

## Projeto PMA-CEAGXXI: exercício metodológico exploratório de análise e estrutura hierárquica de itens e composição sociodemográfica dos perfis de resposta

*José Verdasca, Ricardo Monginho, Ana Cristóvão e Rui Espadeiro*

*CIEP-UÉvora*

### Introdução

O presente relatório incide sobre a análise dos resultados da Prova Global Diagnóstica (PGD\_K3), aplicada em 8 e 9 de junho às sete turmas do 3.º ano de escolaridade envolvidas no projeto 'Promoção de Mudanças na Aprendizagem/Comunidades Escolares de Aprendizagem Gulbenkian XXI' (PMA-CEAGXXI).

A prova é constituída por vinte e oito questões (itens), distribuídas pelas áreas curriculares de Português (sete questões), de Matemática (nove questões) e de Estudo do Meio (doze questões) e foi realizada com recurso ao *Socrative*<sup>1</sup>, uma ferramenta multiplataforma que permite criar uma sala virtual e disponibilizar provas aos alunos, bem como acompanhar a sua aprendizagem em tempo real e, se identificados défices, propor ações imediatas de correção.

A preparação da prova iniciou-se com a construção de uma bolsa de trinta e seis questões, tendo por base as metas curriculares definidas para o 3.º ano de escolaridade. Com o objetivo de promover a transversalidade das práticas avaliativas e de induzir a (auto)regulação colaborativa entre escolas, partilhou-se um instrumento de avaliação cujas questões foram previamente submetidas à apreciação das professoras titulares das turmas do projeto PMA-CEAG XXI, tendo sido selecionadas vinte e oito das trinta e seis questões inicialmente propostas.



---

<sup>1</sup> A prova global diagnóstica pode ser acedida e consultada em:  
[https://dl.dropboxusercontent.com/u/23206103/Publica%C3%A7%C3%B5es%20Site%20Projeto/Quiz\\_provaglobaldediagn%C3%B3sticojunho2015.pdf](https://dl.dropboxusercontent.com/u/23206103/Publica%C3%A7%C3%B5es%20Site%20Projeto/Quiz_provaglobaldediagn%C3%B3sticojunho2015.pdf).

---

As características da prova dependem, entre outros aspetos, das características psicométricas das suas questões, tendo-se procedido ao cálculo dos parâmetros dos itens no âmbito da Teoria Clássica dos Testes (TCT). As respostas foram tratadas como dicotómicas – certas ou erradas – e comparados os valores obtidos<sup>2</sup>. Num primeiro momento, deter-nos-emos nos índices de dificuldade e de discriminação da prova, que abordaremos de forma conjugada; num segundo momento, procuraremos identificar a estrutura hierárquica das respostas dos alunos aos itens da prova; por último, e em jeito de conclusão, deixaremos algumas de reflexão sobre a estrutura composicional dos perfis de desempenho académico com o propósito de mantermos em permanência o debate sobre a responsabilidade sociopolítica e escolar de atores e instituições na construção de dinâmicas curriculares e pedagógicas de promoção da melhoria escolar.

### **Análise da dificuldade e da discriminação dos itens**

O índice de dificuldade de um item está relacionado com a percentagem de respostas erradas dadas a esse item. Na teoria clássica, o índice de dificuldade de um item corresponde à percentagem de alunos que não respondem corretamente a esse item. O seu valor varia entre 0 e 100 pontos percentuais (ou entre 0 e 1), sendo considerado tanto mais difícil quanto maior é o seu índice.

Por sua vez, o índice de discriminação de um item traduz a diferença de sucesso, nesse item, entre os alunos com melhor classificação na prova e os alunos com classificações mais baixas. Os valores de discriminação que se apresentam a seguir traduzem a diferença entre os índices de sucesso dos alunos com classificações mais elevadas pelo critério ‘ordem percentual  $\geq 75\%$ ’ e os índices de sucesso dos alunos do grupo com classificações mais baixas pelo critério ‘ordem percentual  $\leq 25\%$ ’, em cada item da respetiva área curricular. Um índice de discriminação baixo, digamos que igual ou inferior a 0.25, indica que os dois grupos de alunos obtiveram percentagens semelhantes de respostas corretas, no respetivo item. Pelo contrário, índices elevados (igual ou superior a 0.75) revelam diferenças acentuadas entre os índices de sucesso dos alunos melhores classificados na prova e os dos piores classificados.

No diagrama de dispersão projetam-se as 28 questões da prova, através das suas coordenadas referenciadas aos índices de dificuldade e de discriminação, com especificação das respetivas áreas curriculares.

---

<sup>2</sup> Na presente prova não há lugar a respostas e pontuações politómicas (pontuações intermédias) mas exclusivamente dicotómicas, uma vez que as possibilidades de resposta às diversas questões se apresentavam na forma de escolha múltipla ou de verdadeiro ou falso e gerando, conseqüentemente, respostas absolutamente certas ou erradas e a que se fez corresponder pontuações de um ou zero, respetivamente.

---

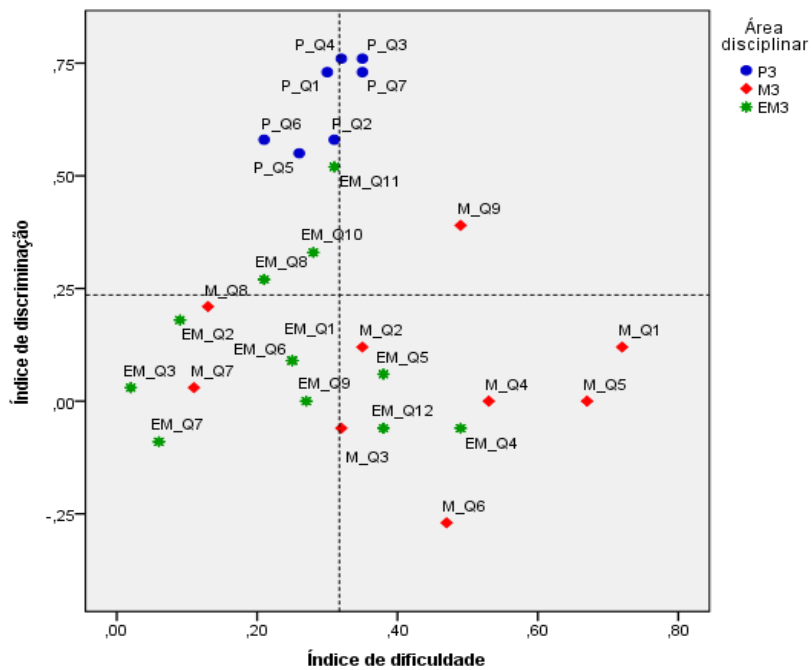


Figura 1: Diagrama de dispersão dos itens da PGD\_K3

As linhas tracejadas na vertical e na horizontal representam as respectivas médias dos índices de dificuldade e de discriminação da totalidade das questões da prova e dividem o diagrama em quatro quadrantes. A dificuldade dos itens aumenta da esquerda para a direita e a discriminação de baixo para cima, pelo que, quanto mais as posições relativas dos itens se aproximarem do canto superior direito do diagrama, tanto maior o seu grau de dificuldade e de discriminação e, do canto inferior esquerdo, tanto menor a sua dificuldade e discriminação.

Os resultados mostram:

- um índice médio de dificuldade inferior a 50%, quer na área curricular do Português, quer da Matemática, quer do Estudo do Meio;
- Existem apenas três itens (questões) com um índice de dificuldade superior a 50% (M\_Q1, M\_Q2 e M\_Q3) e todas na área curricular da Matemática;
- O Estudo do Meio é a área curricular que concentra maior número de questões com menor dificuldade (oito em doze) tomando como referência a média global de dificuldade da prova;
- Em termos de discriminação, no quadro do critério definido, apenas na área do Português se observa a existência de duas questões com índices de discriminação superiores a 0,75 (P\_Q3 e P\_Q4), sendo que, oito dos nove itens de Matemática e oito dos doze de Estudo do Meio, registam índices de discriminação inferiores a 0,25 e, destes, cinco registam mesmo valores negativos de discriminação, significando que, quer os alunos com melhor pontuação na respetiva prova, quer os alunos com piores resultados, obtiveram percentagens semelhantes de respostas corretas.

A observação da posição relativa das 28 questões no diagrama de dispersão sugere a existência de duas 'nuvens' posicionais. Uma, concentrando questões nos quadrantes ímpares ou relativamente próximas dos eixos, mostrando a tendência de concentração das questões ao longo de uma trajetória em diagonal e indiciando a existência de uma elevada correlação positiva entre ambos os índices, tendendo a ser os índices de discriminação dos itens da prova tanto mais elevados, quanto maior o respetivo índice de dificuldade. Sobressai ainda da projeção gráfica uma segunda 'nuvem' de itens, cujas posições relativas ocupam o segundo quadrante mas em zonas relativamente afastadas dos eixos, sugerindo que, ao contrário das questões da primeira nuvem, a um maior grau de dificuldade dos itens que a compõem não corresponde um maior grau de discriminação.

Por outro lado, a aplicação do modelo de regressão pelos mínimos quadrados e o apuramento do respetivo coeficiente de determinação permite confirmar que apenas na área curricular de Português a variação dos índices de discriminação é significativamente explicada pela variação dos índices de dificuldade ( $R^2=0,56$ ), ao contrário das áreas da Matemática e do Estudo do Meio que registam coeficientes de determinação muito baixos.

Em síntese, conclui-se que em termos de qualidade da prova global diagnóstica, apenas a componente de Português regista um bom índice global de qualidade.

### **Análise da estrutura das respostas às questões da prova global diagnóstica K3**

A identificação dos alinhamentos sequenciais das respostas na produção da pontuação global na prova diagnóstica K3 foi realizada com recurso ao algoritmo (EXHAUSTIVE) CHAID (*Chi-square Automatic Interaction Detector*), um método estatístico adequado para identificar, classificar e explorar estruturas complexas (Breyman *et al.*, 1984; Pestana e Gageiro, 2009; IBM-SPSS, 2012).

Na sua base está a criação de uma estrutura hierárquica – designada de árvore de classificação ou árvore de regressão, consoante a variável de resposta é, respetivamente, quantitativa ou qualitativa –, decorrente do fracionamento sucessivo dos dados em subconjuntos mutuamente exclusivos gerando nós intermédios ou nós terminais. A partição dos níveis da árvore faz-se por ordem decrescente de importância das questões. A questão estatisticamente mais significativa define a primeira partição gerando o primeiro nível de profundidade e é avaliada a possibilidade da subdivisão de cada uma das suas categorias de resposta gerando 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 das respostas dos alunos aos itens da prova.

A solução gerada projetou uma árvore com vinte e nove nós, quinze dos quais terminais, distribuídos por cinco níveis de profundidade (ver Figura 2 e Apêndice 1) e que usaremos na

caracterização dos perfis de resposta obtidos nas ramificações sequenciais do nó raiz aos nós terminais da estrutura de classificação gerada.

A estimativa da qualidade do modelo foi obtida mediante validação cruzada, uma alternativa sugerida quando o número de observações é inferior a mil (Pestana e Gageiro, 2009). A proporção de variância explicada pelo modelo é de 85% e, após validação cruzada, de 53% (ver tabela *Risk* em Apêndice 1), sugerindo tais valores que a informação da validação não contraria a aplicação do modelo a outras amostras do mesmo universo.

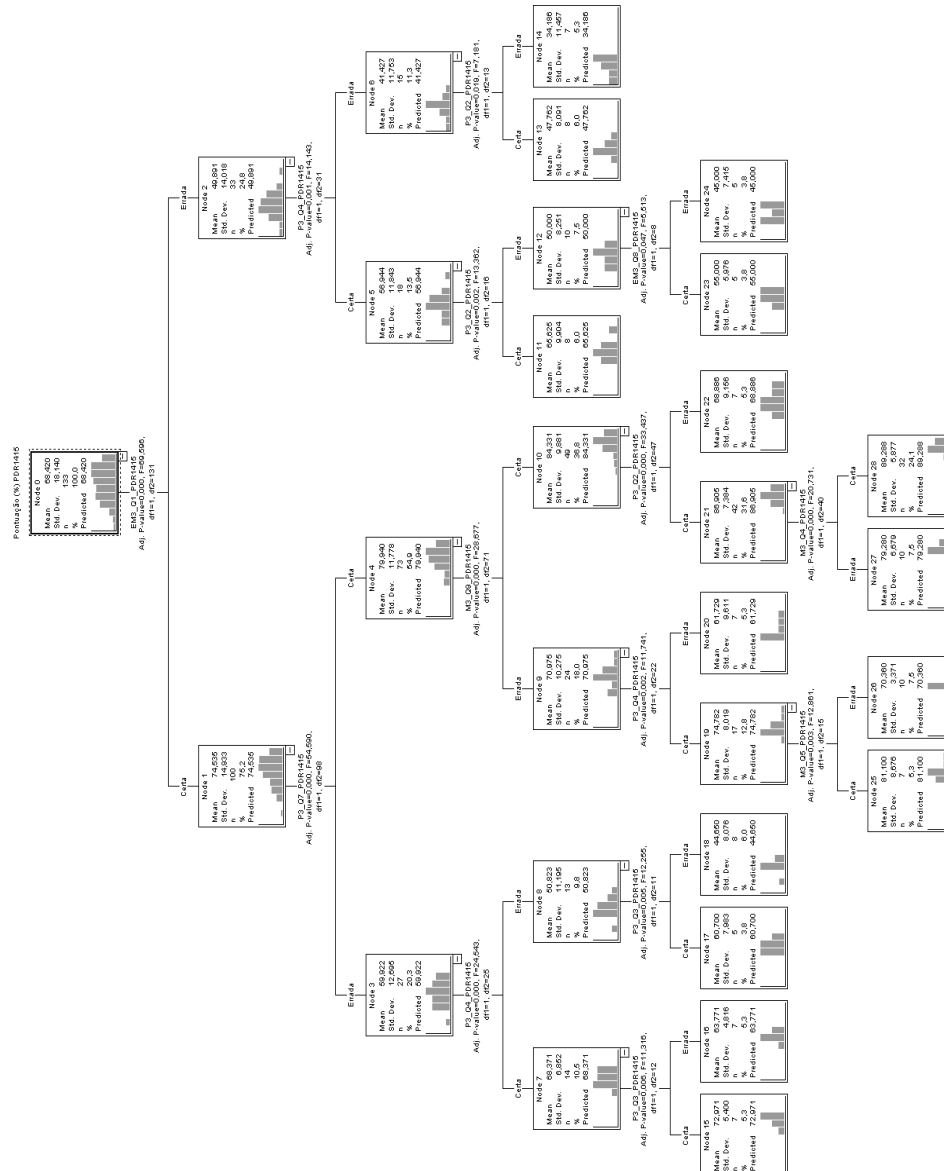


Figura 2: Estrutura hierárquica das respostas da PGD\_K3

Observando a solução arbórea projetada na figura 2, as respectivas especificações e resultados constantes do Apêndice 1, constata-se que:

- i) As estatísticas do nó raiz (nó 0) mostram que a pontuação global obtida pelos 133 alunos que realizaram a prova foi de 68,42%, com um desvio padrão de 18,14 sugerindo tais valores a existência de alguma disparidade dos resultados obtidos pelos alunos na prova (coeficiente de variação = 26,5%);

- ii) Das 28 questões da prova, face aos critérios de análise estabelecidos, apenas nove e algumas das quais recursivamente, se revelam estatisticamente significativas com uma probabilidade de erro de tipo I ( $\alpha$ ) inferior a 0,05 na explicação dos perfis sequenciais de resposta e respetiva estrutura hierárquica da prova, pelo que se conclui que nem todas as questões da prova tiveram a mesma importância na configuração final da estrutura hierárquica. Ter acertado ou errado a questão EM\_Q<sub>1</sub> não é propriamente o mesmo que ter acertado ou errado EM\_Q<sub>2</sub>, por exemplo. Enquanto EM\_Q<sub>1</sub> emerge como a principal determinante na segmentação da amostra geral em dois grandes ramos, EM\_Q<sub>2</sub> nem sequer emerge na solução estrutural obtida com cinco níveis de profundidade. Do mesmo modo, as questões de Português foram mais relevantes no processo de segmentação e desenvolvimento da estrutura do que as questões de Matemática ou de Estudo do Meio, tendo as primeiras interferido em nove segmentações estruturais e as segundas e terceiras apenas em seis.
- iii) É também na componente de Português que se revelam questões recursivas, casos de P\_Q<sub>4</sub>, P\_Q<sub>2</sub> e P\_Q<sub>3</sub>, determinantes em várias segmentações e em diversos níveis de profundidade e gerando estruturas sequenciais recursivas de respostas.

Apresentam-se de seguida, por ordem decrescente, as médias das pontuações percentuais de cada nó terminal. Para facilidade de visualização dos pares de nós de maior contraste face à amostra geral (Verdasca, 2014), optou-se por projetar graficamente (figura 3) a posição de cada nó relativamente ao nó raiz (base 100) com base nos respetivos *index*, indicativos do grau de concentração de respostas certas dos respetivos nós.

Da projeção gráfica e da tabela que lhe corresponde, sobressaem como pares de maior contraste os pares (28, 14), (25, 18) e (27, 24) com acréscimos de pontuações médias dos primeiros elementos de cada par que superam em 30%, 19% e 16%, respetivamente, o desempenho médio global da amostra geral e com défices de pontuação dos segundos elementos dos respetivos pares de -50%, -35% e -34%. Em termos de diferenciais médios absolutos de pontuações entre os elementos dos pares as diferenças são de 55%, 36% e 34%, respetivamente para o primeiro, segundo e terceiro pares.

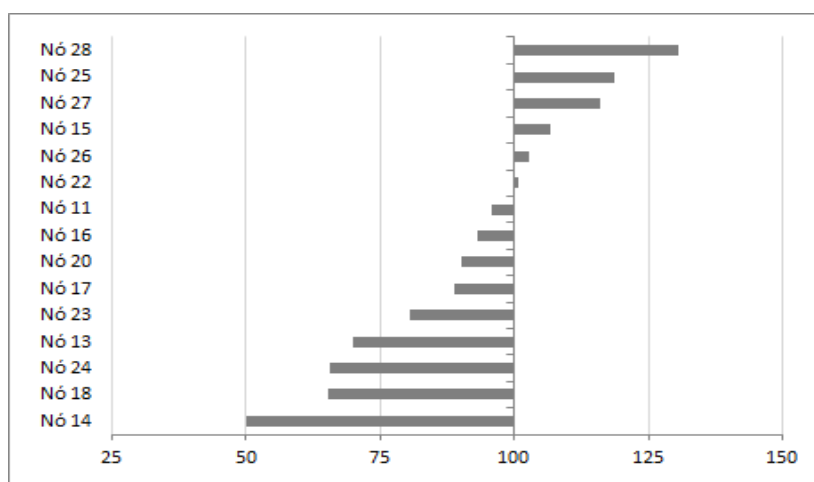


Figura 3: Index dos nós terminais da estrutura hierárquica da PGD\_K3

A análise da composição e das sequências estruturais associadas a cada um dos nós dos três pares de maior contraste permite constatar que:

- Primeiro par (28, 14)

. *Nó 28*: Para além de ser o nó com maior concentração de respostas certas, é também o nó com maior peso amostral, representando com os seus 32 casos quase um quarto dos alunos que realizaram a PGD\_K3 (24,1%) e tendo como perfil de resposta a estrutura sequencial de resposta certa às questões: EM\_Q1, P\_Q7, M\_Q9, P\_Q2 e M\_Q4, ou seja, [(c) EM\_Q1 -> (c) P\_Q7 -> (c) M\_Q9 -> (c) P\_Q2 -> (c) M\_Q4];

. *Nó 14*: É o nó com mais baixa concentração de respostas certas. Integra sete alunos que representam 5,3% da amostra, com uma média de pontuações de apenas 34,2%. Como perfil caracterizador deste nó emerge uma estrutura sequencial de respostas erradas aos itens EM\_Q1, P\_Q4 e P\_Q2, ou seja, [(e) EM\_Q1 -> (e) P\_Q4 -> (e) P\_Q2 ];

- Segundo par (25, 18)

. *Nó 25*: É o segundo nó com maior concentração de respostas certas, sendo constituído por sete alunos que representam aproximadamente 5% da amostra, tendo como perfil de respostas a seguinte estrutura sequencial: [(c) EM\_Q1 -> (c) P\_Q7 -> (e) M\_Q9 -> (c) P\_Q2 -> (c) M\_Q4];

. *Nó 18*: Com a segunda pior média de pontuações (44,65%), os oito alunos que o constituem representam 6% da amostra e a estrutura sequencial de respostas que o caracteriza a seguinte: [(c) EM\_Q1 -> (e) P\_Q7 -> (e) M\_Q4 -> (e) P\_Q3];

- Terceiro par (27,24)

. *Nó 27*: É o terceiro nó com maior concentração de respostas certas (79,28%) e elevada homogeneidade de resultados (cv=8,4%), sendo constituído por dez alunos com um peso amostral de 7,5% e tendo como perfil de respostas a seguinte estrutura sequencial: [(c) EM\_Q1 -> (c) P\_Q7 -> (c) M\_Q9 -> (c) P\_Q2 -> (e) M\_Q4];

. *Nó 24*: É constituído por 5 alunos, com um peso amostral de apenas 4%. É o terceiro nó com menor concentração de respostas certas (44,65%), com baixa heterogeneidade de pontuações (cv=16%), apresentando a seguinte estrutura sequencial de respostas: [(e) EM\_Q1 -> (c) P\_Q4 -> (e) P\_Q2 -> (e) EM\_Q8].

### **Composição sociodemográfica dos perfis de desempenho**

No âmbito dos atributos sociodemográficos dos alunos que realizaram a prova, seleccionámos algumas das variáveis comumente descritas na literatura como fortemente condicionantes do desempenho académico dos alunos e das escolas<sup>3</sup>, caso do capital escolar dos pais, do apoio social escolar, da tipologia familiar do aluno e do local de residência do aluno implicar ou

---

<sup>3</sup> Referindo-nos apenas a alguns dos estudos e relatórios institucionais mais recentes e de abrangência nacional, veja-se, por exemplo: CNE, 2013, 2014; Canto e Castro *et al*, 2014; Justino *et al*, 2014. No que concerne à turma, enquanto unidade organizacional de análise, a literatura é, desta perspectiva, bem mais escassa e os estudos existentes são distanciados no tempo e, de algum modo, circunscritos a um determinado território (Verdasca, 2002).

não deslocação diária deste para frequentar a escola. As duas primeiras variáveis são indicativas do contexto sociocultural e económico familiar dos alunos; a terceira variável, da vivência no seio de família nuclearizada ou não; a quarta variável, é indicativa de um contexto de habitat de maior ou menor ruralidade.

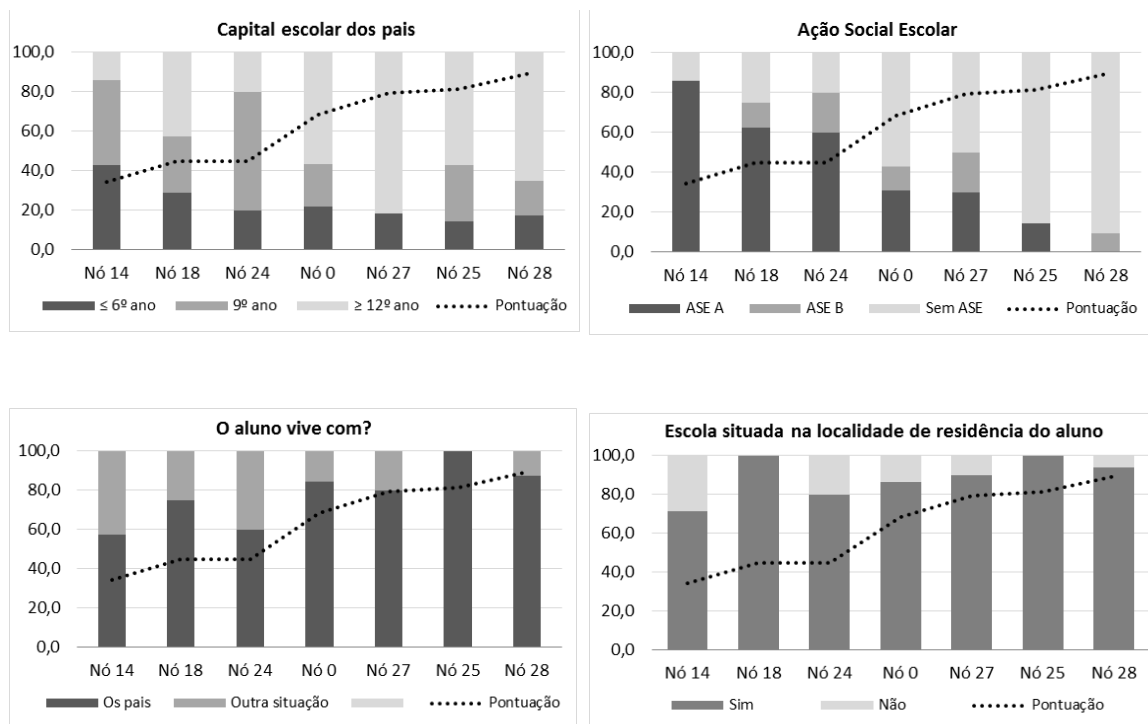


Figura 4: Perfis de desempenho e estrutura composicional sociodemográfica

Mergulhando na configuração contextual de cada um dos nós que integram os três pares de nós de maior contraste, sobressaem claramente linhas tendenciais de diferenciação e segmentação em termos dos níveis de proficiência académica na realização da prova. Esta situação é visível para quaisquer dos fatores contextuais em análise como se se depreende da estrutura de composição de cada um dos perfis (nós) constituídos. Perante perfis com maiores densidades de alunos cujos pais têm graus de escolaridade mais baixos, de alunos com maior grau de carência socioeconómica, de alunos de famílias não nuclearizadas, de alunos com proveniência de habitats de maior ruralidade e isolamento, tanto menor foi o seu índice de desempenho na prova e vice-versa.

Em jeito de síntese e de conclusão, estamos perante sinais claramente reveladores de uma estrutura de desempenho académico dos alunos que continua a ser fortemente marcada pelas próprias estruturas e lógicas de dominância cultural e social, sintomático de processos escolares aparentemente mais permeáveis e reféns de fatores contextuais exógenos do que das dinâmicas educativas internas (Verdasca, 2015) e que importa fazer notar, tendo em conta a responsabilidade sociopolítica e organizacional da escola (pública) e da comunidade na construção e organização de respostas e práticas curriculares e pedagógicas com vista à melhoria da qualidade educativa e à promoção da igualdade de oportunidades e da equidade.



## Referências

- Breiman, L., Friedman, J., Olshen, R. e Stone, C. (1984). *Classification and Regression Trees*. Wadsworth International Group. Belmont, CA.
- Castro, L., Santos, J., Pereira, T. e Vitorino, A. (2014). *Modelos para comparação estatística dos resultados académicos em escolas de contexto análogo: Painel de dados para apoio à avaliação externa das escolas*. Lisboa: MEC-DGEEC.
- CNE (2013). *Estado da Educação 2012*. Lisboa: CNE.
- CNE (2014). *Estado da Educação 2013*. Lisboa: CNE.
- IBM SPSS (2012). *Decision Trees 21*. (<ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/21.0/en/client/Mauals/IBMSPSSDecisionTrees.pdf>)
- Justino, D., Pascueiro, L., Franco, L., Santos, R., Almeida, S. e Batista, S. (2014). *Atlas da Educação – Contextos sociais e locais do sucesso e insucesso: Portugal 1991-2012*. Lisboa: CESNOVA.
- Pestana, M. 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.
- Verdasca, J. (2002). *Desempenho escolar, dinâmicas de evolução e elementos configuracionais estruturantes. Os casos do 2.º e 3.º ciclos do ensino básico nos municípios de Évora e Portel*. Évora: Universidade de Évora.
- Verdasca, J. (2014). *Instrumentos metodológicos e de análise exploratória de dados em educação. (Coletânea de textos e documentos de apoio à disciplina de Administração e Gestão Educacional)*. Évora: Universidade de Évora (polic.).
- Verdasca, J. (2015). *A turma como unidade socio-organizacional de análise: contributos para a construção de um roteiro de monitorização e (auto)regulação escolar (Texto de enquadramento ao modelo 3Q\_CDR)*. CIEP-UEvora (polic.).

## Apêndice 1

### Prova global diagnóstica K3

---

**Model Summary**

Specifications	Growing Method	EXHAUSTIVE CHAID
	Dependent Variable	Pontuação (%) PDR1415
	Independent Variables	M3_Q1_PDR1415, M3_Q2_PDR1415, M3_Q3_PDR1415, M3_Q4_PDR1415, M3_Q5_PDR1415, M3_Q6_PDR1415, M3_Q7_PDR1415, M3_Q8_PDR1415, M3_Q9_PDR1415, P3_Q1_PDR1415, P3_Q2_PDR1415, P3_Q3_PDR1415, P3_Q4_PDR1415, P3_Q5_PDR1415, P3_Q6_PDR1415, P3_Q7_PDR1415, EM3_Q1_PDR1415, EM3_Q2_PDR1415, EM3_Q3_PDR1415, EM3_Q4_PDR1415, EM3_Q5_PDR1415, EM3_Q6_PDR1415, EM3_Q7_PDR1415, EM3_Q8_PDR1415, EM3_Q9_PDR1415, EM3_Q10_PDR1415, EM3_Q11_PDR1415, EM3_Q12_PDR1415
	Validation	Cross Validation
	Maximum Tree Depth	5
	Minimum Cases in Parent Node	10
	Minimum Cases in Child Node	5
Results	Independent Variables Included	EM3_Q1_PDR1415, P3_Q7_PDR1415, P3_Q4_PDR1415, P3_Q3_PDR1415, M3_Q9_PDR1415, M3_Q5_PDR1415, P3_Q2_PDR1415, M3_Q4_PDR1415, EM3_Q8_PDR1415
	Number of Nodes	29
	Number of Terminal Nodes	15
	Depth	5

**Gain Summary for Nodes**

Node	N	Percent	Mean
28	32	24,1%	89,288
25	7	5,3%	81,100
27	10	7,5%	79,280
15	7	5,3%	72,971
26	10	7,5%	70,360
22	7	5,3%	68,886
11	8	6,0%	65,625
16	7	5,3%	63,771
20	7	5,3%	61,729
17	5	3,8%	60,700
23	5	3,8%	55,000
13	8	6,0%	47,762
24	5	3,8%	45,000
18	8	6,0%	44,650
14	7	5,3%	34,186

Growing Method: EXHAUSTIVE CHAID  
 Dependent Variable: Pontuação (%) PDR1415

**Risk**

Method	Estimate	Std. Error
Resubstitution	48,151	6,396
Cross-Validation	153,697	23,404

Growing Method: EXHAUSTIVE CHAID  
 Dependent Variable: Pontuação (%) PDR1415