

WORKING PAPERS

CESNOVA

WORKING PAPER N.º 1 – DEZEMBRO DE 2014 – ISSN: 2182-4991

**Indicadores sintéticos de contexto
socioeconómico dos resultados escolares
em exames nacionais nos concelhos de
Portugal Continental (2009-2014)**

Susana Batista, Luísa Franco e Rui Santos



FICHA TÉCNICA:

TÍTULO : WORKING PAPERS CESNOVA

COORDENADOR

José Alberto Simões

CONSELHO EDITORIAL

Ana Ferreira

Cláudia Urbano

Jordi Nofre

EDITOR

CESNOVA – Centro de Estudos
de Sociologia da Universidade Nova de Lisboa

Av. de Berna, Edifício FCSH-ID, 3º piso, sala 3.14

Endereço Postal: Av. de Berna, 26 C

1069-061 LISBOA - Portugal

E-mail cesnova.wp@fcs.unl.pt

Web <http://cesnova.fcs.unl.pt/>


GRAFISMO

Carolina Basto

ISSN

2182-4991

Este trabalho é financiado por Fundos Nacionais através da FCT – Fundação para a
Ciência e a Tecnologia, no âmbito do PEst-OE/SADG/UI4067/2014



**Indicadores sintéticos de contexto socioeconómico dos
resultados escolares em exames nacionais nos
concelhos de Portugal Continental (2009-2014)¹**

Susana Batista – Mestre em Sociologia, CESNOVA, FCSH/NOVA.
susanabatista@fcsch.unl.pt

Luísa Franco – Doutora em Sociologia, CESNOVA, FCSH/NOVA.
luisa.franco@fcsch.unl.pt

Rui Santos – Doutor em Sociologia, CESNOVA, FCSH/NOVA e
Departamento de Sociologia, FCSH/NOVA.
rsantos@fcsch.unl.pt

¹ Este estudo foi desenvolvido no âmbito dos projetos “ESCXEL – Rede de Escolas de Excelência” (CESNOVA-FCSH/NOVA) e “Atlas da Educação – Desempenho e potencial de sucesso e insucesso escolar do concelho” (EPIS – Empresários pela Inclusão Social e CESNOVA – FCSH/NOVA).

Resumo

Neste documento de trabalho, propomos à discussão uma metodologia de construção de indicadores sintéticos de contexto socioeconómico concelhio para os resultados escolares, construídos com base em modelos de regressão linear múltipla dos resultados médios dos exames dos 2º e 3º ciclos do Ensino Básico (EB) e do Ensino Secundário (ES), sobre variáveis socioeconómicas concelhias. Os modelos mostram um peso considerável dos contextos socioeconómicos à escala do concelho nos resultados dos exames. Assim, os indicadores derivados são medidas fiáveis das condicionantes socioeconómicas dos resultados escolares. Apresentam a vantagem de sintetizarem uma grande variedade de características empíricas em medidas quantitativas únicas, que podem ser usadas para contextualizar a análise dos resultados escolares e como variáveis relevantes para outros estudos. Em função da disponibilidade de dados, o leque de indicadores será futuramente alargado ao 1º ciclo do EB e a outras unidades de análise.

Palavras-chave: Indicadores socioeconómicos; Resultados escolares; Exames Nacionais; Desigualdades sociais; Portugal Continental/ Concelhos (2009-2014).

Indicators for socioeconomic context of school results in national exams in mainland Portugal (2009-2014)

Abstract

In this working paper, we put forward for discussion a methodology for constructing summary indicators of the socioeconomic context of school results at the municipal level, which were built upon multiple linear regression models of average exam results at the end of upper elementary (6th grade), the lower secondary (9th grade) and the upper secondary level (11th and 12th grades) on municipal-level socioeconomic variables. The models display a considerable weight of socioeconomic contexts on exam results. The derived indicators are therefore a reliable measure of the socioeconomic underpinnings of school results. They have the advantage of summarizing a wide variety of empirical characteristics into single quantitative measures, which can be used to contextualize analyses of school results and as relevant variables in other studies. Pending on sufficient data availability, the range of indicators will be extended to the earlier schooling cycle (4th grade), as well as to different units of analysis.

Keywords: Socio-economic indicators; School results; National exams; Social inequality; Mainland Portugal (2009-2014)

1. INTRODUÇÃO

Na sequência de investigação anterior (Justino et al., 2014, 62-77), este trabalho apresenta a construção de indicadores sintéticos do contexto socioeconómico à escala concelhia¹, com impacto sobre os resultados dos exames nacionais do 2º e 3º ciclos do Ensino Básico e do Ensino Secundário, para Portugal Continental. Trata-se de indicadores sociais descritivos (Land e Ferriss, 2007, 521) que medem a variação intermunicipal da combinação de múltiplas características socioeconómicas, em função dos efeitos estatisticamente determinados dessas características sobre a variação dos resultados dos exames em cada ciclo de ensino.

Na última década, os resultados escolares dos exames nacionais dos diferentes ciclos de estudos do Ensino Básico e Secundário em Portugal, têm servido de base à elaboração de *rankings* de escolas por jornais que os publicitam, tendo em conta o interesse que despertam no público. Parte da discussão pública em torno desse procedimento vem apelando à necessidade de os enquadrar na diversidade dos contextos socioeconómicos em que as escolas trabalham para produzir esses resultados. O próprio Ministério da Educação e Ciência (MEC) começou a utilizar indicadores socioeconómicos (a partir do Sistema de Informação, MISI) para contextualizar os resultados escolares, com o propósito de informar o processo de avaliação externa das escolas (Canto e Castro et al., 2011; Grupo de Trabalho para a Avaliação Externa das Escolas [GTAEE], 2011). Recentemente, começaram a ser publicadas análises sistemáticas referentes aos contextos socioeconómicos nas ordenações dos resultados das escolas nos exames nacionais (Viana,

¹ As medidas de síntese deste tipo, que resumem uma pluralidade de indicadores, são também conhecidas como “índices” no vocabulário das ciências sociais e particularmente da Sociologia. Optámos pela expressão equivalente “indicadores sintéticos” (“composite or *summary* social indicators”, Land e Ferriss, 2007, 520, itálico nosso) para, por um lado, distingui-los dos indicadores empíricos que eles resumem, e por outro evitar a confusão com o termo “índice” que empregaremos adiante, a propósito do processamento dos resultados dos exames como índices referidos às médias nacionais.

2012; Portela e Azevedo, 2013; Azevedo e Portela, 2014)², atendendo a que o MEC começou a disponibilizar aos investigadores dados de caracterização socioeconómica da população escolar. Em todos esses estudos, as unidades de análise foram as “unidades orgânicas”, ou seja, os agrupamentos de escolas ou as escolas não agrupadas, sempre do setor público, que por facilidade de expressão passaremos a designar simplesmente por “escolas”.

Canto e Castro et al. (2011), em estudo publicado pela Direção Geral das Estatísticas da Educação e Ciência (DGEEC), seguiram uma metodologia bietápica. Partiram de uma análise de *clusters*, com recurso a variáveis de contexto socioeconómico por escola (escolaridade média das mães e dos pais, proporção de alunos beneficiários de Ação Social Escolar (ASE) em cada escalão) e a variáveis de morfologia das escolas (nomeadamente distribuição dos alunos pelos níveis e modalidades de ensino), na qual a população de escolas foi segmentada em três *clusters*, segundo diferentes perfis dessas variáveis. Seguidamente, realizaram dentro de cada *cluster*, análises de regressão linear múltipla dos resultados de exame por disciplina e das taxas de conclusão, sobre conjuntos de variáveis preditoras apuradas por ano terminal de ciclo de escolaridade (4^o, 6^o, 9^o e 12^o): média das idades dos alunos, percentagem de alunos que não beneficiam de ASE, percentagem de alunos do sexo feminino, média do número de alunos por turma, médias do número de anos de escolaridade das mães e dos pais, percentagem de docentes do quadro. Apuraram assim modelos de regressão diferentes, em que entraram em cada caso apenas algumas das variáveis acima mencionadas consoante o *cluster* a que se aplicavam, o ano escolar considerado, e a variável dependente (resultado no exame por disciplina ou taxa de conclusão). Esta metodologia resulta, em cumprimento dos objetivos do estudo, em múltiplas combinações de indicadores socioeconómicos ajustados para a contextualização dos resultados em distintos segmentos de

² Elaborados pela Universidade Católica do Porto, sob a coordenação de Joaquim Azevedo.

escolas. A classificação prévia das escolas em *clusters* não ordenados, que constitui de facto um indicador sintético que resume características socioeconómicas e organizacionais, assume uma métrica qualitativa nominal.

O relatório final do Grupo de Trabalho para a Avaliação Externa das Escolas (GTAAE, 2011) propõe igualmente o uso de valores esperados obtidos por modelos de regressão linear múltipla como referência para a avaliação dos resultados escolares, usando como variáveis independentes diversos indicadores de caracterização socioeconómica das populações escolares, também retirados da base de dados MISI. Neste caso, contrariamente ao estudo antecedente, os modelos de regressão foram aplicados à população total, sem segmentação prévia, gerando valores esperados para um conjunto de variáveis dependentes (taxas de conclusão e abandono, resultados de avaliação interna e em provas de aferição e exames, entre outras) para os diferentes ciclos de escolaridade (GTAAE, 2011, 93-104). Tal como o antecedente, e cumprindo idênticos objetivos, este estudo resulta em múltiplas combinações de indicadores socioeconómicos segundo os tipos de resultados considerados e o ciclo de ensino; com a diferença de serem aqui comuns a toda a população, sem construir previamente um indicador qualitativo sintético para segmentar a população como, implicitamente, fizeram Canto e Castro et al. (2011).

Os três estudos da Universidade Católica do Porto realizados para o jornal *O Público*, embora não exatamente idênticos entre si, seguiram uma lógica de análise comum, pelo que abordaremos aqui apenas o mais recente (Azevedo e Portela, 2014). As escolas foram repartidas segundo a percentagem de alunos beneficiários do escalão A de ASE (por ordem decrescente) e a média de anos de escolaridade dos pais. O indicador socioeconómico sintético de base foi a média dos percentis da escola nestas duas variáveis de caracterização. A partição da população por tercis segundo a distribuição desse indicador de base, criou três escalões ordinais

designados como “contextos”, do menos ao mais favorável, que assim constituíram a escala do indicador final para efeito de contextualização socioeconómica dos resultados. O denominado “valor esperado do contexto” (VEC), face ao qual foram comparados os resultados obtidos por cada escola, é a média dos resultados de exame obtidos nas escolas de cada um dos três contextos. Embora, ao contrário de Canto e Castro et al. (2011), este indicador socioeconómico sintético constitua uma métrica ordinal comum à totalidade da população, é uma medida pouco discriminante entre escolas, dado que atribui a mesma categoria (e consequentemente o mesmo VEC) a casos extremos muito mais distantes entre si no valor do indicador socioeconómico de base, do que com os casos extremos das categorias contíguas; e, inversamente, categorias ordinais diferentes a casos com maior proximidade entre si do que com a maior parte dos outros casos das respetivas categorias.

Como foi dito, os objetivos dos estudos precedentes ditaram um enfoque na variável dependente, na forma dos denominados “valores esperados” em função das variáveis de contexto socioeconómico utilizadas. O que propomos aqui é um ângulo de abordagem complementar, incidindo sobre as variáveis independentes dos modelos de estimativa, de modo a resumi-los em indicadores sintéticos da sua variação combinada, segundo o peso de cada um na explicação estatística da variação dos resultados. Por outro lado, ao invés dos estudos publicados pelo MEC, e porque o nosso objetivo não é analisar os resultados escolares mas sim sintetizar medidas das características socioeconómicas relevantes, utilizaremos aqui um único indicador dos resultados escolares, a média dos exames nacionais em cada nível de escolaridade.

Adicionalmente, e também devido aos seus objetivos, as unidades de análise consideradas nos estudos acima recenseados foram as escolas. Um ponto fraco nas abordagens a esta escala é o tratamento dos dados omissos, que constituem um problema inevitável no estado presente da informação. O

problema consiste na existência de um grande número de escolas em que essas variáveis enfermam de taxas de não-resposta superiores a um terço, e frequentemente de mais de metade, o que implica a possibilidade de enviesamentos significativos nos resultados obtidos. O estudo da DGEEC esclarece que para diversas variáveis, resultantes de informação recolhida pelas escolas junto dos encarregados de educação, foram consideradas as percentagens “por referência ao total de respostas consideradas válidas” (Canto e Castro et al., 2011, 14-5). Os estudos da Universidade Católica do Porto não esclarecem como foram tratadas as não-respostas, mas os números de agrupamentos para os quais consideram os indicadores socioeconómicos permitem-nos inferir que seguiram o mesmo procedimento. O GTAEE (2011, 37), que explicita esta dificuldade, afirma só ter sido possível completar a análise estatística após a correção dos dados lacunares, através de meios dificilmente acessíveis aos investigadores. Infelizmente, desconhecemos se esses dados corrigidos foram tornados acessíveis a utilizadores externos. De toda a forma, enquanto não forem resolvidas as lacunas na recolha da informação de base do sistema MISI, os resultados obtidos nesta escala de análise serão precários, sobretudo quando individualizados por escolas.

Esse não é, contudo, o único nível de análise pertinente. Por um lado, os estudos empíricos confirmam a hipótese, assente em estudos internacionais (cf. a revisão de literatura em Seabra, 2009, 82, 88), de que os resultados escolares (sejam eles as classificações em exames, as taxas de conclusão ou de abandono) se relacionam com contextos sociais com expressão territorial (Justino et al., 2014; Verdasca, 2013). Por outro, os processos que têm vindo a ser designados de “territorialização” vêm acentuando a importância dos níveis de ação e decisão locais na regulação do sistema educativo (Barroso, 2013). Designadamente, ganha acrescida relevância a contextualização dos resultados escolares à escala concelhia, dada a progressiva transferência de competências educativas para os

municípios (Formosinho e Machado, 2013). Acresce que é neste nível de análise que dispomos de mais indicadores socioeconómicos de base, de melhor qualidade e mais recentes, tendo em conta a realização dos Censos da população em 2011. Para níveis de análise mais desagregados, a informação estatística ou está por enquanto inacessível, à escala dos alunos, ou à escala das escolas é, como dissemos, lacunar. Num outro aspeto, a desigual qualidade dos dados recolhidos junto dessas unidades e a instabilidade da sua composição ao longo dos últimos anos tornam muito difícil a construção de séries de dados plurianuais, tanto sobre os resultados como sobre as características socioeconómicas das suas populações discentes.

Como referimos, este estudo considera os resultados dos exames nacionais dos 2º e 3º ciclos do Ensino Básico e do Ensino Secundário. Para o nível de escolaridade mais baixo (1º ciclo), o número de anos com dados de exames é ainda escasso para a utilização da metodologia, que, como veremos, passa pelo cálculo de médias plurianuais. Por outro lado, não nos foi possível estender a análise aos concelhos das Regiões Autónomas por algumas das fontes disponíveis não fornecerem dados suficientes, sendo omissas variáveis fundamentais para os modelos estatísticos construídos.

Diferentemente de publicações anteriores em que aplicámos idêntica construção das variáveis e análises de regressão múltipla (Batista, Franco e Santos, 2014; Justino et al., 2014), o objetivo deste documento de trabalho é apresentar à discussão a construção dos próprios indicadores sintéticos do contexto socioeconómico e disponibilizar os seus valores. Além da metodologia, que nos seus traços gerais poderá ser estendida a outras escalas e ciclos de estudos, apresentamos em anexo os valores dos indicadores de contexto socioeconómico por ciclo de estudos e concelho. No que toca à análise empírica, a metodologia aqui empregue apenas difere dos nossos estudos anteriores por não trabalharmos agora diretamente com os

valores dos exames, mas sim com índices desses valores às médias nacionais, como abordaremos adiante.

Creemos que a metodologia que seguimos apresenta duas virtualidades. Em primeiro lugar, resulta numa métrica quantitativa que permite comparar todos os elementos da população diretamente entre si, sem os segmentar em categorias discretas. Em segundo lugar, os indicadores resumem num valor único os efeitos conjuntos, estatisticamente comprovados, das múltiplas variáveis de contexto socioeconómico convocadas. Dito de outro modo, estes indicadores medem o conjunto das desigualdades sociais significativas na realização das oportunidades educativas. O que, por um lado, facilita a sua leitura e dá uma expressão facilmente apreensível à noção, frequente no discurso público e sociológico, de (des)favorecimento socioeconómico relativo de determinados segmentos populacionais (neste caso, com expressão territorial) face aos resultados dos exames nacionais. Por outro lado, estes indicadores podem ser utilizados autonomamente, seja em sínteses de caracterização socioeconómica territorial, seja em análises estatísticas subsequentes, em que a substituição de múltiplas variáveis por um indicador único que sintetiza os seus efeitos tem vantagens evidentes.

Os resultados dos exames agregam as escolas públicas e privadas existentes em cada concelho, apesar de utilizarmos uma variável socioeconómica (a percentagem de mães com grau de ensino superior) retirada de uma base de dados oficial onde as escolas privadas não estão representadas. Esta inconsistência de fontes parece-nos preferível à consideração dos resultados exclusivamente das escolas públicas, pois os modelos de análise integram maioritariamente variáveis de âmbito concelhio incidindo sobre a totalidade dos territórios municipais e das suas populações; consideramos ainda que, dado o peso amplamente predominante das escolas

públicas na população escolar de Portugal Continental, esta característica não pode ser ignorada.³

Na próxima secção apresentamos as variáveis de análise, relativas aos resultados escolares e à caracterização socioeconómica, assim como a metodologia de construção dos indicadores sintéticos de contexto socioeconómico. Depois apresentamos os resultados obtidos e, por último, a sua discussão e perspectivas de trabalho futuro.

³ As escolas públicas acolhiam 86% dos matriculados no Ensino Básico e 78% no Ensino Secundário em Portugal Continental em 2012 (Pordata, 2014).

2. METODOLOGIA

2.1. VARIÁVEIS

2.1.1. RESULTADOS ESCOLARES

Os resultados escolares estão expressos como médias dos índices anuais das classificações obtidas pelos alunos das escolas públicas e privadas dos concelhos, que realizaram o exame na primeira chamada ou fase, como alunos internos. As médias respeitam aos últimos seis anos (2009 a 2014) para o 9º ano e o Ensino Secundário; limitando-se por agora, para o 6º ano, à média dos únicos três anos (2012 a 2014) em que se realizaram exames neste nível. Os índices anuais I_{kc} de cada ciclo k em cada concelho c são obtidos pela fórmula:

$$I_{kc} = \frac{k_c}{k_p} \times 100$$

em que k_c é a média do ciclo k no concelho c , e k_p é a média do ciclo k em Portugal continental. O valor nacional corresponde, assim, à base indicial constante de 100. Utilizamos índices em vez de simples médias para eliminar a variação espúria devida a diferenças de dificuldade das provas de exame entre anos (resolvendo a dificuldade apontada pelo GTAEE, 2011, 45, segundo o qual “as provas nacionais não estão calibradas, [...] pelo que] não faz sentido fazer julgamentos acerca da evolução positiva ou negativa [...]”). No mesmo sentido, consideramos médias de seis anos, que retêm as diferenciações gerais, abstraindo de outras flutuações interanuais aleatórias, para os exames do 3º Ciclo do Ensino Básico (9º ano) e do Ensino Secundário (11º e 12º anos). No caso do 2º ciclo, recorreremos a uma média dos três anos de existência destes exames (2012-2014), que estabelecemos como período mínimo para análise. Nesta lógica, o ciclo de escolaridade mais baixo (1º ciclo) não conta ainda com um número de anos de exame suficiente para este efeito.

Quase todos os concelhos com Ensino Secundário no período considerado (245 concelhos) tiveram resultados em exame ao longo dos seis anos, à exceção dos de Armamar (1 ano), Fronteira (1 ano), Pampilhosa da Serra (2), Vila de Rei (4) e Murtosa (5). Como referido, estabelecemos em três o número mínimo de anos, pelo que os concelhos de Armamar, Fronteira e Pampilhosa da Serra não foram incluídos.

As variáveis originais dos resultados estavam expressas em diferentes escalas. No caso do Ensino Secundário, tratava-se da média dos exames realizados em todas as disciplinas, expressa em escala 0-200. No caso do 6º e 9º anos, recorremos aos resultados dos exames de Língua Portuguesa e Matemática na escala de 5 níveis. Ao operarmos a transformação em índices, todas estas escalas foram homogeneizadas numa escala de 100 pontos indiciais⁴.

2.1.2. CARACTERIZAÇÃO SOCIOECONÓMICA

Partimos das variáveis socioeconómicas que têm conhecido efeito nos resultados escolares, tais como variáveis relativas à origem social dos alunos ou aos meios envolventes para além do contexto escolar (Azevedo e Portela, 2014; Canto e Castro et al., 2011; Diogo, 2008; Portela e Azevedo, 2013; Verdasca, 2013; Viana, 2012), e que têm sido mobilizadas para explicar diferenças no sucesso escolar (Justino e Rosa, 2009; Justino et al., 2014). Sobre o amplo leque de variáveis teoricamente relevantes e para as quais existem indicadores empíricos nas fontes estatísticas disponíveis, usámos um procedimento indutivo. Ensaámos sucessivamente modelos de análise fatorial de componentes principais e de regressão linear múltipla, retendo os modelos e, por consequência, as variáveis que, no seu conjunto, apresentaram os melhores resultados, primeiro na diferenciação

⁴ As estatísticas descritivas das variáveis incluídas nos modelos podem ser consultadas no Anexo 1.

socioeconómica dos concelhos e, subseqüentemente, como preditoras dos resultados em exame.

As variáveis utilizadas para caracterizar socioeconomicamente os concelhos são de dois tipos.

Em primeiro lugar, variáveis compósitas, ou “componentes”, que resumem indicadores estatísticos de natureza diversa:

- Indicadores sociodemográficos⁵: densidade populacional, percentagem de jovens na população residente e percentagem de ativos na população residente;
- Indicadores de atividade económica: percentagem das pessoas residentes ao serviço por setor de atividade⁶, taxas de desemprego total e na população com menos de 25 anos⁷ e percentagem de superfície agrícola útil no território do concelho⁸;
- Indicadores de estratificação socioeconómica: percentagem da população residente por classes socioeconómicas e grupos ocupacionais⁹;
- Indicadores de escolarização: percentagem da população segundo o nível de escolaridade¹⁰;
- Indicadores de desvantagem socioeconómica: percentagem de residentes beneficiários de Rendimento Social de Inserção¹¹, percentagem de alojamentos familiares com rendas com apoio social¹².

⁵ Instituto Nacional de Estatística (INE), Censos 2011.

⁶ Ministério da Segurança Social e do Trabalho (MSST) – Classificação Portuguesa de Atividades Económicas (CAE) 3, 2010. As percentagens foram calculadas, a partir das variáveis originais em número, com base no Total de Pessoas ao Serviço nos estabelecimentos.

⁷ Instituto do Emprego e Formação Profissional (IEFP), 2011. As taxas foram calculadas a partir das variáveis originais em número, sobre a População Total do concelho e a População dos 15 aos 24 anos, respetivamente.

⁸ INE, 2009. A percentagem de superfície agrícola utilizada (variável original em Hectares) foi calculada sobre a Área total (Ha) (Fonte: DGA/IPCC, 1998).

⁹ Markttest, referente a 2010.

¹⁰ INE, Censos 2011. Percentagens calculadas sobre a População Total do concelho.

Todos estes indicadores empíricos, procedentes das diversas fontes primárias identificadas, foram colhidos através de Marktest Computing (2012).

Resumimos estas características socioeconómicas dos concelhos num conjunto de novas variáveis por meio de uma Análise de Componentes Principais (Justino et al., 2014, 64). O modelo, com uma adequabilidade média (KMO=0,774) e que resume 77,3% da variância total, reteve cinco componentes com valores próprios superiores a 1, que passamos a descrever sucintamente:

«*Urbano - Atividade Serviços de Qualificação Elevada*»: Nesta componente, que resume 28,2% da variância total, surgem com maior importância aspetos como a elevada densidade populacional, as classes sociais mais elevadas (correlação direta da componente com as classes média e média-alta e correlação inversa com a classe baixa), os níveis de instrução mais elevados (correlação direta com a população com ensino superior e, menos forte, com ensino secundário, e inversa com população sem instrução), bem como o emprego nas profissões mais qualificadas do setor terciário e nos serviços de modo mais amplo. Apesar de relacionada com características socioeconómicas mais favorecidas, verifica-se uma correlação elevada desta componente com a variável das rendas com apoio social, traduzindo em parte a diversidade socioeconómica interna dos grandes centros urbanos. Com *loadings* superiores a 0,4, encontramos ainda a correlação com a componente do peso da população ativa e da população jovem, sendo também de notar a correlação inversa (de menor intensidade: -0,34) com o desemprego jovem.

«*Atividade Industrial de Qualificação Média-baixa*»: Esta componente, que resume 19,2% da variância total, está mais correlacionada positivamente com

¹¹ Instituto de Gestão Financeira da Segurança Social (IGFSS), 2010. Percentagem calculada sobre a População Total do concelho (INE, Censos 2011).

¹² INE, Inquérito à Caracterização da Habitação Social, 2011. Percentagens calculadas, a partir da variável em número, sobre o Total de Alojamentos Familiares.

a percentagem de trabalhadores qualificados, o emprego na indústria transformadora, a classe média baixa, a presença de ativos e jovens, e a população com o nível de ensino básico (correlação próxima de 0,5), e, de maneira inversa, com a população sem instrução e com a classe baixa (correlações próximas de -0,5). A componente tem ainda uma fraca correlação inversa com as pessoas ao serviço no alojamento, restauração e similares (-0,31).

«*Atividade de Turismo e Serviços de Qualificação Média*»: Esta componente, que resume 10,7% da variância total, está mais correlacionada com as pessoas ao serviço no alojamento, restauração e similares e a população com o ensino secundário e o ensino básico. A componente está ainda correlacionada com os empregados dos serviços, comércio e administrativos (0,47) e a classe média (0,46). É nesse sentido que referimos esta componente como de «Turismo e Serviços», pois pode ter uma configuração claramente de hotelaria e restauração, mas corresponde ainda a serviços em centros de pequena ou média dimensão. A componente tem uma correlação inversa (-0,37) com o pessoal ao serviço da indústria transformadora.

«*Desvantagem Socioeconómica (Exclusão)*»: Esta componente resume 10,4% da variância total e é principalmente definida pelos indicadores de desfavorecimento socioeconómico, como a percentagem de desempregados em geral, de desempregados jovens e de beneficiários do RSI, e tem ainda uma correlação mais fraca (0,4) com os alojamentos familiares com arrendamentos de apoio social.

«*Agrícola - Atividade Agrícola de Qualificação Baixa*»: A quinta componente, que resume 8,8% da variância total, é a única que apresenta correlações consideráveis com a Superfície Agrícola Utilizável e com o emprego no setor agrícola, bem como, embora mais fraca (0,41), com a população sem instrução.

Cada uma destas componentes constitui uma nova variável, expressa em cada concelho pelo *score* que lhe é atribuído pela análise fatorial¹³.

Em segundo lugar, retivemos uma variável específica de caracterização da população escolar do concelho, proveniente da base de dados MISI¹⁴, recolhida ao nível das escolas e aqui agregada para o nível concelhio: a percentagem de mães dos alunos matriculados nas escolas públicas do concelho no ano letivo de 2010/2011, que completaram o Ensino Superior. Embora a data de referência desta variável esteja desatualizada, especialmente em relação aos anos dos resultados de exame considerados nos índices do 2º ciclo, não é de supor que esta característica da população escolar, tal como as outras variáveis colhidas nos Censos 2011, se altere substancialmente no curto prazo. Em alguns casos de concelhos para cujos estabelecimentos de ensino esta informação era omissa ou insuficiente, imputámos a percentagem da população feminina residente em 2011 com habilitação escolar superior, dentro do grupo de idades 25-44 anos¹⁵.

¹³ Ver estatísticas descritivas e distribuições dos *scores* de cada uma das componentes no Anexo 1. A matriz da correlação das componentes com os indicadores de base pode ser consultada em Justino et al. (2014, 64).

¹⁴ Base de dados referida ao ano letivo 2010/2011, gentilmente cedida pelo Conselho Nacional de Educação.

¹⁵ INE, Censos 2011. Considerámos como insuficiente a informação dos concelhos em que a variável era conhecida para menos de dois terços dos alunos (7 concelhos). O valor imputado resulta de uma transformação da variável “percentagem de mulheres 25-44 anos com ensino superior no concelho”, multiplicando-a pelo rácio entre a percentagem de mães dos alunos das escolas públicas com ensino superior e a percentagem de mulheres 25-44 anos com ensino superior na população residente, apurado sobre a população total dos concelhos do Continente para que dispomos das duas variáveis. A estatística descritiva da variável final está disponível no Anexo 1.

2.2. CONSTRUÇÃO DOS INDICADORES SINTÉTICOS

Os indicadores sintéticos de contexto socioeconómico foram construídos a partir de análises de regressão linear múltipla, de que retivemos os modelos que apresentam as estimativas mais próximas dos resultados observados. De modo a conseguir o melhor ajustamento possível, usámos um modelo e, subsequentemente, chegámos a um indicador específico para os resultados dos exames de cada ciclo de ensino. Cada modelo e indicador resumem as combinações e pesos distintos das variáveis socioeconómicas que, de entre as disponíveis, deram melhor conta da variação dos resultados dos exames nesse mesmo ciclo.

As variáveis dependentes dos modelos são as médias dos índices dos resultados dos exames nacionais em cada nível de ensino, à escala concelhia. Os modelos usam como preditoras diversas combinações das variáveis de caracterização socioeconómica atrás apresentadas (*scores* dos concelhos nas componentes socioeconómicas e habilitação superior das mães da população escolar).

Cada modelo para o ciclo de escolaridade k produz uma estimativa E_{kc} da média dos índices de resultados de exame para cada concelho c , em função das variáveis socioeconómicas x_1, x_2, \dots, x_n usadas como preditoras no modelo. Usámos a estatística R^2 ajustado como uma medida simples da adequação dessas estimativas aos valores observados. Os valores estimados ¹⁶ somam o peso dos aspetos relevantes do contexto socioeconómico na variação dos resultados escolares observados, ao valor da constante A_k que é comum a todos os concelhos e, portanto, independente

¹⁶ Correntemente designados na linguagem estatística como “valores esperados”. Em outras publicações, destinadas a maior divulgação (Justino et al., 2014; Batista, Franco e Santos, 2014) preferimos evitar essa expressão, devido aos mal entendidos a que dá origem na linguagem comum, na qual frequentemente o termo “esperado” assume um significado normativo (cf., por exemplo, o título dos artigos de jornal de Azevedo e Portela, 2014, e Viana, 2012), e até entre os avaliadores e avaliados nos processos de avaliação externa das escolas (GTAEE, 2011, 37). Por uma questão de consistência, mantivemos aqui essa opção terminológica.

das variáveis de contexto socioeconómico (o valor estimado se o efeito combinado das variáveis socioeconómicas assumisse o valor zero).

$$E_{kc} = A_k + B_{1k}X_c + B_{2k}X_c + \dots + B_{nk}X_c$$

O indicador I_{kc} do contexto socioeconómico concelhio para cada ciclo, que aqui propomos, isola o efeito conjunto dessas variáveis na variação dos resultados escolares, abstraindo do valor constante A_k . Corresponde à resolução para cada concelho apenas da parte variável da equação de regressão múltipla do modelo do ciclo k , que soma os valores das variáveis predictoras ponderados pelos respetivos coeficientes de regressão; o que, simplificando, equivale a subtrair a constante A_k aos valores estimados E_{kc} :

$$\begin{aligned} I_{kc} &= B_{1k}X_c + B_{2k}X_c + \dots + B_{nk}X_c \\ &= E_{kc} - A_k \end{aligned}$$

Por outras palavras, o indicador sintético adota uma das formas mais correntes de construção indicial: corresponde à soma ponderada dos valores dos indicadores empíricos que resume. Neste caso, os fatores de ponderação não são arbitrários ou resultantes de opções teóricas *a priori*, mas sim o peso estatístico da variação unitária de cada variável sobre o comportamento tendencial dos resultados dos exames, tal como empiricamente aferido pelo modelo de regressão.

Esta operação é relevante porque, embora os indicadores I_{kc} estejam, por definição, em correlação perfeita com os valores estimados E_{kc} , diferem substancialmente destes últimos no que respeita às diferenças proporcionais e ao coeficiente de variação, diferença tanto maior quanto maior for a proporção da constante A_k nos valores E_{kc} – que é grande no caso presente, dado que, como veremos, a constante nos três modelos é de quase 90 (isto é, 90% da média nacional). Ou seja, a omnipresença do valor constante nos valores estimados mascara a amplitude das diferenças especificamente dependentes das variáveis de contexto socioeconómico, e portanto das

desigualdades efetivas de contextos socioeconómicos que pretendemos medir. Como se verificará nos resultados abaixo, a subtração da constante aos valores estimados realça particularmente os contextos mais desfavorecidos, cujas características socioeconómicas produzem tendencialmente resultados de exame inferiores à constante; casos em que l_{kc} assume valores negativos¹⁷.

Como é evidente, um modelo de regressão linear simples usando l_{kc} como preditor dá conta de idêntica percentagem da variação das classificações reais,¹⁸ e produz as mesmas estimativas de classificação, do que o modelo de regressão linear múltipla que lhe serviu de base, com a vantagem, como dissemos, de resumir numa só medida todos os efeitos das variáveis de contexto incluídas em cada modelo.

3. RESULTADOS

A Tabela 1 mostra os traços fundamentais dos modelos de regressão mais bem ajustados à variação observada dos resultados médios de exame nos dois ciclos de escolaridade. Os modelos foram computados com o *IBM SPSS Statistics*, versão 20.0.0. Os do 3º ciclo do Ensino Básico e do Ensino Secundário cumprem todos os pressupostos estatísticos; o do 2º ciclo do Ensino Básico falha o da normalidade da distribuição, pelo que foi calculado com *bootstrap*, revelando-se os coeficientes fiáveis para o intervalo de confiança de 95%. Os valores de R^2 ajustado são de 0,350 (35,0% da variação é capturada pelo modelo) para o 2º ciclo, 0,474 (47,4%) para o Ensino Secundário e 0,545 (54,5%) para o 3º ciclo. Os três modelos têm significância

¹⁷ Diferentemente de uma primeira tentativa, com resultados parcialmente publicados em Justino et al. (2014), em que transformámos os valores dos indicadores em positivos, consideramos mais interessante para usos futuros reter os valores negativos, pela razão exposta e também para evitar transformações algébricas que alteram a proporcionalidade das variações.

¹⁸ Tratando-se neste caso de modelos de regressão linear simples, não há lugar ao ajustamento do R^2 em função do número de preditores, pelo que a variância explicada pelos l_{ck} é ligeiramente superior ao R^2 ajustado dos modelos de regressão múltipla em que se baseiam.

estatística para $p < 0,001$ (ou seja, com mais de 99,9% de probabilidade de a relação apurada entre o conjunto das variáveis preditoras e a variável dependente não se dever ao acaso).

Tabela 1. Regressão linear múltipla dos resultados médios de exames (Índices 2012-14 e 2009-14) dos concelhos de Portugal Continental sobre variáveis socioeconómicas

	2º ciclo ⁽¹⁾		3º ciclo		Ensino Secundário	
	Índices 2012-14		Índices 2009-14		Índices 2009-14	
	N=278		N=278		N=242	
Constante (A)	89,821***		89,611***		89,589***	
Variáveis explicativas	B	Beta	B	Beta	B	Beta
MSUP Mães com Ensino Superior (%)	0,528	0,464***	0,557	0,552***	0,438	0,369***
URQ Componente «Urbano Qualificado»	---	---	---	---	1,669	0,238***
IND Componente «Industrial»	1,230	0,183***	1,407	0,236***	2,605	0,366***
TUR Componente «Turismo e Serviços»	-1,218	-0,181***	-1,263	-0,212***	---	---
EXC Componente «Exclusão»	-0,819	-0,122*	-1,022	-0,172***	-0,934	-0,134**
AGR Componente «Agrícola»	-2,164	-0,322***	-2,445	-0,411***	-2,409	-0,301***
R ² ajustado	0,350***		0,545***		0,474***	
F	30,775		67,466		44,474	
	(5,272)		(5, 272)		(5,236)	

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

(1) Regressão com *bootstrap* com correção de enviesamento, para um intervalo de confiança de 95%.

A significância estatística dos modelos e a força das relações sustentam a aplicação das estimativas de regressão na construção dos indicadores de contexto socioeconómico por ciclo, que apresentamos de seguida:

$$I_{2c} = 0,528.MSUP_c + 1,230.IND_c - 1,218.TUR_c - 0,819.EXC_c - 2,164.AGR_c$$

O indicador de contexto socioeconómico dos resultados dos exames do 2º ciclo varia entre o mínimo de -3 e o máximo de 21, e a sua distribuição apresenta um coeficiente de variação de 57,9%¹⁹.

$$I_{3c} = 0,557.MSUP_c + 1,407.IND_c - 1,263.TUR_c - 1,022.EXC_c - 2,445.AGR_c$$

O indicador de contexto socioeconómico dos resultados dos exames do 3º ciclo varia entre o mínimo de -4 e o máximo de 22, e a sua distribuição apresenta um coeficiente de variação de 60,2%²⁰.

$$I_{sc} = 0,438.MSUP_c + 1,669.URQ_c + 2,605.IND_c - 0,934.EXC_c - 2,409.AGR_c$$

O indicador de contexto socioeconómico dos resultados dos exames do Ensino Secundário varia entre o mínimo de -7 e o máximo de 20, e a sua distribuição apresenta um coeficiente de variação de 75,6%²¹.

Os gráficos de dispersão nas Figuras 1, 2 e 3 ilustram as regressões lineares simples dos índices de resultados de exames sobre os indicadores de contexto socioeconómico, por ciclo de ensino²².

¹⁹ Como termo de comparação, os valores estimados E_{2c} variam entre o mínimo de 87 e o máximo de 111 com um coeficiente de variação de 4,2%. Consultar estatísticas descritivas das duas variáveis na Tabela 2 do Anexo 2.

²⁰ Os valores estimados E_{3c} variam entre o mínimo de 86 e o máximo de 111, com um coeficiente de variação de 4,6%. Consultar estatísticas descritivas das duas variáveis na Tabela 2 do Anexo 2.

²¹ Os valores estimados E_{sc} variam entre o mínimo de 83 e o máximo de 109, com um coeficiente de variação de 5,2%. Consultar estatísticas descritivas das duas variáveis na Tabela 2 do Anexo 2.

²² Gráficos computados com o Microsoft EXCEL for Mac 2011, versão 14.2.5. Os mesmos valores de R^2 são obtidos com o IBM SPSS Statistics, versão 20.0.0.

Figura 1. Relação dos resultados médios em índices dos Concelhos de Portugal Continental nos exames nacionais do 2º ciclo (2012-14) com o Indicador de Contexto Socioeconómico do 2º ciclo (I_{2c})

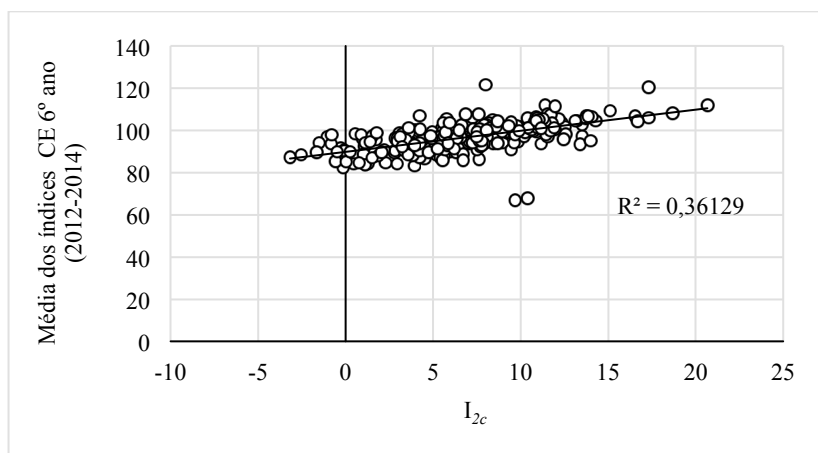


Figura 2. Relação dos resultados médios em índices dos Concelhos de Portugal Continental nos exames nacionais do 3º ciclo (2009-14) com o Indicador de Contexto Socioeconómico do 3º ciclo (I_{3c})

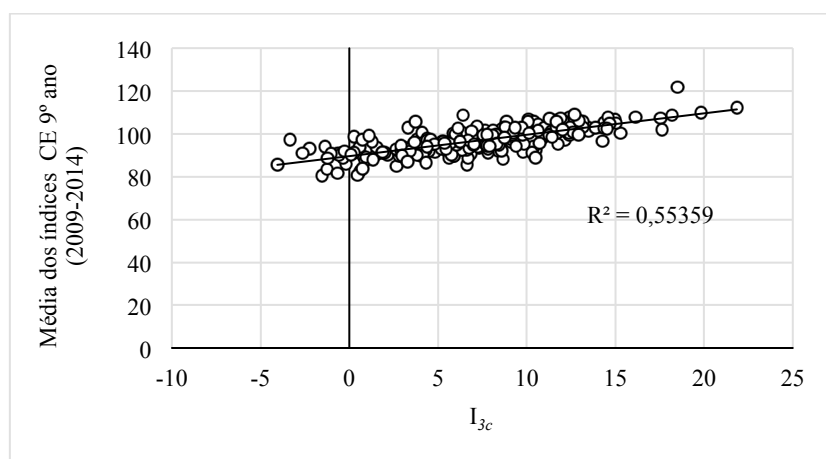
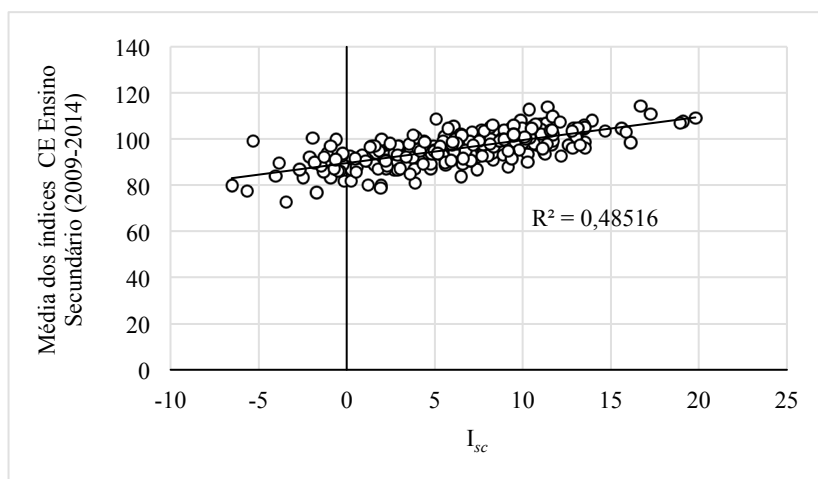


Figura 3. Relação dos resultados médios em índices dos concelhos de Portugal Continental nos exames nacionais do Ensino Secundário (2009-14) com o Indicador de Contexto Socioeconómico do Ensino Secundário (I_{sc})



4. CONCLUSÕES E VIAS DE PROSSEGUIMENTO DO TRABALHO

Consideradas as médias dos índices das classificações de seis anos nos exames (3º ciclo e Ensino Secundário), cada conjunto de variáveis socioeconómicas incluído nos modelos de regressão dá conta de praticamente metade das diferenças observadas entre concelhos; no caso das médias de três anos, o conjunto de variáveis dá conta de mais de 35% das diferenças observadas (6º ano). A utilização das médias dos índices estabiliza tanto mais a variância dos resultados entre concelhos, quanto mais anos forem considerados. Isso explica que se obtenham modelos de regressão mais bem ajustados para o 3º ciclo e o Ensino Secundário, relativamente ao modelo do 2º ciclo.

Visto este grau de ajustamento, e dado que todos os modelos são estatisticamente significativos, os valores estimados pelas equações de regressão constituem uma aproximação fiável ao peso das variações de contextos socioeconómicos sobre os resultados escolares.

Assim enquadrados, os níveis de variabilidade dos resultados escolares atribuíveis às diferenças de contexto socioeconómico apontam para uma forte influência de mecanismos sociais subjacentes, relacionáveis com as diferenças de contextos socioeconómicos sobre as oportunidades escolares, que persiste até nos níveis mais precoces da escolaridade obrigatória para que dispomos de resultados de exames nacionais. De um ponto de vista territorial, revelam desigualdades objetivas de realização das oportunidades educativas entre os concelhos de Portugal Continental, que podemos atribuir às diferentes composições socioeconómicas das suas populações (Seabra, 2009, 88). Vistos de outro ângulo, não é menos verdade que, apesar desse peso persistente das desigualdades socioeconómicas nas oportunidades escolares, metade (ou mais) da variação dos resultados médios concelhios lhes escapa. Tenhamos em conta o carácter incompleto e aproximativo das variáveis preditoras e a probabilidade de haver outros

fatores socioeconómicos que nenhuma delas mede; e, sobretudo, a relevância de fatores que não os de contexto socioeconómico – incluindo as variações do trabalho das próprias escolas, das famílias e de outros atores educativos. Nesse sentido, procuraremos em futuros estudos complementar a análise dos concelhos com a análise das escolas que os constituem e proceder a estudos de caso das escolas identificadas.

Para o objetivo específico deste documento de trabalho, resolvendo as equações de regressão apenas para as parcelas que exprimem os efeitos lineares das variáveis socioeconómicas na variação dos resultados estimados, construímos indicadores sintéticos de contexto socioeconómico que medem as diferenças de contextos socioeconómicos com impacto sobre os resultados escolares dos dois ciclos de ensino analisados, entre os concelhos de Portugal Continental. Disponibilizamos em anexo esses indicadores, que poderão ser usados autonomamente em outros estudos, como uma medida empírica para uma dimensão fundamental das desigualdades sociais na sociedade contemporânea, a da realização das oportunidades educativas, na sua expressão territorial. O facto de termos utilizado médias de seis anos dos resultados indexados às médias nacionais permite esperar um grau razoável de estabilidade dos indicadores do 3º ciclo e do Ensino Secundário, visto que é improvável que se deem flutuações de curto prazo com amplitude suficiente para alterar significativamente essas médias e, por conseguinte, as ponderações dos indicadores de base no indicador sintético. Naturalmente, este pressuposto de estabilidade terá que ser testado ao longo dos próximos anos, e os indicadores terão que ser atualizados em função da variação dos resultados de exames com uma periodicidade máxima de três anos. Já no que respeita ao indicador do 2º ciclo, impõe-se a sua revisão anual até ter igualmente atingido as médias de 6 anos.

Uma primeira limitação dos indicadores sintéticos de contexto socioeconómico aqui apresentados decorre de os resultados escolares serem indicados apenas pelas classificações nos exames nacionais. Inegavelmente, há outras dimensões dos resultados escolares que poderiam ser operacionalizadas por diferentes indicadores de resultados, à semelhança dos estudos publicados pelo MEC e de outros acima revistos (nomeadamente, Verdasca, 2013). Uma dessas dimensões, a do abandono escolar, já foi trabalhada usando basicamente a mesma metodologia aqui apresentada (Justino et al., 2014, 50-61). Tendo em atenção o nosso objetivo presente, consideramos útil reduzir as variáveis dependentes a uma por ciclo de escolaridade, e de entre a informação disponível, os resultados médios nos exames nacionais constituem o indicador singular mais homogéneo, dado o carácter nacional das provas, dos júris e dos critérios de correção. Uma linha de investigação futura, no sentido do aperfeiçoamento destes indicadores, poderá ser a construção de indicadores sintéticos também dos próprios resultados escolares, agregando diferentes dimensões de desempenho além das classificações nos exames. Para tal, seria necessário começar por definir e testar critérios para a seleção e ponderação dos indicadores de base, trabalho que se afigura à partida complexo mas sem dúvida de grande interesse.

Uma segunda limitação será a sua progressiva desatualização, à medida que nos afastamos do ano de referência da última operação censitária (e, mais parcelarmente, dos dados do MEC sobre a escolarização das mães dos alunos). Em parte, esse efeito poderá ser mitigado com informação proveniente de outras fontes, que venha a ser atualizada, bem como da publicação de estatísticas intercalares. Mas, como é inevitável em estudos deste tipo, a caracterização socioeconómica das unidades de análise irá progressivamente perdendo precisão até que se realize nova operação censitária. Importará, também por isso, acompanhar no tempo a fiabilidade preditiva destes indicadores sobre os resultados médios dos exames em anos

consecutivos. À medida que forem atualizadas as médias plurianuais de exames e a informação socioeconómica de base consoante a disponibilidade das fontes, será fundamental a construção de séries temporais dos próprios indicadores, que mostrem a mais longo prazo as tendências de estabilidade ou mudança dos contextos socioeconómicos e do seu impacto sobre os resultados escolares.

Uma outra limitação que esperamos poder ultrapassar, simultaneamente com a atualização dos dados, é a de a variável relativa à escolaridade das mães, nos dados a que tivemos acesso, ser agregada para todos os ciclos de escolaridade. É provável que a sua diferenciação por ciclos de escolaridade possa melhorar a capacidade preditiva desta variável, que é à partida a que tem maior influência líquida na variação dos resultados escolares em todos os ciclos. Uma vez que o mais recente trabalho de contextualização socioeconómica dos resultados de exames já usou as variáveis de caracterização fornecidas pelo MEC desagregadas por níveis de ensino (Azevedo e Portela, 2014), contamos em breve poder fazer mais esta adequação dos modelos.

Progressivamente, à medida que se amplie o número de anos com resultados desses exames, visamos completar esta análise com os resultados médios do 1º ciclo do Ensino Básico (4º ano de escolaridade), assim que nos for possível estabelecer médias de índices com pelo menos três anos, ampliando em seguida progressivamente os anos de referência até ao total de seis, tal como já iniciámos para o 2º ciclo.

Em desenvolvimentos futuros, e na medida do que a disponibilidade e a qualidade dos dados permitirem, convirá explorar efeitos e construir indicadores operatórios a diferentes escalas (agrupamentos, escolas, alunos), como sugerido em Marôco (2011), o que nos permitiria aplicar outras técnicas de análise estatística, como os modelos multinível (Abbott, 2011). Na investigação em curso, já ensaiámos procedimentos semelhantes aos aqui

utilizados para o nível concelhio, à escala das escolas, usando os dados de caracterização socioeconómica extraídos da base MISI. Embora tenhamos obtido resultados estatisticamente significativos, a qualidade desigual dos dados entre escolas, muitas das quais excluídas por excesso de não respostas em variáveis chave, e a inviabilidade prática de obtenção de médias de resultados plurianuais, devido à sucessiva recomposição das unidades orgânicas durante os últimos anos, reduzem a robustez dos indicadores. A superação dessas dificuldades obrigará ainda, a par da esperada melhoria da informação de base, a um trabalho muito mais complexo e aturado sobre as fontes, que esperamos começar a desenvolver num futuro próximo.

REFERÊNCIAS

Abbott, Martin Lee. 2011. *Understanding Educational Statistics Using Microsoft Excel and SPSS*. New Jersey: Wiley.

Azevedo, Joaquim e Conceição Portela. 2014. “Como se Define o Contexto de Uma Escola e Aquilo que se Espera Dela?”. *O Público*, 29 de novembro.

Barroso, João. 2013. “A Emergência do Local e os Novos Modos de Regulação das Políticas Educativas”. *Educação: Temas e Problemas 12-13*: 13-25.

Batista, Susana, Luísa Franco e Rui Santos. 2014. “Uma Abordagem aos Contextos Socioeconómicos Locais e Efeitos nos Resultados Escolares”. In *O Estado da Educação 2013*, 230-238. Lisboa: Conselho Nacional de Educação.

Canto e Castro, Luísa, Joaquim Santos, Tiago Pereira e Ana Vitorino. 2011. *Modelos para a Comparação Estatística dos Resultados Académicos em Escolas de Contexto Análogo – Painel de dados para apoio à avaliação externa das escolas*, Lisboa: Direção Geral de Estatística da Educação e Ciência. <http://www.dgeec.mec.pt/np4/120/>

Diogo, Ana Isabel. 2008. *Investimento das Famílias na Escola: Dinâmicas Familiares e Contexto Escolar Local*. Oeiras: Celta.

Formosinho, João e Joaquim Machado. 2013. “A Regulação da Educação em Portugal: Do Estado Novo à Democracia”. *Educação: Temas e Problemas 12-13*: 27-40.

GTAAE (Grupo de Trabalho para a Avaliação Externa das Escolas). 2011. *Propostas Para um Novo Ciclo de Avaliação Externa das Escolas: Relatório Final*. Lisboa: Ministério da Educação. http://www.ige.min-edu.pt/upload%5CRelatorios/AEE2_GT_2011_RELATORIO_FINAL.pdf

Justino, David e Maria João Valente Rosa. 2009. “Contexto Envolvente e Factores Condicionantes do Trabalho Infantil (Demográficos, Económicos, Sociais e

Culturais)”. In *Infância Interrompida: Caracterização das Actividades Desenvolvidas pelas Crianças e Jovens em Portugal*, org. por Manuel Lisboa, 35-82. Lisboa: Colibri.

Justino, David, Liliana Pascueiro, Luísa Franco, Rui Santos, Sílvia Almeida e Susana Batista. 2014. *Atlas da Educação: Contextos Sociais e Locais do Sucesso e Insucesso: Portugal, 1991/2012*. Lisboa: EPIS. <http://www.epis.pt/downloads/mentores/atlas-da-educacao.pdf>

Land, Kenneth C. e Abbott L. Ferriss. 2007. “The Sociology of Social Indicators”. In *21st Century Sociology*, org. por Clifton D. Bryant e Dennis L. Peck, 518-26. Londres, Thowsand Oaks CA, Nova Dehli: Sage.

Marktest Computing. 2012. *Sales Index v. 6.3.3.1*.

Marôco, João. 2011. *Análise Estatística com o SPSS Statistics*. Pêro Pinheiro: ReportNumber.

Pordata. 2014. *Alunos Matriculados no Ensino Básico: Total e por Subsistema de Ensino – Municípios*.

<http://www.pordata.pt/Municipios/Alunos+matriculados+no+ensino+basico+total++por+subsistema+de+ensino-344> , atualizado em 15/09/2014.

Portela, Conceição e Joaquim Azevedo. 2013. “A Importância de Atribuir um Contexto e um Valor Esperado às Escolas”. *O Público*, 9 de novembro.

Seabra, Teresa. 2009. “Desigualdades Escolares e Desigualdades Sociais”. *Sociologia, Problemas e Práticas* 59: 75-106.

Verdasca, José L. C.. 2013. “Rankings Escolares: ‘A César o que É de César, a Deus o que É de Deus’ ”. *Educação: Temas e Problemas* 12-13: 175-99.

Viana, Clara. 2012. “Metade das Escolas Secundárias Públicas Ficaram Abaixo do Esperado”. *O Público*, 13 de outubro.

Agradecimentos

O início deste estudo foi desenvolvido por Luísa Franco e Susana Batista como parte do trabalho final para a Pós-Graduação em Análise de Dados para as Ciências Sociais do ISCTE-IUL, pelo que devemos um agradecimento a todo o corpo docente desse programa pelas aprendizagens que nos proporcionou e que aqui foram diretamente aplicadas. Em particular, gostaríamos de deixar um especial reconhecimento à Professora Patrícia Ávila, que nos forneceu pistas cruciais para o enriquecimento da análise estatística.

Temos também a agradecer as críticas de um(a) revisor(a) anónimo/a da primeira versão deste documento de trabalho, que muito contribuíram para melhor precisarmos os objetivos e esclarecermos os procedimentos, bem como para uma revisão substantiva da própria metodologia de construção dos indicadores.

Todos os erros e falhas são da nossa responsabilidade.

ANEXOS

Anexo 1 – Estatística descritiva das variáveis de *input* dos modelos de RLM

Tabela 1. Estatística descritiva das variáveis de resultados nos exames e de caracterização socioeconómica dos concelhos¹

	Mínimo	Máximo	Média	Mediana	Desvio padrão	Coef. Variação (%)	N válidos
Índices das classificações nos exames do 6º ano (2012-2014) ^a	66,99	121,59	96,78	96,92	6,72	6,9	278
Índices das classificações nos exames do 9º ano (2009-2014) ^a	80,46	121,70	96,95	96,54	5,95	6,1	278
Índices das classificações nos exames do Ensino Secundário (2009-2014) ^a	72,65	114,33	96,18	96,98	7,15	7,4	242
Componente «Urbano – Atividade de Serviços de Qualificação Elevada»	-1,76	5,04	0,00	-0,20	1,00	--	278
Componente «Atividade Industrial de Qualificação Média-Baixa»	-2,48	2,85	0,00	-0,05	1,00	--	278
Componente «Atividade de Turismo e Serviços de Qualificação Média»	-2,55	5,79	0,00	-0,18	1,00	--	278
Componente «Desvantagem Socioeconómica (Exclusão)»	-1,80	3,67	0,00	-0,13	1,00	--	278
Componente «Agrícola – Atividade Agrícola de Qualificação Baixa»	-1,74	3,15	0,00	-0,34	1,00	--	278
Mães com Ensino Superior (%) ^b	1,40	37,39	13,18	12,06	5,90	44,8	278

¹ Por definição, as médias dos valores concelhios nas componentes extraídas da análise de componentes principais são iguais a 0 e os desvios-padrão iguais a 1.

Fontes: ^a Júri Nacional de Exames; ^b MISI e INE, 2011, cálculos dos autores.

Figura 1. Distribuição dos *scores* das cinco componentes

