalty* we need combinations with and without teacher turnover - around 40% of the combinations show teacher turnover between the 5th and 6th grade.

In (2) each combination and its relation with the teacher value-added is taken as a moment condition. Then, we consider all the terms in the system as a set of moment conditions as $g(\theta_j)$. In order to estimate the different TVAs we follow a minimum distance estimation approach, considering:

$$Min_{\theta_j} Q(\theta_j) = g(\theta_j)' W g(\theta_j) \tag{4}$$

We apply a two step minimum distance estimator. In the first step we take the weighting matrix as a diagonal matrix with the number of students in each combination,μ_c . This way we give more weight to those combinations which have a higher number of individuals, and then provide more precise estimations of the parameters. In the second step we use as weighting matrix the inverse of the moment variance obtained in the first step.

[2]Thus, when $\rho \longrightarrow 1$ the teachers are closer to substitute and $\rho \longrightarrow -\infty$ the teachers show complementarity between them.

Given the structure of the moment conditions, the analytical variance of the estimators obtained through minimum distance estimation cannot be applied. We follow a bootstrap procedure as described in Appendix 1.

3 Data

We use administrative data for Portugal, provided by the Ministry of Education - *MISI*. This data comprises data at the student level from the 1st to the 12th grade in the period between 2007 and 2015. This is a rich database which gathers several information on students and students' family characteristics. Additionally this data provides the link between the student, the class and the teacher. This way we can follow the student across time, identifying the classes he belonged to and the teachers he had in different subjects. We link this information to other database *Juri Nacional de Exames* which comprises information on the standardized tests and exams performed by the portuguese students.

We focus on the students who were enrolled in reading and mathematics in the 5th and 6th grade between the academic year 2007/2008 and 2014/2015. Namely we follow several cohorts of students in the 5th and 6th grade. For example the first cohort corresponds to the students enrolled in the the 5th grade in the academic year 2007/2008 and in the 6th grade in the following academic year, 2008/2009. Below, in table 1, we report the descriptive statistics on this data.

We observe 388,763 students in the reading classes and 405,129 in the mathematics classes. These students are spread across 866 schools in the portuguese mainland and correspond to those students who have not failed in the 5th grade and to those who remained in the same school during these two grades. We also restricted the analysis to those combinations who have at least 10 students.[3]

[3]In the administrative data provided by the portuguese Ministry of education not every school consistently reports the classes the students are allocated to. This does not allow to identify the teachers for all the students in the database. In appendix 2 we show that despite this, the sample we are working with is representative of the student population.

Table 1: Descriptive statistics

	Reading				Mathematics			
	Mean	**S.d**	**Min**	**Max**	**Mean**	**S.d**	**Min**	**Max**
Student variables								
Female	0.50	0.50	0	1	0.50	0.49	0	1
Portuguese mother	0.95	0.22	0	1	0.94	0.23	0	1
Portuguese Father	0.94	0.24	0	1	0.93	0.25	0	1
Mother - college graduated	0.15	0.36	0	1	0.15	0.35	0	1
Father - college graduated	0.09	0.30	0	1	0.09	0.29	0	1
Social support - type A	0.20	0.40	0	1	0.19	0.40	0	1
Social support - type B	0.22	0.41	0	1	0.22	0.41	0	1
Mother unemployed	0.09	0.29	0	1	0.09	0.29	0	1
Father unemployed	0.05	0.21	0	1	0.05	0.22	0	1
Age	10.15	0.48	8.15	14.98	10.15	0.48	8.01	14.98
Access to computer	0.64	0.48	0	1	0.64	0.48	0	1
Access to internet	0.49	0.49	0	1	0.49	0.49	0	1
School variables								
4th grade	3.41	0.32	1.74	4.5	3.44	0.38	1.82	5
Social support. type A (%)	0.20	0.11	0	1	0.19	0.09	0	1
Social support. type B (%)	0.22	0.12	0	1	0.22	0.12	0	1
Mother college graduated (%)	0.15	0.11	0	1	0.15	0.11	0	0.85
Mother Portuguese (%)	0.95	0.08	0	1	0.94	0.08	0	1
	N=388,763				**N=405,129**			

In table 2 we report the distribution of the the discrete score (1-5) in the exams of reading and mathematics in the 4th and 6th grade. We clearly observe a deterioration of the results between the two exams, with the percentage of negative scores (lower than 3) increasing by around 7 p.p in reading and 18 p.p in mathematics.[4]

[4]In Appendix 3 we report the relation between the scores in the 4th grade and the scores in the 6th grade

Table 2: Results in the Reading and Mathematics exam- 4th and 6th grade

	Reading				Mathematics			
	4th grade		**6th grade**		**4th grade**		**6th grade**	
Score	**Number**	%	**Number**	%	**Number**	%	**Number**	%
1	1,301	0.33	1,664	0.43	2,619	0.65	16,813	4.15
2	42,749	11.00	69,036	17.76	61,006	15.06	120,374	29.71
3	170,591	43.88	187,500	48.23	154,330	38.09	150,027	37.03
4	143,931	37.02	115,010	29.58	130,228	32.14	92,674	22.08
5	30,191	7.77	15,553	4.00	56,946	14.06	25,240	6.23
Total	388,763	100	388,763	100	405,128	100	405,128	100

4 Results

We start by reporting the results on the OLS estimation of (1). This allows us to recover the fixed effects of all the teacher combinations, μ_c. We run three linear probability models considering that students have a positive score (> 2 in the discrete scale 1-5), a good score (> 3 in the discrete scale 1-5) or a very good score (> 4 in the discrete scale 1-5). $X_{i,c,s}$ is the set of individual student characteristics, which correspond to the variables in table 1: female, portuguese mother, portuguese father, social support (type A and type B)[5], unemployed mother, unemployed father, age, access to computer and access to internet. $Z_{i,c,s}$ stands for school variables, which as detailed in table 1, corresponds to the mean score in 4th grade exam, as well as the school share of students under social support (type A and B), school share of college graduated mothers and school share of portuguese mothers. Table 5 reports the OLS estimation of (1) for reading and mathematics. Regarding the variables at the student level, we observe the expected signs, namely students with higher previous performance, in families with graduated an employed parents and without social support are more likely to have positive or good scores in the exam. There are some slight variations in the significance of the variables whether we look at different score thresholds, such as the negative impact of the parents' unemployed status tends to vanish as we measure the probability of having an higher score. Concerning the school variables, not all variables show to be significant. We find quadratic significant effects for the mean score in 4th grade exam, in the variable regarding the social support type B and the student's mother college graduation.

[5]The variables Social Support, Type A and B correspond to the means-tested school programs in the portuguese schools. Type A and B are the two levels of eligibility in the program.

This first step allow us to recover the estimation of teacher combination effects, μ_c, whose distribution is reported in table 3 and 4. For the case of reading, we observe 13.335 different combinations which are formed by 8.869 teachers. If we increase the teacher combination value-added by a standard deviation, this leads to an increase in the probability of having a score > 2 in 10 p.p, the probability of having a score > 3 in 12.3 p.p and the probability of having a score > 4 in 4.6 p.p. In the case of mathematics, the 13.194 combinations are composed by 7.894 teachers. For this case, an increase of one standard deviation in the quality of the combination value-added leads to an increase in the probability of having a score > 2 in 13.8 p.p, a score > 2 in 12.3 p.p and the probability of a score > 2 in 5.8 p.p. For both cases, reading and mathematics, we observe similar distributions for the cases of the probabilities of having scores > 2 and > 3, but a lower standard deviation for the case of a the probability of a score > 5.

Table 3: Distribution of the μ_c - Reading

Probability > 2		Probability > 3		Probability > 4			
Mean	**S.D**	**Mean**	**S.D**	**Mean**	**S.D**	N_μ	$N_{Teachers}$
0	0.103	0	0.123	0	0.046	13.335	8.869

Table 4: Distribution of the μ_c - Mathematics

Probability ≥ 2		Probability ≥ 3		Probability ≥ 4			
Mean	**S.D**	**Mean**	**S.D**	**Mean**	**S.D**	N_μ	$N_{Teachers}$
0	0.138	0	0.123	0	0.058	13.194	7.894

In Table 6 we report the results in the minimum distance estimator, which allow us to recover the individual teacher value (TVA) from the combination value-added (VAC). In reading if the TVA of a teacher alone in the combination increases by one s.d, the probability of the student to have: 1. a positive score increases in 16 p.p; 2. a score higher than 3 increases in 18 p.p; 3. a score higher than 4 increases in 10 p.p. In Mathematics, if the TVA of a teacher alone in the combination increases by one standard deviation, the probability of the student to have: 1. positive score increases in 21 p.p; 2. have a score higher than 3 in 19 p.p; 3. have a score higher than 4 in 10 p.p.[6]

[6]In Appendix 4 we report the kernel distributions of the TVAs.

Table 5: OLS results (1)

	Reading						**Mathematics**					
	Probability > 2		Probability > 3		Probability > 4		Probability > 2		Probability > 3		Probability > 4	
	Coef.	**S.D**	**Coef.**	**S.D**	**Coef.**	**S.D**	**Coef.**	**S.D**	**Coef.**	**S.D**	**Coef.**	**S.D**
Exam 4th grade (2)	0.279***	0.0125	0.059***	0.006	0.002	0.001	0.1388***	0.008	-0.0006	0.004	-0.003***	0.001
Exam 4th grade (3)	0.569***	0.0125	0.216***	0.006	0.009***	0.002	0.467***	0.008	0.104***	0.004	-0.002*	0.001
Exam 4th grade (4)	0.732***	0.0126	0.527***	0.007	0.059***	0.002	0.716***	0.008	0.398***	0.005	0.054***	0.002
Exam 4th grade (5)	0.783***	0.127	0.796***	0.007	0.194***	0.003	0.804***	0.008	0.682***	0.005	0.242**	0.004
Female	0.045***	0.001	0.071***	0.001	0.017***	0.001	0.025***	0.0012	0.015***	0.001	0.002***	0.001
PT Mother	-0.004	0.003	0.13***	0.003	0.003*	0.001	-0.003***	0.0032	0.0034	0.003	0.005***	0.002
PT Father	0.004	0.003	0.009***	0.003	0.001	0.001	0.0126***	0.003	0.005*	0.003	-0.0005	0.002
Mother college graduated	0.031***	0.002	0.094***	0.002	0.032***	0.002	0.074***	0.002	0.111***	0.003	0.048***	0.002
Father college graduated	0.037***	0.002	0.072***	0.003	0.03***	0.002	0.037***	0.002	0.09***	0.003	0.057***	0.002
Social support (type a)	0.0114***	0.002	-0.044***	0.002	-0.007***	0.001	-0.0537***	0.002	-0.047***	0.001	-0.012	0
Social support (type b)	-0.027***	0.002	-0.066***	0.002	-0.009***	0.001	-0.098***	0.002	-0.066***	0.026	-0.013	0
Mother unemployed	-0.051**	0.002	-0.008***	0.002	-0.002*	0.001	-0.015***	0.002	-0.102***	0.017	-0.001	0.001
Father unemployed	-0.004***	0.003	-0.014***	0.003	0.001	0.001	-0.02***	0.003	-0.0133***	0.033	-0.003**	0.001
Age	-0.011***	0.002	-0.034***	0.001	0.003***	0	-0.0854***	0.001	-0.027***	0.055	0	0
Computer	0.009***	0.002	0.012***	0.002	0.004***	0.001	0.011***	0.002	0.006	0.002	0	0.001
Internet	0.018***	0.002	0.023***	0.002	0.003***	0.001	0.023***	0.002	0.025****	0.002	0.006***	0.001
School variables												
Mean 4th grade	-0.249***	0.086	0.154	0.098	0.142***	0.041	-0.141***	0.039	0.019	0.053	0.056**	0.026
Mean 4th grade sqr	0.031***	0.012	-0.037***	0.014	-0.024***	0.006	0.007	0.058	-0.018**	0.008	-0.013***	0.004
Social support, type A (%)	0.036	0.061	0.140**	0.069	-0.002	0.029	-0.037	0.045	-0.058	0.061	-0.061**	0.031
Social support, type A sqr (%)	-0.119	0.117	-0.139	0.131	0.069	0.053	-0.035	0.085	0.095	0.108	0.134***	0.058
Social support, type B (%)	0.013	0.045	0.035	0.052	0.000	0.021	0.098***	0.034	0.089	0.045	-0.005	0.024
Social support, type B sqr (%)	-0.054	0.079	-0.157**	0.077	-0.015	0.031	-0.201***	0.058	-0.149**	0.072	-0.002	0.038
Mother college graduated (%)	-0.054	0.047	0.167***	0.054	0.068	0.028	0.053	0.039	0.179***	0.051	0.077***	0.025
Mother college graduated sqr (%)	0.012	0.083	-0.325***	0.093	-0.172***	0.006	0.154**	0.075	-0.250***	0.096	-0.147***	0.051
Mother Portuguese (%)	-0.131	0.087	-0.072	0.133	-0.018	0.047	-0.059	0.086	-0.048	0.085	-0.025	0.052
Mother Portuguese sqr (%)	0.091	0.063	0.052	0.088	0.019	0.035	0.016	0.059	0.013	0.065	-0.008	0.037
Constant	1.701***	0.154	0.206***	0.177	-0.203***	0.074	1.6155***	0.132	0.286***	0.098	-0.111***	0.051
Cohort effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	388,763	-	388,763	-	388,763	-	405,129	-	405,129	-	405,129	-
R2	0.274	-	0.337	-	0.112	-	0.375	-	0.376	-	0.3177	-

Standard errors clustered at the school level.Statistically significant at *10%, *5%, ***1%

Table 6: Distribution of the TVA's

	Reading			**Mathematics**		
	> 2	> 3	> 4	> 2	> 3	> 4
Mean	0	0	0	0	0	0
S.d	0.157	0.183	0.095	0.207	0.188	0.104
Min	-0.524	-0.551	-0.185	-0.744	-0.551	-0.233
Max	0.668	0.846	0.51	0.826	0.938	0.636

In Table 7 are reported the results for the estimation of the parameters ρ, a and *penalty* in (3). For both reading and mathematics the parameter ρ falls as the estimation is performed on the probability of having a higher score. This means that for the student to have a positive score (> 2), if the teacher in the $5th$ grade has a low TVA, part of this negative impact is offset by a teacher with high TVA in the $6th$ grade, or vice-versa. The same is not true when we consider the probability of having a higher scores, since the effect of the teacher with low TVA prevails. Regarding the weight parameter a, as we measure the probability of having a higher score we observe that the weight between the TVA in the 5th and 6th grade becomes more balanced. This implies that for the student to have a positive score (> 2) the teacher in the 6th grade shows to have a higher contribution. However to achieve a higher score the contribution is splited between the two teachers. Finally, to notice the consistent result of the *penalty* parameter to be equal to zero, meaning that, on average, there is any impact in the student's performance, due to the change in the teacher between the 5th and the 6th grade.

Table 7: ρ; a and *penalty* - Reading and Mathematics

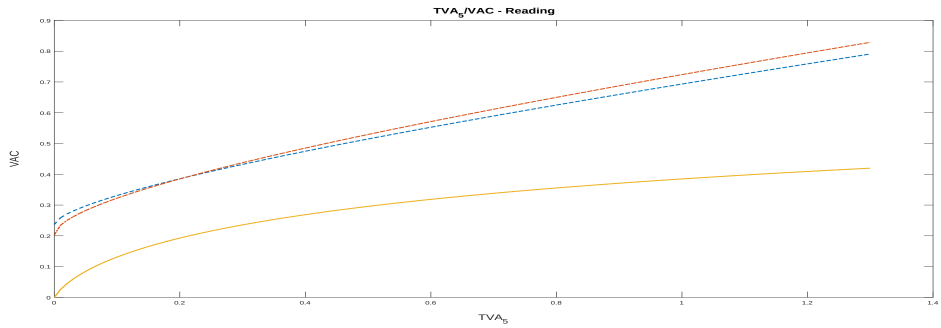
	Reading			**Mathematics**		
	> 2	> 3	> 4	> 2	> 3	> 4
ρ	0.613	0.542	-0.446	0.651	0.373	-0.593
	(0.02)	(0.03)	(0.07)	(0.03)	(0.07)	(0.03)
a	0.385	0.418	0.527	0.318	0.326	0.453
	(0.03)	(0.04)	(0.01)	(0.05)	(0.01)	(0.03)
penalty	0	0	0	0	-0.006	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Standard errors in parentheses, obtained through the bootstrap procedure described in Appendix 1

To better interpret the implications of the results reported in the tables above, we now perform some graphical analysis. For the combinations which do not have the same teacher, and given the non-linearities in the CES function, the marginal effects of increasing a standard deviation in the TVA of the teacher in the 5th and 6th grade is not constant. In order to understand this effect, we plot the TVA of the teacher in the

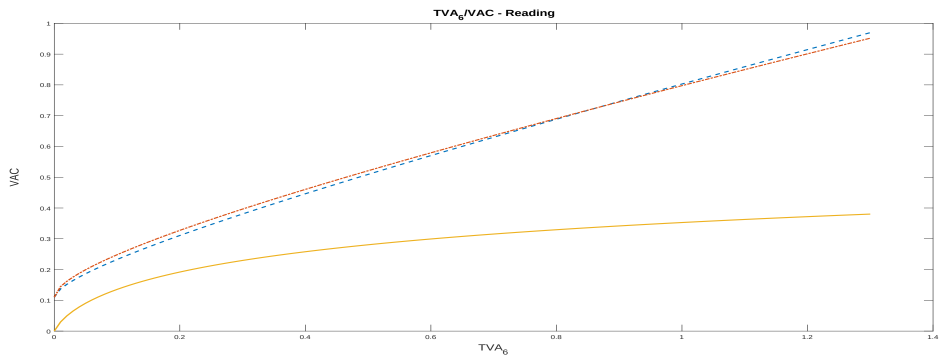
5th grade and the value added of the combination (VAC), this for an average value of the TVA of the teacher in the 6th grade. The same is done, but plotting the TVA in the 6th grade and the value-added of the combination (VAC), keeping constant the TVA of the 5th grade at the average value:

Figure 2: Relation between TVA_5 and VAC - Reading



Blue dashed line - Probability > 2; Red pointed line - Probability > 3,Orange line - Probability > 4

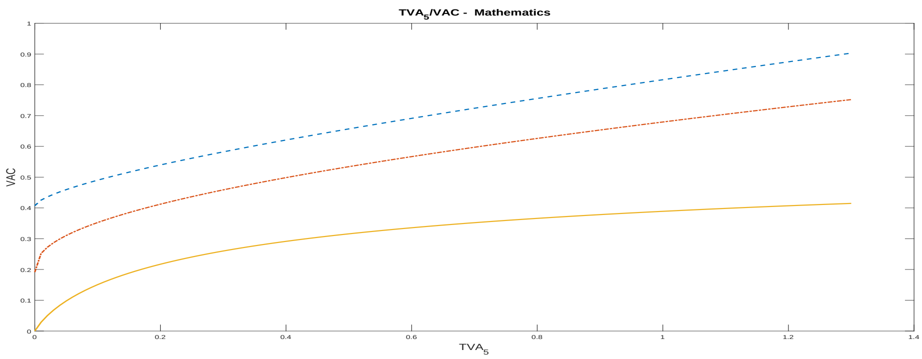
Figure 3: Relation between TVA_6 and VAC - Reading



Blue dashed line - Probability > 2; Red pointed line - Probability > 3,Orange line - Probability > 4

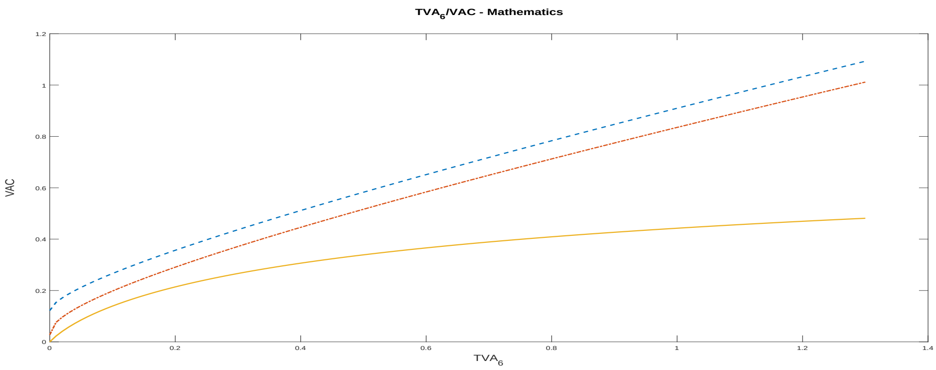
In the graphs above we observe that the cases closer to the linear case are the ones for which the parameter ρ is closer to one in (3). The increase in the TVA in the 5th or 6th grade has a similar impact on the vale added combinations for the functional forms on the probability of the score > 2 and > 3. However, and given the shape which demands a higher complementarity between the teachers, the impact of TVA in the combination value-added is lower when we analyze the probability of having a score > 4.

Figure 4: Relation between TVA_5 and VAC - Mathematics



Blue dashed line - Probability > 2; Red pointed line - Probability > 3,Orange line - Probability > 4

Figure 5: Relation between TVA_6 and VAC - Mathematics

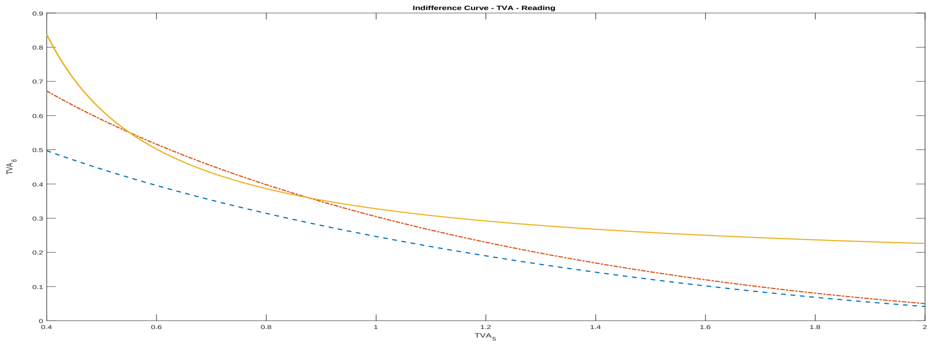


Blue dashed line - Probability > 2; Red pointed line - Probability > 3,Orange line - Probability > 4

Regarding the subject of mathematics, and contrary to what was previously observed for the case of reading, the marginal impact of an increase the teacher's quality in the 5th or 6th grade is different for the probability of having a score > 2 and > 3. As expected, the marginal impact is even lower when ρ decreases and then the complementarity between the teachers increases.

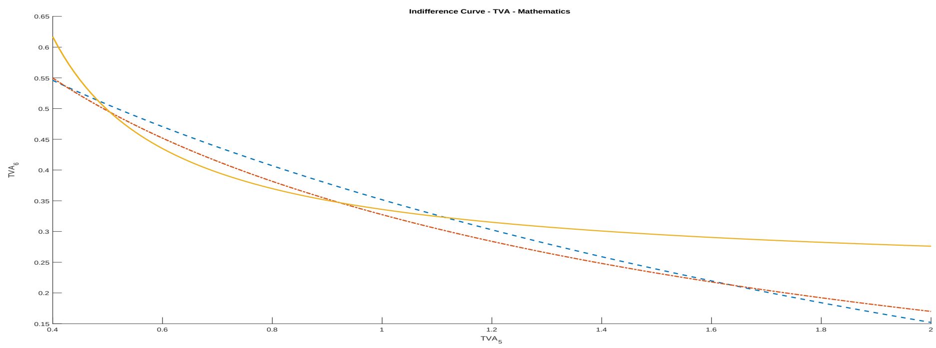
To better understand the implications of the elasticities and weights in the shape of the function, f, we plot the indifference curves of the combination value added. This means, the combinations of the TVA in the 5th grade and the TVA in the 6th grade, which provide a certain given value for the VAC:

Figure 6: Isoquants for the VAC in terms of TVA_5 and TVA_6 - Reading



Blue dashed line - Probability > 2; Red pointed line - Probability > 3,Orange line - Probability > 4

Figure 7: Isoquants for the VAC in terms of TVA_5 and TVA_6 - Mathematics



Blue dashed line - Probability > 2; Red pointed line - Probability > 3,Orange line - Probability > 4

As we can observe, both in figure 6 and 7, as we consider the probability of having a higher score, the indifference curves go from a straight line, to a Leontief shape.

4.1 Simulation Exercise

After estimating the individual Teacher Value Added, we can perform the exercise of how the probability of the students to have the different scores in the reading and mathematics exam would be affected if we increase the teachers' quality. Particularly, we consider what would be the impact of replacing the lowest quartile of the teacher distribution by a median teacher. To perform this exercise, we estimate first what is the predicted VAC when we replace these teachers using the functional in (3). Then and following the linear specification in (1) we observe how the average probability of the students to have a score > 2, > 3 and > 4 changes.

Table 8: Replacing teachers - Reading and Mathematics

	Reading			**Mathematics**		
	> 2	> 3	> 4	> 2	> 3	> 4
Initial probability	0.82	0.34	0.04	0.66	0.29	0.06
After the exercise	0.85	0.37	0.05	0.71	0.33	0.08
Δ	**3 p.p**	**3 p.p**	**1 p.p**	**5 p.p**	**4 p.p**	**1 p.p**

From the results reported in the table above we observe that replacing the worst teachers by a median has an aggregate impact in the probabilities of the different scores between 3 p.p and the 5 p.p in reading and mathematics. We observe that this change does not have the same impact across the different probabilities. This is closely related with the functional form for each case. Since for the probability of the students to have a high score, the degree of complementarity between the teachers increases, when we replace the worst teachers by a median teacher, the impact is smaller than in a more additive model.

4.2 Teacher characteristics

The *MISI* database which gathers the data used in this work also contains several information on teacher characteristics. A common result in literature is that easily observed characteristics show a lower explanatory power of teacher quality. We can observe the following variables regarding teachers: gender, master level, tenure, age, experience.[7]. Below we report the regression of the TVA on these same variables:

Table 9: OLS results TVA - teacher characteristics

	Reading						**Mathematics**					
	Coef.	S.D	Coef.	S.D	Coef.	S.D	Coef.	S.D	Coef.	S.D	Coef.	S.D
Female	0.018	0.005	0.019	0.006	0.002	0.003	0.004	0.007	0	0.007	-0.006	0.004
Tenure	-0.008	0.004	-0.01	0.005	-0.009	0.02	-0.004	0.005	-0.001	0.005	0.002	0.003
Master	0.003	0.006	0.004	0.006	0.003	0.003	-0.011	0.008	0.003	0.007	0.0058	0.004
Age	0	0	0	0	0	0	0	0	0	0	0	0
Exper	0	0	0	0	0	0	0	0	0	0	0	0
Constant	0.5	0.012	0.55	0.015	0.19	0.008	0.76	0.018	0.55	0.017	0.24	0.009
N	8,659	-	8,659	-	8,659	-	7,755	-	7,755	-	7,755	
R2	0.001	-	0.002	-	0.004	-	0	-	0	-	0	

From the estimations above we can observe how the gathered characteristics have any significant relation with the TVA, which is also confirmed by the very low R^2 of the specifications above. This result has important conclusions in terms of policy implication, since it shows that factors such as accumulated experience or a master degree are not the factors that help to explain the teacher quality.

5 Conclusion

When students do not performed standardized tests every academic year, and when there are high levels of teacher turnover between grades, the traditional approaches to measure individual teacher value-added cannot be used. In this work we propose a methodology for these cases. We consider that between two standardized tests what we can observe is the value-added of the combination of teachers (VAC). Considering that each teacher may belong to several different combinations, we use this feature of the data to recover the teacher value-added from the combination value-added.

We use data from Portugal between 2007 and 2015. We look at the students who were enrolled in the 5th and 6th grade, considering the national exam of reading and mathematics in the 4th grade as the baseline measure and as outcome the performance in the national exam in the same subjects at the end of the 6th grade.

When the teacher is the same during the 5th and 6th grade, we observe a higher level of teacher value-added to obtain a positive score (> 2) or a good score (> 3): between 15 p.p and 20 p.p for an increase of one standard deviation in teacher quality. For the probability of having a very good score (> 4), this impact decreases for 10 p.p. When

[7]The descriptive statistics of these same same variables are reported in Appendix 5

the teacher is not the same, this marginal effects becomes non-linear, being lower when we measure the probability of having a higher score. Given the difference between the Teacher Value Added (TVA) and the Combination Value Added (VAC) we recover the technology behind the combination of different teachers in the 5th and the 6th grade. We find that this technology is different considering the probability of the student to have different scores. Particularly, as we are measuring the teacher's impact on the probability of having a higher score, we observe: 1. a higher level of complementarity between TVAs; 2. a more balanced weight between the teachers. This result has relevant policy implications, since when we change the teachers quality, we observe that the impact is not the same on the probabilities of having different scores. Additionally, when we consider the possibility of existing a penalty for those students who change their teacher between the 5th and the 6th grade, we observed that this did not harm the student's performance. This result is notably important for an educational system like the Portuguese which demands low levels of teacher turnover in what concerns the allocation of human resources.

We also used the estimated teacher value-added to observe how they correlate with observable teacher characteristics. We concluded that standard variables on teachers such as tenure, master degree, age or experience have any impact to explain the source of teacher quality.

This is an on-going project. Currently the same methodology is being followed to estimate the TVA of the teachers between the 7th and the 9th grade and later during the high school years, between the 10th and the 12th grade.

Bibliography

M. Caridad Araujo, P. Carneiro, Y. Cruz-Aguayo, and N. Schady. Teacher quality and learning outcomes in kindergarten. *The Quarterly Journal of Economics*, 131 (3): 1415–1453, 2016.

R. Chetty, J. Friedman, and J. Rockoff. Measuring the impacts of teachers I: evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632, 2014a.

R. Chetty, J. Friedman, and J. Rockoff. Measuring the impacts of teachers II: teacher value-added and student outcomes in adultohood. *American Economic Review*, 104(9): 2633–2679, 2014b.

C. Cotfelter, H. Ladd, and J. Vigdor. Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, 41(4):778–820, 2006.

L. Feng and T. Sass. Teacher quality and teacher mobility. *Washington DC: National Center for Analysis of Longitudinal Data in Education Research CALDER Working Paper 57*, 2011.

E. Hanushek and S. Rivkin. The distribution of teacher quality and implications for policy. *Annual review of Economics*, 2012 (4):131–157, 2012.

E. Hanushek, S. Rivkin, and J. Schiman. Dynamic effects of teacher turnover on the quality of instruction. *Economics of Education Review*, 55:132–148, 2016.

H. Horváth. Classroom assignment policies and implications for teacher value-added estimation. 2015.

G. Imbens. Matching methods in practise: Three examples. *Journal of Human Resources*, 50 (2):373–419, 2015.

T. Kane and D. Staiger. Estimating teacher impacts on student achievement: an experimental evaluation. *NBER Working paper*, 14607, 2008.

T. Kane, J. McCaffrey, R. Miller, and D. Staiger. Have we identified effective teachers? Validating measures of effective teaching using random assignment. *MET Project: Bill and Melinda Gates Foundation*, 2013.

S. Rivkin, E. Hanushek, and J. Kain. Teachers, schools, and academic achievement. *Econometrica*, 73 (2):417–458, 2005.

M. Ronfeldt, S. Loeb, and J. Wyckoff. How teacher turnover harms student achievement. *American Educational Research Journal*, 50(1):4–36, 2013.

J. Rothestein. Teacher quality in educational production: Tracking, decay, and student achievment. *Quarterly Journal of Economics*, 125 (1):175–214, 2010.

6 Appendix

6.1 Appendix 1

Given the structure of the moment conditions in the TVA estimation, the variance of the estimators cannot be derived analytical. Then we follow the bootstrap procedure described below:

1. From the specification in (1) estimate the Variance of the Combination value added, Var(μ_c)
2. From N$\sim$(μ_c, Var(μ_c)) we take $K = 2,000$ different draws, $\mu*_{c,k}$
3. Then we replace each of the draws, $\mu*_{c,k}$ in the moment conditions and estimate again the parameters θ
4. Then the variance of the estimators is taken from the distribution of the $K = 2,000$ bootstrapped estimations

6.2 Appendix 2

We show that the sample we are working with is representative of the whole population of students. Since we are working with a large database (more than 300,000 observations) any small difference ends up to be significant in the traditional t-statistics. In order to avoid such case we use the normalized test proposed by [Imbens, 2015], instead of the t-statistic for testing the null hypothesis of equal means between the sample and the population. For a sample, A and a population P the test is given as:

$$\Delta X = \frac{X_P - X_A}{\sqrt{(S^2_{X,P} + S^2_{X,A})/2}} \tag{5}$$

Table 10: Differences between sample used and population

	Population		Reading			Mathematics		
	Mean	**S.d**	**Mean**	**S.d**	**Test**	**Mean**	**S.d**	**Test**
Female	0.50	0.50	0.50	0.50	0	0.50	0.49	0
Portuguese mother	0.94	0.24	0.95	0.22	-0.04	0.94	0.23	0
Portuguese Father	0.93	0.25	0.94	0.24	-0.04	0.93	0.25	0
Mother - college graduated	0.15	0.36	0.15	0.36	0	0.15	0.35	0
Father - college graduated	0.09	0.29	0.09	0.30	0	0.09	0.29	0
Social support - type A (%)	0.20	0.40	0.20	0.40	0	0.19	0.40	0.03
Social support - type B (%)	0.22	0.42	0.22	0.41	0	0.22	0.41	0
Mother unemployed	0.09	0.29	0.09	0.29	0	0.09	0.29	0
Father unemployed	0.05	0.22	0.05	0.21	0	0.05	0.22 6	0
Age	10.15	0.47	10.15	0.48	0	10.15	0.48	0

The absolute values of the normalized test are modest, not exceed the value of 0.3. This shows that, in terms of the student variables, the sample we work with is close to the true population.

6.3 Appendix 3

Below we depict the transition matrices between the score in the 4th grade and the score in the6th grade:

Table 11: Grades - 4th and 6th grade - Reading

Score 4th/6th grade	**1**	**2**	**3**	**4**	**5**	**Total**
1	141	857	598	67	1	1664
2	870	20626	37095	10138	307	69036
3	277	18869	100283	61998	6073	187500
4	12	2361	31371	63462	17804	115010
5	1	36	1244	8266	6006	15553
Total	1301	42749	170591	143931	30191	388763

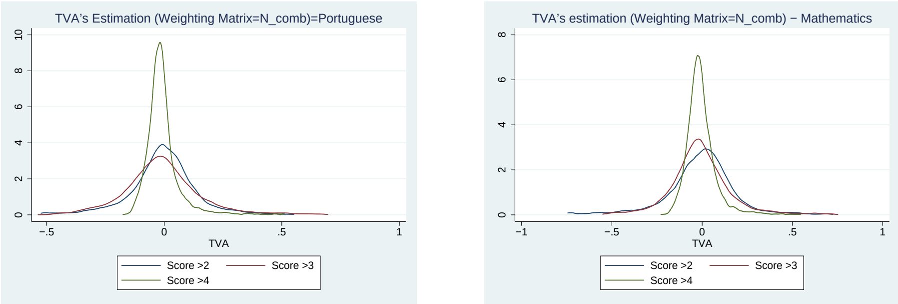
Table 12: Grades - 4th and 6th grade - Mathematics

Score 4th/6th grade	**1**	**2**	**3**	**4**	**5**	**Total**
1	1444	9534	4936	1073	126	16813
2	1335	37768	59545	18949	2777	120374
3	123	12427	70660	53751	13066	150027
4	14	1253	18171	47272	25964	92674
5	3	24	1018	9182	15013	25240
Total	2619	61006	154330	130227	56946	405128

6.4 Appendix 4

Below we show the kernel densities on the teacher value-added estimations:

Figure 8: Distribution Combination Value Added



6.5 Appendix 5

In the following table we report the descriptive statistics of the teacher characteristics used in section 4.2. To note that the number of teachers used to access the relation between the TVA and the teacher characteristics is different than the total number of teachers reported in Table 4. This is due to the imperfect merge within the used database.

Table 13: Teacher characteristics

	Reading				**Mathematics**			
	Mean	S.D	Min	Max	Mean	S.D	Min	Max
Female	0,875	0,33	0	1	0,87	0,32	0	1
Tenure	0,615	0,49	0	1	0,62	0,48	0	1
Master	0,1	0,3	0	1	0,1	0,3	0	1
Age	45,44	8,98	22	70	45,02	8,63	23	68
Exper	18,33	10,32	0	44,36	17,94	9,97	0	44,36
N	8659	-	-	-	7755	-	-	-