

Adjusting for measurement error in the value added model: evidence from Portugal

Maria Eugénia Ferrão · Harvey Goldstein

© Springer Science+Business Media B.V. 2008

Abstract The multilevel value added approach to measuring school effectiveness is now widely used. We propose a method to adjust for measurement error to investigate the extent to which this changes school effect estimates. It is applied to longitudinal data collected in the region of Cova da Beira (NUT III) for 1st, 3rd, 5th, 7th and 8th grades. Three different variance component models are considered, depending on the predictor variables included. Assuming measurement error occurs in explanatory and/or response variables, corrections are made for different values of the coefficient of reliability. Moreover, models are fitted under the assumption of either independent or correlated measurement errors.

Keywords Multilevel model · Value-added · Measurement error · Longitudinal data · Reliability · Correlated measurement errors · School effectiveness

1 Introduction

From the position of thirty years ago that schools made no difference, the School Effectiveness Research movement ([Mortimore et al. 1988](#)) has provided such a great deal of evidence that nowadays there is a general consensus that schools add value to pupils' integral development.

It is now generally agreed that performance comparisons among schools should be made based on the value added by a school to the learning and development of each and every student ([Scheerens 1992](#), [Raudenbush and e Willms 1995](#), [Goldstein and Spiegelhalter 1996](#), [Goldstein 1997](#), ?). This concept that Education borrowed from Economics is defined by ?, p. 24 as an indication of how a school has contributed for the progress of all the students in a set of areas (academic or not) during a limited period of time—e.g., typically from entering

M. E. Ferrão (✉)
Departamento de Matemática da, Universidade da Beira Interior, Rua Marquês d'Ávila e Bolama,
6200-001 Covilha, Portugal
e-mail: meferrao@mat.ubi.pt

H. Goldstein
University of Bristol, Bristol, UK

that school until the final school results are known—in comparison to other schools. [Braun and Wainer \(2007\)](#) define the Value Added Model (VAM) as the “family of statistical models that are employed to make inferences about the effectiveness of educational units, usually schools and /or teachers.”

Numerous studies have aimed at identifying the factors that explain differences across schools and at assessing the magnitude and stability of school contributions to student outcomes. That evidence has had governmental implications on educational policies in several countries since there has been increasing interest in the accountability of public institutions. However, there has been some controversy on how to measure the “quality” of public services. The implementation of any definition of value added implies the use [Goldstein \(1997\)](#) of adjusted comparisons.

The debate focuses on the theoretical but mainly methodological models and procedures to properly quantify the value added, school adjusted comparison ([Goldstein 1997](#)) or school effect ([Raudenbush and e Willms 1995](#)). [Fielding et al. \(2003\)](#), for example, compare VA estimates based on a multilevel model for point scores and a multilevel model for ordered categories. For both kind of models covariates are student prior achievement, gender, age, school type of funding and admission policy and examination board. It is shown that the correlation coefficients and rank correlations between the institution residual estimates from each pair of models are larger than 0.96.

Data for the VAM are drawn from observational studies, so the randomization criterion is not met. Therefore adjustments for prior differences among students across classes/schools are needed for valid inferences. [Goldstein and Spiegelhalter \(1996\)](#) set out the statistical issues involved in making reliable quantitative comparisons between institutions in the fields of education and health and say,

Our use of multilevel models reflects our default assumption that having made suitable adjustments we expect institutions broadly to be similar. Statistically this means that higher level units can be regarded as drawn from a population of units or more technically to be ‘exchangeable’ ([Bernardo and Smith, 1994](#)).

The sets of variables used for *adjustment* purposes vary. In some models the adjustments are based on students demographic characteristics, or based on students test scores histories and some comprise both, possibly including some school characteristics. Thus all VAM estimates are conditional on the set of variables used in the model.

Like in many subject areas where statistical methods are used, variables in VAM are not free of measurement error or misclassification errors for discrete variables. The consequences of ignoring measurement error lead to underestimation of coefficients and biased standard errors. The research on the impact of measurement error on VA estimates has not been extensive. [Ladd and Walsh for example \(2002\)](#) use prior year test scores as an instrumental variable to adjust generally for misspecification error, including measurement errors. [Goldstein et al. \(to appear\)](#) on the other hand develop Bayesian methods for adjusting exclusively for measurement error and misclassification in multilevel models. They show how inferences about both fixed and random effects are changed when measurement error and misclassification are considered.

Two main objectives are pursued in this paper. The first is to assess the impact of different set of adjusting or controlling variables on school value added estimates. The second is to evaluate the impact of measurement error in VA estimates. We present and discuss methodological issues in Sect. 2. In Sect. 3 multilevel model estimates are reported and VA estimates are compared. The paper ends with a brief concluding discussion.

2 Methodology

2.1 Statistical models

Several other authors, such as [Aitkin and Longford \(1986\)](#), [McPherson \(1992\)](#), [Goldstein et al. \(1993\)](#), [Willms and Raudenbush \(1989\)](#), [Rubin et al. \(2004\)](#) pointed out methodological requisites for the use of *value added* and/or *adjusted comparison* in the educational context. In short, such requisites are: (1) to assess baseline knowledge/skills (student prior achievement); (2) to consider the longitudinal nature of *progress*; (3) to consider the multilevel structure of the population under study (reflected in the data collected); (4) to consider extra-school factors that may enhance or slow down progress, such as the socio-economic and cultural level of the students.

It is broadly agreed that multilevel models are appropriate for VA estimates. Typically a two-level model—students nested within schools—is applied to represent the hierarchical structure, and in some cases a three level model for pupils within classes within schools.

In our analysis, three types of variance component models are used: (1) a null model, (2) a traditional value added model (TVA) and (3) a TVA model with extra-school variables. These models are presented in the first part of this section while the second part introduces the statistical framework to deal with measurement errors occurring both in the explanatory variables and response variables of the TVA model. Additionally, random coefficient models were fitted. However, none of the random effects were statistically different from zero.

2.1.1 TVA

The value added model for the 3EM project is estimated annually based on [Goldstein's \(1997\)](#) proposal and adapted to the Portuguese educational and social context. [Goldstein \(1997, p. 383\)](#) defines model (1) for estimating school value added.

$$\begin{aligned}
 y_{ij} &= \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \varepsilon_{ij} \\
 \beta_{0j} &= \gamma_{00} + u_{0j} \\
 \varepsilon_{ij} &\sim N(0, \sigma_e^2) \\
 u_{0j} &\sim N(0, \sigma_{u_0}^2)
 \end{aligned}
 \tag{1}$$

Thus, the value added is quantified by adjusted residuals (\hat{u}_{0j}) at level 2; where \hat{u}_{0j} is the school effect.

The response, y_{ij} , is the observed value (3EMat score) for student i in school j ; x_{1ij} is the prior achievement and x_{2ij} the socio-economic and cultural level of the ij student (parents' education is used as proxy).

2.1.2 TVA+

Other explanatory variables were added to model (1) in order to compare the impact of a different set of adjusting variables on the VA estimates. The variables are gender, special educational needs, whether the pupil repeats a year and whether the pupil is in a mixed age class.

2.1.3 TVA model with measurement errors

Classical test theory (see e.g. Lord and Novick 1968) proposes a measurement model that can be written as

$$\begin{aligned}x_{1ij} &= x_{1ij}^* + m \\ m &\sim N(0, \sigma_m^2)\end{aligned}\quad (2)$$

Where x_{1ij} is the measurement on the ij student (the observed value), x_{1ij}^* is the true score for student ij , and m is the measurement error with variance σ_m^2 . The variance of the true score is $\sigma_{x_{1i}^*}^2$. It is usually assumed that the m are mutually independent. Furthermore, it is assumed that the expectation of m is zero and the measurement error is independent of the true score.

The reliability of x_1 , R , can be defined as the proportion of the total variance of the measurements that is due to the true score variance,

$$R = \frac{\sigma_{x_{1i}^*}^2}{\sigma_{x_{1i}^*}^2 + \sigma_m^2}\quad (3)$$

The variance of x_1 , $\sigma_{x_1}^2$, can be estimated from the sample of values, and assuming σ_m^2 is known, $\sigma_{x_{1i}^*}^2$ can be obtained.

Goldstein et al. (to appear) show how information given by (3) can be used in an MCMC estimation algorithm.

2.2 The 3EM data set

The project *Eficácia Escolar no Ensino da Matemática* (3EM¹) is a school effectiveness research (SER) project which takes place in the region of Cova da Beira, a NUT III Portuguese region. Students enrolled in compulsory education (primary—four years, elementary—two years, and lower secondary, three years) define the target population.

The survey design is longitudinal and consists of three waves—2005/6, 2006/7 and 2007/8, and data are collected at the beginning and at the end of each academic year. In the 1st wave (2005/6) the academic years involved are the 1st, 3rd, 5th, 7th and 8th. In the 2nd wave (2006/7) these students will be monitored, respectively, at the 2nd, 4th, 6th, 8th and 9th years and a new group of students at the 1st, 3rd, 5th, and 7th years will be surveyed. In the 3rd wave (2007/8) all these students will be monitored and a new group of students at the 1st grade will be added. The random sample is representative at the level of county and NUT III region (Vicente 2007).

The Portuguese Council for Data Protection gave permission to run the survey, but conditional on parents' agreement. The initial sample was oversampled in order to take account of parents non-agreement and dropout or attrition, which is a known problem in longitudinal studies. The analysis of data collected at the first occasion show that the observed sample is still representative of the population (Vicente 2007).

The 3EM study includes three cycles of education—primary, elementary, and lower secondary. School organization in primary education is quite different from elementary or secondary, since there is, usually, one teacher per class while in other cycles one class is taught by several teachers and several teachers teach several classes. In general, primary schools are smaller considering the number of students and the number of classes. The 3EM hierarchical

¹ Research project granted by the Ministry of Science, Technology and Higher Education and by the Calouste Gulbenkian Foundation.

Table 1 1st wave sample composition

Grade	Number of		
	Students	Classes	Schools
1	309 ^a	35	35
3	327 ^b	37	37
5	306	19	9
7	287	18	11
8	248	16	11

^a 36% in mixed classes (classes where more than one level is taught)

^b 47% in mixed classes

Table 2 Variables used—1st and 3rd grades

Student level variable	Description
3EMat-score0	Prior achievement: Maths score at the beginning of the year
Gender	Male/Female
Parent's education	1: more than 12 years; 2: 12 years; 3: 9 years; 4: 6 years; 5: 4 years; 6: less than 4 years of schooling; "Don't know" was coded as missing
Special educational needs	Whether pupil was statemented in 2005/6
Repeated+1	Whether pupil was not promoted 1 or more years.
Kindergarten	Whether pupil attended kindergarten
School level variable	Description
Mixed class	Whether pupil attend a mixed class or not

structure is in accordance with this. Table 1 presents the number of students, classes and schools involved.

As we can see, in the first cycle there only is one class per school. On average, the number of students per class² is 17 (5 in mixed classes) at grade 1 and 16 (5 in mixed classes) at grade 3. In the target population, the class size is, on average, 19 students (13 in mixed classes).

Student outcome variable is achievement in Mathematics—3EMat (Ferrão et al. 2006). The 3EMat instrument is based on the Portuguese mathematics core curriculum. Contextual, input and process questionnaires were developed based on a theoretical SER model and were applied to students, families, teachers and headteachers.

2.2.1 Variables used in the analysis

The analyses have been computed for Mathematics test scores, 3EMAT, at the end of the year as response variable. Independent variables are presented in the Tables 2, 3 and 4. For the purpose of this paper 3EMat scores and parent's education have been standardized.

The largest dropout rate is 4.8% at the 8th grade. In primary education classes the rate is less than 1%. The action of teachers and head teachers strongly contributed to keep the rate at a low level.

² Students authorised by their parents to participate in the study are 85% of all students.

Table 3 Variables used–5th grade

Student level variable	Description
3EMat-score0	Prior achievement: Maths score at the beginning of the year
Gender	Male/Female
Parent's education	1: more than 12 years; 2: 12 years; 3: 9 years; 4: 6 years; 5: 4 years; 6: less than 4 years of schooling; "Don't know" was coded as missing
Special educational needs	Whether pupil was stated in 2005/6
Kindergarten	Whether pupil attended kindergarten
Repeated+1	Whether pupil was not promoted 1 or more years during the 1st cycle of primary education
School level variable	Description
Mixed	Whether pupil attended or not a mixed class during the primary education

Table 4 Variables used–7th and 8th grades

Student level variable	Description
3EMat-score0	Prior achievement: Maths score at the beginning of the year
Gender	Male/Female
Parent's education	1: more than 12 years; 2: 12 years; 3: 9 years; 4: 6 years; 5: 4 years; 6: less than 4 years of schooling; "Don't know" was coded as missing
Special educational needs	Whether pupil was stated in 2005/6
Repeated+1	Whether pupil was not promoted 1 or more years during the 1st or 2nd cycle of education

The dropout and missing responses, mainly due to parent's education variable, reduce the number of cases by 6.1%, 5.7%, 10.0%, 8.1% and 10.3% at each grade, respectively.

For the purposes of parameter estimation, missing responses are assumed "missing at random" (Little and Rubin 2002).

3 Modelling results

3.1 Multilevel model estimates

3.1.1 Variance component (null) model

The null model estimates are presented in Table 5. Level one and level two variances allow us to compute the variance partition coefficient (VPC) (Goldstein 2003, p. 17; Goldstein et al. 2002). The results suggest that VPC is higher in primary education than in elementary or lower secondary. The Math score variability at the end of academic year due to differences between schools represents around 0.15 in primary education and varies from 0.08 to 0.10 in elementary or lower secondary.

Table 5 Null model estimates

	Grade				
	1st	3rd	5th	7th	8th
Intercept	-0.019 (0.092)	-0.122 (0.088)	-0.005 (0.086)	-0.038 (0.091)	-0.003 (0.100)
Random parameters					
σ_u^2	0.155 (0.067)	0.164 (0.066)	0.077 (0.045)	0.080 (0.048)	0.096 (0.057)
σ_e^2	0.848 (0.073)	0.857 (0.072)	0.918 (0.079)	0.911 (0.080)	0.897 (0.086)
-2(log-likelihood)	823.01	879.929	816.399	763.993	647.909
VPC	0.15	0.16	0.08	0.08	0.10
Number of cases	297	316	291	273	232

Table 6 VAM model estimates

	Grade				
	1st	3rd	5th	7th	8th
Intercept	-0.014 (0.093)	-0.046 (0.076)	-0.039 (0.048)	-0.042 (0.077)	-0.008 (0.088)
Prior score	0.468 (0.051)	0.443 (0.052)	0.656 (0.047)	0.599 (0.056)	0.374 (0.063)
Parent's education	-0.113 (0.054)	-0.209 (0.054)	-0.117 (0.047)	-0.037 (0.057)	-0.113 (0.065)
Random parameters					
σ_u^2	0.179 (0.068)	0.112 (0.048)	0.005 (0.013)	0.054 (0.034)	0.037 (0.035)
σ_e^2	0.636 (0.057)	0.632 (0.055)	0.479 (0.043)	0.642 (0.059)	0.770 (0.079)
-2(log-likelihood)	701.896	735.763	553.206	614.581	543.728
VPC	0.22	0.15	0.01	0.08	0.05
R ²	0.19	0.27	0.51	0.31	0.19
Number of cases	279	298	262	251	208

3.1.2 Traditional value added model

Based on the TVA model, fixed parameter estimates for maths prior achievement show the relevance of that variable for maths achievement measured at the end of the year. The coefficient is statistically different from zero for all levels of education varying from 0.37 at grade 8 to 0.66 at grade 5 (Table 6).

Concerning the effect of parents' education on Maths scores the model results show a negative relationship³ ($\alpha = 0.1$ for 8th grade) with the exception of 7th grade. Note that where the prior achievement is included the other coefficients are interpreted as effects on progress.

Adjusting for student's prior achievement and parent's education, VPC tends to increase in primary education and to decrease in elementary and lower secondary. The variance of score in Maths is strongly explained by these two variables.

3.1.3 Traditional value added model with additional variables

The results of the model TVA+ are presented in Table 7. They support the findings described above. Moreover, the model suggests that, at grade 1 and 3, students statemented get reduced scores compared with their colleagues. The attendance at kindergarten is positive and statistically significant ($\alpha = 0.1$) at the 1st grade. On average, at the 3rd grade, when students

³ Parents' education scale is inverted.

Table 7 VAM model with additional variables

	Grade				
	1st	3rd	5th	7th	8th
Intercept	-0.484 (0.265)	-0.044 (0.206)	0.087 (0.135)	-0.194 (0.135)	0.138 (0.164)
Prior score	0.429 (0.052)	0.429 (0.054)	0.618 (0.050)	0.579 (0.060)	0.348 (0.063)
Parent's education	-0.099 (0.055)	-0.149 (0.056)	-0.100 (0.049)	-0.025 (0.125)	-0.077 (0.067)
Special needs	-0.748 (0.318)	-0.682 (0.254)	-0.322 (0.203)	-0.388 (0.243)	-0.076 (0.250)
Kindergarten	0.120 (0.065)	0.141 (0.179)	0.147 (0.124)	0.193 (0.125)	0.015 (0.157)
Mixed class in primary education	-0.005 (0.183)	-0.267 (0.152)	0.027 (0.094)	-	-
Boy	0.144 (0.100)	0.189 (0.100)	0.110 (0.086)	0.064 (0.103)	0.002 (0.124)
Repeated 1+	-	-0.305 (0.172)	-0.286 (0.118)	-0.015 (0.134)	-0.393 (0.140)
Random parameters					
σ_u^2	0.159 (0.063)	0.086 (0.042)	0.016 (0.017)	0.048 (0.032)	0.027 (0.031)
σ_e^2	0.621 (0.056)	0.624 (0.056)	0.456 (0.042)	0.632 (0.059)	0.745 (0.076)
$-2(\log\text{-likelihood})$	683.384	688.741	533.074	607.320	535.306
R^2	0.22	0.30	0.53	0.31	0.22
Number of cases	275	284	256	250	208

go to a mixed class their results are reduced on average 0.27 standard deviations, comparing with students who go to a single year class. According to these results, the advantage of boys in Maths is statistically significant only at the 3rd grade.⁴

With the exception of 7th grade, having repeated at least one year penalises student achievement in Maths. The marginal effect varies from -0.29 to -0.43 .

3.1.4 Traditional value added model with measurement errors

The TVA model was fitted adjusting for measurement error as explained in the previous section. The analyses were carried out using the software described in Goldstein et al. (to appear). Tables 8–12 show the estimates for different assumptions about the error measurement variance. First the model, in column [1], assumes that both prior achievement and parent's education have error with reliability of 0.9 and that the correlation between their respective error terms is zero. The second model assumes that correlation to be 0.5. The third model is similar to the first but with reliability of 0.6. In the fourth the reliability of both measures is 0.6 and it is assumed that the response variable is also measured with a reliability of 0.9. Finally, the last model, assumes the correlation between the error terms to be 0.5.

In general, the results corroborate those reported by Goldstein et al. (to appear). The prior achievement coefficients are greatly increased when the lowest reliability is assumed, with also a very large increase in standard error and the level 1 variance is reduced. The same happens when the model allows for measurement error in the response variable, but the impact is not so severe. It can be seen in column [2] and [5] results that the effect of the correlation between the errors of the measurement is revealed mainly in the parent's education coefficient.

⁴ The hypothesis test for the difference of prior achievement at year 1 between boys and girls was statistically significant.

Table 8 VAM model estimates with measurement error

Grade 1			
	R(x1) = R(x2) = 0.9 Corr(m1,m2) = 0.5	R(x1) = R(x2) = 0.6 R(y) = 0.9	R(x1) = R(x2) = 0.6 R(y) = 0.9
Intercept	-0.021 (0.099)	-0.0168 (0.1003)	-0.0115 (0.0968)
Prior score	0.5155 (0.0583)	0.7699 (0.0976)	0.7341 (0.1273)
Parent's education	-0.1158 (0.0606)	-0.1012 (0.1027)	0.0780 (0.1393)
Random parameters			
σ_u^2	0.2002 (0.0862)	0.2022 (0.0864)	0.1844 (0.0811)
σ_e^2	0.6218 (0.0862)	0.4922 (0.0658)	0.5272 (0.0651)

Table 9 VAM model estimates with measurement error

Grade 3			
	R(x1) = R(x2) = 0.9 Corr(m1,m2) = 0.5	R(x1) = R(x2) = 0.6 R(y) = 0.9	R(x1) = R(x2) = 0.6 R(y) = 0.9
Intercept	-0.0462 (0.0776)	-0.0433 (0.0791)	-0.0329 (0.0824)
Prior score	0.4883 (0.0603)	0.7520 (0.1284)	0.7924 (0.2733)
Parent's education	-0.2150 (0.0625)	-0.1695 (0.1298)	0.0630 (0.2848)
Random parameters			
σ_u^2	0.1156 (0.0538)	0.1146 (0.0534)	0.1083 (0.0545)
σ_e^2	0.6118 (0.0561)	0.4848 (0.0644)	0.5259 (0.0594)

Table 10 VAM model estimates with measurement error

Grade 5			
	$R(x1) = R(x2) = 0.9$	$R(x1) = R(x2) = 0.9$ $Corr(m1,m2) = 0.5$	$R(x1) = R(x2) = 0.6$ $R(y) = 0.9$
Intercept	$-0.0485 (0.0504)$	$-0.0473 (0.0495)$	$-0.0654 (0.0594)$
Prior score	$0.7414 (0.0551)$	$0.7603 (0.0588)$	$0.9364 (0.1026)$
Parent's education	$-0.0971 (0.0551)$	$-0.0463 (0.0590)$	$-0.0294 (0.1083)$
Random parameters			
σ_u^2	$0.0110 (0.0127)$	$0.0105 (0.0126)$	$0.0107 (0.0122)$
σ_e^2	$0.4356 (0.0432)$	$0.4403 (0.0448)$	$0.4009 (0.0427)$
			$R(x1) = R(x2) = 0.6$ $Corr(m1,m2) = 0.5$
			$-0.0679 (0.0661)$
			$1.0235 (0.1442)$
			$0.1734 (0.1534)$
			$0.0093 (0.0109)$
			$0.3578 (0.0530)$

Table 11 VAM model estimates with measurement error

Grade 7			
	$R(x1) = R(x2) = 0.9$	$R(x1) = R(x2) = 0.9$ $Corr(m1,m2) = 0.5$	$R(x1) = R(x2) = 0.6$ $R(y) = 0.9$
Intercept	$-0.0432 (0.0806)$	$-0.0452 (0.0826)$	$-0.0484 (0.0894)$
Prior score	$0.6788 (0.0656)$	$0.6934 (0.0691)$	$0.9230 (0.1187)$
Parent's education	$-0.0134 (0.0647)$	$0.0283 (0.0678)$	$0.0830 (0.1184)$
Random parameters			
σ_u^2	$0.0661 (0.0457)$	$0.0647 (0.0443)$	$0.0636 (0.0447)$
σ_e^2	$0.6136 (0.0613)$	$0.6156 (0.0628)$	$0.4794 (0.0686)$
			$R(x1) = R(x2) = 0.6$ $R(y) = 0.9$
			$-0.0429 (0.0932)$
			$1.0462 (0.1703)$
			$0.3085 (0.1725)$
			$0.0636 (0.0424)$
			$0.4898 (0.0751)$

Table 12 VAM model estimates with measurement error

Grade 8					
	$R(x1) = R(x2) = 0.9$	$R(x1) = R(x2) = 0.9$ Corr(m1,m2) = 0.5	$R(x1) = R(x2) = 0.6$	$R(x1) = R(x2)$ = 0.6 $R(y) = 0.9$	$R(x1) = R(x2) = 0.6$ $R(y) = 0.9$ Corr(m1,m2) = 0.5
Intercept	-0.0106 (0.0789)	-0.0069 (0.0789)	-0.0113 (0.0825)	-0.0123 (0.0886)	-0.0125 (0.0870)
Prior score	0.4165 (0.0734)	0.4160 (0.0756)	0.6249 (0.1239)	0.5599 (0.1155)	0.6314 (0.1856)
Parent's education	-0.1175 (0.0740)	-0.0956 (0.0748)	-0.1055 (0.1246)	-0.1146 (0.1208)	0.0699 (0.1942)
Random parameters					
σ^2	0.0412 (0.0421)	0.0413 (0.0416)	0.0423(0.0418)	0.0442 (0.0462)	0.0468 (0.0471)
σ^2_{η}	0.7781 (0.0815)	0.7809 (0.0819)	0.6881(0.0872)	0.6484 (0.0824)	0.6536 (0.0847)

Table 13 Correlation between VA estimates

	Grade				
	1st	3rd	5th	7th	8th
Null-TVA	0.88	0.80	0.61	0.83	0.94
Null-TVA+	0.88	0.79	0.59	0.81	0.91
TVA-TVA+	0.99	0.88	0.97	0.99	0.96

Table 14 Correlation between ranks

	Grade				
	1st	3rd	5th	7th	8th
Null-TVA	0.90	0.78	0.56	0.72	0.93
Null-TVA+	0.91	0.77	0.54	0.68	0.89
TVA-TVA+	0.99	0.85	0.93	0.94	0.97

3.2 Comparison between VA estimates

Value added estimates are school-level residuals, as defined in Sect. 2.1. They were generated from the null, TVA and TVA+ models and also from TVA with measurement error. The correlation between VA estimates based on null model, TVA and TVA+ is presented in Table 13.

The correlation between VA estimates generated from the Null and TVA models varies from 0.61 to 0.94 depending on the grade. With the exception of the 3rd grade, the values of the correlation between TVA and TVA+ estimates are equal or larger than 0.96.

Another way to illustrate the impact of different VA models on estimates is comparing the position of each institution at the rank produced by the respective model. The correlations are in Table 14. In general, the above conclusions are corroborated.

We have computed the correlation between TVA model estimates with and without considering measurement error. The correlation is larger than 0.97 for every model/grade.

4 Discussion

In this paper value added models are applied to a dataset collected in a representative sample of students at the 1st, 3rd, 5th, 7th and 8th grades in Cova da Beira, a Portuguese NUT III region. Depending on the adjusted variables included, three different models are considered—null model, traditional value added model (TVA, which includes SES and prior achievement) and TVA adding variables such as gender, whether the student is stated as special needs, if he/she attended the kindergarten, type of class in primary education, and repetition. In addition, the influence of measurement error in value added estimates was explored in the TVA model.

It was shown that the set of adjusting variables in a VAM changes the VA estimates, particularly SES and prior achievement variables. Additional variables have fairly small effects on VA estimates. Including the measurement error influences greatly the fixed parameter estimates and their standard errors, but not the value added estimates. However, and despite the fact that we know that measurement error is not constant along the score scale, and that

there can be unobserved patterns of measurement error variation across schools, these aspects were not considered here. More research work is needed on this issue.

From the educational point of view, all models make clear the importance of prior achievement as a requirement for success in mathematics development. Moreover, when the student was not promoted to the next grade at least once, his/her results in maths are lower than their colleagues.

The statistical models fitted suggest that the variance partition coefficient is larger in primary education than in elementary or lower secondary.

References

- Aitkin, M., Longford, N.: Statistical modelling issues in school effectiveness studies. *J. Roy. Stat. Soc. A* **149**, 1–43 (1986)
- Braun, H., Wainer, H.: Value-Added Modeling. In: Rao, C.R., Sinharay, S. (eds.) *Handbook of Statistics 26, Psychometrics*, pp. 475–501. Elsevier (2007)
- Ferrão, M.E., Costa, P., Dias, V., Dias, M.: Medição da competência dos alunos do ensino básico em Matemática: 3EMat, uma proposta. [Measuring math skills of students in compulsory education: 3EMat, a proposal]. In: *Actas da XI Conferência Internacional de Avaliação Psicológica Proceedings of the XI International Conference on Psychological Evaluation*, Braga, Portugal, 5–8 October 2006
- Fielding, A., Yang, M., Goldstein, H.: Multilevel ordinal models for examination grades. *Stat. Model.* **3**, 127–153 (2003)
- Goldstein, H.: Methods in school effectiveness research. *School Effectiv. School Improv.* **8**, 369–395 (1997)
- Goldstein, H.: *Multilevel Statistical Models*, 3rd edn. Edward Arnold, London (2003)
- Goldstein, H., Spiegelhalter, D.J.: League tables and their limitations: statistical issues in comparison of institutional performance. *J. Roy. Stat. Soc. A* **159**, 385–443 (1996)
- Goldstein, H., Browne, W., Rasbash, J.: Partitioning variation in multilevel models. *Understand. Stat.* **2**, 223–232 (2002)
- Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttal, D., Thomas, S.: A multilevel analysis of school examination results. *Oxford Rev. Educ.* **19**, 425–433 (1993)
- Goldstein, H., Kounali, D., Robinson, A.: Modelling measurement errors and category misclassifications in multilevel models. *Statistical Modelling*. (to appear)
- Ladd, H.F., Walsh, R.P.: Implementing value-added measures of school effectiveness: getting the incentives right. *Econ. Educ. Rev.* **21**, 1–17 (2002)
- Little, R.J.A., Rubin, D.B.: *Statistical Analysis with Missing Data*, 2nd edn. Wiley, New York (2002)
- Lord, F.M., Novick, M.R.: *Statistical Theories of Mental Test Scores*. Addison-Wesley Publishing, Reading, Massachusetts (1968)
- McPherson, A.: *Measuring Added Value in Schools (NCE briefing 1)*. National Commission on Education, London (1992)
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D., Ecob, R.: *School Matters*. Open Books, Wells (1988)
- Raudenbush, S.W., e Willms, J.D.: The estimation of school effects. *J. Educ. Behav. Stat.* **20**(4), 307–335 (1995)
- Rubin, D.B., Stuart, E.A., Zanutto, E.L.: A potential outcomes view of value-added assessment in Education. *J. Educ. Behav. Stat.* **29**(1), 103–116 (2004)
- Scheerens, J.: *Effective Schooling: Research, Theory and Practice*. Cassell, London (1992)
- Vicente, P.: Plano amostral do projecto 3EM—*Eficácia Escolar no Ensino da Matemática* In: Ferrão, M.E., Nunes, C., Braumann, C. (eds.) *Estatística: Ciência Interdisciplinar, Proceedings of the XIV Annual Conference of the Portuguese Statistical Society*. SPE, Lisboa (2007)
- Willms, J.D., Raudenbush, S.W.: A longitudinal hierarchical linear model for estimating school effects and their stability. *J. Educ. Meas.* **26**, 209–232 (1989)