

Misclassification error in language proficiency and pupil achievement in England

Lorraine Dearden, Francesca Foliano,
Alfonso Miranda, Sophia Rabe-Hesketh

ADMIN Node

Outline

- ▶ Why it is important to look at the English proficiency classification of pupils
- ▶ What are the incentives that schools have and the implications for measurement errors
- ▶ Econometric model to extract information from linked data

Motivation

- ▶ Importance of using administrative data linked to survey data
- ▶ The proportion of pupils who are labelled English as additional language (EAL) has constantly increased over time in schools in England, reaching 13.50% of the all population of pupils in 2008
- ▶ EAL is a proxy for English proficiency
- ▶ EAL may be subject to misclassification
 - ▶ Schools / teachers assign the EAL label NOT the children or the parents
 - ▶ LEAs pays schools a different “price” for EAL and non EAL pupils
 - ▶ Schools can “play’ the rules’ of Contextualised Value Added (CVA) using EAL as a *control variable*

What is Contextualised Value Added?

Contextualised value added in school j is the **average difference between actual y_{ij} and expected \hat{y}_{ij} attainment** for pupils $i = \{1, \dots, n_j\}$ in school j .

$$\text{CVA}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} (y_{ij} - \hat{y}_{ij}) \quad (1)$$

- ▶ CVA_j is interpreted as the *value* school j “adds” over what pupils are “expected” to archive
- ▶ \hat{y}_{ij} takes into account observable characteristics at pupil and school level, including pupil’s previous attainment
- ▶ CVA is used in England for:
 - ▶ Performance Tables
 - ▶ Self-evaluation and target setting
 - ▶ Trigger school inspections

Do schools know how CVA works?

Key Stage 2 (Prior attainment finely graded point scores) Notes

To calculate these scores click [HERE](#)

	Difference from Average	Model Adjustment	
English	28.2	-0.86	28.2
Mathematics	29.22	0.16	29.22
Science	29.76		29.76
Average Point Score	29.06	+330.23	29.06 +328.29

Pupil Characteristics Notes

Gender: Female **-14.57**

SEN: Non-SEN **+0.00**

Date of Joining (DDMMYYYY): 01092004 **+0.00**

Date of Birth (DDMM): 1004 **+1.40**

In care whilst at school: No **+0.00**

IDACI:

First Language (for given Prior Attainment): English or believed to be English **+0.00**

Ethnicity: Irish **-1.40**

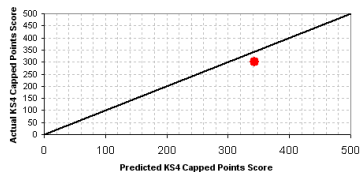
FSM (for given Ethnicity): No **+0.00**

School Characteristics Notes

Prior Attainment:

Spread of Prior Attainment:

Ready Reckoner KS2-4 2006 - Mainstream



Enter Pupil Outcome (Capped Point Score)

Pupil Predicted Score

Pupil Predicted Score after Ceiling/Floor Adjustments Notes

Contextual Value Added Score

Reset Form

So...

Even when English is the first language, children from ethnic minorities may be classified as EAL if that helps the school with the CVA or additional funding

Why is misclassification important?

This misclassification is a *confounder* of the relationship between English proficiency and achievement. . . And we would like to identify the *true relationship*

Research questions

- ▶ Is there a true relationship between English proficiency and achievement?
- ▶ Is this relationship stable at different stages of education?

Data: NPD and LSYPE linked

- ▶ The National Pupil Database (NPD) contains yearly records for each pupil enrolled in a state school in England, covering both the pupil's characteristics and their examination results.
- ▶ The Longitudinal Survey of Young People in England (LSYPE) is a large-scale panel study collecting information about young people in England aged 13 to 14 in 2004. Amongst the main topics covered by the study are household and demographic information, languages spoken in the home, attitudes towards school and involvement in education, parental expectations and aspirations, family activities, etc

EAL definitions

- ▶ **EAL_{school}**: first language is not (or is believed not to be) English
- ▶ **EAL_{child}**: English is not the only or first/main language yp speaks
- ▶ **EAL_{parents}**: English is not the only or first/main language spoken at home

Table: EAL status reported by parents and by the school

		EAL in PLASC 2006		
		0	1	Total
EALp	0	2,017 66.28%	1,026 33.72%	3,043 100%
	1	201 8.08%	2,287 91.92%	2,488 100%
Total		2,218 40.1%	3,313 59.9%	5,531 100%

Table: EAL status reported by the yp and by the school

		EAL in PLASC 2006		
		0	1	Total
EALc	0	1873 50.32%	1849 49.68%	3722 100%
	1	345 19.07%	1464 80.93%	1809 100%
Total		2218 40.1%	3313 59.9%	5531 100%

The set up

- ▶ The main response y_c is a continuous variable that represents achievement (KS3 and GCSEs score)
 - ▶ There are three measures of EAL
 - ▶ School (EAL_s)
 - ▶ Parent (EAL_p)
 - ▶ Child (EAL_c)
- \mathbf{x} : $1 \times K$ vector of predictors for achievement y_c
- \mathbf{z} : $1 \times L$ vector of predictors for EAL

The set up

- ▶ We have a measurement model that relates the three fallible measures to the unobservable truth that is a latent variable in the model for achievement
- ▶ We start by assuming that the misclassification probability is unrelated to achievement
- ▶ We will then relax this assumption for the school measure

Model 1: Non-differential ME + independence

We specify a simple linear model for the main response

$$y_c = \mathbf{x}'_c \boldsymbol{\beta} + \lambda_2 \eta_{1c} + v_c \quad (2)$$

and a set of latent variable models for the three observed measures of EAL

$$\text{EAL}_s^* = \gamma_s + \lambda_{1s} \eta_{1c} + \epsilon_s \quad (3)$$

$$\text{EAL}_p^* = \gamma_p + \lambda_{1p} \eta_{1c} + \epsilon_p \quad (4)$$

$$\text{EAL}_c^* = \gamma_c + \lambda_{1c} \eta_{1c} + \epsilon_c \quad (5)$$

with $\text{EAL}_j = 1(\text{EAL}_j^* > 0)$; $j = \{s, p, c\}$. η_{1c} represents the *true unobserved* dummy (i.e., a latent discrete variable), and $v_c, \epsilon = \{\epsilon_s, \epsilon_p, \epsilon_c\}$ represent error terms. Finally, $\boldsymbol{\gamma} = \{\gamma_s, \gamma_p, \gamma_c\} \in \mathbb{R}^3$, $\boldsymbol{\lambda}_1 = \{\lambda_{1s}, \lambda_{1p}, \lambda_{1c}\} \in \mathbb{R}^3$ and $\lambda_2 \in \mathbb{R}$ are all free parameters.

Outstanding questions

- ▶ What is the exact mechanism behind the schools gaming?