

# Efeitos da composição da turma no desempenho dos alunos das escolas públicas portuguesas

Ana Catarina Pimenta <sup>1</sup>

<sup>1</sup> Universidade do Minho

Mestrado em Economia

Trabalho efetuado sob a orientação de Carla Sá e Miguel Portela  
Março de 2019

# Estrutura da apresentação

- Motivação
- Objetivo da dissertação
- Construção e tratamento da base de dados
- Metodologia econométrica
- Resultados empíricos
  - ▶ Matemática
  - ▶ Português
- Discussão
- Investigação futura

# Motivação

- Com quem o estudante partilha a sala de aula pode determinar a quantidade de tempo que presta atenção ao professor, o tipo de aulas a que é exposto ou até mesmo os pares com quem vai interagir fora da escola (Firmino *et al.*, 2016).
- A forma de organização dos alunos pelas turmas mais frequentemente debatida na literatura prende-se com o **desempenho prévio**, considerado como *proxy*, ou indicativo, da capacidade do estudante (e.g. Booij *et al.*, 2017; Collins e Gan, 2013; Ding e Lehrer, 2007; Duflo *et al.*, 2011; Ferrão, 2012; Hattie, 2002 e Slavin, 1990).

# Motivação

- Para além disso, diversos autores relacionam o desempenho dos alunos com **características da turma** como:
  - ▶ O **contexto socioeconómico** (e.g. Carneiro, 2008; Firmino *et al.*, 2016; Kirjavainen, 2012; Lee e Barro, 2001; Pereira e Reis, 2012 e Sousa, 2016);
  - ▶ A **composição em termos de género** (e.g. Firmino *et al.*, 2016; Hill, 2017; Pereira, 2010; Schone *et al.*, 2017 e Sousa, 2016);
  - ▶ A **fração de repetentes** (e.g. Lavy *et al.*, 2012);
  - ▶ A **idade média** (e.g. Firmino *et al.*, 2016 e Sousa, 2016);
  - ▶ A **fração de alunos com computador e/ou internet em casa** (e.g. Firmino *et al.*, 2016).

# Motivação

- Adicionalmente, ao longo das últimas décadas, a **dimensão da turma** tem sido um dos fatores mais discutidos, possivelmente por ser normalmente resultado de uma decisão governamental.
  - ▶ Existem estudos que comprovam a **eficácia de turmas pequenas** (e.g. Chetty *et al.*, 2011; Finn, 1998; Krueger, 1999; Lee e Barro, 2001 e Rivkin *et al.*, 2005);
  - ▶ Estudos com **resultados mistos** (e.g. Jepsen e Rivkin, 2009 e Woessmann e West, 2006);
  - ▶ E estudos em que **alterações no tamanho das turmas não surtem qualquer efeito no desempenho dos alunos** qualquer que seja o tipo de estudante (e.g. Hanushek, 1998; Hanushek, 1999 e Hoxby, 2000a).

# Objetivo da dissertação

- Quando comparada com os estudos internacionais, a literatura sobre os efeitos da composição da turma no desempenho dos alunos para o contexto português é ainda muito limitada e os resultados estão longe de serem consensuais.
- O presente estudo pretende contribuir para a literatura que analisa os efeitos da composição da turma no desempenho dos alunos, procurando dar resposta à seguinte questão:
  - ▶ **Qual a dimensão e composição das turmas que favorece o desempenho dos alunos nos exames nacionais de Matemática e Português?**

# Construção e tratamento da base de dados

Todas as análises empíricas serão implementadas com recurso a uma nova base de dados, construída a partir da informação recolhida de duas fontes do Ministério da Educação:

- **MISI** (Sistema de Informação do Ministério da Educação) gerida pela Direção Geral de Estatísticas da Educação e Ciência (DGEEC);
  - ▶ A MISI contém **informação ao nível do aluno**, tal como, género, data de nascimento, ano letivo, ano de ensino, turma, escola, classificações internas nas diversas disciplinas, existência de computador e internet em casa, elegibilidade para apoio social escolar e escolaridade dos pais/encarregado de educação.

# Construção e tratamento da base de dados

- **JNE** (Júri Nacional de Exames – Direção Geral da Educação).
  - ▶ Os dados fornecidos pelo JNE incluem informação acerca das **classificações obtidas pelos alunos nos exames nacionais**.

Uma vez que os estudantes são identificados com o mesmo número anónimo em ambas as bases de dados é possível relacionar a informação e utilizar o **carácter longitudinal dos dados** para seguir a trajetória do aluno ao longo do tempo.

Dada a disponibilidade de dados da MISI, optou-se por seguir **duas coortes de alunos** ao longo do tempo.



	2008/2009	2009/2010	2010/2011	2011/2012
<b>9º ano</b>	Nunca reprovaram	Reprovou 1 vez	Reprovou 2 vezes	Reprovou 3 vezes
<b>10º ano</b>		Nunca reprovaram	Reprovou 1 vez	Reprovou 2 vezes
<b>11º ano</b>			Nunca reprovaram	Reprovou 1 vez
<b>12º ano</b>				Nunca reprovaram

Figura 1: 1ª coorte de alunos analisados.

	2007/2008	2008/2009	2009/2010	2010/2011
<b>9º ano</b>	Nunca reprovaram	Reprovou 1 vez	Reprovou 2 vezes	Reprovou 3 vezes
<b>10º ano</b>		Nunca reprovaram	Reprovou 1 vez	Reprovou 2 vezes
<b>11º ano</b>			Nunca reprovaram	Reprovou 1 vez
<b>12º ano</b>				Nunca reprovaram

Figura 2: 2ª coorte de alunos analisados.

# Construção e tratamento da base de dados

- Ambas as coortes pretendem seguir os alunos **desde o 9º até ao 12º ano de escolaridade**;
- Em ambas as coortes assumiu-se que os alunos podem apresentar **até 3 reprovações**;
- Cada linha na base de dados contém **informação microeconómica ao nível do aluno** num determinado ano letivo e ano de ensino;
- Foram consideradas apenas as disciplinas de **Português e Matemática**;
- As classificações internas dos alunos correspondem às notas obtidas no **3º Período**;

# Construção e tratamento da base de dados

- Foram consideradas apenas as **escolas públicas de Portugal continental**;
- Foram mantidos apenas os alunos de **Ciências e Tecnologias**;
- Para cada disciplina, caso o aluno tenha realizado mais do que uma fase, foi considerada **a melhor nota de exame**;
- Para cada disciplina, foram considerados apenas os alunos que possuem **todas as notas internas** (do 9º até ao 12º ano de escolaridade) e **os dois exames nacionais** (do 9º e do 12º ano).

Tendo em consideração as duas coortes utilizadas, a base de dados restringe-se aos seguintes casos ...

# Construção e tratamento da base de dados

- ❶ Se o aluno realiza o exame nacional de 12º ano em 2011/2012 (1ª coorte) ou em 2010/2011 (2ª coorte), estará no 11º ano em 2010/2011 ou anteriormente, no 10º ano em 2009/2010 ou anteriormente e realiza o exame nacional de 9º ano em 2008/2009 ou 2007/2008.
- ❷ Se o aluno realiza o exame nacional de 9º ano em 2008/2009 (1ª coorte) ou em 2007/2008 (2ª coorte), estará no 10º ano em 2008/2009 ou posteriormente, no 11º ano em 2009/2010 ou posteriormente e realiza o exame nacional de 12º ano em 2010/2011 ou em 2011/2012.

# Construção e tratamento da base de dados

- **Combinando as duas restrições, para possuir tanto o exame de 9º como o de 12º ano, um aluno terá de frequentar o:**
  - ▶ 9º ano em 2007/2008 ou 2008/2009;
  - ▶ 10º ano em 2008/2009 ou 2009/2010;
  - ▶ 11º ano em 2009/2010 ou 2010/2011;
  - ▶ 12º ano em 2010/2011 ou 2011/2012.

# Metodologia econométrica

- A **variável dependente** utilizada nos diferentes modelos corresponde ao **logaritmo da nota de exame** dos alunos no 9º e no 12º ano – **medida externa da qualidade escolar**.
- A base de dados possui **duas linhas por cada identificador anonimizado do aluno**.
  - ▶ Uma primeira linha é referente à nota de exame de 9º ano e às características do estudante e da turma no ano de realização deste exame.
  - ▶ A segunda linha possui a nota de exame de 12º ano do mesmo aluno, bem como informação relativa às características do estudante e da turma no ano de realização deste exame nacional.
- Apenas foram considerados os alunos que realizaram o exame nacional e a componente interna da disciplina no mesmo ano letivo.

# Metodologia econométrica

Apesar da riqueza da base de dados permitir a inclusão de diversas variáveis de controlo, existem dois problemas que têm sido frequentemente debatidos na literatura:

- 1 Existem **características não observáveis** (tal como a capacidade inata do aluno). Se esta variável estiver correlacionada com alguma das variáveis explicativas incluídas no modelo – **endogeneidade**.
- 2 Os alunos podem **autosseleccionar-se para determinadas escolas**. Ao utilizar efeitos fixos da escola é possível controlar para tudo o que seja específico da escola.

- Em termos conceptuais, o modelo **HDFE** (*High-Dimensional Fixed Effects*) – **efeitos fixos do aluno e da escola** – parece ser mais adequado:
  - ▶ Controla para a heterogeneidade não observada ao nível do aluno;
  - ▶ E para as especificidades da escola, tal como, o tipo e o número de alunos que uma determinada escola atrai, e as políticas específicas relacionadas com a formação das turmas dentro das escolas.



## Resultados empíricos: Matemática

- A **base de dados para Matemática possui 3338 alunos** com informação tanto para o ano de realização do exame de 9º como de 12º ano (6676 observações).
- Estes alunos encontram-se inseridos em **2668 turmas de 813 escolas**, localizadas em Portugal Continental.
- **56,1% dos alunos mudam de escola** entre o 9º e o 12º ano.

	9º ano	12º ano
2007/2008	408	
2008/2009	2930	
2010/2011		397
2011/2012		2941
Total	3338	3338

Fonte: Cálculos próprios com base na MISI/JNE.

Figura 3: Estrutura da base de dados de Matemática utilizada nos vários modelos.

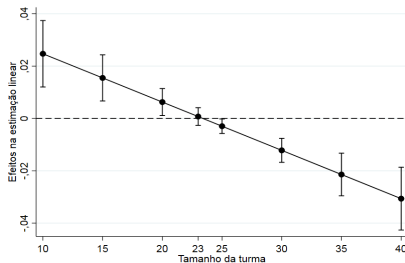
	OLS 4 (robusto)	RE (robusto)	FE (robusto)	HDFE (robusto)
Ln nota interna	0,9877*** (0,0185)	0,9877*** (0,0180)	0,9188*** (0,0269)	0,9157*** (0,0324)
Mulher	-0,0151** (0,0059)	-0,0151*** (0,0058)		
Internet	0,0048 (0,0079)	0,0048 (0,0078)	-0,0030 (0,0136)	-0,0128 (0,0155)
Beneficiário SASE	-0,0239*** (0,0083)	-0,0239*** (0,0083)	0,0023 (0,0185)	0,0283 (0,0192)
Mãe com Ensino Superior	0,0051 (0,0073)	0,0051 (0,0074)	-0,0463 (0,0436)	0,0036 (0,0456)
Tamanho da turma	0,0239*** (0,0058)	0,0239*** (0,0059)	0,0234*** (0,0082)	0,0432*** (0,0105)
Tamanho da turma <sup>2</sup>	-0,0005*** (0,0001)	-0,0005*** (0,0001)	-0,0006*** (0,0002)	-0,0009*** (0,0002)
% Mulheres na turma	-0,0581*** (0,0169)	-0,0581*** (0,0170)	-0,0418* (0,0218)	-0,0528** (0,0263)
% Repetentes na turma	-0,0930*** (0,0312)	-0,0930*** (0,0311)	-0,2093*** (0,0420)	-0,1614*** (0,0502)
% Alunos com internet na turma	-0,0641*** (0,0117)	-0,0641*** (0,0116)	-0,1252*** (0,0182)	-0,1437*** (0,0222)
% Beneficiários SASE na turma	-0,0242 (0,0156)	-0,0242 (0,0154)	0,0283 (0,0254)	0,0516 (0,0319)
% Bons alunos (p90) na turma	0,2720*** (0,0109)	0,2720*** (0,0109)	0,2465*** (0,0173)	0,1931*** (0,0228)
% Mães com E. Sup. na turma	0,0375*** (0,0130)	0,0375*** (0,0133)	0,0144 (0,0216)	-0,0020 (0,0272)
R <sup>2</sup> within   R <sup>2</sup> between   R <sup>2</sup> overall	0,58	0,51   0,63   0,58	0,52   0,58   0,55	
RMSE (Root Mean Square Error)	0,22	0,22	0,16	0,21
$\rho$		0,00	0,36	
Número de observações	6676	6676	6676	6366
F (significância global)	451,36***			

Notas: Níveis de significância: \*\*\*  $p < 0,01$ ; \*\*  $p < 0,05$  (\*\*) e \*  $p < 0,1$ .

**Figura 4:** Resultados da estimação utilizando o logaritmo das notas do exame nacional de Matemática como variável dependente - modelos para dados em painel.

## Resultados empíricos: Matemática

- Até ao momento em que os efeitos marginais são positivos, aumentar o tamanho da turma aumenta o desempenho médio dos alunos no exame nacional de Matemática. Contudo, a ritmos decrescentes.
- O **intervalo ótimo de alunos** encontra-se compreendido entre os **21** e os **24** alunos.



Fonte: Cálculos próprios com base na M|S|/JNE.

**Figura 5:** Efeitos marginais médios do tamanho da turma a Matemática com um intervalo de confiança de 95%.

## Resultados empíricos: Português

- A **base de dados para Português possui 4279 alunos** com informação tanto para o ano de realização do exame de 9º como de 12º ano (8558 observações).
- Estes alunos encontram-se inseridos em **3041 turmas de 874 escolas**, localizadas em Portugal Continental.
- **57,7% dos alunos mudam de escola** entre o 9º e o 12º ano.

	9º ano	12º ano
2007/2008	523	
2008/2009	3756	
2010/2011		507
2011/2012		3772
Total	4279	4279

Fonte: Cálculos próprios com base na MISI/JNE.

Figura 6: Estrutura da base de dados de Português utilizada nos vários modelos.

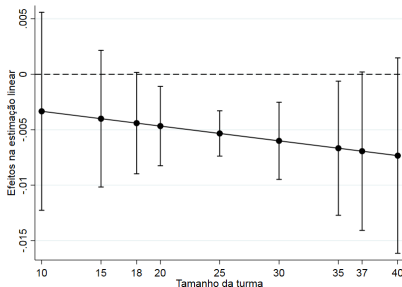
	OLS 4 (robusto)	RE (robusto)	FE (robusto)	HDFE (robusto)
Ln nota interna	0,6178*** (0,0122)	0,5884*** (0,0127)	0,1562*** (0,0230)	0,1485*** (0,0273)
Mulher	0,0363*** (0,0047)	0,0385*** (0,0049)		
Internet	0,0082 (0,0063)	0,0083 (0,0064)	0,0036 (0,0098)	-0,0075 (0,0103)
Beneficiário SASE	-0,0115* (0,0063)	-0,0140** (0,0064)	-0,0382*** (0,0110)	-0,0300** (0,0117)
Mãe com Ensino Superior	0,0281*** (0,0058)	0,0299*** (0,0062)	-0,0433 (0,0319)	0,0101 (0,0342)
Tamanho da turma	0,0029 (0,0041)	0,0033 (0,0041)	0,0045 (0,0052)	-0,0020 (0,0074)
Tamanho da turma <sup>2</sup>	-0,0001 (0,0001)	-0,0001* (0,0001)	-0,0002** (0,0001)	-0,0001 (0,0001)
% Mulheres na turma	-0,0054 (0,0129)	-0,0037 (0,0127)	0,0053 (0,0156)	-0,0042 (0,0177)
% Repetentes na turma	-0,0542*** (0,0204)	-0,0575*** (0,0205)	-0,0636** (0,0303)	-0,0679** (0,0340)
% Alunos com internet na turma	-0,0428*** (0,0093)	-0,0479*** (0,0092)	-0,0830*** (0,0124)	-0,0889*** (0,0145)
% Beneficiários SASE na turma	-0,0377*** (0,0122)	-0,0346*** (0,0124)	0,0178 (0,0175)	0,0191 (0,0218)
% Bons alunos (p90) na turma	0,3155*** (0,0109)	0,3178*** (0,0108)	0,3056*** (0,0142)	0,2255*** (0,0177)
% Mães com E. Sup. na turma	0,0066 (0,0110)	0,0071 (0,0111)	0,00001 (0,0161)	-0,0028 (0,0200)
R <sup>2</sup> within   R <sup>2</sup> between   R <sup>2</sup> overall	0,40	0,16   0,52   0,40	0,22   0,25   0,23	
RMSE (Root Mean Square Error)	0,20	0,18	0,13	0,17
ρ		0,13	0,52	
Número de observações	8558	8558	8558	8274
F (significância global)	458,59***			

Notas: Níveis de significância: \*\*\*  $p < 0,01$ ; \*\*  $p < 0,05$  (\*\*) e \*  $p < 0,1$ .

**Figura 7:** Resultados da estimação utilizando o logaritmo das notas do exame nacional de Português como variável dependente - modelos para dados em painel.

## Resultados empíricos: Português

- Ao contrário do obtido para Matemática, aumentar o tamanho da turma **não produz um efeito estatisticamente diferente entre si** em Português.



Fonte: Cálculos próprios com base na M|S|/JNE.

**Figura 8:** Efeitos marginais médios do tamanho da turma a Português com um intervalo de confiança de 95%.

# Discussão

- Tanto em Matemática como em Português, a **nota interna encontra-se indexada à nota obtida no exame nacional.**
- As **mulheres possuem um desempenho médio inferior aos homens em Matemática, mas superior em Português.**
  - ▶ Em linha com o obtido por Pereira (2010), apesar de não ser consensual na literatura.
- Controlando para a heterogeneidade não observada ao nível do aluno e da escola (estimação por *HDFE*), os alunos **beneficiários do Sistema de Ação-Social Escolar possuem um desempenho médio inferior em cerca de 3% relativamente aos não beneficiários em Português, mantendo tudo o resto constante.**
  - ▶ Este resultado encontra-se em linha com o obtido por Sousa (2016) e por Firmino *et al.* (2016).

- O intervalo de dimensão ótima da turma encontra-se compreendido entre os 21 e os 24 alunos em Matemática e não produz um efeito diferente de zero em Português.
  - ▶ Sousa (2016) concluiu que a dimensão ótima da turma ronda os 22 a 23 alunos em Matemática, não sendo estatisticamente significativa em Português.
- Parece existir uma **externalidade negativa associada ao aumento de mulheres na turma no desempenho dos alunos em Matemática**, porém **não estatisticamente significativa em Português**.
  - ▶ Existem estudos que defendem que aumentar a proporção de alunas é propício ao sucesso escolar (e.g. Kirjavainen, 2012 e Pereira, 2010), estudos que argumentam que depende do tipo de aluno (e.g. Schone *et al.*, 2017) e estudos onde não foram encontrados efeitos estatisticamente significativos (e.g. Hill, 2017).



- Parece existir uma **externalidade negativa associada ao aumento de repetentes na turma** no desempenho dos alunos em ambas as disciplinas.
  - ▶ Este resultado encontra-se em consonância com o obtido por Lavy *et al.* (2012) e Pereira (2010).
- E uma **externalidade positiva associada ao aumento de bons alunos na turma** em ambas as disciplinas.
  - ▶ Este resultado parece ser consensual na literatura (e.g. Ding e Lehrer, 2007; Hanushek *et al.*, 2003 e Hoxby, 2000b).

# Investigação futura

- Avaliar se as variáveis da composição da turma afetam o desempenho dos estudantes nos exames nacionais de forma heterogénea, consoante o tipo de aluno.
  - ▶ Para o efeito, será desenvolvido um modelo alternativo com interações entre as variáveis da composição da turma e as variáveis referentes a características individuais do estudante.

*Muito obrigada pela vossa atenção!*