

Literacia da leitura, a prioridade necessária nos dois primeiros anos de escolaridade

José Verdasca

1. Apesar do enorme progresso registado nos indicadores de escolarização e de qualidade de sucesso educativo, os dados escolares continuam a evidenciar vulnerabilidades comprometedoras da realização plena do imperativo constitucional da equidade educativa e da igualdade de oportunidades de sucesso escolar para todos os alunos. A comprová-lo está o elevado número de alunos que continua anualmente em elevado risco de retenção e o facto de a sua incidência tender a ocorrer nos grupos de alunos provenientes de meios socioeconómicos e culturais mais desfavorecidos.

Estas manifestações ocorrem nos percursos escolares dos alunos desde muito cedo, como o evidenciam as taxas de retenção do 2º ano de escolaridade e, por vezes, repetidamente. O sentido dos resultados das provas de aferição de Português do 2º ano realizadas pela primeira vez em 2016/17, foi indicativo de fragilidades, designadamente na literacia da leitura, com quase 40% dos alunos a revelarem dificuldades na leitura e cerca de 60% na escrita e, simultaneamente, apesar de se estar no início da escolaridade, as turmas com maiores densidades de alunos provenientes de meios familiares economicamente carenciados e com habilitações académicas mais baixas tenderem a registar piores desempenhos na prova de aferição nos domínios da leitura e da escrita do que as turmas com menores densidades de alunos com estes perfis socioculturais e económicos. (Estudos PNPSE, 2017).

2. A título ilustrativo, exploremos algumas das soluções obtidas quando fazemos interagir os domínios e subdomínios das provas de aferição do 2º ano (K2_2017) de Português com os de Matemática. O que sugerem os dados obtidos? Como interpretar as interações observadas entre os diversos domínios e subdomínios de ambas as disciplinas e que decisões tomar em cada contexto para assegurar a progressiva melhoria das aprendizagens dos alunos?

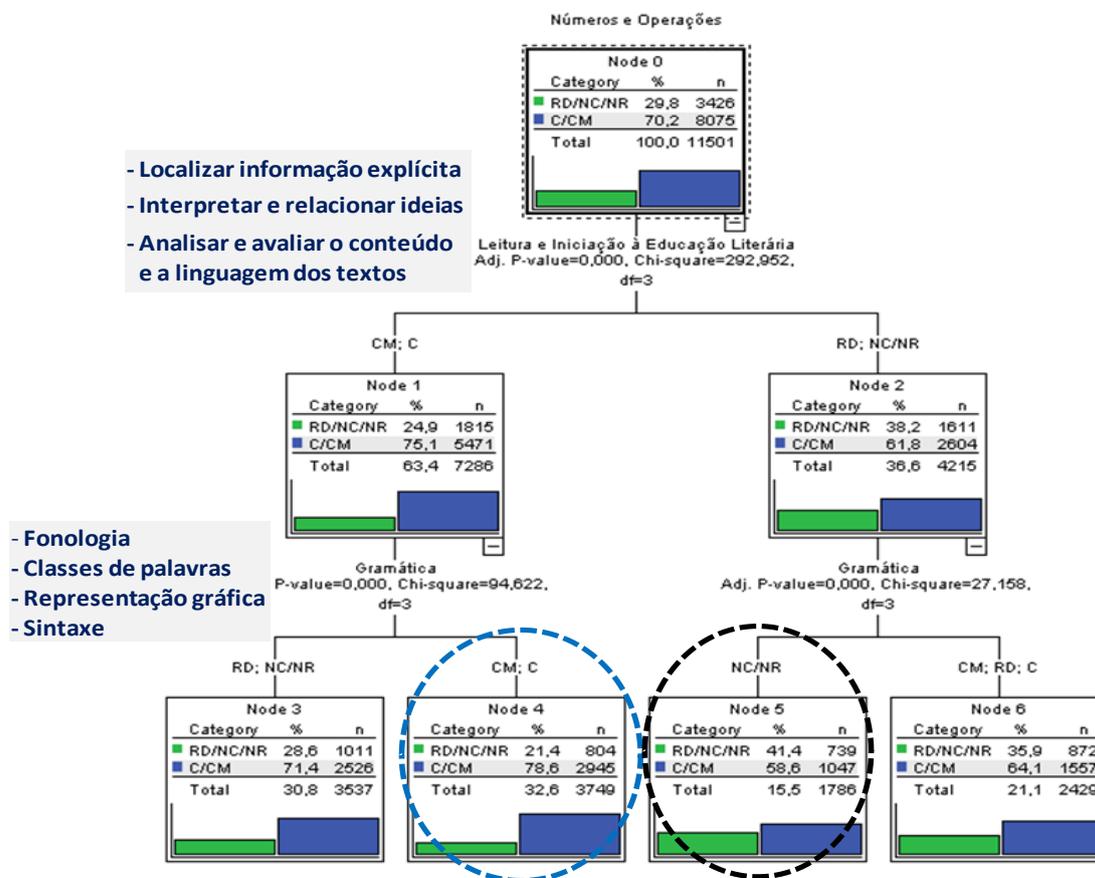
Para respondermos às questões anteriores tomámos por referência uma amostra aleatória de 11501 alunos K2 de escolas do continente (cerca de 12% do universo dos alunos) que realizaram as provas de aferição em 2017. Estabelecemos como variáveis antecedentes (explicativas) os quatro domínios da prova de Português (*Compreensão do Oral, Leitura e Iniciação à Educação Literária, Gramática, Escrita*) e com variáveis consequentes (explicadas) os domínios da prova de Matemática (*Números e Operações, Geometria e Medida, Organização e Tratamento de Dados*). Neste breve exercício, por questões de economia de espaço, e uma vez que as interações obtidas são similares para os restantes domínios da Matemática, apresentamos apenas as soluções geradas para 'Números e Operações' e 'Organização e Tratamento de Dados'.

Dada a dimensão da amostra, o número de variáveis simultaneamente em análise e respetivas escalas de medida e, sobretudo, as vantagens que num exercício com cunho exploratório as abordagens heurísticas proporcionam, optamos pelo método das árvores de classificação, um método estatístico adequado para identificar, classificar e explorar estruturas complexas. As estruturas arbóreas geradas, que designamos de '*diagrama da estrutura hierárquica das*

interações', foram obtidas com recurso ao algoritmo *QUEST (Quick, Unbiased, Efficient Statistical Tree)*, um algoritmo que otimiza o comportamento de vários fatores com várias categorias em estruturas de segmentação binária, utilizando a medida do qui-quadrado e a escala de medida nominal na variável explicada (Pestana e Gageiro, 2009). Na sua base está a criação de uma estrutura hierárquica, designada de árvore de classificação, decorrente do fracionamento sucessivo dos dados em subconjuntos mutuamente exclusivos gerando nós intermédios e nós terminais. A partição dos níveis da árvore faz-se por ordem decrescente de importância das variáveis explicativas ou predictoras. A variável explicativa estatisticamente mais significativa define a primeira partição, gerando o primeiro nível de profundidade, e assim sucessivamente, sempre que ocorra a possibilidade da subdivisão de cada uma das suas categorias de resposta em novas segmentações. Este processo é aplicado recursivamente até que a análise esteja terminada. A obtenção de divisões sucessivas dos dados de modo a permitir definir grupos ou classes com base em segmentações significativamente diferentes relativamente à variável dependente constitui o principal objetivo do algoritmo. Qualquer um dos níveis de profundidade da estrutura arbórea pode conter nós terminais, sendo estes a principal base de referência para a identificação das sequências estruturais até aos nós terminais ou finais. A estrutura arbórea é constituída pelo nó inicial, por nós intermédios e por nós finais. Estes estão representados através de caixas que contém informação sobre as frequências absolutas e relativas percentuais dos alunos nas diversas menções qualitativas utilizadas. Em cada caixa, a categoria modal prevista da variável de resposta aparece sombreada a cinzento.

3. Detendo-nos na solução arbórea gerada, no nó inicial (nó 0) está assinalada a sombreado a categoria 'C/CM' (conhece/conhece mas), mostrando que na amostra geral há uma maior probabilidade de ocorrerem menções 'C/CM' ($8075/11501 \cdot 100 = 70,2\%$) do que menções 'RD/NC/NR' (revela dificuldades/não conhece/não responde) no domínio 'Números e Operações' ($3426/11501 \cdot 100 = 29,8\%$).

Figura 1: Diagrama da estrutura hierárquica das interações entre os resultados nos domínios de Português e os resultados em 'Números e Operações'



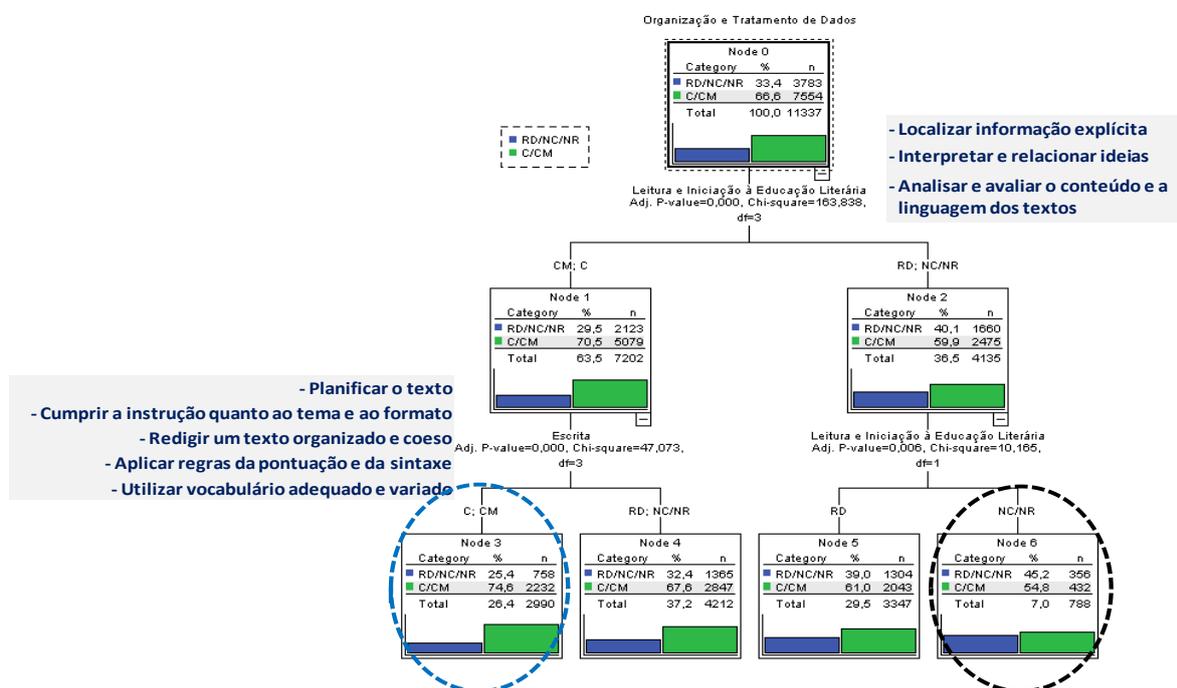
O primeiro nível de profundidade da árvore é determinado pelo nível de proficiência dos alunos no domínio 'Leitura e Iniciação Literária', revelando que, do total dos 11501 alunos da amostra, 7286 (63,4%) obtiveram menções de desempenho 'C/CM' na 'Leitura e Iniciação Literária'. Ao gerar a primeira partição da amostra geral em duas categorizações separando as categorias de desempenho positivo das de desempenho negativo no domínio 'Leitura e Iniciação Literária', concluiu-se ser este, de entre os quatro domínios da literacia da leitura, o preditor mais significativo do desempenho em 'Números e Operações'. Apesar de no primeiro nível de profundidade não existirem nós terminais, mas apenas nós intermédios, retira-se do nó 1, que o algoritmo QUEST lhe atribui 75,1% de probabilidade de pertencer à categoria de alunos com desempenho 'C/CM' no domínio 'Números e Operações'. Retira-se ainda deste nó que 68% (5471/8075*100) da totalidade dos alunos da amostra geral com desempenho positivo neste domínio da Matemática se deve a alunos com desempenho positivo no domínio da 'Leitura e Iniciação Literária'.

No segundo nível de profundidade emergem como preditores estatisticamente mais significativos na subsegmentação dos ramos do primeiro nível a 'Gramática' e surgem os nós terminais 3, 4, 5 e 6, respetivamente com 3537, 3749, 1786 e 2429 casos e com uma probabilidade de 71%, 79%, 59% e 64% de incluir alunos com desempenho positivo em Matemática (Números e Operações).

O nó 4 é, de todos os restantes, o que regista maior 'ganho', pois contém 36% (2945/8075*100) da amostra de alunos da categoria alvo ('C/CM'); por outro lado, concentram-se na categoria alvo deste nó mais 12% de alunos do que na categoria alvo da amostra geral (78,6%/70,2%*100). A contrastar, o caso do nó 5, com apenas 13% desses alunos e um index de apenas 83 (58,6/70,2*100). Por último, as respetivas sequências estruturais destes nós, que podem ser descritas do seguinte modo: Nó 4 - alunos com desempenho positivo na prova de Português nos domínios da 'Leitura e Iniciação Literária' e da 'Gramática' apresentam uma probabilidade de 78,6% de desempenhar positivamente em 'Números e Operações'; Nó 5 - alunos com desempenho negativo (RD/NC/NR) na prova de Português no domínio da 'Leitura e Iniciação Literária' e que obtiveram menções qualitativas de desempenho apenas ao nível do 'NC/NR' na 'Gramática' apresentam uma probabilidade de apenas 58,6%, ou seja, menos 20% do que os seus pares com desempenhos positivos naqueles domínios, de desempenhar positivamente em 'Números e Operações'.

Projetemos agora a solução arbórea para o domínio de 'Organização e Tratamentos de Dados', deixando ao leitor o desafio da análise das interações geradas.

Figura 2: Diagrama da estrutura hierárquica das interações entre os resultados nos domínios de Português e os resultados em 'Organização e Tratamento de Dados'



4. O que nos mostram as interações entre os diversos domínios de Português e de Matemática?

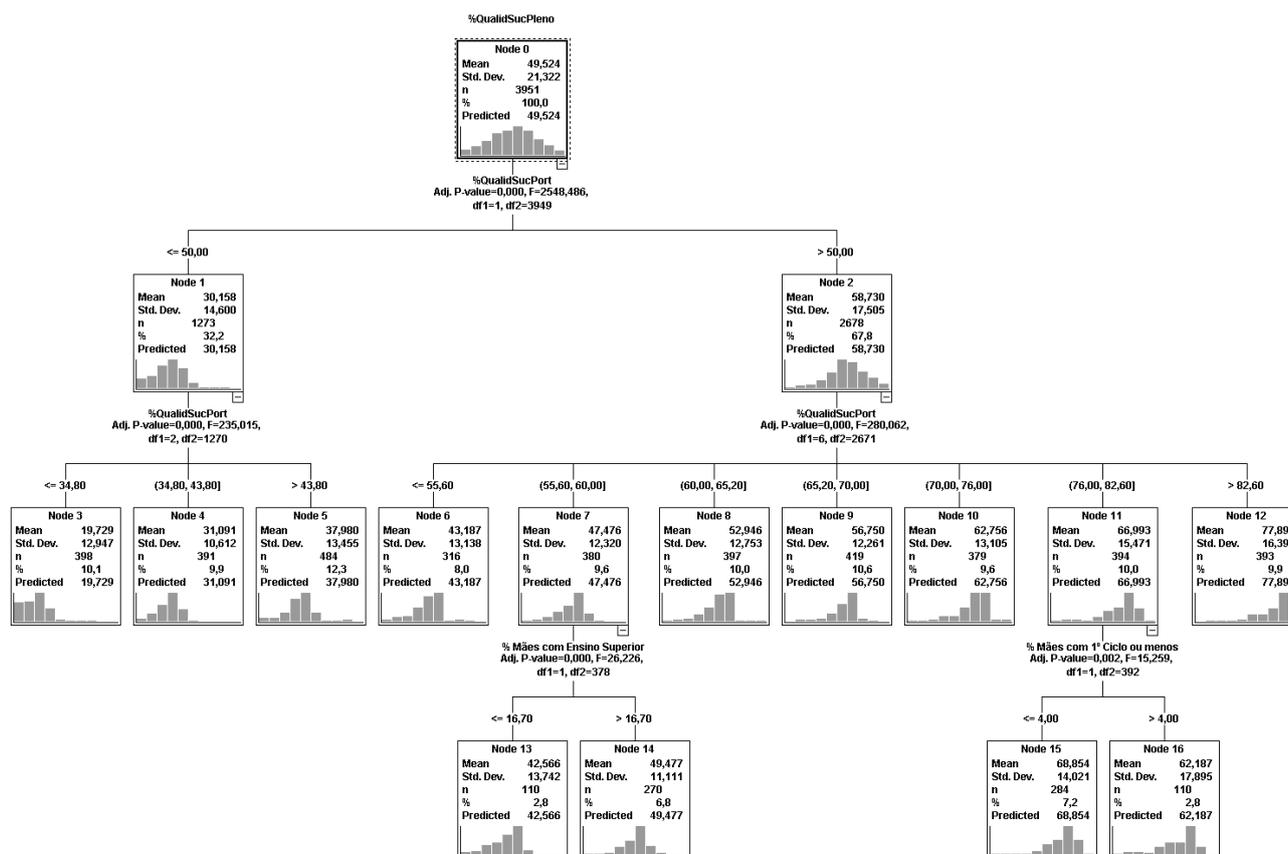
Em ambos as soluções apresentadas, os resultados sugerem que maiores competências na literacia de leitura, nomeadamente, relacionadas com a capacidade de localizar informação, interpretar e relacionar ideias e de analisar e avaliar o conteúdo e a linguagem dos textos e,

consequentemente, a elas associadas uma maior capacidade na compreensão e interpretação oral e escrita, introduzem incrementos estatisticamente significativos no desempenho em 'Números Operações' e 'Organização e Tratamento de Dados'. Estes incrementos de desempenho ampliam-se no primeiro caso quando cumulativamente os alunos demonstram também maiores competências na 'Gramática' e, no segundo caso, quando conseguem melhores níveis de proficiência na 'Escrita', designadamente, ao nível da capacidade de redigir um texto de forma organizada e coesa e da utilização de vocabulário adequado e variado.

5. Em jeito de triangulação, vejamos agora o que nos sugerem outras fontes de dados, como sejam por exemplo os resultantes dos registos efetuados por professores do 1º ciclo no final do 1º período do presente ano letivo relativamente a cada uma das suas turmas. É certo que, no que respeita sobretudo ao desempenho escolar, não são dados finais, mas dados intercalares. Ainda assim, permitem-nos reflexões complementares, de natureza pedagógico-didática, organizacional, contextual, quando se fazem interagir simultaneamente métodos de leitura utilizados, horas de apoio educativo, modalidades organizativas prevalentes, características composicionais das turmas e qualidade de desempenho em Português e em todas as áreas curriculares.

Nos dois exercícios exploratórios realizados com recurso ao método das árvores de classificação e regressão junto de uma amostra de 3951 turmas de tipologia 'pura' do 1º e do 2º ano do ensino básico, concluiu-se que a percentagem de mães com ensino superior na turma, seguida da percentagem de alunos do escalão A da ação social escolar emergem recursivamente como as principais variáveis explicativas da qualidade do sucesso em Português, medida através da percentagem de alunos com menção de Bom ou Muito Bom, sendo por isso as determinantes mais relevantes da estrutura arbórea e dos respetivos nós terminais (perfis) apurados. Na segunda análise, e é desta que apresentamos as soluções estatísticas apuradas, procurou-se verificar se a introdução da variável qualidade de sucesso em Português como hipotética variável explicativa da plena qualidade do sucesso (% de alunos da turma com desempenho Bom ou Muito Bom em todas as áreas curriculares) se sobrepunha às metodologias e modalidades organizativas declaradas e às características socioeconómicas e culturais das turmas ou se, pelo contrário, a estrutura composicional das turmas, nomeadamente no que respeita às densidades de elevada escolaridade e condição socioeconómica dos alunos, continuava a prevalecer como determinantes explicativos. Os resultados são claros: a qualidade de sucesso em Português emerge como primeira determinante da segmentação da árvore de regressão nos dois primeiros níveis de profundidade, seguindo-se no terceiro nível o capital escolar das mães (Figura 3 e Apêndice 1).

Figura 3: Diagrama da estrutura hierárquica das interações com a elevada qualidade do sucesso em todas as áreas curriculares em turmas do 1º e do 2º ano do ensino básico



Numa brevíssima análise destes resultados comparemos os dois pares de perfis mais extremos, ou seja, os pares constituídos pelos perfis 12 e 3 e pelos perfis 15 e 4. Abrangem 1466 turmas, representando cerca de 37% do total. As turmas do perfil 12, com níveis de bom ou muito bom em Português superiores a 82,6% apresentam uma média de rendimento académico pleno que supera em cerca de quatro vezes o rendimento das turmas do perfil 3 com uma qualidade de desempenho em Português inferior a 35%. O contraste entre as turmas do segundo par são ligeiramente menos acentuados, ainda assim, o perfil 15, cujas turmas têm como principal característica composicional apresentarem uma densidade de mães com 1º ciclo ou menos que é no máximo inferior ou igual a 4% e, simultaneamente, com desempenho de bom ou muito bom em Português entre os 76% e 82,6% , regista um rendimento académico pleno de mais do dobro do do perfil 4 e em que as turmas que o constituem registam um desempenho em Português compreendido apenas entre os 34,9% e os 43,8%.

6. Procedamos agora à conversão dos 12 nós terminais (perfis) em unidades de observação e observemos como se comportam as diferentes variáveis e projetam os resultados por perfil quando submetidas à análise pelos mínimos quadrados parciais (PLS), um método de regressão considerado particularmente útil na construção de equações de predição quando se está em presença de várias variáveis resultado e explicativas correlacionadas entre si e se

dispõe de um reduzido número de casos (Pestana e Gageiro, 2009; Silveira et al, 2012; Verdasca, 2015).

Tabela: Matriz de correlação

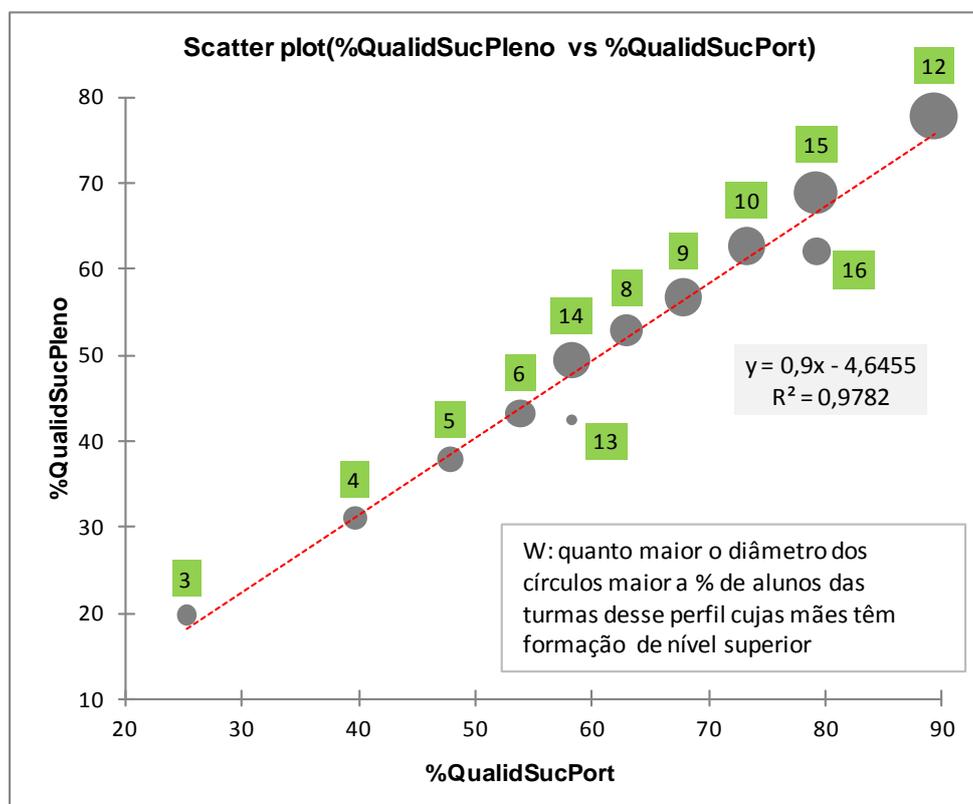
Variáveis	Média de Alunos da Turma	%Alunos declarados com dificuldades de aprendizagem	%Alunos ASE A	% Mães com Ensino Superior	% Mães com 1º Ciclo ou menos	Horas de apoio educativo	%Alunos com Bom ou Muito Bom em Português	%Alunos com Bom ou Muito Bom em todas as disciplinas
Média de Alunos da Turma	1,00	-0,75	-0,86	0,86	-0,89	-0,30	0,57	0,66
%Alunos declarados com dificuldades de aprendizagem	-0,75	1,00	0,85	-0,71	0,57	0,38	-0,88	-0,89
%Alunos ASE A	-0,86	0,85	1,00	-0,89	0,83	0,63	-0,86	-0,91
% Mães com Ensino Superior	0,86	-0,71	-0,89	1,00	-0,77	-0,43	0,70	0,79
% Mães com 1º Ciclo ou menos	-0,89	0,57	0,83	-0,77	1,00	0,45	-0,47	-0,56
Horas de apoio educativo	-0,30	0,38	0,63	-0,43	0,45	1,00	-0,66	-0,67
%Alunos com Bom ou Muito Bom em Português	0,57	-0,88	-0,86	0,70	-0,47	-0,66	1,00	0,99
%Alunos com Bom ou Muito Bom em todas as disciplinas	0,66	-0,89	-0,91	0,79	-0,56	-0,67	0,99	1,00

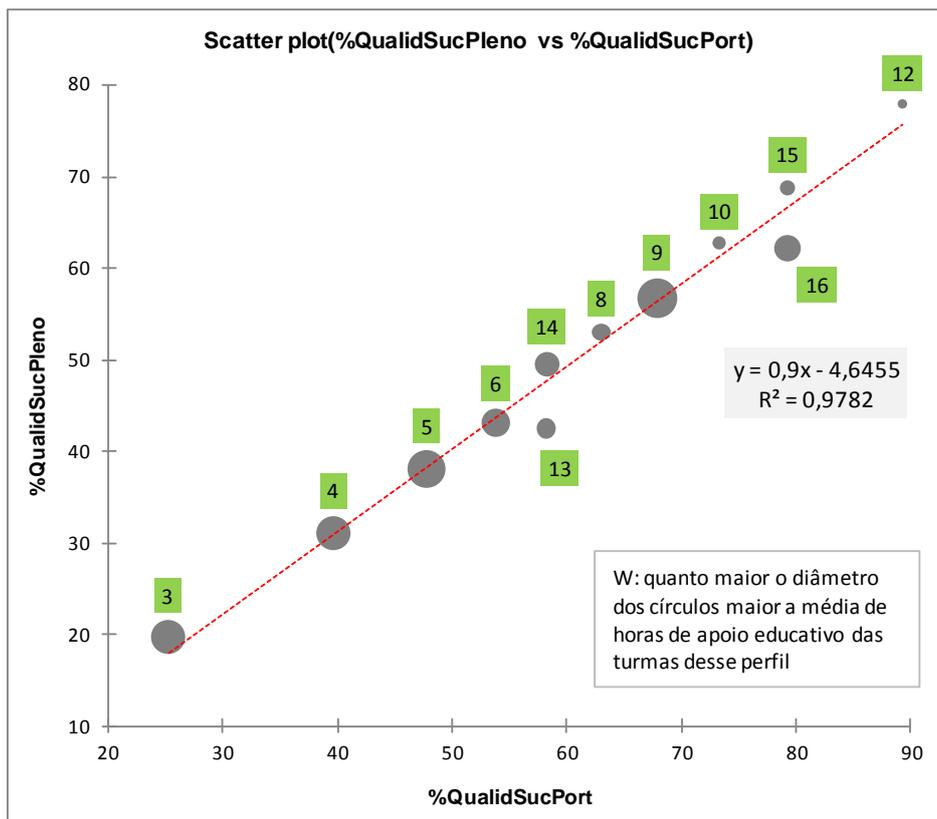
Observando os resultados constantes da matriz de correlação (Tabela 1), retiremos conclusões das intercorrelações registadas entre algumas das variáveis explicativas e as variáveis de desempenho académico:

- i) os perfis com maior número médio de alunos por turma são constituídos por turmas que tendem a ter menores densidades de alunos declarados com dificuldades de aprendizagem e também menores densidades de alunos socioeconomicamente carenciados, com baixo capital escolar familiar e menor média de horas de apoio educativo;
- ii) as turmas com maiores densidades de alunos com dificuldades de aprendizagem são também as que tendem a ter menor desempenho académico, a beneficiar de maior número de horas de apoio e a ter maiores densidades de alunos do escalão A e de mães com escolaridade muito reduzida;
- iii) os perfis constituídos por turmas mais carenciadas, com escolaridades das mães mais baixas e com pior rendimento académico são os que tendem a dispor de um maior número médio de horas de apoio educativo, o que sugere a existência de uma distribuição das horas de apoio por parte das lideranças escolares compaginável com diagnósticos adequados dessas necessidades;
- iv) a correlação quase perfeita de sentido positivo entre a qualidade de desempenho em Português e a qualidade de desempenho em todas as áreas curriculares, superando em intensidade todas as outras, corrobora a ideia da importância estratégica que temos vindo a sustentar - uma aprendizagem de elevada qualidade na literacia da leitura nos dois primeiros anos do ensino básico será fortemente potenciadora do rendimento académico nos anos escolares seguintes e base essencial da concretização do princípio da igualdade de oportunidades de sucesso educativo para todos e cada um dos alunos por via da realização no tempo próprio das aprendizagens essenciais e significativas nos ensinos básico e secundário.

7. Os diagramas de dispersão dos perfis de desempenho acadêmico com pelo menos nível bom em Português e em todas as áreas curriculares evidenciam estas conclusões. Para além de projetarem as posições relativas dos aglomerados de turmas de acordo com os seus perfis de qualidade de desempenho académico, a representação destes perfis através de bolhas (círculos) de diferentes dimensões mostra alguns casos de desalinhamento.

Figuras 4A e 4B: Diagramas de dispersão com projeção da densidade do capital escolar das mães e das horas semanais de apoio educativo por perfil de desempenho académico através do tamanho dos círculos





Nos diagramas, a posição relativa dos perfis 13 e 16 na linha de ajustamento tendencial sugeriria comprimentos dos diâmetros similares aos dos seus vizinhos mais próximos. Quer no caso do perfil 13, quer no caso do perfil 16, ambos com 110 turmas, os resultados em Português sugeriam maiores percentagens de alunos com bom ou muito bom em todas as áreas curriculares; mas, por outro lado, tendo em conta as características contextuais destes perfis, no que concerne ao capital escolar das mães dos alunos destas turmas, os desempenhos académicos acabam por superar o que seria expectável tendo em conta que as turmas de ambos os perfis têm bastante menor densidade de mães com ensino superior que as turmas dos perfis vizinhos. Já no que respeita à media de horas de apoio educativo das turmas desses perfis, em particular o tamanho do círculo correspondente ao perfil 13 é indicativo de um volume de horas de apoio semanal inferior à maioria dos restantes perfis apesar das características de composição menos favorecidas das turmas que integram esse perfil. Por que territórios se distribuem e a que escolas pertencem é, naturalmente, uma questão a explorar.

8. Em jeito de conclusão, e perante tais evidências, parece ganhar consistência a ideia da importância estratégica de nos dois primeiros anos de escolaridade fazer incidir o principal foco da ação pedagógica na literacia de leitura, sob pena de ficar e continuar comprometida não só a qualidade e evolução das aprendizagens dos alunos em Português como a qualidade de desempenho nas outras áreas curriculares. Priorizar a leitura e escrita e estabelecer como principal objetivo e compromisso educacional alcançar com todos e cada um dos alunos níveis

de elevada proficiência, independentemente dos seus perfis sociais, requer um compromisso sério e ousado em cada escola e território e a necessidade de envolver nessa construção local os professores enquanto líderes pedagógicos mas também outras entidades e atores da comunidade, definindo localmente estratégias de ação baseadas em abordagens curriculares abertas e enriquecidas e agindo em contexto escolar e fora dele de forma articulada, convergente e complementar.

Referências

- Pestana, M. H. e Gageiro, J. (2009). *Análise Categórica, Árvores de Decisão e Análise de Conteúdo em Ciências Sociais e da Saúde com o SPSS*. Lisboa: Lidel, Edições Técnicas.
- Silveira, F., Chaves, L., Silva, F. e Costa, L. (2012). Avaliação da regressão PLS (Partial Least Square) considerando diferentes estruturas de correlação entre as covariáveis. *Revista da Estatística, UFOP*, Vol. II, p. 85.
- Verdasca, J. (2015). Monitorização e (auto)regulação escolar: exercício metodológico para a comparação e projeção de resultados escolares em turmas de estrutura composicional semelhante. CIEP-Publicações (<https://dspace.uevora.pt/rdpc/handle/10174/20880>).

Apêndice 1

Tabela de especificações, ganhos por nó terminal (perfil) e risco

Model Summary

Specifications	Growing Method	EXHAUSTIVE CHAID
	Dependent Variable	%QualidSucPleno
	Independent Variables	%QualidSucPort, Nº de Alunos da Turma, % Alunos NEE, %Alunos NEE_ReduzTurma, % Alunos ASE A, %Alunos ASE B, % Mães com Ensino Superior, % Mães com 1º Ciclo ou menos, Habilitação Média Mães (em anos), Modalidade de apoio prevalente, Método de leitura prevalente, Horas Apoio
	Validation	None
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	100
	Minimum Cases in Child Node	50
Results	Independent Variables Included	%QualidSucPort, % Mães com Ensino Superior, % Mães com 1º Ciclo ou menos
	Number of Nodes	17
	Number of Terminal Nodes	12
	Depth	3

Gain Summary for Nodes

Node	N	Percent	Mean
12	393	9,9%	77,894
15	284	7,2%	68,854
10	379	9,6%	62,756
16	110	2,8%	62,187
9	419	10,6%	56,750
8	397	10,0%	52,946
14	270	6,8%	49,477
6	316	8,0%	43,187
13	110	2,8%	42,566
5	484	12,3%	37,980
4	391	9,9%	31,091
3	398	10,1%	19,729

Growing Method: EXHAUSTIVE CHAID
 Dependent Variable: %QualidSucPleno

Risk

Estimate	Std. Error
175,697	6,171

Growing Method:
 EXHAUSTIVE CHAID
 Dependent Variable: %
 QualidSucPleno

% de variância explicada pelo modelo: 61%

Apêndice 2

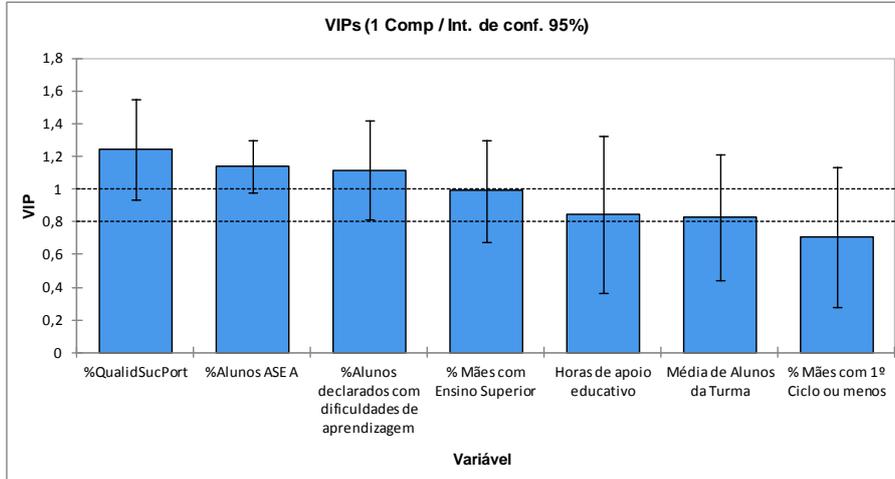
Método PLS: Anotações complementares

Qualidade do modelo

Estatística	Comp1
Q ² acum	0,849
R ² Y acum	0,875
R ² X acum	0,734

Importância de cada variável no ajustamento

A medida da importância de cada variável no ajustamento dos scores de X e de Y (VIP) deve ser inferior a 0.8 e a retenção do número de fatores latentes a que maximize a proporção acumulada de variância explicada (Wold, 1994, citado por Pestana e Gageiro, 2009).



Coefficientes padronizados

