



UNIVERSIDADE DE ÉVORA

ESCOLA DE CIÊNCIAS SOCIAIS

Mestrado em Ciências da Educação

Especialização em Avaliação Educacional

Dissertação

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

Maria Florentina Alves Gomes Lopes de Andrade

Orientadora:

Professora Doutora Isabel José Botas Bruno Fialho

Évora, 2012

UNIVERSIDADE DE ÉVORA

ESCOLA DE CIÊNCIAS SOCIAIS

Mestrado em Ciências da Educação

Especialização em Avaliação Educacional

Dissertação

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

Maria Florentina Alves Gomes Lopes de Andrade

Orientadora:

Professora Doutora Isabel José Botas Bruno Fialho

AGRADECIMENTOS

Este trabalho não teria sido desenvolvido e concluído sem os preciosos contributos de alguns familiares, colegas, amigos e docentes. A breve referência aos seus nomes não expressará toda a gratidão sentida, mas deixa registado o meu reconhecimento.

À minha Orientadora, Professora Doutora Isabel Fialho, pelas imprescindíveis, sábias e eficazes orientações. Sem o seu conhecimento, esta investigação não teria passado do plano das intenções.

À minha amiga Su, pelo seu ombro, pelos seus ouvidos pacientes, por toda a ajuda que deu num assunto do meu interesse, usando o seu tempo, habilidade informática e sabedoria.

À Bellzé, por uma amizade já tão longa e pelo incentivo e apoio constantes.

Ao Hugo, pela disponibilidade demonstrada na ajuda do desenvolvimento empírico deste trabalho e pela partilha da sua experiência em investigação.

Aos Diretores das Escolas que gentilmente se disponibilizaram para a cedência dos documentos necessários e facilitaram a aplicação dos questionários.

Aos meus *amigos de sempre*, pela força, alento, motivação, paz e conforto que todos os dias me ajudaram na caminhada para a concretização desta obra.

Ao Paulo, companheiro de todos os dias, pela sua sensibilidade, compreensão e força, sempre presentes, ao longo deste trabalho.

RESUMO

Enquadrado na problemática dos exames nacionais, este estudo teve como principal objetivo verificar se a tipologia dos itens interfere nos resultados do exame nacional de Biologia e Geologia. Procurou-se, também, conhecer as perceções dos alunos relativamente ao grau de dificuldade dos itens dos exames nacionais realizados em 2011 e as razões dessas dificuldades.

A recolha de dados foi feita por aplicação de um questionário a 106 alunos de quatro escolas do distrito de Évora e com base em documentos do GAVE, onde constam os resultados por item, de cada escola.

Os dados revelam que os itens de resposta restrita são os que se apresentam com resultados médios mais baixos, sendo também percecionados como os mais difíceis. Os melhores resultados correspondem aos itens de associação/correspondência. As principais razões de dificuldade prendem-se com relacionar os conhecimentos com as informações fornecidas (textos/tabelas/gráficos/esquemas) e a complexidade da matéria.

Palavras-chave: Biologia e Geologia; Exames; Tipologia de item; Nível de dificuldade.

National Exams: the influence of the type of items in the results of Biology and Geology exam.

ABSTRACT

Framed on the issue of national exams, this study aimed to verify if the types of items affect the results of the national exam of Biology and Geology. We tried, also, to know the perceptions of students in relation to the degree of difficulty of the items of national exams conducted in 2011, and the reasons for these difficulties.

Data collection was performed by administering a questionnaire to 106 students from four schools in the district of Évora and based on documents of GAVE, which contains the results by item, for each school.

Data reveal that restricted response items had lower average results, and are also the most difficult in students' perceptions. The best results correspond to items of association/correspondence. The main reasons of difficulty have to do with relate knowledge with the information provided (text/tables/charts/diagrams) and the complexity of the subject.

Keywords: Biology and Geology; Exams, item type, difficulty level.

ÍNDICE GERAL

LISTA DE ABREVIATURAS / SIGLAS	viii
ÍNDICE DE FIGURAS	ix
ÍNDICE DE QUADROS	x
ÍNDICE DE TABELAS	xi
ÍNDICE DE APÊNDICES	xiv
INTRODUÇÃO	1
CAPÍTULO 1 – A AVALIAÇÃO EXTERNA DAS APRENDIZAGENS: OS EXAMES	5
1.1. A avaliação ao longo da história	5
1.1.1. Da Antiguidade às primeiras contestações à avaliação	5
1.1.2. A Docimologia: aparecimento e evolução	8
1.1.3. Do paradigma docimológico a uma avaliação pertinente	12
1.2. Os exames: suas funções	15
1.3. Os exames: seus efeitos	17
1.4. Os exames do ensino secundário: suas características	20
CAPÍTULO 2 – A AVALIAÇÃO EXTERNA DAS APRENDIZAGENS EM PORTUGAL ..	22
2.1. Resenha histórica dos exames em Portugal.....	22
2.2. Entidades envolvidas nos exames em Portugal.....	32
2.3. A problemática dos exames em Portugal	36
2.3.1. <i>Rankings e accountability</i>	41
2.4. As informações dadas, nos últimos anos, pelos exames.....	46
CAPÍTULO 3 – A CONSTRUÇÃO DE UM EXAME	52
3.1. Das modalidades de avaliação às técnicas de testagem. Os testes	52

3.1.1. Tipos de testes	54
3.1.2. As qualidades de um bom exame (teste standardizado)	57
3.1.3. Tipos de itens	72
3.1.4. Considerações sobre as vantagens e desvantagens do uso dos itens objetivos e dos itens subjetivos	78
3.1.5. O número de itens a propor num teste/exame	80
3.1.6. A ordenação dos itens numa prova.....	81
3.1.7. Determinação do formato de uma prova	81
3.1.8. Procedimento do GAVE na divulgação das características da prova de exame nacional do ensino secundário na disciplina de Biologia e Geologia, código 702, entre 2006 e 2011	82
3.1.9. Análise das provas de exame nacional de Biologia e Geologia, código 702, entre 2006 e 2011.....	84
3.1.10. O formato dos exames nacionais entre 2006 e 2011	88
CAPÍTULO 4 – METODOLOGIA DA INVESTIGAÇÃO	90
4.1. Opções metodológicas e sua justificação	90
4.1.1. Natureza da investigação.....	90
4.1.2. Fases da investigação	91
4.2. Critérios da escolha das escolas em estudo.....	92
4.3. Instrumento de recolha de dados: Inquérito por questionário	93
4.5. Recolha e tratamento dos dados.....	98
4.5.1. Caracterização dos alunos.....	99
CAPÍTULO 5 – APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS.....	103
5.1. Resultados obtidos, pelos alunos internos, na prova de exame nacional de Biologia e Geologia (Código 702), em 2011	103
5.2. Perceção dos alunos quanto ao grau de dificuldade dos itens	109

5.2.1. Resultado das variáveis «Item de escolha múltipla», «Item de ordenação», «Item de resposta restrita (aberta)» e «Item de associação / correspondência»	109
5.3. Análise comparativa dos resultados obtidos no exame com as perceções dos alunos quanto ao grau de dificuldade dos itens	115
5.4. Perceção dos alunos quanto às razões que contribuem para o grau de dificuldade das perguntas (itens)	115
5.4.1. – Resultado das variáveis para «Item de escolha múltipla»	116
5.4.2. – Resultado das variáveis para «Item de ordenação»	124
5.4.3. – Resultado das variáveis para «Item de resposta restrita»	130
5.4.4. – Resultado das variáveis para «Item de associação»	137
CONSIDERAÇÕES FINAIS	144
REFERÊNCIAS BIBLIOGRÁFICAS	149
LEGISLAÇÃO	157
APÊNDICES	158

LISTA DE ABREVIATURAS / SIGLAS

DES	Departamento do Ensino Secundário
DGE	Direção Geral de Educação
EMEC	Editorial do Ministério da Educação e Ciência
GAVE	Gabinete de Avaliação Educacional
IGEC	Inspeção-Geral da Educação e Ciência
JNE	Júri Nacional de Exames
LBSE	Lei de Bases do Sistema Educativo
PGA	Prova Geral de Acesso
QEM	Questionários de escolha múltipla

ÍNDICE DE FIGURAS

Figura 1. Distribuição dos alunos participantes por sexo.	99
Figura 2. Distribuição dos alunos participantes por idade.	101
Figura 3. <i>Output</i> do teste de Kruskal-Wallis	114
Figura 4. <i>Output</i> das comparações múltiplas	114

ÍNDICE DE QUADROS

Quadro 1. Critérios de seriação dos candidatos ao ensino superior, de acordo com o Decreto-Lei n.º 189/92, de 3 de setembro.....	28
Quadro 2. Prováveis moderadores inválidos de dificuldade	66
Quadro 3. Sugestões para aumentar ou diminuir a dificuldade dos itens	67
Quadro 4. Tipologia de itens das provas de Biologia e Geologia (1.ª fase), entre 2006 e 2011	85

ÍNDICE DE TABELAS

Tabela 1. Número de questionários distribuídos em cada escola e respetiva taxa de retorno	98
Tabela 2. Distribuição do sexo dos alunos participantes por escola.....	100
Tabela 3. Distribuição das idades dos alunos participantes por escola.....	101
Tabela 4. Distribuição dos alunos participantes por fase de realização da prova de Biologia e Geologia.....	102
Tabela 5. Resultados médios e cotação nula por tipo de item: Escola A.....	104
Tabela 6. Resultados médios e cotação nula por tipo de item: Escolas B+C	104
Tabela 7. Resultados médios e cotação nula por tipo de item: Escola D	105
Tabela 8. Resultados médios e cotação nula para os itens de escolha múltipla	105
Tabela 9. Resultados médios e cotação nula para os itens de ordenação	106
Tabela 10. Resultados médios e cotação nula para os itens de resposta restrita	106
Tabela 11. Resultados médios e cotação nula para os itens de correspondência/ associação	107
Tabela 12. Valores médios e cotação nula, para cada tipo de item.....	107
Tabela 13. Respostas e respetivas percentagens, por tipo de item, para cada escola...	110
Tabela 14. Média e moda por tipo de item e por escola.....	111
Tabela 15. Teste à normalidade da amostra.....	112
Tabela 16. Teste à homogeneidade de variâncias.....	113
Tabela 17. Comparação dos resultados médios reais e o grau de dificuldade percebido pelos alunos	115
Tabela 18. Razões que contribuem para o grau de dificuldade dos itens de escolha múltipla	116
Tabela 19. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de escolha múltipla	117

Tabela 20. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de escolha múltipla – Escola A	119
Tabela 21. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de escolha múltipla – Escolas B+C.....	121
Tabela 22. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de escolha múltipla – Escola D	122
Tabela 23. Razões que contribuem para o grau de dificuldade dos itens de ordenação	124
Tabela 24. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de ordenação.....	125
Tabela 25. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de ordenação – Escola A.....	126
Tabela 26. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de ordenação – Escolas B+C.....	127
Tabela 27. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de ordenação – Escola D.....	128
Tabela 28. Razões que contribuem para o grau de dificuldade dos itens de resposta restrita	130
Tabela 29. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de resposta restrita	131
Tabela 30. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de resposta restrita – Escola A	133
Tabela 31. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de resposta restrita – Escolas B+C.....	135
Tabela 32. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de resposta restrita – Escola D	136
Tabela 33. Razões que contribuem para o grau de dificuldade dos itens de associação	138
Tabela 34. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de associação.....	138

Tabela 35. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de associação – Escola A.....	140
Tabela 36. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de associação – Escolas B+C.....	141
Tabela 37. Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de associação – Escola D.....	142

ÍNDICE DE APÊNDICES

Apêndice A. Questionário (Versão final)	159
Apêndice B. Pedido de autorização à DGIDC para aplicação do questionário	163
Apêndice C. Autorização da DGIDC para aplicação do questionário	164
Apêndice D. Autorização do Diretor da Escola A para aplicação do questionário	165
Apêndice E. Autorização do Diretor da Escola B para aplicação do questionário	166
Apêndice F. Autorização do Diretor da Escola C para aplicação do questionário	167
Apêndice G. Autorização do Diretor da Escola D para aplicação do questionário	168

INTRODUÇÃO

1. Justificação do Estudo

“A investigação realizada em Portugal no domínio da avaliação das aprendizagens parece ser relativamente escassa” (Fernandes, 2009).

Tendo em conta que a problemática da avaliação externa das aprendizagens dos alunos já vem de há vários anos, considera-se oportuno e necessário analisar e saber o que se passa com este tipo de avaliação nas suas variadas dimensões para que se possa desta forma contribuir para melhorar o sistema educativo.

Neste âmbito, parece fazer sentido desenvolver estudos relacionados com a avaliação externa das aprendizagens, nomeadamente sobre os exames nacionais. Os exames nacionais do ensino secundário têm constituído um dos temas mais polémicos e centro de grandes debates, neles participando professores, investigadores, estudiosos e a sociedade em geral, o que fomentou o interesse pelo desenvolvimento de um estudo empírico fundamentado em conceitualizações teóricas que possam conduzir a uma análise do que se passa, com base na auscultação e perceção dos próprios alunos. É importante que se proceda a investigações que se possam conjugar com esforços associados às práticas, que permitam encontrar linhas orientadoras devidamente fundamentadas e que ajudem a melhorar o sistema educativo.

2. Identificação do problema

No âmbito dos debates em que se confrontam diferentes perspetivas de educação e de sistemas educativos, na sequência de conversas no seio das escolas, com professores, e principalmente com alunos, foi sendo consolidada a perceção de que os exames nacionais do ensino secundário parecem ser vistos como um tipo de avaliação com grandes impactos aos mais variados níveis: nas vidas pessoal e académica dos alunos, naquilo que é ensinado e na forma como é desenvolvido e trabalhado o currículo, nas

práticas de ensino adotadas pelos professores visando a preparação dos alunos para o exame, e na forma como as próprias escolas se organizam.

3. Objetivos da investigação

Com esta investigação pretendeu-se, num primeiro momento, verificar quais são as causas que levam ao fraco desempenho dos alunos, em determinados tipos de itens, na prova de exame nacional do ensino secundário de Biologia e Geologia¹ (Código 702), no ano de 2011.

Face à problemática do estudo definiram-se os seguintes objetivos:

- (I) saber quais os resultados obtidos no ano de 2011, em diferentes tipos de itens;
- (II) conhecer as perceções que alunos do 12.º ano de escolaridade têm quanto ao grau de dificuldade de cada tipo de item e compará-las com os resultados obtidos nas provas de exame;
- (III) identificar as razões que contribuem para o grau de dificuldade das perguntas (itens) dos exames nacionais de Biologia e Geologia.

Neste contexto, levantaram-se algumas questões que, eventualmente, possam ser responsáveis por esses resultados:

Poderão os resultados obtidos nos diferentes tipos de itens da prova ser influenciados:

- Pela tipologia do item?
- Pelo número de itens de um determinado tipo de itens?
- Pelo grau de complexidade das competências testadas?
- Por obstáculos encontrados pelos alunos à mobilização de conceitos, modelos e teorias?
- Por dificuldades de explicação de contextos em análise?

¹ Sendo a disciplina de Biologia e Geologia bienal, a avaliação sumativa externa (exame) é realizada no 11.º ano. (Cfr. n.º 5 do artigo 17.º da Portaria n.º 244/2011, de 21 de junho - *Os exames finais nacionais realizam-se nos termos definidos no artigo 11.º do Decreto -Lei n.º 74/2004, de 26 de Março, alterado pelos Decretos -Leis n.º 24/2006, de 6 de Fevereiro, 272/2007, de 26 de Julho, 4/2008, de 7 de Janeiro, e 50/2011, de 8 de Abril, e incidem [...] sobre os programas relativos à totalidade dos anos de escolaridade em que a disciplina é leccionada [...].*

Considera-se que a presente investigação poderá dar uma resposta a um interesse individual, mas também constituir-se como uma base de trabalho para o desenvolvimento de futuros estudos no âmbito da avaliação sumativa externa das aprendizagens.

4. Estrutura do trabalho

O estudo que ora se apresenta encontra-se organizado por vários capítulos em que se procurou dar unidade de conceção entre a teorização e a observação empírica, evitando a sua divisão em partes (Azevedo, 2009).

Desta forma, a conceção da sua estrutura ficou assim definida:

- ❖ **Introdução**

Apresenta o problema objeto de estudo, que de forma simples e objetiva procura justificar, identificar e caracterizar a investigação.

- ❖ **Capítulo 1 – A avaliação externa das aprendizagens: os exames**

É feita uma referência às questões relacionadas com a avaliação ao longo da história com destaque para o aparecimento e desenvolvimento da docimologia. Posteriormente faz-se uma abordagem aos exames, considerando-se as suas funções, os seus efeitos e as suas características.

- ❖ **Capítulo 2 – A avaliação externa das aprendizagens em Portugal**

Neste capítulo é feita uma resenha histórica dos exames em Portugal seguida de uma abordagem à problemática dos exames, nomeadamente as questões relacionadas com a elaboração dos *rankings*.

Por último, são referenciados alguns estudos sobre os resultados dos exames nacionais que têm fornecido um conjunto de informações sobre os saberes dos alunos portugueses.

- ❖ **Capítulo 3 – A construção de um exame**

São abordados vários tipos de testes e os vários tipos de itens e suas características, bem como as questões relacionadas com a validade e a fiabilidade dos itens e dos exames.

❖ **Capítulo 4 – Metodologia da investigação**

Apresenta-se as opções metodológicas que estiveram subjacentes ao estudo empírico realizado, a natureza e as fases da investigação, os critérios de escolha dos participantes, a opção de escolha do instrumento de recolha de dados – inquérito por questionário: forma; conteúdo; validação do instrumento; estudo piloto ou preliminar efetuado; autorização e aplicação do questionário. Aborda, no tratamento dos dados, os programas informáticos utilizados, a caracterização dos participantes e os procedimentos gerais para análise dos dados.

❖ **Capítulo 5 – Apresentação e discussão dos resultados**

Este capítulo integra a apresentação, a descrição e a interpretação dos resultados obtidos no estudo empírico.

❖ **Considerações finais**

Apresenta-se as conclusões do estudo face aos objetivos definidos, seguidas das limitações encontradas ao longo do mesmo e, por último, sugestões de novas linhas de investigação que poderão vir a ser desenvolvidas em estudos futuros.

❖ **Referências Bibliográficas**

Referencia-se toda a bibliografia consultada e referenciada (livros; artigos científicos; revistas científicas; publicações disponíveis em linha).

❖ **Legislação**

Discrimina-se os diplomas normativo-legais consultados e referenciados no corpo da dissertação.

❖ **Apêndices**

Dado que devem fazer parte integrante do trabalho e não vão ser encadernados num volume em separado (Azevedo, 2009), apresenta-se os elementos que foram utilizados na investigação, mas que não mereceram integrar o corpo central da dissertação.

CAPÍTULO 1 – A AVALIAÇÃO EXTERNA DAS APRENDIZAGENS: OS EXAMES

Tendo em conta que esta investigação se debruça sobre uma prova de exame nacional do ensino secundário - Biologia e Geologia, inserida na avaliação externa das aprendizagens dos alunos, considera-se importante fazer uma referência às questões relativas a este tipo de avaliação, numa sequência temporal, nomeadamente as razões que conduziram ao aparecimento e posterior desenvolvimento da docimologia. Na sequência deste assunto é feita uma abordagem às funções e aos efeitos dos exames, em diferentes países, com destaque para os do ensino secundário, por ser este o âmbito do presente estudo.

1.1.A avaliação ao longo da história

1.1.1. Da Antiguidade às primeiras contestações à avaliação

Desde a Antiguidade que se tem vindo a criar e a usar procedimentos *instrutivos*. Nesse período da História, já os professores utilizavam referentes, embora de forma implícita, para valorizar, diferenciar e selecionar os estudantes.

Na Idade Média, introduziu-se os exames nos meios universitários: eram exames orais e públicos, em presença de um tribunal (júri).

No século XVIII, à medida que aumentava o acesso à educação, acentuou-se a necessidade de comprovar os méritos individuais, e as instituições educativas foram elaborando e introduzindo normas sobre a utilização de exames escritos (Gil, citado em Escudero Escorza, 2003).

Chegado o século XIX, estabeleceram-se os sistemas nacionais de educação e apareceram os diplomas de graduação, tendo surgido os exames do estado. Estes, segundo Barbier (citado em Escudero Escorza, 2003), destinavam-se a comprovar uma preparação específica para satisfazer uma sociedade hierárquica e burocratizada, encaminhando, desta forma, a avaliação para uma função mais administrativa, em que a seleção e a certificação constituíam aspetos centrais do sistema. Como referem Pinto e Santos (2006), “o exame, ou os seus substitutos, são a expressão da ideia de medida, o gesto avaliativo” (p.18), adquirindo um peso significativo socialmente e na vida de cada um. O exame passou a ser um sinónimo de controlo de qualidade.

Entretanto, nos Estados Unidos, em 1845, começou a utilizar-se as primeiras técnicas avaliativas do tipo “teste” escrito, as quais abriram caminho à existência de referentes mais objetivos e explícitos, apesar de pouco fiáveis.

Nos finais do século XIX, em 1897, despertou-se um grande interesse pela medição científica do comportamento humano, o que está relacionado com o movimento renovador das ciências humanas, que assumiram o positivismo das ciências físico-naturais. Em consequência, a avaliação recebeu as mesmas influências que outras disciplinas pedagógicas relacionadas com processos de medição, como a pedagogia experimental e a diferencial (Cabrera, citado em Escudero Escorza, 2003). Esta utilização da medida estaria na base do nascimento da psicometria, no início do século XX.

A atividade avaliativa viu-se, assim, condicionada e impulsionada a nível do seu desenvolvimento por diversos fatores:

- (I) o aparecimento das *correntes filosóficas positivistas e empíricas*, que apoiavam a observação e a experimentação - apareceu a exigência do rigor científico e da objetividade na medida do comportamento humano (Planchard, citado em Escudero Escorza, 2003) e potenciou-se as provas escritas como meio de combater a subjetividade dos exames orais (Ahman & Cook, citados em Escudero Escorza, 2003);
- (II) a influência das teorias evolucionistas de Darwin e das probabilidades de Galton, apoiando a medição das características dos indivíduos e as diferenças entre eles;
- (III) o desenvolvimento de métodos estatísticos que favoreciam a decisão métrica (Nunnally, citado em Escudero Escorza, 2003);

- (IV) o desenvolvimento de uma sociedade industrial que potenciava a necessidade de encontrar mecanismos de acreditação e seleção de alunos segundo os seus conhecimentos;
- (V) a extensão da escolaridade obrigatória, aliada à constatação de que muitas crianças não percorriam a escolaridade com sucesso;
- (VI) a articulação mais estreita entre profissão e diplomas escolares.

No seio deste contexto, marcado por preocupações de natureza científica e de transformações sociais, e na escola em particular, abriu-se caminho para o desenvolvimento da avaliação.

Com o virar do século recomeçou-se, então, a dar ênfase à avaliação dos traços da personalidade e, seguidamente, ao desenvolvimento dos alunos. Em 1905, Binet e Simon lançaram os fundamentos da psicometria. Estes serviriam de modelo aos testes utilizados no campo pedagógico (Pinto & Santos, 2006).

Assim, segundo Escudero Escorza (2003), entre os finais do século XIX e o início do século XX, desenvolveu-se uma atividade avaliativa conhecida por *testing*, que se definia por apresentar as seguintes características:

- (I) medição e avaliação eram termos intercambiais;
- (II) o objetivo consistia em detetar e estabelecer diferenças individuais dentro do modelo dos tipos de personalidade; este caracterizava as elaborações psicológicas da época, pretendendo determinar a posição relativa do sujeito dentro da norma grupal;
- (III) os testes de rendimento, sinónimo de avaliação educativa, elaboravam-se para discriminar os indivíduos, esquecendo-se em grande medida a representatividade e a congruência com os objetivos educativos.

De facto, o *testing* informava algo sobre os alunos mas não sobre os programas com os quais esses alunos se tinham formado, ou seja, a avaliação e a medida tinham pouca relação com os programas escolares. Neste âmbito, destaca-se, ainda, a partir de 1910, o surgimento de alguns instrumentos no campo educativo, como as escalas de ortografia, de leitura e de aritmética (Pinto & Santos, 2006).

A década de 1920 a 1930 marcou o ponto mais alto da utilização do *testing* para medir todo o tipo de destrezas escolares com referentes externos objetivos e explícitos, baseados em procedimentos de medida da inteligência e de fácil utilização com um grande número de estudantes. Estas aplicações estandardizadas foram muito bem acolhidas no meio educativo e foi proposto ao professor que construísse as suas próprias provas objetivas.

O fervor do *testing* decresceu a partir dos anos quarenta e começaram, então, a surgir movimentos hiper-críticos destas práticas. Criticava-se, essencialmente, a tecnicidade do papel do avaliador enquanto provedor de instrumentos de medida.

Segundo Guba e Lincoln (citados em Escudero Escorza, 2003, p.14), esta “geração da medida” ainda permanece viva, pois existem publicações e textos que utilizam de forma indissolúvel avaliação e medida.

1.1.2. A Docimologia: aparecimento e evolução

Paralelamente ao desenvolvimento do *testing*, iniciou-se em França, nos anos vinte do século XX, uma corrente independente que preconizou a primeira verdadeira abordagem à avaliação educativa – a docimologia. Esta corrente criticava sobretudo o divórcio entre o que era ensinado e as metas de instrução.

Na época, um conjunto de estudos realizados por investigadores (Piéron, Laugier, Weinberg) veio pôr em causa o valor dos exames como índice de aptidão escolar, o seu estatuto eliminatório e a sua pouca validade e fiabilidade. Assim, num desses estudos, foram distribuídos vários testes, a vários alunos, em várias escolas, com o objetivo de verificar o valor dos testes. Posteriormente, comparou-se os resultados com as notas escolares, tendo-se encontrado discrepâncias significativas. Um outro estudo foi realizado sobre a estabilidade das notas atribuídas por dois corretores a um conjunto de provas de acesso a uma escola superior e, mais uma vez, foram encontradas discrepâncias entre as duas classificações. Desta forma, como refere Bonboir (1976), concluiu-se que “as notas são pouco dignas de confiança: não só variam com os examinadores e, talvez com os estabelecimentos como, para um mesmo examinador, se modificam com o tempo” (p. 122).

Até 1930, vários estudos foram sendo feitos confirmando os anteriores, tendo-se posto em evidência a instabilidade das avaliações sob o ponto de vista quer das diferenças intra e interindividuais quer da validade e da precisão, sendo que estas evidenciavam uma convergência sobre a pouca fiabilidade dos exames. Poderia pensar-se que atualmente já não é assim; contudo, num estudo realizado por Merle sobre a perícia dos julgamentos avaliativos dos professores, foram encontrados resultados que apontam no mesmo sentido (Pinto & Santos, 2006). Neste estudo, até as situações de multcorreções mostram que, tal como no passado, continua a verificar-se divergências entre os avaliadores.

Foi nesse quadro que Piéron propôs o termo “docimologia” que, segundo o autor, significa “o estudo sistemático dos exames (atribuição de notas, variabilidade interindividual e intraindividual dos examinadores, factores subjectivos, etc)” (Miranda, 1982, p. 39). Este significado é similar ao apresentado por Landsheere (1979), que considera que “a *docimologia* é uma ciência que tem por objeto o estudo sistemático dos exames, em particular do sistema de atribuição de notas e do comportamento dos examinadores e dos examinandos” (p. 13). A docimologia surgiu, assim, como uma crítica aos métodos utilizados nos exames e nos concursos com fins de seleção.

A docimologia vai evoluindo e, desta perspetiva, que inicialmente foi considerada tradicional ou negativa, emergiu uma perspetiva positiva, que se debruçou sobre duas vertentes: o aperfeiçoamento da avaliação e a “docimologia experimental” de Bonniol (Miranda, 1982, p. 47). Por um lado, e no âmbito do aperfeiçoamento da avaliação, a docimologia, segundo Bomber, Reuchlin e Landsheere, preocupou-se com o desenvolvimento de técnicas de construção de instrumentos de avaliação (como provas de exame e provas normalizadas de conhecimentos), a análise das suas qualidades métricas, a sua adequabilidade a determinadas situações e os métodos de moderação das classificações (Miranda, 1982); por outro lado, considerou a linha da “docimologia experimental”, a qual se centra na avaliação como um comportamento, isto é, a avaliação como sendo uma resposta global a uma situação. Esta nova perspetiva da docimologia, conduziu ao desenvolvimento de estratégias no intuito de reduzir as inúmeras divergências até então encontradas e aumentar a validade e a precisão das avaliações.

Perante tal facto, os investigadores apresentaram várias explicações para os erros encontrados a nível da avaliação. Refere-se, por exemplo, Bacher (citado em Miranda, 1982), que considera que os erros da avaliação podem estar relacionados com o avaliador, com as provas e com os alunos:

- (I) com o avaliador, na medida em que, quando este faz a sua apreciação, fá-la na sua perspetiva; além do mais, como refere Piéron (1974, p.11) “muitos examinadores escrupulosos declararam uma preocupação por vezes angustiante em face das suas responsabilidades e, ao mesmo tempo, o espanto por verem alguns dos colegas não levarem a sério a sua missão”;
- (II) no que diz respeito ao conteúdo da prova, o tema testado pode não ser representativo dos conhecimentos adquiridos ao longo de um determinado período de aprendizagem;
- (III) com o aluno, porque o seu desempenho pode ser influenciado por fatores ocasionais, de natureza externa ou interna.

Assim, e no sentido de minorar as divergências (intra e interindividuais) encontradas, surgiram diversas sugestões que passavam por:

- diminuir os escalões das escalas de classificação;
- utilizar escalas do tipo qualitativo;
- uma intervenção na composição dos júris, no sentido de equilibrar a severidade-indulgência;
- fixar os parâmetros comuns de apreciação, no sentido de balizar os aspetos a ter em conta aquando da avaliação.

Dada a falta de confiança nas notas escolares, era necessário encontrar um modo de avaliação objetivo e válido.

A docimologia positiva propôs, então, o teste de aquisições que, segundo Bonboir (1976, p.141), é “um instrumento de «medida»² que responde às exigências da objetividade e da standardização das condições, da validade, da constância, da aferição”. Estes testes apresentam, no entanto, vantagens e desvantagens. Tecnicamente, em comparação com as provas tradicionais, permitem examinar um grande número de indivíduos em pouco tempo, garantindo uma maior amostragem de conteúdo e de objetivos, devido à variedade de questões que podem contemplar. No aspeto métrico, são testes objetivos, cuja validade é teoricamente assegurada durante a sua construção, assumindo,

² “medir significa «atribuir números aos objetos ou aos acontecimentos» de acordo com certas regras” (Brown, citado em Bonboir, 1976, p. 141).

igualmente, uma maior objetividade de correção. No entanto, a sua preparação é demorada e favorecem saberes muito superficiais, o que prejudica os indivíduos criativos (Bonboir, 1976).

No âmbito da docimologia experimental, trabalhos realizados por Noizet e Boniol, em 1969, e por Caverni, em 1975 (Miranda, 1982), tiveram em vista a análise experimental do comportamento da avaliação. As explicações das divergências de avaliação passaram pela análise experimental da situação, o que permitiu determinar, experimentalmente, os mecanismos intervenientes na decisão avaliativa e os fatores de distorção desses mesmos mecanismos.

Apesar do trabalho válido da docimologia, alguns autores têm posto em causa os estudos docimológicos. Como refere Landsheere (1979), a maior parte dos trabalhos docimológicos “não têm sido mais do que a elaboração de sistemas aperfeiçoados para continuar a fazer mal as coisas. É que com efeito não se alicerçam sobre uma revisão completa do ensino” (p. 8). Também Rodrigues (1992, p. 17) questiona “se os estudos docimológicos, que equipararam a *avaliação à medida*, terão verdadeiramente estudado a avaliação do ponto de vista pedagógico, conferindo à docimologia um estatuto educativo sólido e um espaço real no interior da ciência da educação.” Perrenoud refere ainda que “no âmbito dos estudos docimológicos restrito ao estudo dos exames, os resultados pouco esclareceram sobre a avaliação contínua e o seu significado” (Rodrigues, 2002, p. 34). Nesta sua linha de pensamento, Perrenoud (1992) acrescenta ainda “sabemos o quanto este modo de avaliação é frágil no plano docimológico” (p. 168).

Apesar das opiniões referidas, e em termos conclusivos, considera-se que os estudos no âmbito da docimologia, como referem Santos e Pinto (2006, p. 21), “vieram pôr em evidência os problemas da avaliação enquanto medida” e “vieram mostrar a debilidade e a falta de rigor científico nos processos de avaliação.” Constata-se que a “Docimologia surgiu, pois, como uma crítica aos métodos tradicionais utilizados, com fins de seleção, nos exames e nos concursos” (Miranda, 1982, p. 47), compreendendo-se, deste modo, “as recomendações propostas pela docimologia para resolver os problemas constatados, entre as quais se contam a proposta de elaboração de provas *normalizadas-estandardizadas* de conhecimentos e a harmonização dos critérios de avaliação” (Rodrigues, 1992, p.17).

Acrescenta-se ainda, que a docimologia se apresentou com uma evolução significativa, contribuindo para que da perspetiva redutora dos testes psicométricos se passasse para

um conceito muito mais alargado de avaliação, conduzindo ao seu progressivo melhoramento, nomeadamente, ao nível dos instrumentos de avaliação, dos critérios de apreciação e das provas standardizadas.

1.1.3. Do paradigma docimológico a uma avaliação pertinente

Neste ponto da dissertação apresentar-se-á, de forma particular, a perspetiva de Hadji relativamente à problemática da avaliação, por se considerar uma abordagem conducente à reflexão.

Na sequência das questões ligadas à problemática da objetividade da avaliação no âmbito da docimologia, Hadji (1994) alertou para a cilada colocada no *objetivismo*, defendendo uma perspetiva em que a preocupação da avaliação deverá estar centrada não na classificação, mas na informação clara e útil que a mesma poderá fornecer. Ultrapassou, deste modo, largamente, a questão docimológica.

Considera Hadji (1994) que “tornar segura a avaliação é, pois, antes de mais, permitir a quem trabalha no terreno do «ensino» fazer correctamente aquilo que há a fazer neste âmbito” (p. 95). De acordo com o paradigma docimológico, a avaliação foi reduzida a uma classificação e considerou-se que o problema do avaliador era a pesquisa de procedimentos científicos para a operacionalização dessa classificação. Assim, na opinião de Hadji (1994), o importante é “vencer a cilada do objetivismo, antes de mais e sempre, ao interrogarmo-nos sobre o que é necessário conhecer objetivamente” (p. 96).

A docimologia constatou erros nos resultados da avaliação que não surpreendem, se se considerar algumas fontes de distorção de natureza individual, social ou relacionadas com o próprio processo de avaliação. A este propósito Bonboir (1976, p.128) refere que “devemos lamentar que ninguém se tenha lembrado de pensar na formação para a avaliação que daria à classificação, à predição (prognóstico) e à previsão [das cotações escolares] uma garantia de ordem científica”.

Hadji (1994) numa linha semelhante à de Bacher (citado em Miranda, 1982) identificou três fontes de distorção nas classificações:

- (I) a subjetividade do avaliador e classificador;
- (II) a influência de fatores sociais;
- (III) as circunstâncias em que o ato avaliativo decorre.

No que respeita à primeira fonte de distorção mencionada, é verificável através da distribuição da mesma produção por vários corretores. Se as classificações atribuídas por estes não forem iguais, significa que os corretores não controlaram a dimensão subjetiva da sua reação ao produto que classificaram. Neste ponto, importa não esquecer, que o estado de espírito e a sensibilidade de cada um, por exemplo, podem ser responsáveis por flutuações nas classificações.

Relativamente à segunda fonte de distorção, há que estar ciente de que a apreciação do professor / avaliador pode ser determinada, embora inconscientemente, por preconceitos sociais. Por exemplo, no caso de um exame, o saber do aluno (característico de um determinado grupo social), o seu nível cultural e, conseqüentemente, a forma como se exprime podem, ou não, ir ao encontro do avaliador / classificador. Nesta perspetiva, o exame é uma prova de reconhecimento social. Um outro fator social poderá, por exemplo, prender-se com a localização geográfica em que decorre o ato de avaliação e classificação: as notas são relativas ao contexto e traduzem as características de uma escola ou de uma região, sendo alvo de adaptação pelos avaliadores / classificadores às características socioculturais em questão.

Finalmente, no que concerne à terceira fonte de distorção, os autores suprarreferidos colocaram em causa circunstâncias relacionadas com o próprio ato de avaliar enquanto tarefa psicológica. Neste âmbito, Hadji (1994) refere que Noizet e Caverni constataram que não só a ordem pela qual as produções estão colocadas no lote de que fazem parte, como também as informações de que se dispõe sobre quem as produziu influenciam as flutuações encontradas. No primeiro caso, uns exercícios podem modificar a perceção em relação às produções seguintes; no segundo caso, um estudante considerado “bom aluno” é melhor classificado e um considerado “pior aluno”, pior classificado.

Mas, questiona Hadji (1994): impedirá tudo isto que os avaliadores continuem a trabalhar no sentido de classificarem de forma mais justa e mais fiável? O próprio conhecimento dos fatores de distorção pode levar a uma procura de outros procedimentos que os minimizem. Assim, perante as constatações efetuadas, surgiu um conjunto de procedimentos que poderiam ser utilizados na luta contra as divergências encontradas. Destes, destaca-se à posteriori a multicorreção com o objetivo de diminuir a incerteza da nota e, à priori, e no sentido de prevenir as distorções, a harmonização dos pontos de vista e das expectativas pela classificação, determinando diferentes linhas de leitura, o valor e a importância a atribuir a cada aspeto, pretendendo-se, desta forma, dar mais ênfase ao significado da classificação do que à própria classificação. Mas, apesar das

boas intenções, tais procedimentos foram alvo de críticas, pois continuava a haver sempre uma apreciação da produção do aluno, a não ser que se suprimisse o avaliador.

Esta supressão poderia ser efetivada através de questionários de escolha múltipla [QEM] ou de testes de conhecimento, dado, por um lado, a sua classificação ser automática e, por outro, apresentarem a vantagem de suprimir ambiguidade(s) e permitir chegar a um resultado único. Mas o que é que se classifica com os QEM? De acordo com Hadji (1994), estes limitam-se a permitir a reprodução de conhecimentos, não avaliando atividades de nível superior, como a capacidade de síntese, nem comportamentos sociais, nem criações complexas, nem as capacidades de observar e formular hipóteses, entre outras. Resolver-se-á o problema da subjetividade da avaliação suprimindo o avaliador? Na realidade, só desaparece a parte do avaliador que atribui a nota, sendo substituída por uma máquina. A montante o avaliador existe, pois foi ele que delimitou os campos de questionamento e que definiu os critérios de análise. Há, portanto, um trabalho à priori sobre o objeto de avaliação, sobre o que é preciso avaliar, sobre as tarefas que permitirão manifestar o domínio das qualidades requeridas, e é só por estes motivos que a classificação se pode mecanizar. O resultado traduz apenas o grau de expetativa do objetivo. Como referem Noizet e Caverni “a única conclusão a tirar é que convém ajustar o processo de avaliação ao objetivo pretendido” (Hadji, 1994, p. 103). O interesse dos testes QEM está no facto de levarem a uma reflexão sobre os conteúdos, os objetivos e os métodos de avaliação e fazerem compreender que “o problema essencial para o avaliador-classificador não é o de encontrar o verdadeiro valor do produto, mas de poder comunicar, de uma forma clara, um juízo construído sobre o grau de realização de um projeto preciso” (Hadji, 1994, pp. 103-104).

Reconhece-se, assim, que o principal problema da avaliação objetiva não é a subjetividade do corretor, apesar de ser importante gerir o tal estado de espírito ou as ideologias de que se falou anteriormente. Então como pode tornar-se o corretor num recurso fiável? Como se pode ter confiança no avaliador? Esta fiabilidade e esta confiança terão de resultar, necessariamente, do sentido de justiça e da integridade do avaliador.

Cumulativamente, será ainda imperativo que se estabeleça uma comunicação entre o avaliador e o avaliado que clarifique o que está em causa, quais as referências que estão em jogo no processo de avaliação, porque “avaliar não é pesar um objeto que pudéssemos isolar no prato de uma balança. É apreciar este objecto *em relação a outra coisa* que não ele próprio” (Hadji, 1994, p.106).

Considera-se assim que, para avaliar objetivamente, se deve (i) determinar a questão em relação à qual a avaliação terá de fornecer respostas precisas e (ii) explicitar a resposta dada a cada questão, justificando-a, ou seja, explicitando critérios (linhas de leitura correspondentes a domínios de referência) e estabelecendo indicadores para cada critério. Deste modo, o avaliado, sabendo como é questionado, poderá beneficiar do que lhe é dito. Nesta perspetiva, fala-se, então, de uma avaliação pertinente. A avaliação deve, de facto, fornecer ao aluno informação que ele possa compreender e que lhe seja útil, ou seja, garantir um diálogo permanente que permita ao aluno gerir da melhor forma as suas aprendizagens.

1.2. Os exames: suas funções

Historicamente, as formas de avaliação mais dominantes são os testes estandardizados, os exames externos e os testes e exames conduzidos por professores. Comprova-se que, desde o início, os exames fizeram parte de um processo de classificação e de seleção, com interesses, por vezes, particulares. Como refere Broadfoot (citado em Hargreaves, 1996), nas sociedades industriais ocidentais, os exames surgiram para assegurar a entrada para certas profissões de elite, em que só os associados teriam esse benefício.

Segundo Kellaghan, Madaus e Keeves (citados em Fernandes, 2008b), presentemente, quase todos os países têm um sistema de exames. Até mesmo os países que não tinham tradição de exames públicos externos, como é o caso de países da ex-União Soviética e da Europa Oriental, estão neste momento a introduzi-los, principalmente no final do ensino secundário.

Tendo em conta os vários sistemas educativos, constata-se que os exames podem ser realizados a nível de escola, a nível regional ou a nível nacional. Nuns casos assumem unicamente a função de certificação, noutros só apresentam funções de seleção e noutros assumem as duas funções, a de certificar e a de selecionar, como é o caso de Portugal, no que aos exames nacionais do ensino secundário respeita.

Salienta-se ainda que, em alguns sistemas educativos, como o português, a certificação e a seleção podem ter em conta a combinação entre as classificações obtidas no exame nacional e as que são obtidas na escola. Outros sistemas há em que só se considera ou

as classificações dos exames, ou as classificações internas (resultantes de uma avaliação contínua).

De forma a tornar mais compreensíveis as funções dos exames, procede-se, seguidamente, a uma clarificação das mesmas. Assim, segundo Fernandes (2008b), podem constituir funções dos exames:

- a certificação, que comprova que, ao fim de certo tempo, o indivíduo possui um determinado conjunto de aprendizagens;
- a seleção, que controla a progressão escolar do aluno, nomeadamente ao nível do acesso ao ensino superior. Como refere Cardinet (1993, p. 25), “políticos há, por exemplo, que não ocultam o seu desejo de, por este processo [de seleção], manterem a proporção desejada de população, nas profissões manuais, indispensáveis à sobrevivência da sociedade”;
- o controlo, que permite que os governos, através de exames standardizados por si controlados, assegurem que conteúdos semelhantes sejam lecionados nas escolas do país de acordo com o previsto no currículo nacional;
- a monitorização, associada à prestação de contas por parte das escolas e, em alguns países, dos professores, promovendo-se, normalmente, a publicação de *rankings* das escolas de acordo com o seu desempenho nos exames nacionais (este tema será abordado no capítulo seguinte);
- a motivação, quando os exames não têm qualquer efeito na vida escolar do aluno (quando têm efeitos no progresso escolar do aluno, podem tornar-se desmotivadores, nomeadamente para quem os considera difíceis).

1.3. Os exames: seus efeitos

Broadfoot Hargreaves, Earl & Ryan (2001) referem que, segundo Broadfoot, “a avaliação tem um efeito retroativo sobre o programa de ensino e sobre os processos de ensino e de aprendizagem que ocorrem dentro do mesmo” (p. 143). Dado o poder da avaliação para moldar o programa curricular, o ensino e a aprendizagem, os desequilíbrios na avaliação poderão criar desequilíbrios no programa curricular e sobre o ensino e a aprendizagem. Alguns destes efeitos são atribuídos a exames escritos e a testes estandardizados (Hargreaves et al., 2001).

No mesmo sentido, Fernandes (2008a) menciona autores como Barksdale-Ladd e Thomas, Bishop, Crooks, Natriello que referem as avaliações externas com funções de certificação e seleção (como os exames) determinam de forma muito relevante o que os alunos *devem saber e o que e como* os professores devem ensinar.

Ao longo da sua história, os exames têm sido sempre alvo de críticas, pois, apesar de apresentarem vantagens, também apresentam desvantagens. Refere-se, em seguida, por se considerar curioso, o conteúdo de um documento relativo aos exames, apresentado pelo Ministério da Educação, em Inglaterra, em 1911, no qual são apresentados efeitos positivos e efeitos negativos dos exames sobre os professores e os alunos.

Argumentos a favor e contra os exames públicos:

Para o aluno, os efeitos positivos são os seguintes:

- fazem com que o aluno (*sic*) trabalhe, tomando em consideração o tempo, ao exigirem-lhe que alcance um grau de conhecimento estabelecido, numa data fixada;
- incitam-no a moldar o conhecimento em forma reproduzível;
- fazem-no trabalhar em partes de um estudo que, embora importantes, podem ser desinteressantes ou aversivas, para ele pessoalmente;
- treinam o poder de criar um tema para um propósito definido, embora possa não parecer necessário recordá-lo de seguida-um treino que é útil por partes do dever profissional do advogado, do administrador, do do jornalista, do homem (*sic*) de negócios e da secretária;
- em alguns casos, encorajam uma certa perseverança no trabalho, através de um período longo de trabalho; e
- permitem ao aluno medir os seus conhecimentos reais através:
 - de um critério requerido por examinadores externos;
 - da comparação com os conhecimentos dos seus colegas; e

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

- da comparação com os conhecimentos dos colegas de outras escolas.

Por outro lado, os exames podem ter um efeito negativo sobre a mente do aluno ao:

- valorizarem o poder da mera reprodução de ideias de outras pessoas ou dos seus métodos de apresentação, afastando desse modo a energia do processo criativo;
- recompensarem formas de conhecimento provisórias (rápidas a desaparecer);
- favorecerem um tipo de mente, de certo modo, passivo;
- fornecerem uma vantagem injusta àqueles que, ao responderem a questões no papel, podem inteligentemente tirar o maior proveito de conhecimentos escassos;
- induzirem o aluno, na sua preparação para o exame, a apontar para a absorção de informação transmitida pelo professor, em vez de procurar a formação de um juízo independente sobre os conteúdos de que recebeu instrução; e
- estimularem o espírito competitivo (e, inclusivamente, mercenário) na aquisição do conhecimento.

Os efeitos positivos de exames bem conduzidos sobre o professor consistem no facto de estes.

- os induzirem a trabalhar a sua disciplina cuidadosamente;
- os incitarem a organizar as suas lições de modo a cobrir com minúcia intelectual um programa de estudos determinado, dentro de limites temporais atribuídos;
- os impelirem a prestar atenção não só aos seus melhores alunos, mas também aos que ficam na retaguarda e aos mais lentos, entre os que estão a ser preparados para o exame; e
- os familiarizarem com o nível que tanto os seus alunos como outros são capazes de alcançar na mesma disciplina, noutros lugares da educação.

Por outro lado os efeitos dos exames no professor são negativos na medida em que:

- o forcem a observar os pontos fracos do examinando e a notar as suas idiossincrasias(ou a tradição do exame) no sentido em que possam prejudicar os seus alunos com o tipo de conhecimento requerido para lidar eficazmente com as questões que provavelmente lhes serão colocadas;
- limitam a liberdade do professor na escolha da forma como irá trabalhar a sua disciplina;
- o encorajam a assumir trabalho que deveria ser deixado aos esforços, largamente não apoiados, dos seus alunos, levando-o a transmitir a informação de uma forma excessivamente digerida ou a seleccionar para eles grupos de

factos ou aspectos da disciplina que deveriam ser recolhidos ou examinados por cada aluno;

- predispõem o professor a sobrevalorizar, entre os seus alunos, o tipo de desenvolvimento mental que assegura o sucesso dos exames;
- conduzem o interesse do professor para a parte puramente examinável do seu trabalho profissional, afastando-o daquelas outras partes da educação que não podem ser testadas pelo processo do exame.

(Hargreaves, Earl & Ryan, 2001, pp. 149-151)

Os autores realçam que este documento foi escrito numa época em que o conjunto de estratégias de avaliação era consideravelmente mais limitado do que atualmente.

Mais recentemente, muitas outras avaliações se têm feito sobre os efeitos dos exames, mas substancialmente pouco acrescentaram.

Neste âmbito, Hargreaves et al. (2001) apresentam uma abordagem, manifestando alguma preocupação, relativamente ao efeito dos exames no enviesamento sobre os programas de ensino, sobre o professor e sobre o aluno. No que aos programas diz respeito, os exames podem conduzir, por um lado a um estreitamento dos mesmos, na medida em que só parte dos conhecimentos e das competências previstas podem ser medidas através de uma prova escrita, por outro lado, os exames, podem ainda colocar em destaque e dar ênfase, às disciplinas testadas. Como refere Broadfoot (citado em Hargreaves et al., 2001), apesar de os exames ocorrerem, na maioria das vezes, no final do percurso escolar dos estudantes, têm sido criticados por se verificar que influenciam o programa de ensino muito antes deste ponto terminal. Ao estreitarem desta forma o programa de ensino, os padrões tradicionais de avaliação diminuem as possibilidades de sucesso e tendem a criar um programa de ensino direcionado ao académico, distante do quotidiano e desmotivante para os “menos capazes”.

Relativamente ao efeito que os exames têm sobre os professores, parece que têm um efeito negativo. Estudos realizados permitem verificar que as práticas que os professores atribuem aos constrangimentos dos exames (por exemplo, ensinar a partir do quadro e as revisões) debilitam a inovação na sala de aula. No entanto, pelo facto de os exames serem uma fonte de motivação para os alunos, também os professores se sentem motivados e reconhecidos pelo seu trabalho (Hargreaves et al., 2001).

Relativamente ao efeito dos exames sobre os estudantes, Hargreaves et al. (2001) mencionam alguns autores que referem que os exames não promovem o pensamento

crítico desejado. Corre-se, assim, o risco de os exames incentivarem atitudes de natureza calculista e instrumental em relação à aprendizagem. Os próprios alunos poderão induzir os professores a desenvolverem um ensino orientado para o exame. No entanto, outros professores reconhecem vantagens nos exames, vendo nestes uma fonte de motivação para os alunos (Sikes, Measor & Woods, citados em Hargreaves et al., 2001). Outro dos efeitos possíveis sobre os estudantes é poder desencadear fenómenos de ansiedade, sendo que estes, como se sabe, podem influenciar o desempenho do aluno. Por outro lado, os exames tanto podem ser fonte de motivação para muitos, pois são recompensados com notas, como podem ser desmotivadores para outros com capacidades mais reduzidas, levando-os a atitudes de desistência.

Mas a verdade é que, na procura da justiça num sistema onde prevalece a igualdade de oportunidades formal, cada vez há mais estudantes a submeterem-se a exames, testes e outras avaliações em níveis cada vez mais elevados para maximizar as suas hipóteses de sucesso. Quanto maior for o número de indivíduos a alcançar o sucesso, maiores são as exigências de quem os contrata, no sentido de diminuir o número de candidatos e, posteriormente, selecionar os de maior qualidade. Como referem Hargreaves *et. al* (2001), a certificação exerce uma influência cada vez mais vasta, tornando-se num dos objetivos mais poderosos da avaliação educativa.

1.4. Os exames do ensino secundário: suas características

Fernandes (2008b), segundo Kellaghan & Madaus, refere que os exames apresentam um conjunto de características que, na sua maioria, são comuns a um elevado número de países, sendo estas as seguintes:

- a) Os exames são externos, isto é, são preparados e controlados por uma ou mais entidades externas às escolas a que os alunos pertencem e por departamentos do Ministério da Educação;
- b) A administração dos exames é normalmente controlada pelo Ministério da Educação, com a colaboração das escolas;
- c) São concebidos e elaborados, no caso de Portugal, por um departamento do Ministério da Educação – o Gabinete de Avaliação Educacional [GAVE];

- d) As provas de exame são iguais para todos os alunos, administradas segundo procedimento standardizado, fora do ambiente normal da sala de aula e realizadas obrigatoriamente por todos os alunos;
- e) As suas funções principais são a certificação dos saberes adquiridos pelos alunos, o controlo do desenvolvimento e cumprimento do currículo, a seleção de alunos que pretendem prosseguir estudos e a avaliação do sistema educativo;
- f) Os conteúdos, os critérios de correção e os resultados dos exames são tornados públicos;

Registam-se, no entanto, algumas diferenças no seu modo de operacionalização (Fernandes, 2008b), que a seguir se referenciam:

1. Os exames ocorrem, geralmente, no final do ensino secundário. Contudo há países que têm exames em níveis mais baixos de escolaridade, como é o caso de Portugal, onde há exames nacionais ou provas de aferição no final de cada ciclo do ensino básico;
2. O número de entidades envolvidas no processo de exames varia consoante o país;
3. O peso dos exames, para efeitos de certificação ou de seleção, varia entre os países. Na Irlanda, em França e em Inglaterra, por exemplo, a avaliação no final do secundário é totalmente ou quase totalmente externa. Noutros países, há exames que são administrados e corrigidos pelos professores nas escolas com algum sistema de moderação externa. Em Portugal, a ponderação é distribuída pela classificação final do ensino secundário e a classificação do exame;
4. O número de exames que os alunos têm de fazer é variável consoante o país;
5. O formato dos exames também pode variar tanto no número de partes que o constituem como na natureza e número das questões. Em muitos sistemas educativos, as questões tipo ensaio têm sido predominantes mas também fazem uso de questões de resposta curta e de questões objetivas.

Face ao exposto, pode constatar-se que as modalidades, os formatos e formas de controlo variam em função do sistema educativo de cada país, dependendo das finalidades a que se destinam.

CAPÍTULO 2 – A AVALIAÇÃO EXTERNA DAS APRENDIZAGENS EM PORTUGAL

O tema em epígrafe surge por se considerar relevante dar uma perspetiva dos exames em Portugal e de alguns aspetos relacionados com a problemática dos mesmos. Assim, far-se-á, num primeiro momento, uma abordagem normativo-legal dos exames, desde a sua implementação até à atualidade. Segue-se uma descrição das várias entidades envolvidas na consecução dos exames e respetivas missões. Posteriormente, são apresentadas algumas perspetivas sobre a pertinência deste tipo de avaliação e de *rankings*. Por último, apresenta-se um conjunto de estudos sobre os exames realizados pelo Ministério da Educação (GAVE), que evidenciam o tipo de informação fornecida pela avaliação externa, nomeadamente os saberes dos alunos.

2.1. Resenha histórica dos exames em Portugal

Nóvoa (2005) aponta os meados do século XIX como o período em que, em Portugal, os exames se tornaram no principal dispositivo para regular as políticas educativas e as práticas de ensino. Aludindo ao mesmo período temporal, Fernandes (2008b) considera relevantes as datas de 1836 e 1860, correspondentes, respetivamente, à criação, por Passos Manuel, dos liceus oficiais e do “primeiro plano sistematizado de estudos secundários, integrando aspectos curriculares, pedagógicos e de administração escolar” (Adão, 1997, p. 30), e ao ano em que Fontes Pereira de Melo, através da publicação do Regulamento para os Liceus Nacionais, deu um “passo digno de destaque na concretização de um ensino secundário de características uniformes em todo o país” (Fernandes, 2008b, p. 101). Apesar de esta uniformização não se ter estendido à

definição dos conteúdos dos exames, a qual era da responsabilidade dos conselhos dos liceus, a tendência para a centralização sobressaía na “obrigação dos exames propostos a nível de cada liceu serem submetidos à aprovação do Conselho Geral de Instrução Pública” (Beato, 2011, p. 25).

Em 1895, na reforma introduzida por Jaime Moniz, apresentava-se os exames liceais como um dos cinco planos que conformavam o currículo. Sendo os exames um momento de avaliação formal, considerava-se que os mesmos, mais do que um instrumento de certificação de conhecimentos, constituíam um instrumento de pressão sobre os alunos, incentivando-os ao trabalho (Ó, 2009). Embora os exames realizados não fossem iguais em todo o país, a sua implementação contribuía para regular o funcionamento do ensino liceal (Fernandes, 2008b).

Os exames de âmbito nacional, e iguais para todos os alunos, viriam a ser implementados trinta e cinco anos depois, aquando da publicação do Decreto 18024, de 1 de março de 1930. Este decreto, que fixava o regime de exames liceais, seria posteriormente complementado com a Circular aos reitores de 23 de junho de 1930, na qual eram apresentadas provas-modelo, e com o Decreto 18884 de 27 de setembro de 1930, que estabelecia o novo sistema de exames liceais (Ó, 2009) e que consignava, entre outros,

- a) a redução significativa da importância das provas orais; b) a separação das funções de examinador e de professor; c) a concepção de provas orientadas mais para a avaliação da inteligência do que da memória; e d) a instituição do regime de anonimato dos alunos nas provas escritas (Fernandes, 2008b, p. 102).

O sistema implementado em 1930 vigoraria até 1944, ano em que, com a publicação do Decreto 34053, de 21 de outubro, os exames passaram a ser organizados em cada liceu, sendo constituídos por uma prova escrita, uma prova oral e, nas disciplinas que tinham trabalhos práticos, uma prova prática. Além disso, era abolida o regime de anonimato dos alunos (Ó, s.d.).

Três anos volvidos, o ensino liceal foi alvo de reforma, tendo-se registado novas alterações no que aos exames respeita. Assim, o Decreto 36507, de 17 de setembro de 1947, consagrou um sistema de exames nacionais obrigatórios, cuja correção obedecia ao regime de anonimato. Os exames decorriam numa época única e eram constituídos por provas escritas, orais e práticas, sendo que a elaboração das primeiras era da

responsabilidade do Ministério da Educação Nacional, mais precisamente da então criada Inspeção do Ensino Liceal, secundada por professores nomeados pelo ministro (Decreto 36508, de 17 de setembro de 1947).

As características essenciais do sistema de exames de 1947 perdurariam “até ao início do regime democrático, altura em que o processo de transformação do sistema educativo acabou por conduzir à abolição dos exames” (Fernandes, 2008b, p. 103). Neste âmbito, Reis (2009) assinala que, entre 1974 e 1976, os sucessivos governos provisórios procuraram estabelecer a ligação entre *educação* e *democracia*, o que conduziu a que, por pressão de estudantes e professores, se suspendesse quer os exames nacionais quer as provas de admissão à universidade.

No ensino secundário, ter-se-ia de esperar até à década de noventa para que os exames nacionais, com carácter de obrigatoriedade para a conclusão desse ciclo de estudos, fossem reintroduzidos em Portugal; já a prestação de provas para ingresso na universidade seria retomada ainda nos anos 70.

No ano letivo de 1977/78, o Decreto-Lei n.º 491/77, de 23 de novembro, instituiu “o Ano Propedêutico do ensino superior oficial”, que funcionaria “na dependência da Direcção-Geral do Ensino Superior” (Artigo 1.º). Neste regime, os alunos eram avaliados através de dois conjuntos de provas de exame, compostos por cinco provas escritas, correspondentes às cinco disciplinas em que estavam inscritos. A aprovação no Ano Propedêutico estava exclusivamente dependente das classificações obtidas nos dois conjuntos de provas (Portaria n.º 210/78, de 15 de abril), tendo estas, portanto, inerente a função de certificação. A esta estava aliada a função de seleção, na medida em que a classificação obtida no Ano Propedêutico tinha a ponderação de 50% na média de acesso à universidade, como o demonstra a fórmula

$$\frac{\frac{MCC + MDN}{2} + NAP}{2}$$

em que *MCC* corresponde à “Média do curso complementar do ensino secundário”, *MDN* à “Média arredondada das classificações obtidas no curso complementar do ensino secundário nas disciplinas nucleares” e *NAP* à “Classificação do Ano Propedêutico” (Portaria n.º 210/78, de 15 de abril, Artigo 20.º).

Em 1980, com a criação do 12.º ano de escolaridade, que passou “a constituir o ano terminal dos cursos complementares do ensino secundário” (Decreto-Lei n.º 240/80 de 19 de julho, Artigo 1.º), extinguia-se o Ano Propedêutico, bem como as provas para ingresso no ensino superior.

Em 1986, a Lei de Bases do Sistema Educativo (Lei n.º 46/86, de 14 de Outubro) [LBSE] dotou o sistema educativo de um quadro jurídico geral, consagrador de orientações políticas consonantes com a Constituição da República de 1976. Tendo estabelecido princípios de reorganização para o sistema de ensino português, eliminou “a figura do exame como momento de conclusão, certificação e selecção dos alunos no sistema educativo” (Rosário, 2007, p. 22). No entanto, estabelecia como condição para ingresso no ensino superior a realização de provas “de âmbito nacional e específicas para cada curso ou grupo de cursos afins” (Lei n.º 46/86, de 14 de Outubro, Artigo 12.º), o que, segundo Pires (1997), permitia corrigir as desigualdades que se tinham registado desde 1980 e introduzir “uma unidade objectiva de avaliação da capacidade (científica e outras) para a frequência do ensino superior” (p. 57).

Nos anos seguintes, e até à reintrodução dos denominados exames nacionais em 1993, essas provas não só teriam diferentes ponderações no cálculo da classificação de acesso ao ensino superior, como assumiriam diferentes características e denominações. Enquadradas pelo artigo 12.º da LBSE, foram, assim, implementadas a Prova Geral de Acesso [PGA], a que se sucedeu a Prova de Aferição, e as Provas Específicas.

A primeira regulamentação da LBSE no que ao acesso ao ensino superior concerne ocorreu em 1988, com a aprovação do novo regime de acesso ao ensino superior. Este foi concretizado pelo Decreto-Lei n.º 354/88, de 12 de outubro, e estipulava a

- a) Realização, por todos os candidatos, de uma prova geral de acesso, destinada a avaliar o seu desenvolvimento intelectual, o seu domínio da língua portuguesa ao nível da compreensão e da expressão e a sua maturidade cultural;
[...]
- c) Realização, pelas instituições de ensino superior, se o entenderem necessário, de provas específicas para a seriação dos candidatos a cada um dos seus cursos.

(Decreto-Lei n.º 354/88, de 12 de outubro)

De acordo com o mesmo normativo, a PGA consistia numa prova escrita, de natureza não eliminatória e, não incidindo sobre conteúdos específicos de qualquer das disciplinas

do ensino secundário, era comum a todos os cursos deste nível de ensino. A classificação obtida na PGA não poderia ter peso inferior a 30% no cálculo da média de acesso ao ensino superior, cabendo a cada universidade ou instituto politécnico a definição e a aprovação, para cada curso, dos critérios de seriação dos candidatos. Por seu lado, as Provas Específicas eram de carácter facultativo, sendo que cada instituição de ensino superior decidia sobre a necessidade, ou não, da sua realização para a candidatura a cada um dos seus cursos. Caso as considerasse necessárias, a instituição era responsável pela divulgação da sua natureza, dos conteúdos sobre os quais incidiriam e do calendário da sua realização. Sempre que possível, as instituições que optassem pela realização de Provas Específicas deveriam coordenar-se no sentido de estas serem comuns. Nos critérios definidos para o cálculo da nota de acesso ao ensino superior, a classificação obtida na(s) Prova(s) Específica(s) não poderia ter peso superior ao atribuído à PGA.

O regime de acesso definido pelo Decreto-Lei n.º 354/88, de 12 de outubro, viria a ser aplicado apenas no ano letivo de 1988/1989. No início de 1990, o Decreto-Lei n.º 33/90, de 24 de janeiro, introduzir-lhe-ia algumas alterações, designadamente no que às Provas Específicas e aos critérios de seriação dos candidatos respeitava. Deste modo, as Provas Específicas, cuja necessidade de realização continuava a ser determinada por cada instituição de ensino superior, passaram a ser “de âmbito nacional, sendo a coordenação da sua organização e execução assegurada conjuntamente pelo Conselho de Reitores das Universidades Portuguesas e pelo Conselho Coordenador dos Institutos Superiores Politécnicos” (Decreto-Lei n.º 33/90, de 24 de janeiro, Artigo 26.º). A definição dos critérios de seriação dos candidatos continuava, igualmente, a ser da responsabilidade de cada instituição de ensino superior, mas eram (re)definidos limites mínimos para o peso de cada critério: 20% para a classificação obtida na PGA e 30% para a obtida na(s) Prova(s) Específica(s), quando existentes. Cumulativamente estipulava-se que os limites máximos seriam “fixados anualmente para cada um dos critérios, através de deliberação conjunta do Conselho de Reitores das Universidades Portuguesas e do Conselho Coordenador dos Institutos Superiores Politécnicos” (Decreto-Lei n.º 33/90, de 24 de janeiro, Artigo 24.º). As alterações introduzidas podem, assim, ser sintetizadas em quatro pontos:

- (I) carácter nacional das Provas Específicas;
- (II) diminuição do peso mínimo atribuído à classificação obtida na PGA (de 30% para 20%);

- (III) valorização da classificação obtida na(s) Prova(s) Específica(s) em detrimento da auferida na PGA, na definição de limites mínimos;
- (IV) introdução da definição anual de limites máximos, pelo ensino superior, para cada um dos critérios de seriação dos candidatos.

Em 1992, os alunos do ensino secundário mobilizaram-se na contestação à PGA, “salientando a sua ambiguidade e subjectividade e os critérios aleatórios da classificação” (Seixas, 2005, p. 195). Após um longo período de contestação, a PGA acabou por ser abolida. Com a publicação do Decreto-Lei n.º 189/92, de 3 de setembro, estabeleceu-se um novo regime de acesso ao ensino superior, baseado nas classificações finais obtidas pelos alunos no decurso do ensino secundário e nas classificações das Provas Específicas e da Prova de Aferição. Esta última era apresentada como um “exame nacional, escrito e não eliminatório, sobre matéria do curso do ensino secundário”, tendo como objetivo “assegurar a homogeneidade das classificações atribuídas no ensino secundário” (Decreto-Lei n.º 189/92, de 3 de setembro, Artigo 7.º). As Provas Específicas, provenientes do anterior regime de acesso ao ensino superior, eram alvo de nova regulamentação. Reiterava-se o seu carácter nacional e não eliminatório, mas introduzia-se a obrigatoriedade da sua realização para ingresso em qualquer par estabelecimento / curso. Além disso, determinava-se que a sua seleção pelas instituições de ensino superior, em número de uma ou duas para cada par estabelecimento / curso, estaria confinada a um elenco definido pelo Ministério da Educação. A sua classificação, expressa na escala de 0 a 100, tinha um peso de 50% nos critérios de seriação dos candidatos. Os restantes 50% destinavam-se à classificação global do ensino secundário, a qual era calculada considerando:

- a) A classificação final do 10.º / 11.º anos de escolaridade, com o peso de 60%;
- b) A classificação final do 12.º ano de escolaridade, com o peso de 20%;
- c) O resultado da prova de aferição, com o peso de 20%.

(Decreto-Lei n.º 189/92, de 3 de setembro, Artigo 29.º)

Desta forma, e em síntese, a nota de candidatura era calculada segundo os critérios e as ponderações constantes do Quadro 1.

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

Quadro 1

Critérios de seriação dos candidatos ao ensino superior, de acordo com o Decreto-Lei n.º 189/92, de 3 de setembro

Critérios	Ponderação
Provas Específicas	50%
Classificação final dos 10.º / 11.º anos de escolaridade	30%
Classificação final do 12.º ano de escolaridade	10%
Prova de Aferição	10%

Da análise dos critérios de seriação dos candidatos conclui-se que, neste regime, as provas de avaliação externa (Provas Específicas e Prova de Aferição), embora mantivessem o seu carácter não eliminatório, passavam a ter um peso superior às classificações finais obtidas nos três anos do ensino secundário. Constata-se, igualmente, que, com a consideração da classificação obtida na Prova de Aferição para efeitos de cálculo da média / classificação global do ensino secundário, estavam lançadas as bases para a introdução dos exames nacionais com as funções de certificação e seleção, o que viria a acontecer em 1993.

O Despacho Normativo n.º 338/93, de 21 de outubro, surgiu na sequência da reforma curricular do ensino secundário, que tinha sido aprovada em 1989, e preencheu uma lacuna desde então existente: “a aprovação de um novo regime de avaliação dos alunos deste nível de ensino” (Despacho Normativo n.º 338/93, de 21 de outubro). Este regime distinguia três modalidades de avaliação: formativa, sumativa e aferida. A respeito da avaliação sumativa referia-se que consistia

na formulação de um juízo globalizante sobre o grau de desenvolvimento dos conhecimentos e competências, capacidades e atitudes do aluno, no final de um período de ensino e de aprendizagem, tomando por referência os objectivos fixados para o ensino secundário e para as disciplinas que o integram.

(Despacho Normativo n.º 338/93, de 21 de outubro, Ponto 20)

A avaliação sumativa processava-se através de duas formas – interna e externa – e a classificação dela resultante expressava-se na escala de 0 a 20 valores. A avaliação sumativa externa, da responsabilidade do Ministério da Educação, visava “contribuir para a homogeneidade nacional das classificações do ensino secundário” (Despacho Normativo n.º 338/93, de 21 de outubro, Ponto 30) e era operacionalizada através da

realização de exames finais nacionais. As provas eram escritas e a sua elaboração cabia ao Departamento do Ensino Secundário [DES].

Ficavam sujeitas a exame nacional as disciplinas de 12.º ano, excepcionando-se Educação Física, Educação Moral e Religiosa e Desenvolvimento Pessoal e Social e, nos cursos orientados para o prosseguimento de estudos, as disciplinas da “componente de formação técnica, tecnológica ou artística” (Despacho Normativo n.º 338/93, de 21 de outubro, Ponto 31). A classificação obtida nos exames tinha um peso de 40% na classificação final da disciplina, sendo esta calculada através da fórmula

$$\frac{3 CI + 2CE}{5}$$

em que *CI* (classificação interna) correspondia à “Média aritmética simples, arredondada às unidades, das classificações obtidas na avaliação interna, referente aos anos em que a disciplina é ministrada” e *CE* à “Classificação em exame final” (Despacho Normativo n.º 338/93, de 21 de outubro, Ponto 42). A admissão a exame estava condicionada à obtenção de classificação interna igual ou superior a dez valores. No entanto, os exames nacionais podiam, igualmente, assumir o carácter de exame de equivalência à frequência em situações em que os alunos não tivessem reunido as condições para ser admitidos a exame ou em que quisessem obter aprovação em disciplinas em que não haviam estado matriculados.

O regime de avaliação constante do Despacho Normativo n.º 338/93, de 21 de outubro, entrava em vigor no ano letivo de 1993/94, aplicando-se nesse ano aos alunos que frequentavam o 10.º ano de escolaridade, pelo que os primeiros exames nacionais se realizariam no ano letivo de 1995/96, correspondente ao ano em que aqueles alunos deveriam frequentar o 12.º ano. Em consequência, durante o ano letivo de 1995/96, foi aprovado e publicado o primeiro regulamento dos exames do ensino secundário e foram introduzidas alterações no regime de acesso ao ensino superior.

Assim, em setembro de 1995, através do Despacho Normativo n.º 55/95, de 19 de setembro, surgiu o regulamento suprarreferido, que acrescentou algumas indicações relativas à elaboração dos exames: estes incidiriam “sobre o programa do 12.º ano, podendo avaliar conteúdos dos restantes anos que com ele estejam directamente relacionados” (Despacho Normativo n.º 55/95, de 19 de setembro, Ponto 16.1.1). Até ao

final do 1.º Período, o DES teria de divulgar a estrutura dos exames, os objetivos e os conteúdos sobre os quais incidiriam e as instruções para a sua realização.

As alterações registadas ao nível do acesso ao ensino superior prenderam-se, essencialmente, com a abolição da Prova de Aferição (Decreto-Lei n.º 318/95, de 28 de novembro) e das Provas Específicas (Decreto-Lei n.º 28-B/96, de 4 de abril), tendo ambas sido substituídas pelos exames nacionais, no sentido de evitar a duplicação na realização de exames. Assim, na fórmula utilizada para calcular a nota de candidatura ao ensino superior, o peso anteriormente concedido à classificação da Prova de Aferição (10%) passou a ser atribuído à obtida no exame da disciplina base do curso do ensino secundário frequentado pelo aluno, enquanto os 50% atribuídos à classificação da(s) Prova(s) Específica(s) passaram a ser conferidos à auferida no exame da(s) disciplina(s) específica(s) exigida(s) para ingresso em cada par estabelecimento / curso (Decreto-Lei n.º 28-B/96, de 4 de abril). Desta forma, reiterava-se a importância fulcral das classificações obtidas em exame nacional para efeitos de candidatura ao ensino superior, especialmente no que à disciplina base do curso do ensino secundário e à(s) disciplina(s) específica(s) respeitava.

Invertendo a tendência de valorização progressiva dos exames finais nacionais do ensino secundário, no ano letivo seguinte (1996/97), o peso da sua classificação foi diminuído para efeitos quer de conclusão do 12.º ano quer de acesso ao ensino superior.

O Despacho Normativo n.º 45/96, de 31 de outubro, deu nova redação ao Ponto 42 do Despacho Normativo n.º 338/93, de 21 de outubro, passando a classificação do exame nacional a ter o peso de 30% (em vez dos anteriores 40%) na classificação final das disciplinas de 12.º ano (Despacho Normativo n.º 45/96, de 31 de outubro, Ponto 1).

Em abril de 1997, também a fórmula para cálculo da nota de candidatura ao ensino superior sofreu alterações. Assim, o Decreto-Lei n.º 75/97, de 3 de abril, determinou, que a referida nota resultava da ponderação de 50% da classificação final do ensino secundário, sendo os restantes 50% atribuídos à classificação do exame nacional da(s) disciplina(s) fixada(s) como específica(s) para acesso a cada par estabelecimento / curso. Desta forma, a ponderação de 10% atribuída ao exame da disciplina base do curso do ensino secundário no ano letivo de 1995/96 era agora transferida para a classificação final do ensino secundário.

Se, no que respeita à ponderação das classificações dos exames nacionais na determinação da classificação final das disciplinas do ensino secundário não se

registriam alterações até à atualidade, o mesmo não sucedeu relativamente à sua ponderação para efeitos de candidatura ao ensino superior.

Em 1998, o Decreto-Lei n.º 28-B/96, de 4 de abril, e o Decreto-Lei n.º 75/97, de 3 de abril, foram revogados pelo Decreto-Lei n.º 296-A/98, de 25 de setembro, que regulamentava o novo regime de acesso e ingresso no ensino superior. Relativamente aos anteriores normativos, e no que se reporta à utilização dos exames nacionais para candidatura ao ensino superior, o Decreto-Lei n.º 296-A/98, de 25 de setembro, introduzia alterações significativas, designadamente a utilização de novos critérios para seriação dos candidatos, os quais passavam a ser estabelecidos por cada instituição e contemplavam

- a) A classificação final do ensino secundário, com um peso não inferior a 50%;
- b) A classificação da ou das provas de ingresso, com um peso não inferior a 35%;
- c) A classificação dos pré-requisitos de seriação, quando exigidos, com um peso não superior a 15%.

(Decreto-Lei n.º 296-A/98, de 25 de setembro, Artigo 26.º)

Para efeitos do estabelecido na alínea b), as instituições de ensino superior poderiam elaborar e realizar as suas próprias provas de ingresso ou continuar a utilizar os exames nacionais realizados no final do ensino secundário. Embora desta forma cessasse a obrigatoriedade de as disciplinas específicas para a candidatura a qualquer par estabelecimento / curso serem selecionadas de entre o elenco dos exames nacionais do ensino secundário, esta medida nunca viria a ser aplicada, pelo que as provas de ingresso continuariam a realizar-se através dos exames nacionais (Portaria n.º 505-A/99, de 15 de julho), situação que ainda se mantém.

Com a publicação da Portaria n.º 505-A/99, de 15 de julho, ficava definida, no sistema de ensino português, a relevância dos exames nacionais do ensino secundário quer para conclusão deste nível de ensino, quer para efeitos de candidatura ao ensino superior. Não obstante ter sido implementada, em 2004, a reforma curricular do ensino secundário (Decreto-Lei n.º 74/2004, de 26 de março), esta nada de novo trouxe ao sistema de exames, exceção feita à inclusão, em 2006, das disciplinas bienais da componente de formação específica no elenco das disciplinas sujeitas a exame (Decreto-Lei n.º 24/2006, de 6 de fevereiro). Neste contexto, o primeiro exame nacional de Biologia e Geologia (Código 702) seria realizado no ano letivo de 2005/2006.

2.2. Entidades envolvidas nos exames em Portugal

Em Portugal, as entidades ou instituições envolvidas no processo de exames nacionais são: o Júri Nacional de Exames [JNE], a Direção Geral de Educação [DGE], o Gabinete de Avaliação Educacional, a editorial do Ministério da Educação e Ciência [EMEC], a Inspeção Geral da Educação e Ciência [IGEC], as forças de segurança (PSP e GNR) e, nas regiões autónomas, a Direção Regional Educação da Região Autónoma da Madeira e a Direção Regional da Educação e Formação do Governo Regional Açores. Seguidamente explicita-se as funções e/ou as competências de cada uma das entidades supra referidas:

A) Júri Nacional de Exames

O JNE encontra-se integrado na DGE (cuja orgânica foi aprovada pelo Decreto-Lei n.º 14/2012, de 20 de janeiro), tendo por missão coordenar e planificar o processo de avaliação externa da aprendizagem. No que respeita aos exames do ensino secundário, o JNE tem por competências:

- a) Coordenar e planificar a [sua] realização [...];
- b) Propor os normativos legais de suporte à realização das provas de exame e organizar a logística inerente à sua classificação, reapreciação e reclamação;
- c) Promover os mecanismos de apoio à prestação [...] de exames finais nacionais por parte dos alunos com necessidades educativas especiais;
- d) Validar as condições de acesso dos alunos à realização [...] de exames finais nacionais e consequente certificação dos seus currículos;
- e) Disponibilizar os dados estatísticos referentes à avaliação externa da aprendizagem.

(Despacho normativo n.º 6/2012, de 10 de abril, n.º 5.1 do Anexo I)

B) Direção Geral de Educação

No Decreto-Lei n.º 125/2011, de 29 de dezembro (que, entre outros, estipula a missão e as atribuições da DGE), refere-se que cabe a esta direção-geral “coordenar a planificação das provas finais, dos exames nacionais e equivalentes” (Artigo 12.º, Ponto 2, alínea c).

C) Gabinete de Avaliação Educacional

A missão do GAVE é “desempenhar, no âmbito da componente pedagógica e didáctica do sistema educativo, funções de planeamento, coordenação, elaboração, validação, aplicação e controlo de instrumentos de avaliação externa da aprendizagem” (Decreto-Lei n.º 213/2006, de 27 de outubro, Artigo 15.º). Neste âmbito, é-lhe atribuído o planeamento de todo o processo de elaboração e validação dos exames nacionais, a organização dos sistemas de informação necessários à sua produção e a supervisão da sua correção. Além disso, deve colaborar com a DGE no processo de realização dos exames (Decreto-Lei n.º 213/2006, de 27 de outubro, Artigo 15.º).

D) Editorial do Ministério da Educação e Ciência

A EMEC foi criada pelo Decreto-Lei n.º 648/76, de 31 de julho, sendo um organismo do Ministério da Educação com autonomia administrativa e financeira. A EMEC é a entidade responsável pelo processo de logística dos exames nacionais (impressão e embalagem), estando obrigada a assegurar a sua confidencialidade.

E) Inspeção Geral da Educação e Ciência

De acordo com o Decreto-Lei n.º 125/2011, de 29 de dezembro, entre as atribuições da IGEC encontra-se consagrada a contribuição

para a qualidade do sistema educativo no âmbito da educação pré-escolar, dos ensinos básico e secundário e da educação extra-escolar, designadamente através de acções de controlo, acompanhamento e avaliação, propondo medidas que visem a melhoria do sistema educativo e participando no processo de avaliação das escolas de ensino básico e secundário e das actividades com ele relacionadas (Artigo 11.º, n.º 2, alínea c).

É no âmbito das ações de controlo que a IGEC

efetua todos os anos uma intervenção inspetiva [...] à realização dos exames nacionais dos ensinos básico e secundário, em estabelecimentos do ensino público e do ensino particular e cooperativo, em que estão incluídas, prioritariamente, as escolas:

- onde foram detetadas, em anos anteriores, situações anómalas ou desvios;

- sinalizadas pelo Júri Nacional de Exames (JNE) como necessitando de controlo.

(IGEC, 2012)

Esta atividade tem como objetivos:

- Verificar a adequação das medidas e dos procedimentos adotados pelos agrupamentos de escolas e escolas não agrupadas do ensino público e pelos estabelecimentos do ensino particular e cooperativo, face aos normativos e aos contextos específicos em que [...] os exames nacionais decorrem;
- Contribuir para a melhoria da qualidade da organização dos agrupamentos de escolas e escolas não agrupadas do ensino público e pelos estabelecimentos do ensino particular e cooperativo, no que respeita a todo o serviço inerente [...] aos exames nacionais.

(IGEC, 2012)

A metodologia utilizada pelos inspetores contempla “a observação direta, a análise documental e a realização de entrevistas” (IGEC, 2012), sendo sempre objeto de observação as

- Medidas organizativas da competência do responsável do Agrupamento/Escola;
- Atividades da competência do Secretariado de Exames;
- Atividades da competência dos Professores Coadjuvantes;
- Atividades da competência dos Professores Vigilantes.

(IGEC, 2012)

Caso verifique a existência de situações anómalas ou estas lhe sejam comunicadas, a IGEC procede de imediato à sua análise, dando-lhes tratamento sequencial. No final da atividade de controlo, a IGEC elabora um relatório global.

F) Forças de segurança (PSP e GNR)

As forças de segurança (PSP e GNR) são responsáveis pela distribuição das provas de exame pelas escolas do país, garantindo que a mesma decorra nos prazos estipulados e em segurança. Após a realização das provas pelos alunos, compete, igualmente, às forças de segurança transportar os exames das escolas para os respetivos agrupamentos de correção e, uma vez corrigidas, destes para as escolas.

G) Direção Regional Educação da Região Autónoma da Madeira

A Direção Regional Educação da Região Autónoma da Madeira “é um serviço executivo das políticas públicas definidas pelo Governo Regional da responsabilidade da Secretaria Regional de Educação e Cultura” (Decreto Regulamentar Regional n.º 9/2008/M, de 30 de abril, Artigo 3.º, Ponto 1). No domínio dos exames do ensino secundário, esta direção regional superintende “os júris de exame, que em virtude da lei, se tornem necessários criar”, desde que “sem prejuízo das competências próprias do júri nacional de exames do Ministério da Educação” (Decreto Regulamentar Regional n.º 9/2008/M, de 30 de abril, Artigo 3.º, Ponto 1, alínea g).

H) Direção Regional da Educação e Formação do Governo Regional Açores

De acordo com a orgânica da Secretaria Regional da Educação e Formação, à Direção Regional da Educação e Formação do Governo Regional Açores compete conceber, orientar, coordenar e avaliar o sistema educativo. Neste sentido, a Direção de Serviços de Inovação e Desenvolvimento Curricular, uma das suas unidades orgânicas, planeia “o processo de elaboração e validação dos instrumentos de avaliação externa das aprendizagens” (Decreto Regulamentar Regional n.º 25/2011/A, de 25 de novembro, Artigo 19.º, alínea n) e assegura, em colaboração com a Divisão de Desenvolvimento Curricular, “as acções respeitantes à realização de provas de avaliação externa e exames nacionais” (Decreto Regulamentar Regional n.º 25/2011/A, de 25 de novembro, Artigo 19.º, alínea o).

Em termos conclusivos regista-se que há um conjunto alargado de entidades envolvidas no processo dos exames, que de acordo com as missões que lhes estão consignadas, garantem a sua realização.

2.3. A problemática dos exames em Portugal

O tema em epígrafe suscita, naturalmente, a convergência das atenções da sociedade, nomeadamente de investigadores e estudiosos portugueses, pelas implicações que esta modalidade de avaliação externa pode ter tanto a nível pedagógico como a nível social. É claro, que os diferentes olhares sobre os exames se enquadram nas diferentes concepções de educação, de avaliação e de sistema educativo de que cada um partilha.

Delimita-se esta breve abordagem ao período que se iniciou em 1996, ano a partir do qual, com o (re)aparecimento dos exames obrigatórios no final do ensino secundário, surgiu, novamente, a necessidade de reflexão sobre a pertinência e o significado pedagógico da realização de uma avaliação externa das aprendizagens com função de certificação e seleção no sistema educativo português.

Refere-se, ainda, que serão apresentadas opiniões de algumas figuras políticas do país, bem como de investigadores e estudiosos portugueses.

A realização de exames nacionais obrigatórios a partir de 1996 decorreu da implementação da reforma curricular do ensino secundário, aprovada pelo Decreto-Lei n.º 286/89, de 29 de agosto, e da consequente aprovação de um novo regime de avaliação dos alunos desse nível de ensino (Despacho Normativo n.º 338/93, de 21 de outubro). A propósito das alterações introduzidas por este regime, Pacheco (2005) refere que, apesar de se verificar a reintrodução de exames obrigatórios, estes se revestem de moldes e finalidades diferentes das que existiam no tempo dos liceus. Nessa altura, este sistema de avaliação era utilizado para controlar os professores relativamente ao cumprimento dos currículos. Os exames eram obrigatórios para todos os alunos (exceto para os que deles dispensassem), tendo um carácter autónomo relativamente à progressão, fechando um ciclo de validação. Já o controlo proposto pelo Despacho Normativo n.º 338/93, de 21 de outubro, tinha a intenção de homogeneizar as classificações do ensino secundário, não obstante ter implícita uma valorização da avaliação sumativa externa com função de certificação e as consequências que dela poderão advir.

Face ao impacto produzido pelo ressurgimento deste sistema de avaliação externa, Leal (1997) escreveu um artigo que tinha em vista duas finalidades: por um lado, pretendia suscitar uma reflexão sobre a problemática dos exames e, por outro, manifestar as razões pelas quais discordava deste processo avaliativo. Para o efeito, a autora rebateu

os argumentos utilizados, por muitos, a favor dos exames. Apresentar-se-á, de forma resumida, a análise crítica da autora, em função de cada um desses argumentos.

- (I) **Qualquer avaliação sumativa diferencia os alunos hierarquizando-os de acordo com uma escala de valores.** Para Leal, a hierarquização resultante dos exames não revela as diversas capacidades do aluno, sendo meramente informativa e assumindo um caráter apenas social.
- (II) **Os exames promovem a equidade, uma vez que os alunos são sujeitos a uma mesma situação.** A autora põe claramente em causa esta situação de igualdade, uma vez que os alunos a quem se dirige o exame para além de, por natureza, já apresentarem diferentes capacidades e características, são também possesores de experiências educativas diferentes.
- (III) **Os exames garantem a qualidade do ensino.** Um ensino de qualidade terá de promover o desenvolvimento, por exemplo, de capacidades e de competências nos alunos. Sendo os exames aplicados apenas no fim do processo de aprendizagem, não têm qualquer influência nesse mesmo processo. Mesmo considerando a sua utilidade sobre os alunos vindouros, isso pressuporia uma análise, por exemplo, da qualidade dos itens e das principais dificuldades dos alunos e o que se faz é uma leitura simplista e, por vezes, distorcida da realidade.
- (IV) **Os exames dão credibilidade ao sistema, a qual está associada à garantia de objetividade.** A avaliação é sempre uma atividade humana, que envolve uma apreciação e que, portanto, entra na dimensão da subjetividade. Querer separar a avaliação do avaliador é olhar para a avaliação como sendo uma medida. Como refere Chevallard (1989, p.19) citado em Leal (1997,s/p) “Contudo, este avaliador que queremos esquecer, lembra-se de nós. Porque o avaliador, felizmente, não faz outra coisa do que se enganar.”
- (V) **A existência de exames conduz ao cumprimento do programa.** Considerando que os programas da reforma curricular de 1989 contemplam objetivos de conhecimento, capacidades e atitudes, um exame, por exemplo, do 12.º ano não permite avaliar as capacidades e as atitudes. Desta forma está a enfatizar-se o saber, desvalorizando o saber fazer e o ser.

A autora alerta ainda para uma outra problemática relacionada com o efeito dos exames ao nível dos possíveis desvios causados no sistema educativo, lembrando uma expressão de Romberg e Zarinnia: “a avaliação tem um impacto directo quer naquilo que se ensina, quer no modo como se ensina” (Leal, 1997, s/p). Nesta linha, refere alguns efeitos que os exames poderão ter ao nível da escola, do professor e do aluno:

- (I) **Escola** - Defende-se a ideia de que a escola é tanto melhor quanto mais elevados forem os resultados dos exames dos seus alunos, o que põe em causa os objetivos educativos da LBSE, pois leva a uma diferenciação nos mesmos, dada a tendência de valorizar mais os que se relacionam com o prosseguimento dos estudos em detrimento dos objetivos que visam o desenvolvimento pessoal e a preparação para a inserção na vida ativa.
- (II) **Professor** - Os exames irão ter um efeito no professor e, conseqüentemente, a nível das suas aulas e das suas práticas. O professor terá uma maior tendência para trabalhar mais as competências mais quantificáveis, em detrimento de competências de nível superior, como por exemplo o raciocínio e a comunicação, por serem “difíceis de serem consideradas numa prova individual de papel e lápis” (Perrenoud, citado em Leal, 1997, s/p).
- (III) **Aluno** - Relativamente ao aluno, é de esperar que, apropriando-se este do tipo de exame a que será sujeito, a sua tendência será para canalizar o seu trabalho para os assuntos que considera poderem ser testados. Desinteressar-se-á de tarefas que contribuam para o desenvolvimento da sua autonomia e da sua responsabilidade. Poderá, ainda, ser estimulado no aluno um espírito de individualismo e de competitividade que pode ser questionável na sociedade que se pretende desenvolver. Como refere Perrenoud, “O sistema de avaliação clássico favorece uma relação utilitária com o saber. Os alunos trabalham para a nota” (Leal, 1997, p. 3).

Em suma, na perspetiva de Leal (1997), os exames são uma forma de avaliação nada pertinente que representa um regresso à avaliação tradicional, podendo vir a ser responsável por enviesamentos no sistema escolar.

Fernandes (2007a), na mesma linha de Leal (1997) e Perrenoud (1992), refere, que um exame escrito de duração limitada dificilmente poderá avaliar as várias dimensões do currículo nacional, isto é, o saber estar, o saber ser e o saber utilizar, e as respetivas

competências, uma vez que estas incluem a mobilização de saberes, a investigação, a pesquisa e a comunicação oral.

No entanto, os exames, apesar de não serem um sistema de avaliação que tenha como objetivo a melhoria das aprendizagens dos alunos e de terem limitações aos níveis pedagógico e da validade e fiabilidade, assumem um peso determinante na certificação dos alunos, já que as sociedades não conseguem encontrar alternativas credíveis (Fernandes, 2007b).

Embora os procedimentos utilizados nas fases de conceção, aplicação e classificação dos exames possam garantir a equidade e a justiça entre os alunos, sabe-se que são as avaliações internas as que podem dar informações mais detalhadas sobre os conhecimentos dos alunos. As razões para que estas avaliações não sejam exclusivas na certificação das aprendizagens prendem-se com questões relativas a consistência e validade ou com dificuldades, como a da complexidade em sintetizar a diversidade de processos de aprendizagem que ocorrem na sala de aula. Assim, quando é necessário refletir sobre os saberes que o sistema de ensino português desenvolve nos alunos, recorre-se aos resultados das avaliações externas. No entanto, constata-se, que estas, em Portugal, têm vindo a dar mais relevância a competências no domínio do raciocínio, da resolução de problemas e da comunicação e que as provas incluem itens do tipo resposta aberta, mais adequados a avaliar processos complexos do pensamento (GAVE, 2006). Não obstante as informações proporcionadas pela avaliação externa serem importantes, há questões que se levantam e que devem ser alvo de análise e reflexão, como por exemplo: Que currículos e saberes são avaliados pelos exames? Haverá alunos favorecidos ou desfavorecidos pelos exames? A correção tem níveis de consistência aceitáveis, isto é, há fiabilidade entre os corretores? Que consequências se retiram da análise dos resultados dos exames ao nível dos saberes dos alunos e ao nível do funcionamento do sistema educativo? (Fernandes, 2008a)

Fernandes (2009) considera que pouco se tem feito no âmbito da avaliação externa das aprendizagens, o que constitui

uma lacuna relevante tendo em conta a necessidade de se compreender uma diversidade de questões críticas relativas à avaliação externa dos alunos (e.g., exames, análise de resultados, análise de itens, grau de dificuldade, fiabilidade, validade, consistência curricular, natureza das questões, relações com a avaliação interna, relações com o ensino, efeitos nas práticas do ensino e da avaliação) (p. 89).

A realidade é que uma boa avaliação, sendo ela um bom processo de regulação e de melhoria, ajuda a compreender melhor uma dada realidade, pode contribuir para a melhorar, mas tem sempre de se reconhecer os seus limites e perceber a relevância da utilização que fazemos dos seus resultados (Fernandes, 2007b).

David Justino, ex-ministro da educação (2002-2004), no seu livro *Difícil é Educá-los*, apresenta algumas reflexões sobre os problemas da educação em Portugal entre as quais os relativos à avaliação externa das aprendizagens.

Considera Justino (2010), que a introdução das avaliações externas surgiu na sequência das avaliações internas, pelo facto de a avaliação interna se caracterizar por ser muito personalizada na relação professor-aluno e de se revestir de grande subjetividade relativamente aos objetivos e às exigências da aprendizagem visando aferir os níveis de desempenho entre os alunos e as escolas. Com efeito, como sublinha Fernandes (2008a), estas avaliações externas possuem propriedades psicométricas que permitem comparações entre escolas e regiões do país. Para Justino (2010), as avaliações externas são, por natureza, sumativas, por decorrerem num determinado momento do percurso educativo, e são restritas, por cobrirem apenas uma parte dos conhecimentos e competências adquiridos pelos alunos, de forma a avaliar se o nível de desempenho está de acordo com os objetivos de aprendizagem previstos. Para além da sua função de aferição, a avaliação externa ainda expressa um referencial a atingir, podendo funcionar como motivação e, ainda, como um instrumento de regulação das aprendizagens, especialmente num sistema de ensino que visa impor determinados padrões de qualidade.

A verdade é que, de acordo com Justino (2010), desde o 25 de abril até 1996, não houve uma avaliação externa que cumprisse as funções supramencionadas, tendo-se dado primazia à “avaliação contínua”, o que “foi um erro de consequências dramáticas para o sistema de ensino em Portugal” (Justino 2010, p. 86). O facto de os primeiros exames terem sido realizados a nível do 12.º ano, manifesta a necessidade de aferir classificações para uma maior justiça no acesso ao ensino superior. Justino (2010) reconhece, tal como Leal (1997) e Fernandes (2007a), que os testes nacionais e os exames apresentam alguns inconvenientes, nomeadamente, o facto de orientarem as aprendizagens para que haja sucesso nos exames, o facto de incidirem em determinados conteúdos e competências e a forma como são elaborados os testes. Mas o que importa é ter uma visão integrada da avaliação e compreender que os vários instrumentos da

avaliação, apesar de terem sempre as suas limitações, se completam e se esclarecem mutuamente e que as avaliações interna e externa se complementam.

Castilho (2011), a propósito da pertinência dos exames nacionais, refere, que não retirando à avaliação formativa a importante função de acompanhar e regular os processos de ensino e aprendizagem, os exames são um elo de ligação e de certificação entre a escola e a sociedade.

O que fazemos no remanso da sala ou da escola tem de ser validado exteriormente, face à sociedade que nos paga e tem o direito de exigir a adoção de instrumentos de relativização de juízos tão díspares de escola para escola, de professor para professor, num mesmo país. É para isso também, que devem servir os exames nacionais. Com inconvenientes, como tudo na vida; são incontornáveis (Castilho, 2011, p.41).

Na sequência das abordagens feitas, apraz dizer-se que as opiniões sobre a pertinência dos exames nacionais não são partilhadas de igual modo, dependendo, naturalmente, da perspetiva que cada um terá de educação e de sistema educativo.

2.3.1. *Rankings e accountability*

No âmbito da problemática dos exames nacionais não poderia ignorar-se a temática dos *rankings*, um dos assuntos de polémico debate no seio da opinião pública. Para esta situação em muito têm contribuído os objetivos e as metodologias que lhes subjazem, bem como as consequências que os mesmos poderão ter tanto a nível da educação como a nível social.

Em Portugal, a partir do ano 2000, quase duas décadas depois de outros países, iniciou-se a publicitação anual de *rankings*, um processo de hierarquização de escolas secundárias relacionado com a “cultura de avaliação.” (Santiago, Correia, Tavares & Pimenta, 2004).

A multiplicidade de reações aos *rankings* levou à publicação de um livro de apreciação crítica sobre o tema, em 2004, no qual os respetivos autores referem algumas das razões

que põem em causa os *rankings*. A este propósito, apresentaram diversos argumentos, tais como: “Não existe um modelo explicativo que suporte a seriação feita”; “existem erros matemáticos que afectam o resultado”, “chegaram a uma seriação mas poderiam ter chegado a outras seriações bastante diferentes percorrendo caminhos similares”. (Santiago et al., 2004, p. 40).

De acordo com Santiago et al. (2004), o suporte argumentativo para a publicitação dos *rankings* foi o de permitir:

- (I) formar uma ideia mais clara acerca da “qualidade” da educação;
- (II) a escolha, pelas famílias, das escolas mais adequadas à concretização dos seus objetivos;
- (III) que as escolas monitorizassem melhor os seus processos e os seus resultados, tornassem a sua informação mais transparente e cumprissem o seu dever de responsabilidade social perante as famílias.

As publicitações têm suscitado reações divergentes na sociedade, havendo opiniões favoráveis e desfavoráveis às mesmas.

As reações defensoras dos *rankings* consideram que estes são uma ajuda, tanto para os alunos como para os pais, para a escolha da escola. São, ainda, um meio de fornecer às escolas informação para melhorarem a sua qualidade e para que prestem contas sobre a forma como gastam o dinheiro público (Korsten, Visscher & Jong, citado em Santiago et al. 2004).

De igual modo, Matos, Lopes, Nunes e Venâncio (2006) referem que os *rankings*, tecnicamente, são apenas ordenações das escolas, em função da média das respetivas classificações nos exames, acabando por ser, na prática, considerados indicadores de qualidade ou de eficácia das escolas.

Esta perspetiva foi partilhada por alguns atores políticos e educativos, tais como o ex-ministro da educação, David Justino, quando refere “há o direito dos cidadãos à informação”, “a minha preocupação foi proporcionar o máximo de informação para contextualizar os resultados dos exames” (Fernandes & Leiria, 2002, s/p).

Na mesma linha de pensamento, encontra-se o ex-ministro da educação Marçal Grilo, ao referir que

a apresentação pública dos resultados é importante a dois títulos: porque, no mesmo ano, as escolas se podem comparar entre si. Mas também para que cada escola, ao longo dos anos, se possa analisar a si própria. Esta análise deve ser um elemento de reflexão, em particular para a escola e para todos os seus responsáveis pedagógicos, professores, pais, estudantes (Grilo, 2003, s/p).

Já as reações desfavoráveis à publicitação dos *rankings* consideram que os dados fornecidos dão origem a leituras muito simplistas.

Como refere Silva (2000)

a ordenação de escolas pela retenção de um único e sumário indicador (no caso do ensino secundário, as notas dos alunos) empobrece e distorce a informação. A educação não é um campeonato de escolas. O desempenho de uma escola não se verifica numa só dimensão: a qualidade dos processos de ensino, da gestão e do ambiente organizacional é determinante para o desenvolvimento pessoal dos estudantes e o trabalho dos professores. O mérito da escola não se mede apenas pelas notas dos alunos que leva a exame: mede-se ainda pela quantidade dos alunos que leva a exame, pela capacidade de combater o abandono e a exclusão, pela oferta de novas oportunidades de ensino, pela diversificação dos cursos, pela ligação à comunidade e ao mundo profissional, etc., etc... (s/p).

Após um breve relance sobre estes diferentes pontos de vista, constata-se que o tipo de informação prestada pelos *rankings* é, de facto, parcelar e, portanto, redutora, uma vez que está canalizada para os resultados académicos, excluindo tantos outros aspetos do contexto escolar que influenciam o processo de ensino e de aprendizagem e que, por esta mesma razão, não podem ser descurados.

Estas situações conduzem à reflexão sobre os contextos em que os *rankings* são publicitados, os objetivos que presidem a esta publicitação e os processos de tratamento de dados.

Com base na perspetiva de Santiago *et. al* (2004), os *rankings* poderão apresentar um conjunto de consequências a nível educacional e social, entre as quais:

- (I) a segregação das escolas menos escolhidas pelas famílias, em resultado da hierarquização daquelas;
- (II) levar as escolas a trabalhar mais para os resultados (performances cognitivas dos alunos);
- (III) incentivar os alunos à competitividade;
- (IV) conduzir a uma estigmatização social dos “maus” alunos, resultante da agregação social das escolas;
- (V) levar os professores a olhar os *rankings* como uma forma de ser avaliado o seu trabalho.

Para alguns professores, os *rankings* servem para adotar estratégias pedagógicas que permitam melhorar as posições das escolas em que trabalham nesses *rankings*, mas outros estarão mais preocupados com a “imagem que é projetada sobre a instituição de ensino e, também, a sua própria imagem enquanto docentes” (Melo, citado em Afonso, 2011, p. 92).

Nas famílias a publicitação dos *rankings* conduz a uma desigualdade na escolha da escola pelas mais desfavorecidas. No entanto, as famílias das classes mais favorecidas terão vantagem em todo este processo, pois usam melhor a informação proporcionada pelos *rankings*.

De acordo com um estudo feito por Matos, Lopes, Nunes e Venâncio (2006) “há uma inconsistência inerente à forma como os *rankings* são construídos (médias pesadas), que é agravada pela utilização dos valores brutos das classificações, devido à grande variação, inter e intra disciplinas, do grau de dificuldade dos exames” (s/p).

O estudo aponta ainda para a necessidade de não sobrevalorizar os *rankings*, não deixando, no entanto, de considerar que a informação que fornecem é útil, permitindo que as escolas se situem e permitindo que se comparem situações comparáveis (e.g. escolas privadas de Lisboa).

Os estudos sobre a problemática dos *rankings* têm sido fundamentais para perceber algumas questões subjacentes às políticas educativas nos últimos anos. No entanto, segundo alguns autores, o debate sobre os *rankings* em Portugal, ao contrário do que ocorreu noutros países, terá sido induzido mais por “uma reivindicação em nome da

necessidade da prestação de contas e da transparência” do que por uma “opção inequívoca por uma orientação das políticas educativas para o mercado” (Mendes, Costa & Ventura, 2003, s/p).

Neste contexto, Afonso (2009, 2011) coloca-nos perante duas formas de *accountability*³ no nosso sistema educacional: uma sustentada nos exames nacionais, estandardizados, que, tal como em outros países, se constitui como uma forma de controlo por parte do Estado e uma outra centrada nos *rankings* das escolas, da iniciativa da sociedade e do mercado, impulsionada pela comunicação social.

Estas formas parcelares de *accountability* funcionam de forma desarticulada entre si, não existindo, assim, um sistema de *accountability* no nosso país. Esta situação poderá estar relacionada com as pressões sociais de ordem ideológica que se têm vindo a sentir, nos últimos anos, sobre os professores, impedindo uma intervenção aberta e adequada que pudesse promover a articulação entre a avaliação, a prestação de contas e a responsabilização, contribuindo para a consecução de uma verdadeira *accountability*. No entanto, há orientações políticas, normativos legais e expectativas sociais e práticas em curso que poderão vir a evoluir e a contribuir para a existência da mesma.

A sociedade tem direito à informação e a saber o que ocorre nas escolas. No entanto, considera Afonso (2009, 2011), que não é com a avaliação externa, nomeadamente os exames com funções redutoras e de controlo social, que se caminha para a interligação da avaliação, da prestação de contas e da responsabilização adequadas aos objetivos, às estruturas e aos atores do nosso sistema educativo.

Certamente que os exames nacionais poderão cumprir alguns objetivos importantes, mas não estes. Como referiu Darling-Hammond, “os testes são uma fonte de informação para um sistema de *accountability*, mas não são em si mesmos esse sistema” (Afonso, 2009, p.67).

Segundo Abrantes (2010), verifica-se, desta forma, a centralidade dos exames nacionais na configuração das políticas educativas atuais aproveitando-se da divulgação pública dos resultados gerais das provas.

³ O conceito de *accountability* é polissémico e denso. Porém, o autor pretende fazer uma breve contextualização teórica-conceptual que permita dar sentido à perspetiva que adotou, ou seja, associar o conceito a três dimensões essenciais autónomas, mas fortemente articuladas: a avaliação, a prestação de contas e a responsabilização (Afonso, 2011).

2.4. As informações dadas, nos últimos anos, pelos exames

Parece oportuno, no âmbito deste trabalho, fazer referência ao conjunto de informações que tem sido fornecido, nos últimos anos, pelos exames nacionais.

Neste âmbito, Fernandes (2008b) apresenta alguns estudos sobre os exames nacionais que foram realizados pelo Ministério da Educação e pela então Diretora do GAVE, no sentido de obter informações sobre os saberes dos alunos portugueses. Tendo em conta o âmbito deste trabalho, far-se-á referência, unicamente, aos estudos relativos aos exames do ensino secundário.

Em 2002, o GAVE realizou um estudo destinado a investigar as competências manifestadas pelos alunos, nos exames de 12.º ano realizados entre 1999 e 2001. Foram analisados os resultados nas seis das disciplinas com um maior número de alunos: Biologia, Física, Matemática, Psicologia, Química e Introdução ao Desenvolvimento Económico e Social.

A análise, feita por item, baseou-se nas competências definidas nos programas das disciplinas, nas classificações obtidas e em quatro indicadores destinados a categorizar o nível de desempenho dos alunos em cada item. Concluiu-se que:

1. Os alunos apresentam um desempenho médio ou mesmo bom quando o que está em causa são saberes que exigem pouca elaboração cognitiva, tais como o conhecimento de conceitos, de procedimentos, de factos ou a mera reprodução de informação.
2. Os alunos apresentam um desempenho fraco ou muito fraco quando lhes é pedido que apliquem conhecimentos que integram e mobilizem saberes em situações novas ou que elaborem raciocínios de natureza demonstrativa.

(Fernandes, 2008a, p. 285)

De acordo com Fernandes, (2008a), o próprio autor efetuou uma análise sumária dos resultados obtidos em 2006 nas disciplinas acima referidas a qual mostrou que havia razões para crer que os problemas suprarreferidos persistiam, em particular nas disciplinas de Matemática, Física, Química e Biologia, cujas médias se encontravam abaixo de 10 valores ou muito próximas desta classificação.

Na mesma linha de estudo, há também a salientar o trabalho de Ramalho, (Fernandes, 2008a), que, com base em relatórios nacionais e internacionais, apresentou uma síntese

e posterior análise dos resultados dos alunos portugueses em avaliações externas nacionais e internacionais. A nível nacional, Ramalho debruçou-se sobre (I) os exames nacionais do ensino secundário de Matemática, realizados entre 2000 e 2001 e os exames de Biologia, Física e Química, realizados entre 1999 e 2001; (II) provas de avaliação externa de Língua Portuguesa e Matemática para os 4.º e 6.º anos de escolaridade e (III) avaliações integradas das escolas, da responsabilidade da então IGE (atual IGEC), realizadas entre 1999 e 2000. Este trabalho permitiu apontar algumas tendências quanto ao desempenho dos alunos, tendo sido evidenciados problemas de aprendizagem consistentes com os referidos nos estudos realizados pelo GAVE. Foi o caso da resolução de problemas, da aplicação de conhecimentos a novas situações e da análise e interpretação de informação, em que o desempenho dos alunos era de facto fraco. No que concerne à reprodução de informação, ou no trabalho com textos narrativos e com tarefas rotineiras, o desempenho dos alunos apresentou uma tendência para ser médio ou bom (Fernandes, 2008a).

No âmbito das informações fornecidas pelos resultados das provas realizadas a nível nacional, Ramalho, enquanto diretora do GAVE, divulgou alguns dos processos envolvidos na elaboração das provas de exame produzidas por aquele gabinete, explicitando a finalidade do estudo do comportamento de alguns itens constituintes das referidas provas e de que forma os resultados da aplicação daqueles instrumentos deram indicações importantes sobre a situação do ensino e da aprendizagem em Portugal (Ramalho, 2002).

Relativamente às provas de exame de 12.º ano, Ramalho (2002) referiu que, para a sua elaboração, era tido como base os programas em vigor, sendo considerado, de entre as competências e os conteúdos, a relevância que as mesmas tinham no articulado programático. Além disso, havia ainda a preocupação que as provas assegurassem diferentes níveis de conhecimento relativamente ao objeto de avaliação e que uma classificação de 10 valores indiciasse uma aquisição de conhecimentos e competências fundamentais no domínio em causa.

Neste sentido, na elaboração dos exames, o GAVE seguia duas linhas de inquirição (Ramalho, 2002):

- (I) o estudo do comportamento de todos os itens integrados nas provas da 1ª e da 2ª chamadas (lançou-se e analisou-se os dados correspondentes a um

conjunto de provas previamente definido que representavam o universo das instituições em que tiveram lugar os exames nacionais);

- (II) o estudo antecipado de itens regularmente elaborados e que foram testados numa amostra da população dos alunos do 12.º ano.

Essa inquirição permitiu ao GAVE averiguar o comportamento de cada um dos itens incluídos nos exames, criando condições para uma reflexão mais rigorosa sobre as aquisições e as dificuldades dos alunos nas suas respostas aos itens. Desta forma foi possível obter informações importantes sobre a situação do ensino e da aprendizagem dos alunos, nas várias disciplinas (Ramalho, 2002).

Na sequência da apresentação dos estudos acima referidos, considera-se relevante e pertinente referenciar a principal informação veiculada pelos relatórios dos exames nacionais, da responsabilidade e da autoria do GAVE, resultantes de uma recolha e análise de informação obtida com base nos resultados dos exames nacionais.

Para o efeito considerar-se-á os relatórios publicados entre 2009 e 2011 e somente as informações respeitantes aos exames nacionais do ensino secundário.

Assim, a partir da implementação dos desenhos curriculares do ensino secundário definidos pelo Decreto-Lei 74/2004, de 26 de março, o GAVE iniciou um trabalho de recolha e análise de informação centrado nos resultados dos exames nacionais. Embora este trabalho tenha tido início em 2006, o primeiro relatório apenas foi divulgado em 2009, por só neste ano se ter considerado que estavam reunidas as condições necessárias para o efeito. Seguiram-se os relatórios dos anos subsequentes, 2010 e 2011.

Ainda que os três relatórios se focalizem no desempenho dos alunos nos exames nacionais, a estrutura adotada e os aspetos versados sofreram alterações. Enquanto o relatório de 2009 (GAVE, 2010b) se limitava a apresentar uma identificação de itens com melhor ou pior desempenho, por área disciplinar, o relatório de 2010 (GAVE, 2011) já se encontrava estruturado em duas partes distintas. Na 1.ª parte, procedia-se, por disciplina, a uma caracterização do teste, a partir da qual se identificava os itens com melhor ou pior desempenho e se apresentava sugestões de intervenção didática; na 2.ª parte, apresentava-se uma comparação dos resultados georreferenciados ao nível do NUTIII (unidades territoriais estatísticas de Portugal), que incluíam 7 regiões do país (Norte, Centro, Lisboa, Alentejo, Algarve, Açores e Madeira), divididas em 30 sub-regiões. O

relatório de 2011 (GAVE, 2012) apresentava uma estrutura similar ao do ano de 2010. No entanto, disponibilizava, ainda, informação quantitativa, por disciplina e por região, bem como o enquadramento de cada região em relação aos resultados médios e à variação dos resultados de 2009 a 2011, tendo por referência o valor médio nacional.

As informações divulgadas nos três relatórios foram extraídas de uma análise continuada de indicadores de dificuldade e de discriminação dos itens que integram os exames. Visam evidenciar algumas das razões explicativas do nível de desempenho dos alunos, pretendendo que as escolas possam refletir e agir no sentido de encontrar estratégias que contribuam para uma melhoria das aprendizagens de forma a melhorar o sucesso. Neste âmbito, o GAVE tem disponibilizado, ainda, a cada escola (desde 2009), os respetivos resultados por exame e por item, para que cada organização se possa situar face ao referente nacional, e para que os professores de cada escola possam adequar as estratégias educativas à sua realidade.

Ao longo dos diversos relatórios, o GAVE procurou sensibilizar todos os intervenientes do processo educativo para os efeitos positivos que podem advir de avaliação externa e do contributo que esta pode dar para a melhoria sustentada do desempenho dos alunos (GAVE, 2010b, 2011, 2012).

Tendo em conta as finalidades desta dissertação, considera-se pertinente particularizar as informações relativas à disciplina de Biologia e Geologia - 702, no ano de 2011, as quais são seguidamente apresentadas.

Desta forma, considerando o relatório do GAVE (GAVE, 2012), para a análise dos resultados da prova escrita de Biologia e Geologia (1.^a fase), foram tidas em conta as respostas de 28348 alunos (internos). A média nacional foi de 11,0 valores.

A prova refletiu uma visão integradora dos conteúdos previstos no programa da disciplina, tendo contemplado a mesma ponderação nas componentes de Biologia e de Geologia. Era constituída por 4 grupos, que tinham como suporte textos e esquemas.

Relativamente à tipologia de itens, a prova apresentava itens de seleção, incluindo itens de escolha múltipla, itens de associação e itens de ordenação, e itens de construção de resposta restrita. Os itens permitiram a avaliação da aprendizagem nos domínios concetual e procedimental.

Os alunos revelaram melhor desempenho em três itens de seleção de escolha múltipla, “com valores de classificação média em relação à cotação total de, respetivamente, 84%,

95% e 90,8%” (GAVE, 2012, p. 25). Um dos itens avaliava, simultaneamente, a mobilização e a utilização de dados, conceitos, modelos e teorias; outro avaliava a interpretação de procedimentos experimentais fornecidos e um último avaliava o conhecimento e a compreensão de dados, conceitos, modelos e teorias.

Já os itens em que os alunos revelaram pior desempenho foram os de resposta restrita (dois itens) mas também em itens de seleção de escolha múltipla (um item), “com valores de classificação média em relação à cotação total de, respetivamente, 25,9%, 24,4% e 16%” (GAVE, 2012, p. 26). Nos itens de resposta restrita, num avaliava-se a explicação de contextos em análise, enquanto no outro “as dificuldades parecem ter estado relacionadas não tanto com os conteúdos programáticos, mas com fragilidades de carácter transversal, nomeadamente, o estabelecimento de relações entre conceitos, a interpretação de dados de um suporte e a aplicação a novas situações” (GAVE, 2012, p. 26). Relativamente ao item de escolha múltipla, avaliava o conhecimento e a compreensão de dados, conceitos, modelos e teorias.

Foi assim possível concluir que, de um modo geral,

os baixos níveis de desempenho estão muitas vezes associados a fragilidades no domínio da compreensão e da interpretação de suportes, ao estabelecimento de relações entre conceitos, à aplicação de conceitos, modelos e teorias a novas situações, à síntese de resultados e estabelecimento de conclusões e à produção de texto escrito organizado, com correção formal e de conteúdo (GAVE, 2012, p. 26).

Perante as fragilidades detetadas, foram apresentadas propostas de intervenção didática, as quais passavam

por um reforço do ensino e da aprendizagem de conteúdos, requerendo a reconceptualização de pré-conceitos dos alunos e apostando na diversificação de experiências educativas, na resolução de atividades que promovam a análise de situações-problema e na construção de textos a partir da seleção e análise de informação (GAVE, 2012, p. 26).

Foi ainda sugerido

um reforço nas metodologias que permitam uma maior consecução dos objetivos programáticos relacionados com a aplicação dos conhecimentos adquiridos a novos contextos e a novos problemas; o desenvolvimento de capacidades de seleção, de

análise e de avaliação crítica e a realização de inferências. Transversalmente, será de insistir em tarefas que envolvam a comunicação de resultados de investigação ou de pesquisas (GAVE, 2012, p. 26).

Em termos conclusivos, esta abordagem, realizada a partir de estudos feitos sobre os resultados das avaliações externas, nomeadamente dos exames nacionais, parece permitir a familiarização com questões que se prendem, por um lado, com os saberes dos alunos em Portugal e, por outro, com as dificuldades por eles apresentadas.

CAPÍTULO 3 – A CONSTRUÇÃO DE UM EXAME

Considerando o propósito desta investigação - saber qual a influência da tipologia dos itens nos resultados do exame de Biologia e Geologia de 2011 -, é imprescindível referir alguns aspetos a ter em conta na construção de uma prova de exame. Neste intuito, salientam-se os aspetos relativos à validade, destacando-se as questões relativas ao nível de dificuldade dos itens, e à fiabilidade de um exame, com relevo para o que é feito em Portugal no âmbito do controlo desta última característica.

Mais adiante referem-se os diferentes tipos de itens a utilizar numa prova de exame e as respetivas características, não descurando aspetos como o número de itens que um exame deve ter, a sua ordenação e a determinação do formato de um exame relativamente ao seu perfil (objetivo, subjetivo e misto). Nesta sequência é feita uma referência ao procedimento que o GAVE tem na divulgação das características dos exames nacionais de Biologia e Geologia, por ser o exame desta disciplina o objeto de estudo.

Por fim, e na perspetiva de verificar a evolução das provas de exame nacionais de Biologia e Geologia, é efetuada uma análise comparativa dessa provas desde a sua primeira realização.

3.1. Das modalidades de avaliação às técnicas de testagem. Os testes

Quando se fala em avaliação, fala-se de um processo que se pode efetuar de acordo com diferentes modalidades. Por um lado, pode ser o próprio indivíduo a avaliar os seus níveis de conhecimentos através de um processo de autoavaliação – avaliação interna;

ou pode ser o professor a emitir um juízo e a tomar uma decisão em relação à aprendizagem do aluno - avaliação externa. Neste contexto, fala-se em avaliação espacial. Por sua vez, o tipo de avaliação realizada pelo professor, poderá decorrer em vários momentos, isto é, no início do processo - avaliação diagnóstica, durante o processo - avaliação formativa ou no fim do processo - avaliação sumativa. Neste caso fala-se em avaliação temporal (Pacheco, 1994).

A avaliação diagnóstica poderá ocorrer, no início do ano letivo, ou no início de uma unidade de ensino ou de um determinado conteúdo, permitindo um levantamento dos conhecimentos do aluno, constituindo-se como um indicador para o professor.

A avaliação formativa tem a função de determinar qualitativamente o processo de aprendizagem e fornecer *feedback* para a sua própria regulação, permitindo fazer ajustamentos e melhoramentos em função das necessidades.

A avaliação sumativa está ligada à medição do grau de consecução do aluno no final de um processo e tem a finalidade de classificar e certificar. Neste âmbito, o desempenho do aluno pode ser visto em termos da posição relativa que ocupa em relação a um determinado grupo, não esquecendo que ao se considerar este modelo de avaliação por referência à norma, não deve ser esquecido que, como refere Ferraz et al. (1994, s/p), “as normas variam em função dos grupos que as definem e das reinterpretações pessoais”. Por outro lado, pode avaliar-se o aluno em função de objetivos previamente definidos – avaliação por referência a um critério -, em que o nível de realização do aluno se situa num determinado ponto de um conjunto de tarefas específicas do ensino, razão pela qual se lhe atribui uma nota que expressa o seu rendimento (Pacheco, 1994; Santos et al., 2010).

As perspetivas normativa e criterial da avaliação são abordagens complementares, ou por vezes alternativas, dependendo das finalidades do processo avaliativo, como referem Ferraz et al. (1994).

Ainda em relação à avaliação sumativa, refere Fernandes (2011), que esta pode ter dois propósitos: “recolher informação acerca do que os alunos sabem e são capazes de fazer, que será utilizada para lhes atribuir uma classificação” e “num dado momento, fazer um balanço do que os alunos aprenderam, mas sem quaisquer preocupações classificatórias” (p. 94). Defende o autor, que quando utilizada com este último propósito, a avaliação sumativa se aproxima da formativa, podendo constituir um meio de melhoria e de regulação ou autorregulação das aprendizagens.

Independentemente das perspetivas consideradas, a avaliação está dependente das técnicas, dos instrumentos e dos procedimentos que se utiliza para obter informação em relação a um dado processo.

Deste modo, sendo as técnicas bastante diversificadas, impõe-se que os professores as selecionem de acordo com o que consideram mais adequado às necessidades dos alunos e aos propósitos da avaliação.

Na área cognitiva, de entre os instrumentos utilizados, destacam-se as técnicas de testagem que incluem o teste escrito (Domingos, Neves & Galhardo, citado em Pacheco, 1994).

Nesta sequência far-se-á, no ponto seguinte, uma alusão aos diferentes tipos de testes.

3.1.1. Tipos de testes

De acordo com Tenbrink (2010), o teste é uma das técnicas utilizadas para obter uma informação avaliativa, proporcionando, geralmente, uma informação mais exata e mais eficaz que qualquer outra técnica. Estas vantagens prendem-se com o facto de os especialistas considerarem, que os testes apresentam um conjunto de características responsáveis pela sua standardização e pela objetividade na sua pontuação.

Os testes são instrumentos que se caracterizam por apresentarem: (I) uma situação comum a que todos os indivíduos respondem, (II) um conjunto de instruções comuns que orientam todas as respostas, (III) um conjunto de regras comuns para pontuar as respostas e (IV) uma descrição numérica de cada realização de cada indivíduo.

Quanto à decisão para a escolha do teste pressupõe-se que se tenha tido em conta: (a) o tipo de informação necessária, (b) a quantidade de informação precisa e a (c) quantidade de tempo disponível (Tenbrink, 2010).

De entre os testes destacam-se os que são elaborados pelos professores e os testes standardizados.

Testes elaborados pelo professor

Os testes elaborados pelo professor enquadram-se no tipo de testes referentes a critérios, uma vez que o que está em causa é o progresso dos alunos e o desejo de se tornarem proficientes nas aprendizagens que lhes são propostas (Ribeiro, 1990).

De acordo com Tenbrink (2010), podem utilizar-se para medir todos os níveis de âmbito cognoscitivo: conhecimento, compreensão, aplicação, análise, síntese e avaliação. No entanto, em alguns casos, tornam-se redutores por se limitarem a medir o nível do conhecimento, fornecendo, inquestionavelmente, uma informação incompleta do desempenho dos alunos. É claro que, independentemente dos desempenhos avaliados por qualquer teste, a informação recolhida deve ser complementada com a informação conseguida através do recurso a outras técnicas de informação complementares, não devendo estas ser descuradas no processo avaliativo.

Um outro aspeto destes testes é que se apresentam muito versáteis, podendo ser aplicados em qualquer momento do processo avaliativo, permitindo, deste modo, formular juízos sobre o desempenho dos alunos, mesmo em unidades didáticas pequenas. Consequentemente, qualquer situação de erro poderá ser desde logo corrigida, ou não, dependendo do uso que o professor fizer com a informação recolhida pela aplicação do instrumento. Neste sentido, Harlen, citada por Domingos (2011), defende que

é complicado utilizar os dados produzidos no âmbito da avaliação sumativa para efeitos formativos, uma vez que as avaliações sumativas: a) se destinam a sintetizar as aprendizagens dos alunos com base em critérios gerais; b) não têm normalmente o propósito de identificar e de orientar os passos a dar a seguir; e c) ocorrem com pouca frequência (p. 91).

Testes estandardizados

De acordo com Tenbrink (2010), os testes, para serem considerados estandardizados, têm de apresentar um conjunto de características:

- (I) os itens constantes das provas devem ser os mesmos e devem ser aplicados num formato equivalente a todos os examinandos;
- (II) as condições de administração da prova devem ser comuns, pelo que as instruções de aplicação devem ser as mesmas, exceto nos casos particulares de alunos que apresentam limitações visuais, auditivas ou motoras;

- (III) o procedimento e o processo de revisão e atribuição de pontuações devem ser uniformes, não havendo lugar a discriminação nos critérios a aplicar;
- (IV) a forma como os resultados das provas são divulgados deverá ser uniforme.

A medição em educação tem sido marcada por uma procura constante da possibilidade de comparar, de forma válida, os resultados entre indivíduos, entre escolas e entre países, levando, deste modo, ao desenvolvimento das condições estandardizadas das provas.

Uma das expressões atuais mais representativas das dimensões e evolução destas provas são as provas de estudos comparativos internacionais como o TIMSS e o PISA (Froemel, 2009).

Os testes estandardizados são instrumentos destinados a obter muitos tipos de informação em condições *standard*. A informação fornecida por estes testes é particularmente valiosa em situações de tomada de decisões importantes como, por exemplo, a seleção de indivíduos, o diagnóstico de problemas ou a predição de êxito Tenbrink (2010). Em termos da sua construção são muito laboriosos.

Na sequência do referido, será importante saber que informações fornecem os testes estandardizados. De acordo com Tenbrink (2010) a informação obtida a partir de um teste estandardizado apresenta três características: (i) está claramente definida, (ii) tem parâmetros conhecidos, e (iii) obtém-se em condições ideais.

Relativamente à primeira característica, e de acordo com o autor, a informação está organizada de acordo com uma tabela de especificações. Nesta, encontram-se claramente definidos o número de itens necessários para medir conhecimentos ou comportamentos, em função de um conjunto de objetivos específicos claramente definidos e de um constructo teórico determinado cuidadosamente.

Em relação à segunda característica, há a referir, que os testes estandardizados são aplicados a grandes amostras de indivíduos distintos, com as mais variadas características, isto é, diferentes idades, diferentes cursos, níveis socioculturais distintos, regiões geográficas diferentes, entre outros. A partir dos resultados obtidos, pode analisar-se parâmetros como o rendimento médio de cada grupo em função da idade, a estimação da validade, a estimação da consistência interna, etc. Cada um destes

cálculos serve para fazer uma aproximação aos resultados que se esperaria obter se se aplicasse o teste a um grupo semelhante de pessoas em condições parecidas.

No que concerne à terceira e última característica, os responsáveis pelos testes estandardizados asseguram que as condições sejam as ideais para quem os realiza, garantindo, por exemplo, a clareza das instruções e orientações para controlo do tempo.

De entre os testes estandardizados, destaca-se dois tipos: o teste de rendimento e o teste de aptidão. Para que se deve usar um tipo concreto de teste estandardizado?

(I) Os testes de rendimento

A maior parte das decisões em educação centram-se no rendimento do aluno. Ora, um teste de rendimento desenha-se para medir o nível atual dos conhecimentos do aluno, as suas realizações e as suas habilidades. Existe uma ampla variedade de testes deste tipo como por exemplo teste de diagnóstico, teste para uma única matéria, teste para várias matérias, sendo escolhido o que melhor se adequará à informação que se pretende obter concretamente.

No entanto, o professor não se deve esquecer de que os resultados destes testes devem complementar os resultados de outros testes já realizados (Tenbrink, 2010).

(II) Testes de aptidão

Este tipo de teste é adequado para formular juízos preditivos. Quando se pretende selecionar alunos para uma tarefa concreta, é necessário conhecer a sua aptidão para essa tarefa. Ao permitirem uma estimativa da aptidão do aluno, estes testes tornam-se valiosos na tomada das decisões educativas. Os testes estandardizados de aptidão são, deste modo, uma ajuda para avaliar os indivíduos, os programas educativos e para melhorar o ensino (Tenbrink, 2010).

3.1.2. As qualidades de um bom exame (teste estandardizado)

As qualidades de um bom instrumento de medição, como, por exemplo, um exame, passam pela sua validade e pela sua fiabilidade (Goring, 1981; Tenbrink, 2010).

3.1.2.1. A validade: conceitos

A validade corresponde ao grau de precisão com que um instrumento de avaliação mede o que pretende medir. É uma característica pertencente não ao instrumento de avaliação, mas sim aos resultados duma medição; quer isto dizer, que os resultados podem ter uma certa validade para um determinado uso, mas não para outro (Goring, 1981; Ribeiro, 1990).

De igual modo, Fernandes (2008) e Froemel (2009) consideram que tradicionalmente um teste é válido se avalia aquilo para que foi construído.

Na literatura, encontramos referências a vários tipos de validade, a saber: (I) a validade de previsão, isto é, em que medida um teste indica futuros desempenhos do indivíduo; (II) validade concorrente, isto é, em que medida os resultados de um teste se correlacionam com os resultados de outro teste ou da avaliação das mesmas ou outras aprendizagens; (III) a validade de critério, que possibilita ver em que medida o teste permite prever o desempenho relativo a um critério; (IV) a validade do constructo, ou seja, em que medida o teste avalia de forma adequada o constructo, a competência subjacente ou se o teste se apoia numa teoria psicológica ou educativa consistente; e (V) a validade de conteúdo, aquela que mais interessa ao professor nas suas avaliações educativas e que permite ver em que medida um teste contém uma amostra significativa do(s) conteúdo(s) que foram ensinados.

Apesar desta diversidade de tipos de validade, a verdade é que os investigadores se referem a um ou dois tipos, principalmente a validade de conteúdo.

Constatou-se, no entanto, que este conceito tem evoluído, que a validade é assumida, atualmente, como “um conceito unitário em que o constructo (a competência ou a aprendizagem subjacente) é o tema unificador.” (Gipps, citado em Fernandes, 2008b, p. 113).

Esta ideia é reforçada por Gronlund e Linn (citados por Fernandes, 2008b), uma vez que estes consideram que a validade não está apenas confinada ao teste em si mesmo, mas tem ainda que traduzir as interpretações que se fazem dos resultados e ter em conta as consequências destes resultados. Há, portanto, uma procura destes autores em contextualizar o conceito de validade.

Nesta mesma linha, Messick (citado em Fernandes, 2008b) amplifica o conceito de validade, considerando que o mesmo integra um valor social, uma vez que tem de ter em

conta a natureza e o conteúdo das inferências feitas pelo corretor a partir dos resultados do exame e as consequências das decisões avaliativas.

Este conceito de Messick é reforçado com os resultados obtidos numa versão *estandard* para a medição em psicologia e educação de entidades profissionais norte-americanas, AREA⁴, APA⁵; e NCME⁶, na qual se mencionou que “existem cinco categorias de evidência acerca da validade a ser consideradas ao construir um argumento” (Froemel, 2009, p. 15).

Assiste-se, deste modo, a uma evolução do conceito de validade, uma vez que se considera que a validade não deve estar relacionada apenas com o exame em si mesmo ou com os resultados, mas com a interpretação desses mesmos resultados, com a sua utilização e com as consequências das decisões avaliativas. De um conceito de natureza psicométrica caminha-se para um conceito de natureza ética e social (Fernandes, 2008b).

Técnicas e critérios para garantir a validade

Para conseguir resultados com um grau aceitável de validade, quem elabora uma prova deverá garantir que tanto os conteúdos e suas proporções, como os comportamentos medidos estejam de acordo com o que foi tratado ao longo do curso (Fernandes, 2008b; Goring, 1981). O controlo destes aspetos é assegurado pela existência de objetivos previamente definidos.

No sentido de garantir a maior validade possível, deve ter-se em conta um conjunto de técnicas que evitam a diminuição da validade e um conjunto de critérios para a elaboração dos itens, os quais se apresentam em seguida (Goring, 1981):

Técnicas para evitar a diminuição da validade

(I) Formular claramente as instruções.

As instruções para responder às perguntas devem ser claras e compreensíveis, pois se o aluno não entender o procedimento correto para a resposta, os resultados não traduzirão o seu grau de aprendizagem, interferindo na validade do exame.

⁴ American Educational Research Association (Associação [Norte-] Americana de Investigação Educacional).

⁵ American Psychological Association (Associação [Norte-] Americana de Psicologia).

⁶ National Council on Measurement in Education (Conselho Nacional dos EUA] para a Medição em Educação).

- (II) Evitar a existência de “chaves” que não correspondem ao objeto dos itens (Este aspeto será abordado no ponto seguinte – critérios para a elaboração dos itens).
- (III) Usar um vocabulário adequado (Este aspeto será abordado no ponto seguinte – critérios para a elaboração dos itens).
- (IV) Conseguir um nível ótimo de dificuldade.
Os instrumentos de medição devem ter um nível de dificuldade apropriado para medir o que pretendem. Se um exame destinado a certificar for muito fácil ou muito difícil, a diferença de pontuações será baixa, o que não permitirá diferenciar os alunos. Portanto, para que um exame tenha resultados válidos para certificar e qualificar os alunos, deve discriminar positivamente entre os alunos que responderam bem a mais itens e os que resolveram bem menos itens, ou seja, é necessário que haja diferenças significativas entre as pontuações mais altas e as mais baixas. (Goring, 1981; Tenbrink, 2010).
- (V) Não contaminar a avaliação.
Deve ser assegurado que o exame inclua só os fatores que se relacionam com a matéria correspondente. Por exemplo, não deverá ser incluído na atribuição da nota questões relacionadas com mau comportamento ou com questões de má caligrafia. A avaliação destes aspetos é lógica, mas deve ser feita à parte, pois, caso contrário, converter-se-á num fator “contaminador”.
- (VI) Administrar o instrumento com destreza.
É necessário verificar se o tempo disponível para a realização do exame é suficiente para garantir que o aluno o termine, pois, caso contrário, impede que se consiga medir o que se pretendia, o que iria afetar a validade do exame.
No que diz respeito às instruções orais dadas no início do exame, devem ser dadas a todos os alunos para que todos tenham acesso às mesmas condições de trabalho. Uma informação dada particularmente colocará em vantagem um determinado aluno, o que não deve acontecer.
No que diz respeito ao espaço físico onde decorre o exame, deve ser adequado e favorável, garantindo um espaço entre os alunos que impeça que copiem, pois neste caso as respostas não medirão os conhecimentos próprios de cada um.

Considerando que o nível de dificuldade da prova dependerá do nível de dificuldade dos itens, este assunto será desenvolvido na sequência dos critérios para elaboração de itens.

Critérios para a elaboração dos itens (Goring, 1981):

- (I) Definir um problema específico que exige uma resposta exata.
O aluno, ao ler o item, deve compreender o que lhe está a ser pedido, que destreza ou aprendizagem lhe exigem, pois se o aluno não entender o que se lhe pede, responderá mal e os resultados não irão indicar o seu grau de aprendizagem.
- (II) Não premiar a simples memorização.
A memorização apenas deverá ser tida em conta, quando for necessária para atingir um fim mais complexo, como permitir a aplicação a novas situações. A memorização deverá ser vista como um meio e não como um fim.
- (III) Evitar a interdependência entre os itens de forma a que o aluno, para acertar um, tenha de acertar o outro, pois os itens poderão medir operações mentais diferentes que não poderão ser avaliadas.
- (IV) Evitar as chaves reveladoras.
Os elementos internos do item não podem revelar de forma alguma o caminho para responder corretamente à pergunta. O aluno poderá acertar, mas não por ter realizado a aprendizagem, ameaçando desta forma a validade da prova.
- (V) Escolher o tipo de item apropriado.
A escolha do item terá de ter em conta o tipo de aprendizagem que se pretende medir e a aplicabilidade e limitações dos vários tipos de itens.
- (VI) Agrupar os itens de um só tipo e dar instruções escritas para lhes responder.
Sendo possível aparecer num exame vários tipos de itens, é conveniente que se agrupem de acordo com o tipo e que se tenha o cuidado de, junto a cada tipo, colocar as instruções de forma muito clara (Goring, 1981; Vallejo, 1979).
- (VII) Usar um vocabulário adequado.
Os itens devem redigir-se com um vocabulário que esteja ao alcance dos alunos, porque, caso contrário, poderão levar ao erro do aluno e, portanto, o

resultado não mostrará o que o aluno é capaz de responder dentro daquela matéria.

(VIII) Não recorrer a trivialidades.

No âmbito da gestão do grau de dificuldade nos itens e na necessidade de discriminar os alunos, não deve, quem elabora a prova, recorrer a trivialidades. Estas podem levar o aluno à memorização de informação insignificante, não indicando aprendizagens de qualidade. Devem ser medidos conhecimentos de princípios, capacidade de tirar conclusões, estas sim com valor educativo.

(IX) Sondar todas as áreas da matéria a testar

É necessário que um exame contemple itens que se relacionem com as várias partes da matéria que se pretende examinar e que o número de itens de cada área a examinar seja proporcional à importância que se lhe deu no ensino.

O nível de dificuldade dos itens

De acordo com Tenbrink (2010), o grau de dificuldade dos itens influencia a validade: por exemplo, um teste demasiado fácil levará muitas pessoas a acertarem os itens não oferecendo informação apropriada para a tomada de decisões.

Para se calcular o nível de dificuldade (ND) de um item, terá de se determinar a proporção de alunos que acertaram a resposta; se todos os alunos a acertaram, então, o nível de dificuldade é 1; se só acertaram metade, será de 0,50. Verifica-se, assim, que quanto mais fácil é a pergunta, mais alto é o nível de dificuldade, uma vez que o nível de dificuldade está determinado para a proporção de alunos que resolvem o item corretamente.

O nível de dificuldade de um item de um teste deve, desde logo, ter em conta o tipo de decisão que se pretende tomar. Se a finalidade passa por uma decisão de seleção em função de uma norma, os itens devem aproximar-se dos 50% de dificuldade, isto é, 50% dos alunos responderiam corretamente, o que permitirá distinguir as suas realizações, ou seja, discriminar os alunos.

Segundo Cheng (2006), está bem patente o papel da dificuldade dos itens na medição das aprendizagens, o qual é enfatizado por Stenner, quando refere “se não sabemos a

razão pela qual uma pergunta é mais difícil do que outra, então é porque não sabemos o que estamos a medir” (Cheng, 2006, p. 2). Há de facto uma preocupação imensa em escrever itens que meçam o constructo que está a ser medido, mas pouco se pensa na dificuldade dos itens em relação a esse constructo.

Mas que importância tem controlar a dificuldade de um item? Por um lado, contribui para discriminar os alunos; por outro, em situações em que é necessário construir testes equivalentes, no sentido de manter o rigor e os padrões de avaliação em anos subsequentes, a capacidade de fazer variar a dificuldade dos itens é crucial.

A distribuição dos níveis de dificuldade num ano deve ser comparável à de outros anos. Além disso, no caso dos testes com itens de opção, os respetivos autores devem assegurar que esses itens tenham nível de dificuldade comparável, pois só assim se garante a fiabilidade dos resultados.

Mas onde se situa a dificuldade de um item? Segundo Pollitt (citado em Cheng, 2006), a dificuldade de um item situa-se no processo de pergunta-resposta. Existem várias fontes de dificuldade nas cinco fases do processo de pergunta-resposta, isto é, na aprendizagem, na leitura da pergunta, na busca do conhecimento a mobilizar, na adequação dos modelos de pergunta e do conhecimento a testar e na construção e escrita da resposta.

Para Osterlind (citado em Cheng, 2006), a dificuldade de um item poderá residir em quatro aspetos que advêm do seu conceito de item:

um item de teste numa testagem de atributos mentais é uma unidade de medida com um estímulo e uma forma prescritiva de resposta; pretende produzir uma resposta de um examinando a partir da qual se possa inferir um desempenho num constructo psicológico, como, por exemplo, conhecimento, capacidade, predisposição ou atributo (p. 2).

Com base na interpretação desta definição, a dificuldade do item pode situar-se: (I) no conteúdo avaliado, (II) no estímulo, (III) na tarefa a realizar e (IV) na resposta esperada.

Relativamente à dificuldade do conteúdo, reporta-se à dificuldade do conhecimento ou da matéria testados. No que diz respeito à testagem dos conhecimentos, a dificuldade de um item situa-se nos vários elementos do conhecimento, de factos, de conceitos, de princípios e de procedimentos. Estes elementos podem ser considerados básicos, adequados ou avançados. Os básicos são aqueles cuja aprendizagem se situa em níveis

menos complexos e não deverão colocar dificuldades à maioria dos examinandos; os adequados são os que estão centrados no currículo, fáceis ou difíceis, dependendo do grau de preparação dos examinandos, apresentando, geralmente, um grau de dificuldade moderado; os avançados correspondem a níveis do conhecimento mais complexos e, portanto, são mais periféricos em relação ao currículo central, pelo que o examinado pode não ter tido oportunidade de os aprender, tornando-se, assim, difíceis para a maioria dos examinandos.

A dificuldade do conteúdo pode variar em função da quantidade de elementos do conhecimento testados, ou seja, itens que tenham como objeto dois ou mais elementos do conhecimento são geralmente mais difíceis do que aqueles que incidem sobre um único elemento do conhecimento (Ahmed, Pollitt, Crisp & Sweiry, citados em Cheng, 2006).

No que concerne à dificuldade de estímulo, está ligada à dificuldade que os examinandos sentem ao tentar compreender as expressões usadas no item e os seus suportes, como diagramas, tabelas ou gráficos. Itens que contenham apenas palavras ou expressões de compreensão simples e direta são mais fáceis do que os que requerem um entendimento mais cuidadoso e técnico e obrigam o examinando a selecionar e a analisar uma grande quantidade de informação; quer isto dizer, que a maneira como a informação é disponibilizada também afeta o grau de dificuldade do item.

Relativamente à dificuldade de tarefa, refere-se à dificuldade que os examinandos têm ao construir a resposta. Na generalidade dos itens, o examinando tem de percorrer etapas para construir a sua resposta. Itens que obriguem a percorrer mais etapas apresentam-se, geralmente, mais difíceis. A dificuldade da tarefa também pode ser medida através da quantidade de informação fornecida, ou seja, os itens que apresentam etapas guiadas são mais fáceis do que aqueles em que o examinando terá de definir as etapas. A dificuldade do item também pode ser afetada pelo processo cognitivo requerido. As taxonomias de processos cognitivos sugerem, que estes existem numa hierarquia cumulativa, em que os mais complexos integram os mais simples. Logo, os itens que requerem processos mais complexos, como análise e síntese, podem revelar-se mais difíceis comparativamente aos que recorrem à memorização e à compreensão. De igual modo, os itens que avaliam capacidades mais complexas, como a aplicação e a improvisação, tornam-se, em geral, mais difíceis do que os que avaliam só replicação e padronização.

Por último, a dificuldade da resposta esperada está relacionada com os critérios de classificação, sendo apenas aplicável aos itens de construção. Pode considerar-se aqui três situações: um item em que sejam esperados mais pormenores na resposta torna-se mais difícil do que se se esperarem poucos; em relação à estrutura da resposta, se o item pedir um estabelecimento de relações simples entre ideias é mais fácil do que se solicitar ligações significativas entre as partes e o todo. Por último, também a distribuição da cotação pode afetar o grau de dificuldade da resposta esperada, isto é, se a distribuição da cotação for equitativa ou lógica (e.g. três pontos para a listagem de três aspetos), a atribuição da pontuação será mais fácil do que em itens em que a distribuição da cotação não é clara (e.g. vinte pontos para a discussão de um conceito, sem haver referência a qualquer indicação sobre os aspetos a que o aluno deve aludir e com que profundidade os deve abordar). Este aspeto relativo à dificuldade da resposta esperada afeta a dificuldade do item, na medida em que os examinandos, ao não saberem o que é esperado da sua resposta, poderão não produzir a quantidade de informação esperada, não por falta de conhecimentos, mas por a questão não conter as orientações suficientes.

Perante o exposto, os vários aspetos de dificuldade de um item, também designados por moderadores de dificuldade (Ahmed et al. citado em Cheng, 2006), podem ser válidos ou inválidos, sendo os válidos os que contribuem para a medição do constructo e os inválidos os que confundem ou impedem essa medição. Estes últimos obstam, por um lado, que o examinando faça o que se pretende que faça e que revele as suas capacidades ou competências; por outro lado, impedem os examinadores de atingir o objetivo da avaliação que estabeleceram.

Neste sentido, Cheng, contempla um conjunto de prováveis moderadores inválidos de aprendizagem, referenciados no Quadro 2.

Quadro 2

Prováveis moderadores inválidos de dificuldade (Cheng, 2006, p. 4)

Content Testing of obscure concepts or facts (e.g., facts that are hardly mentioned in major textbooks) Testing of unimportant facts that are not central to learning outcomes and objectives Testing of advanced concepts which candidates have little opportunity to learn
Stimulus Inaccuracy or inconsistency in data or information given Insufficient information Meaning of words unknown or unclear Question asked is not the one that examiners want candidates to answer Grammatical errors in the question that can cause misunderstanding Unclear resources (e.g., badly drawn / printed diagram, inappropriate graph, unconventional table) Dense presentation (too many important points packed in a certain part of the stimulus) Demand on reading comprehension when reading comprehension is irrelevant to the construct measured
Task Illogical order of parts of the items Mark allocation is unclear or illogical Level of detail required in an answer is unclear Context is unrelated / unnatural to the task that candidates have to do Details of a context distract candidates from recalling the right bits of knowledge Interference from a previous question Insufficient space or insufficient time allocated for responding
Expected Response Mark scheme and questions are incongruent – mark scheme spells out expectations to a slightly different question, not the actual question Answer is indeterminable Large number of plausible alternative answers Rigid mark scheme Demand on producing a written answer when producing a written answer is not part of the construct measured

Da leitura do quadro, verifica-se que, no âmbito dos quatro tipos de moderadores inválidos de dificuldade (conteúdo, estímulo, tarefa e resposta esperada), são apresentados exemplos concretos dos mesmos.

Assim, a nível do conteúdo, temos por exemplo, as testagens de (I) factos ou conceitos que dificilmente são referenciados nos manuais, (II) factos que não são centrais aos objetivos ou aos resultados da aprendizagem; (III) conceitos complexos que os examinandos não tiveram oportunidade de aprender.

No que ao estímulo diz respeito, temos, por exemplo, (I) imprecisão ou insuficiência da informação fornecida; (II) utilização de vocabulário desconhecido para o aluno ou dúbio; (III) suportes pouco claros (e.g. gráfico inadequado, diagrama mal impresso); (IV) apresentação muito densa, isto é, muitos aspetos encaixados numa determinada parte do estímulo; (V) a pergunta formulada não é aquela que se quer a que o aluno responda;

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

(VI) erros gramaticais na formulação da questão, o que pode provocar falhas de compreensão.

No que se refere à tarefa, temos como exemplos: (I) a pouca clareza relativamente ao grau de pormenor requerido na resposta; (II) o contexto forçado ou sem relação com a tarefa a desempenhar; (III) pormenores do contexto distratores da mobilização de conhecimentos adequados à resposta; (IV) tempo insuficiente para a resposta; (V) interferência de uma questão anterior.

Por último, em relação à resposta esperada, podem surgir situações relacionadas com (I) incongruência entre os critérios de classificação e a pergunta, isto é, os critérios apontam para uma pergunta que não corresponde totalmente à que foi colocada; (II) inflexibilidade dos critérios; (III) elevado número de respostas alternativas plausíveis.

Questiona-se então: como podem os autores dos itens gerir a maior ou menor dificuldade de um item sem pôr em causa a medição do constructo? Cheng (2006) apresenta algumas sugestões neste sentido, referenciadas no Quadro 3.

Quadro 3

Sugestões para aumentar ou diminuir a dificuldade dos itens (Cheng, 2006, p. 5)

Location	Demands that increase difficulty	Supports that decrease difficulty
Content	<ul style="list-style-type: none"> Test knowledge in the curriculum that requires deep learning and understanding Test two concepts / topics that are rarely combined 	<ul style="list-style-type: none"> Connect knowledge tested with basic level knowledge or knowledge learnt at lower level Reduce the number of concepts / topics tested
Stimulus	<ul style="list-style-type: none"> Use relevant technical terms, without elaboration or clarification, in the item; Remove references to the concept tested in the item; Use novel or foreign contexts that are appropriate Pack more information than needed (if it is appropriate to test selection of information); Present information in such a way that requires candidate to do some re-organisation 	<ul style="list-style-type: none"> Highlight or emphasize terms that require careful comprehension; State the topic or concept tested in the form of a heading to help candidates recall or focus at the right bits of knowledge; Use contexts that are closely related to the task that candidates have to do Improve the physical layout of the item; Replace words that may mislead some candidates; Provide a glossary of command words or replace command words with a simply and clearly stated demand; Remove irrelevant or redundant information / words in the item; Tailor the resources to the task that candidates have to do.
Task	<ul style="list-style-type: none"> Increase the number of steps in executing task; Without cues and leaders, present task that requires candidates to devise steps to execute the task Require candidates to use process skills in uncommon ways 	<ul style="list-style-type: none"> Decrease the number of steps in executing task; Break up the task into a few steps (sub-questions); Order the steps such that they provide the scaffolding for subsequent steps Test lower level process skill as a precursor to the testing of higher level process skill

Da leitura do Quadro 3 pode constatar-se, que (I) a nível do conteúdo, para aumentar a dificuldade de um item, se deve seleccionar partes do currículo que requerem aprendizagens ou compreensão mais profundas ou testar dois conceitos que raramente são combinados; já para diminuir a dificuldade, poderá procurar-se relacionar o conhecimento testado com níveis simples do conhecimento ou ainda reduzir o número de conceitos a testar; (II) a nível do estímulo, no sentido de aumentar a dificuldade, deve usar-se terminologia técnica relevante, usar contextos novos que sejam apropriados, incluir mais informação do que a necessária (se se quiser testar a seleção) e apresentar a informação de modo a que o examinando a tenha de reorganizar; para diminuir a dificuldade, pode destacar-se a terminologia que requeira compreensão cuidada, redigir o tópico sob a forma de cabeçalho para que o examinando se foque no conhecimento necessário, usar contextos intimamente relacionados com a tarefa a executar, retirar do item informações redundantes, ajustar os suportes às tarefas a executar, melhorar o *layout* do item ou substituir palavras que possam induzir o aluno em erro; (III) no âmbito da tarefa, e no sentido de aumentar a dificuldade, deve elevar-se o número de etapas a executar ou conduzir a inferências a partir de uma tarefa apresentada; para diminuir a dificuldade, deve reduzir-se o número de etapas na tarefa solicitada, dividir a tarefa em subperguntas ou testar capacidades processuais de menor complexidade para impulsionar a testagem de capacidades processuais mais complexas.

No final desta abordagem relativa à dificuldade dos itens, refere-se que o enquadramento da dificuldade dos itens apresentado não é, ainda, uma teoria explicativa, pois não consegue explicar por que razão itens de nível elementar que testam conhecimentos específicos se revelam muitas vezes mais difíceis do que itens que exigem diversos conceitos e diversas etapas. Salienta-se, ainda, que o importante é garantir que um item meça aquilo que deve medir e variar a sua dificuldade de forma a medir toda a amplitude do que se pretende, contribuindo, deste modo, para a validade do constructo na avaliação em educação (Cheng, 2006).

3.1.2.2. A Fiabilidade/Fidelidade: conceito

Segundo Goring (1981) “A fidelidade é o grau de exatidão com que um instrumento mede o que na verdade mede” (pp. 25-26). Por outras palavras, a fidelidade corresponde a uma coincidência de resultados, em aplicações sucessivas, aos mesmos indivíduos (Fernandes, 2008b).

Relacionando a fiabilidade com a validade de um exame, pode dizer-se que a fiabilidade é necessária para que o mesmo seja válido, mas o facto de o exame ser fiável não garante que seja válido.

Na fiabilidade integram-se três componentes: a estabilidade, que garante que os resultados sejam idênticos cada vez que se repita a medição do mesmo indivíduo (ou grupo) com o mesmo instrumento; a consistência, que tem a ver com a concordância dos resultados entre as duas metades de um exame feitas pela mesma pessoa, ou seja, a pessoa que teve a nota mais alta numa das metades terá de ser a mesma que o conseguiu na outra metade (se isso não acontecer, então o exame na sua totalidade não mede o mesmo, pondo assim em causa a confiança no mesmo); e a homogeneidade, que se refere à concordância que existe entre os resultados dados por cada item e as pontuações totais (Goring, 1981).

De acordo com Fernandes (2008b), no caso dos exames em Portugal, cuja função de seleção é muito marcada, já que seleciona indivíduos para ingressarem no ensino superior, a fiabilidade assume um papel muito relevante. A questão é saber se, na situação de replicação do mesmo exame, os candidatos selecionados se mantêm os mesmos ou não. Caso os resultados sejam diferentes nas duas avaliações, pode inferir-se que há fatores a influenciar os resultados sem serem os conhecimentos e as aprendizagens dos alunos, o que seria de facto problemático.

De acordo com Kellaghan e Madaus há quatro fatores que podem afetar a fiabilidade de um exame

(I) Os alunos podem ter desempenhos diferentes em momentos de resolução diferentes; (II) os desempenhos dos alunos podem ser influenciados por condições externas no próprio exame; (III) os desempenhos dos alunos podem ser diferentes, de acordo com a variação das questões que têm de resolver e por último, (IV) as correcções dos exames podem variar sensivelmente de corrector para corrector, principalmente em questões não objectivas, de resposta aberta.

(Fernandes, 2008b, p. 114)

Estas ameaças à fiabilidade de um exame podem sempre ser minimizadas, estandardizando as condições de administração, detalhando e clarificando o mais possível os critérios de correção, minimizando as suas possíveis ambiguidades e moderando os procedimentos dos corretores de modo a assegurar a uniformização na aplicação dos critérios de correção. Neste sentido, é hábito dar-se aos corretores provas

já corrigidas e as explicações para as pontuações que se atribuem. Perspetiva similar têm, como refere (Fernandes, 2008b), Gipps e Stobart, uma vez que consideram que a fiabilidade das correções pode ser sempre melhorada com processos de formação, critérios de correção muito detalhados e tarefas estandardizadas.

Na perspetiva de Goring (1981), a fiabilidade, tal como a validade, é uma qualidade que existe em graus e não em formas absolutas e, portanto, não se pode dizer que uma avaliação é confiável, mas sim que os resultados de certos instrumentos não são confiáveis.

A seguir, e de acordo com Goring (1981), enunciam-se os fatores que deverão ser tidos em conta para aumentar a fiabilidade de um exame:

(I) Fazer medições objetivas.

Esta situação prende-se com a correção dos exames. É imperativo, que, independentemente de quem corrija um exame empregando uma chave (esquema de respostas corretas), se obtenha sempre a mesma pontuação. Esta objetividade consegue-se se a elaboração dos itens for de forma a garantir que as respostas aos mesmos sejam breves e definitivas, impedindo, desta forma diferentes interpretações.

(II) Incluir um grande número de itens.

De acordo com Goring (1981), para que um exame seja apropriado para qualificar os alunos, terá de incluir um grande número de itens. Tenbrink (2010) partilha da mesma opinião, referindo que quanto maior for o teste, mais alta será a sua fiabilidade.

(III) Dificuldade dos itens.

O nível de dificuldade dos itens influencia a fiabilidade. Os itens devem apresentar um grau de dificuldade de modo a que só os alunos que tenham alcançado melhor os objetivos respondam acertadamente a uma grande quantidade de itens. É necessário, portanto, que os itens permitam discriminação entre os examinados.

Nesta sequência, parece ser pertinente apresentar os procedimentos adotados em Portugal no sentido de controlar a fiabilidade dos exames.

Procedimentos tidos em Portugal para controlar a fiabilidade dos exames

Em Portugal, o GAVE, enquanto entidade acreditada pelo Conselho Científico Pedagógico da Formação Contínua, promove ações de formação para professores dos ensinos básico e secundário no domínio da avaliação educacional. Anualmente são promovidas ações de formação para supervisores da classificação, cujo objetivo é o de “aumentar a fiabilidade da classificação das Provas, promovendo padrões de qualidade na avaliação educacional, uma vez que se pretende garantir uma aplicação rigorosa dos critérios de classificação definidos” (GAVE, s.d.a, s.p.). Relativamente à classificação das provas de exame nacional do ensino secundário, o GAVE iniciou, em 2010/2011, a constituição de uma bolsa de professores classificadores. Os professores nela incluídos deverão realizar formação, sendo que esta se distribuirá por um período de quatro anos. Em 2011/2012, foi prevista a realização do segundo ano desta formação e iniciou-se um novo ciclo de formação, com o objetivo de integrar uma nova disciplina (Filosofia) e de reforçar as disciplinas que já integravam a bolsa.

O procedimento de supervisão da classificação de provas inicia-se após a realização das provas. O GAVE promove uma reunião com os supervisores para tomarem conhecimento da metodologia a seguir com os classificadores, analisarem “a prova e os respetivos critérios de classificação”, aplicarem “os critérios a alguns exemplos de itens de respostas dos alunos” e concertarem “classificações, tendo em conta a aplicação dos respetivos critérios” (GAVE, s.d.b). Das competências dos supervisores da classificação faz parte:

- 1.- Promover, junto dos classificadores, a análise e discussão da prova e dos critérios de classificação;
- 2.- Promover a classificação de alguns exemplos de resposta;
- 3.- Discutir as divergências e concertá-las;
- 3.- Alertar os classificadores para os fatores que interferem na fiabilidade;
- 4.- Esclarecer dúvidas aos classificadores até à entrega das provas já classificadas;
- 5.- Promover a concertação das classificações;
- 6.- Solicitar ao GAVE os esclarecimentos necessários;
- 7.- Classificar cerca de 15 provas.

(GAVE, s.d.b)

No âmbito do processo de classificação dos exames nacionais do ensino secundário, o GAVE disponibiliza o “Documento GAVE”, elaborado, por disciplina, em conjunto com os professores formadores. Este documento técnico

visa esclarecer as dúvidas suscitadas pelas respostas dadas por todos os alunos em todas as provas analisadas pelos professores classificadores.

O número de dúvidas em qualquer processo de classificação é proporcional ao número de alunos que realiza o exame e, também, à natureza e complexidade das questões da prova. Por exemplo, questões em que os alunos constroem a sua resposta, elaboram textos mais ou menos extensos, efetuam cálculos, requerem um trabalho de reflexão que tem de ser partilhado e que, a bem da equidade, tem de gerar procedimentos tão homogêneos e uniforme quanto possível.

A informação transmitida nos «Documentos GAVE» visa, assim, diminuir a subjetividade inerente a qualquer processo de classificação que, como é o caso dos exames de algumas disciplinas, envolve centenas de professores classificadores que analisam dezenas de milhares de respostas.

(GAVE, s.d.c, s.p)

Salienta-se que os documentos supracitados

não constituem critérios adicionais, aditamentos ou alterações aos critérios publicados no dia de realização das provas [e] são enviados a todos os professores envolvidos no processo de classificação e remetidos também ao Júri Nacional de Exames para serem facultados aos professores que posteriormente, sejam chamados ao processo de reapreciação das provas em situação de recurso (GAVE, s.d.c, s.p).

3.1.3. Tipos de itens

Relativamente aos itens constata-se que ao longo do tempo as suas tipologias nem sempre têm sido integradas nos mesmos grupos e categorias de itens. Refere-se a título de exemplo a nomenclatura apresentada por Ribeiro e Ribeiro (1990). De acordo com estes autores, os itens integram-se em dois grupos, os de tipo objetivo, que requerem uma resposta curta, e os de composição que solicitam uma resposta longa. No grupo dos itens objetivos integram-se os itens em que o aluno dá ou completa a resposta, sendo estes os itens de resposta curta e os itens de completação, respetivamente; e os itens em que o aluno seleciona a resposta, onde estão incluídos os itens de verdadeiro/falso, associação e escolha múltipla. No grupo de composição ou de resposta longa, é exigido ao aluno uma resposta extensa, organizada, com uma dada estrutura de

desenvolvimento, em que terão de ser reveladas, ainda, as suas próprias ideias e capacidade de expressão. Dentro deste grupo localizam-se os itens de resposta orientada e os itens de resposta livre.

Considerando Tenbrink (2010), um teste pode apresentar itens do mesmo tipo ou de tipos diferentes. Os itens categorizam-se em duas dimensões: a primeira, de acordo com o método de pontuação, originando os *testes objetivos*, e a segunda, de acordo com a liberdade de resposta que se pretende dar ao aluno, originando os *testes subjetivos*.

Ainda de acordo com este autor, o tipo de item a usar deverá ter em conta quatro fatores: (I) o nível ou tipo de resultados de aprendizagem que se mede, (II) o modo como se utilizará os resultados do teste; (III) as características dos alunos que realizam o teste; e (IV) o tempo disponível para construir, administrar e pontuar o teste.

De acordo com o GAVE, a nomenclatura dos itens, desde 2011, prevê itens de seleção e itens de construção, os quais substituíram, respetivamente, os itens de resposta fechada e itens de resposta aberta.

Nos itens de seleção o aluno “escolhe a resposta a dar a partir de várias hipóteses dadas no item” (GAVE, s.d.d). Neste tipo de item incluem-se os itens de escolha múltipla, os itens de associação/correspondência e os itens de ordenação. Nos itens de construção é o aluno que produz a resposta. Incluem-se neste tipo de item os itens de resposta curta, resposta restrita e resposta extensa.

“Alguns dos itens de escolha múltipla, de associação/correspondência e de resposta curta, podem apresentar-se sob a forma de tarefas de completamento assim como alguns itens de resposta curta e de resposta restrita podem apresentar-se sob a forma de tarefas de transformação” (GAVE, s.d.d).

Princípios gerais para a formulação de itens de provas objetivas.

A objetividade é uma das condições essenciais para que se atinja a fiabilidade de uma prova da qual dependerá, por sua vez, a validade da mesma.

Para se garantir a objetividade de uma prova, há duas características a ter em conta: a brevidade da resposta a um item objetivo, que pode consistir, entre outros, num número, numa palavra ou numa frase, e a exatidão de correção, a qual pressupõe que, perante uma prova objetiva, qualquer pessoa que corrija a prova com uma chave das respostas consiga o mesmo resultado (Goring, 1981).

3.1.3.1. Os itens de seleção: características

Neste ponto serão apresentadas as características dos itens de: emparelhamento / associação / correspondência; de resposta alternativa, Verdadeiro-Falso / Sim-Não; de escolha múltipla; de completamento e, de ordenação, na perspetiva de diferentes autores.

Item de emparelhamento / associação / correspondência

Esta tipologia de item permite a avaliação de associações de conceitos de forma muito compacta (Goring, 1981).

A pergunta apresenta um conjunto de elementos organizados em duas colunas paralelas, sendo os elementos de uma relacionados com os elementos da outra. Os elementos da coluna da esquerda designam-se premissas e os da direita respostas. As instruções dadas no início da pergunta indicam ao aluno a maneira como a associação deve ser feita (Ribeiro, 1990; Nova, 1997).

Segundo Tenbrink (2010), são itens apropriados para obter informação sobre o conhecimento (factos, termos, conceitos) de um indivíduo. Medem a capacidade do aluno para associar partes de informação, como por exemplo: uma causa a um efeito, um termo e a sua definição, um lugar e um acontecimento.

Na perspetiva de Ribeiro e Ribeiro (1990), as grandes vantagens deste tipo de item é permitir, no mesmo item, avaliar um conjunto de aprendizagens relacionadas entre si e, ainda, dado o número de combinações possíveis entre os elementos das duas colunas, o aluno não poder “adivinhar” a resposta. Consideram ainda os autores, tal como Tenbrink (2010) que são itens fáceis de construir e de aplicar. No entanto, apenas servem para a avaliar as aprendizagens que se situam nos níveis mais baixos das taxonomias de objetivos educacionais.

Item de resposta alternativa: Verdadeiro – Falso / Sim – Não

Este tipo de item não integra os itens utilizados, atualmente, pelo GAVE. No entanto, optou-se por apresentar as suas características.

Neste tipo de item, o aluno é obrigado a escolher entre duas opções opostas para responder a um problema. Geralmente as duas opções são verdadeiro e falso (Goring, 1981; Ribeiro, 1990).

Para Tenbrink (2010) este tipo de item é frequentemente usado pelo professor, essencialmente por duas razões: por um lado, porque os itens de resposta alternativa parecem ser fáceis de construir e, por outro, porque parecem ser úteis para obter muita informação sobre as aprendizagens realizadas pelos alunos em pouco tempo. As duas razões são questionáveis porque (I) para serem bons itens são difíceis de construir, uma vez que é difícil construir perguntas que sejam claramente verdadeiras ou falsas; (II) o facto de só haver duas alternativas de resposta dá ao aluno 50% de probabilidades de acertar por adivinhação. Quer isto dizer que, para se ter uma medida suficientemente fiável e se poder formar um juízo sólido, teria de ser feito um grande número de perguntas, o que não é fácil, uma vez que este tipo de item, ao medir só níveis baixos de aprendizagem, aspeto partilhado por Ribeiro e Ribeiro (1990), estreita a quantidade de matéria a ser testada, limitando a possibilidade de serem feitas muitas perguntas que sejam, de facto, verdadeiras ou falsas. Como refere Choppin, citado em Ribeiro e Ribeiro (1990), esta modalidade de item não é geralmente recomendada. Os testes compostos por este tipo de item tendem a ser muito pouco fiáveis e são suscetíveis de contaminação pela resposta dada ao acaso.

Item de escolha múltipla

Esta tipologia de item consta de um “tronco” em que a questão é apresentada com uma lista de possíveis respostas, entre as quais se encontra a alternativa correta, solicitada ao aluno, sendo as alternativas incorretas, embora plausíveis, consideradas distratores (Goring, 1981; Ribeiro, 1990; Ribeiro & Ribeiro, 1990; Tenbrink, 2010).

De entre os itens do tipo objetivo, estes são os mais versáteis, pois podem ser utilizados para obter resultados de aprendizagem desde o nível mais simples ao mais complexo (Goring, 1981; Nova, 1997; Tenbrink, 2010). De acordo com Ribeiro e Ribeiro (1990) são itens que permitem recolher com exatidão a informação que o professor determinou, uma vez que não há flutuações de resposta e, simultaneamente, diagnostica possíveis fontes de erro, se as mesmas constarem nas alternativas incorretas. Para estes autores, esta tipologia de item é um bom instrumento de diagnóstico, apropriado tanto para uma avaliação diagnóstica, como formativa. Consideram, ainda, que a probabilidade de “adivinhação” pode ser reduzida, por exemplo, aumentando o número de alternativas de resposta, ou o número de itens por objetivo. Todavia, não são os mais apropriados quando a memorização é um fator importante, nem medem com eficiência alguns níveis

superiores de aprendizagem, como por exemplo, propor uma solução criativa perante um problema, aspeto este partilhado pelos autores acima referenciados.

Item de completamento

Este tipo de item não integra os itens utilizados, atualmente, pelo GAVE. No entanto, optou-se por referir as suas características.

Apresenta-se na forma de um enunciado com um espaço em branco para o aluno preencher com um palavra ou uma frase apropriada de forma a completar o juízo (Goring, 1981; Tenbrink, 2010; Ribeiro, 1990). Exige uma resposta breve e definida (Goring, 1981).

São itens muito simples, fáceis e muito limitados. São adequados para obter informação relativa aos níveis mais baixos da aprendizagem (Tenbrink, 2010).

Item de ordenação

Segundo Nova (1997), através deste tipo de itens, o aluno é solicitado a ordenar, numa sequência lógica ou cronológica correta, um conjunto de fenómenos, de operações, de acontecimentos, de situações, etc, que lhe é apresentado desordenadamente.

3.1.3.2. Itens de construção: características

Neste ponto serão apresentadas as características dos itens de: resposta curta; de resposta restrita e de resposta extensa, na perspetiva de diferentes autores.

Item de resposta curta

Este tipo de item faz uma simples pergunta e pede-se que o aluno dê uma resposta curta e concreta (uma frase, uma expressão) (Goring, 1981; Tenbrink, 2010). Tal como o item de preenchimento de espaço, limita-se a níveis inferiores de aprendizagem. Não são itens difíceis de construir, mas exigem muito cuidado na sua construção, pois podem ser muito ambíguos. É considerado um item de transição entre o item de preenchimento de espaços e o item de ensaio de resposta restrita (Tenbrink, 2010).

Itens de resposta restrita

Esta tipologia de itens, assim designada por Gronlund (1997), corresponde aos itens de resposta orientada (Ribeiro & Ribeiro, 1990).

Neste tipo de item são indicados ao respondente determinados parâmetros (conteúdos, relações, estrutura) que a resposta pode satisfazer (resposta orientada). Na opinião de Tenbrink (2010) os itens de resposta restrita, apontam uma determinada direção de resposta, o que tenderá a aumentar a fiabilidade dos resultados, o que os torna mais úteis na tomada de decisões importantes. Estas questões tornam-se, por este motivo, um pouco mais difíceis de construir do que as de resposta extensa. Nesta perspetiva, Goring (1981) acrescenta que por a resposta ser limitada, este tipo de item coloca limites explícitos quanto à extensão e ao género das respostas, encaminhando o aluno “numa direção definida pelas especificações que acompanham a mesma prova” (p.121). Desta forma, permite que o aluno demonstre certos tipos de aprendizagens explícitas. Devido às restrições implícitas que o item pode conter, o aluno terá que estar mais atento às mesmas.

Esta tipologia de item permite avaliar a capacidade de expressão escrita, bem como aprendizagens de nível complexo, como a organização e síntese de ideias, análise crítica de suportes de texto e capacidade de resolução de problemas. Pode ser utilizada tanto no domínio afetivo, como na expressão de atitudes, valores e opinião.

Apresentam alguns aspetos desvantajosos, nomeadamente na oscilação que se verifica entre classificações atribuídas, tanto por professor diferentes como até pelo mesmo professor em diferentes momentos, motivo pelo qual esta tipologia só deverá ser utilizada quando se pretende avaliar aprendizagens que não o podem ser através de itens objetivos (Ribeiro & Ribeiro, 1990). Como refere Gronlund (1977), a classificação desta tipologia de item é difícil, pondo em causa a fiabilidade da mesma, uma vez que as respostas se podem apresentar mais ou menos coerentes e mais ou menos claras.

De acordo com Tenbrink (2010), os itens de resposta restrita permitem ao aluno mostrar a informação que pode recordar de memória, mas como a resposta é muito restrita não é muito apropriada para medir resultados de aprendizagem como os que requerem uma solução criativa de um problema ou a defesa de uma opinião. Este tipo de item está mais adaptado à memorização de factos e à enumeração de acontecimentos, tendendo a definir o processo a usar na resposta. Se se deseja formular juízos relativos à norma,

este tipo de item permite assegurar que todos os alunos interpretam e respondem a cada pergunta *da mesma maneira*.

Segundo Nova (1997), neste tipo de item (resposta orientada), a resposta pedida é menos longa comparativamente ao tipo de item de resposta livre e são indicados ao aluno alguns parâmetros a que a resposta deve obedecer.

Item de resposta extensa

Esta tipologia de itens integra os itens de resposta livre (Ribeiro & Ribeiro, 1990) ou de resposta extensa (Gronlund, 1997). Esta tipologia apresenta uma questão em que o respondente expõe as suas ideias livremente. Neste sentido, e na perspectiva de Tenbrink (2010), os itens de resposta extensa permitem ao aluno uma grande amplitude na sua resposta.

Este tipo de item permite medir a capacidade criativa e de organizar as ideias ou defender uma posição. As respostas extensas são mais apropriadas do que as restritas para medir as capacidades de análise, de síntese e de avaliação e medem melhor a criatividade e a originalidade. Segundo Goring (1981) este tipo de item não limita explicitamente o campo de ação dentro do qual o aluno pode atuar. Quanto maior liberdade de resposta for dada ao aluno, mais informação se conseguirá sobre o seu processo de pensamento, mas menos fiáveis tendem a ser os resultados. Portanto, quanto menos restrições, menos objetividade e menos fiabilidade.

3.1.4. Considerações sobre as vantagens e desvantagens do uso dos itens objetivos e dos itens subjetivos

Na opinião de Gronlund, citado em Ribeiro (1990, p. 162),

as desvantagens de um teste de composição são tão graves que provavelmente levariam a pô-lo completamente de parte como instrumento de avaliação se não fosse o facto de avaliar aprendizagens relevantes que não podem ser avaliados por outros meios[...]só devem ser utilizadas perguntas de resposta longa quando as aprendizagens não podem ser avaliadas através de perguntas do tipo objectivo.

Na sequência do que se acaba de referir, poderá dizer-se que os itens de resposta extensa devem ser utilizados, por exemplo, para avaliar competências de comunicação escrita, tal como refere Estanqueiro (2010).

De acordo com Fernandes (2008b), as questões abertas, do tipo problemas, tarefas e ensaios, são recomendadas pelos autores construtivistas de avaliação por apresentarem um conjunto de vantagens, coincidentes com as que Ribeiro e Ribeiro (1990), Gronlund (1997) e Tenbrink (2010) defendem. Tenbrink (2010) considera que os itens de resposta aberta: (I) permitem analisar os processos que os alunos utilizam na resolução de novos problemas; (II) permitem avaliar as capacidades dos alunos para integrarem, relacionarem, aplicarem e organizarem os conhecimentos; (III) permitem avaliar as capacidades dos alunos para analisarem, sintetizarem e avaliarem a informação existente numa grande diversidade de textos ou de situações problemáticas; e (IV) permitem ainda avaliar aspetos originais e criativos do pensamento dos alunos. No entanto, estes itens, no contexto de um exame nacional, com consequências no progresso dos alunos e na certificação, apresentam alguns problemas, tais como: (a) são difíceis de elaborar; (b) os critérios de correção são difíceis de elaborar; (c) demoram muito tempo a ser corrigidos; (d) a fiabilidade dos corretores tende a baixar e (e) a fiabilidade do exame tende a baixar.

No entanto, Fernandes (2008a) considera que esta tipologia de itens é necessária nos exames pela vantagem que traz na avaliação de aprendizagens, mas evidencia alguma preocupação relativamente à fiabilidade entre os corretores, uma vez que esta pode interferir na equidade dos exames. Realça, no entanto, alguns procedimentos que têm sido postos em prática pelas entidades responsáveis no sentido de minimizar este problema, como o maior cuidado que tem sido posto na elaboração dos critérios de correção, a promoção de formação para corretores e as reuniões entre corretores para aferirem os critérios. Apesar de todo este esforço, Fernandes (2008a) considera que é necessário fazer mais, sendo importante que se desencadeiem mecanismos mais apertados de controlo da correção e que a partir de amostras aleatórias de provas se determine quais são os níveis de fiabilidade das correções.

Recorde-se, que foi por questões relacionadas com as discrepâncias verificadas nas classificações dos exames, que a docimologia investiu os seus esforços no sentido de resolver esta problemática no início do século XX. Em consequência acabaram por se afirmar os testes objetivos, de escolha múltipla, pois, para além de permitirem avaliar um grande número de indivíduos em pouco tempo, a sua fiabilidade era máxima. Para Kellaghan e Madaus (citado em Fernandes 2008b), as vantagens destes testes devem-se

ao facto de (I) apresentarem uma correção objetiva, (II) serem facilmente administrados a grandes grupos de alunos e serem corrigidos rapidamente, (III) poderem abranger uma parte significativa do conteúdo a avaliar em pouco tempo, (IV) serem fiáveis, (V) serem considerados uma boa e credível fonte de informação e (VI) as classificações dos testes serem facilmente entendidas pelas pessoas. No entanto, Kellaghan e Madaus, citados por Fernandes (2008), apontam algumas limitações a este tipo de testes de escolha múltipla: (I) são, por vezes, ambíguos; (II) não fornecem informação sobre as razões que estão na origem dos alunos acertarem ou não, e não informam sobre os processos utilizados pelos alunos na respetiva resposta; (III) não fornecem, em geral, informação acerca dos processos de pensamento mais complexos dos alunos (é mais fácil construir itens de escolha múltipla para avaliar o conhecimento de factos do que itens que avaliem processos mentais mais complexos que se desenvolvam em várias etapas); e (IV) se estes testes tiverem efeitos sobre a progressão académica do aluno, podem exercer uma influência negativa no currículo.

Na sequência da descrição feita relativamente aos tipos de itens, contacta-se que os níveis de aprendizagem de nível inferior se medem melhor com o uso de itens de seleção. No que se refere aos níveis de aprendizagem superiores, considera-se necessário usar itens com características mais subjetivas. No entanto, independentemente do tipo de itens utilizados, o importante é que estejam bem construídos e que meçam os resultados das aprendizagens a que se destinam.

É indiscutível que todas as precauções são poucas de forma a assegurar a maior validade nos resultados de um exame. A meta é que qualquer item seja elaborado de tal forma que o aluno saiba responder acertadamente e mostrar se realmente aprendeu o que o item pretende e que uma resposta errada indique que o aluno não atingiu aquela aprendizagem. Se os itens caminharem para esta meta, os resultados do exame tenderão para a validade (Goring, 1981).

3.1.5. O número de itens a propor num teste/exame

Genericamente, quanto mais perguntas tiver um teste, maior é a probabilidade de obter informações sobre o universo que o mesmo pretende avaliar; por outro lado, por razões pedagógicas e de eficácia de aplicação, aconselha-se a elaboração de testes mais

curtos. É, portanto, necessário encontrar uma situação de equilíbrio, através de uma ponderação dos diferentes fatores envolvidos, como a finalidade do teste, a área que vai avaliar, o tempo necessário para as respostas, entre outros (Ribeiro, 1990).

Relacionada com a extensão do teste está a sua fiabilidade, uma vez que quanto maior for o teste, mais alto será o grau de exatidão da informação e a sua fiabilidade, como já foi referido (Tenbrink, 2010). No entanto, “é preciso não esquecer que os alunos devem dispor do tempo necessário para responder a todos os itens” (Nova, 1997, p.87).

3.1.6. A ordenação dos itens numa prova

Os itens devem ser ordenados no teste de forma lógica. Devem ser agrupados de acordo com a sua tipologia para que o aluno não tenha de mudar o seu tipo de resposta de uma pergunta para a outra. Uma outra regra passa por sequenciar os itens do mais fácil para o mais difícil, para que os indivíduos que tenham mais dificuldade mantenham uma atitude positiva ao longo da prova e para que os mais lentos possam confrontar-se com perguntas mais complexas que seriam incapazes de responder no início do teste. Também se poder organizar os itens de acordo com a sequência dos conteúdos lecionados, o que permite ao aluno pensar sobre o grupo de conceitos que estudou e que estão relacionados entre si. Por último, pode ordenar-se os itens de acordo com o nível de aprendizagem, segundo, por exemplo, a taxonomia de Bloom. Este método permite que o corretor, através da pontuação, saiba o nível de rendimento que adquiriu o aluno. Este é um modo valioso de ordenar os itens de um teste em relação a um critério (Tenbrink, 2010).

3.1.7. Determinação do formato de uma prova

De acordo com Nova (1997), pode constatar-se, que, dependendo das características dos itens que integram um teste, assim será o seu grau de objetividade. Assim, se um teste incluir apenas itens objetivos, será um teste com perfil objetivo. Por outro lado, se um teste integrar itens com características subjetivas, como os itens de resposta aberta, atualmente itens de construção, o teste adquirirá um perfil subjetivo. O teste de perfil

subjetivo poderá conduzir a divergências na classificação, mesmo quando se trata do mesmo avaliador, pondo em causa a fiabilidade do teste.

Finalmente salienta-se ainda um tipo de teste de perfil intermédio - o teste misto, que inclui itens de resposta objetiva e itens de resposta longa, subjetivos.

Considerando as características dos diferentes tipos de itens, já referenciadas nos pontos 3.1.3.1. e 3.1.3.2, o teste misto permite avaliar um conjunto abrangente de níveis de aprendizagem, desde os mais simples aos mais complexos das taxonomias (Nova, 1997).

3.1.8. Procedimento do GAVE na divulgação das características da prova de exame nacional do ensino secundário na disciplina de Biologia e Geologia, código 702, entre 2006 e 2011

De acordo com Fernandes (2008b, p. 110), “são múltiplas as consequências que resultam da utilização de um teste com determinadas características”. Em Portugal, tem havido uma tentativa de diversificar, nas provas de exame, a tipologia dos itens, no sentido de equilibrar as exigências de validade, de fiabilidade, de equidade e de capacidade de comparação, com as exigências de um ensino e de uma aprendizagem em que a resolução de problemas, a integração dos saberes ou as competências metacognitivas das pessoas têm um papel determinante.

O GAVE, desde 2005, que publica um documento relativo à Informação da prova de exame nacional de Biologia e Geologia, código 702, 11.º ou 12.º ano de escolaridade (Decreto-Lei n.º 74/2004, de 26 de março), o qual é enviado para um conjunto de entidades, a saber: Direção-Geral de Inovação e de Desenvolvimento Curricular⁷, Inspeção-Geral de Educação⁸, Direções Regionais de Educação, Secretaria Regional de Educação da Madeira, Secretaria Regional de Educação dos Açores, Escolas com Ensino Secundário, Estabelecimentos de Ensino Particular e Cooperativo com Paralelismo com Ensino Secundário, Centro de Informação e Relações Públicas do Ministério da Educação - CIREP, Federação Regional de Lisboa das Associações de Pais

⁷ Atualmente, Direção Geral de Educação.

⁸ Atualmente Inspeção-Geral da Educação e da Ciência.

- FERLAP e Confederação Nacional das Associações de Pais - CONFAP (GAVE, 2005; 2006; 2008a; 2008b; 2009; 2010a).

O documento visa, essencialmente, divulgar as características da prova de exame. Dá a conhecer as aprendizagens e as competências que são objeto de avaliação, as características e a estrutura da prova, o material a utilizar e a duração da prova. São ainda apresentados os critérios gerais de classificação da prova.

O documento "Informação Exame" esclarece que a avaliação sumativa externa, realizada através de uma prova escrita de duração limitada, só permite avaliar parte das aprendizagens e das competências enunciadas no programa. No entanto, a resolução da prova pode implicar a mobilização de outras aprendizagens e competências inscritas no programa mas não expressas no objeto de avaliação identificado neste documento. É ainda referido que, nas provas de exame desta disciplina, o grau de exigência decorrente do enunciado dos itens e o grau de aprofundamento evidenciado nos critérios de classificação estão balizados pelo respetivo Programa e são adequados ao nível de ensino a que o exame diz respeito e ao escalão etário dos examinandos. No documento está patente a importância de dar a conhecer aos alunos a informação nele constante e que a mesma deve ser com eles analisada, para que fiquem devidamente informados sobre as características da prova de exame nacional que irão realizar (GAVE, 2005, 2006, 2008a, 2008b, 2009, 2010a).

3.1.9. Análise das provas de exame nacional de Biologia e Geologia, código 702, entre 2006 e 2011

Como já foi referido no capítulo 2 – ponto 2.1, após, em 2004, ter sido implementada a reforma curricular do ensino secundário (Decreto-Lei n.º 74/2004, de 26 de março), em 2006 foi incluída no sistema de exames a disciplina de Biologia e Geologia, ao abrigo do Decreto-Lei n.º 24/2006, de 6 de fevereiro, tendo sido então realizado o primeiro exame nacional de Biologia e Geologia (Código 702) no ano letivo de 2005/2006.

No âmbito deste trabalho, considerou-se fazer sentido proceder a uma análise das provas de exame do ensino secundário da disciplina de Biologia e Geologia, código 702, desde o ano letivo de 2005/2006 até ao de 2010/2011.

Dessa análise regista-se que houve características que se mantiveram idênticas nas diversas provas, mas outras houve que sofreram alterações.

Características que se mantiveram idênticas entre 2006 e 2011

1. as provas refletem uma visão integradora dos diferentes conteúdos programáticos da disciplina;
2. as duas componentes da disciplina (Biologia e Geologia) têm uma ponderação idêntica (100 pontos para a componente de Biologia e 100 pontos para a componente de Geologia);
3. o domínio procedimental tem uma cotação que pode variar entre 20 e 40 pontos;
4. as provas de exame estão organizadas em grupos de itens (os itens podem ter como suporte um ou mais documentos, como, por exemplo, textos, tabelas, gráficos, mapas, fotografias, esquemas);
5. as provas incluem itens de resposta fechada / seleção (predominantemente de escolha múltipla) e itens de resposta aberta / construção;
6. alguns dos itens/grupos de itens podem envolver a mobilização de aprendizagens relativas a mais do que um dos temas/unidades do programa;
7. a sequência dos itens pode não corresponder à sequência de apresentação dos Temas/Unidades no programa da disciplina.

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

Características das provas de exame que sofreram alterações entre 2006 e 2011

As características indicadas com os algarismos 1, 2 e 3 foram alvo de análise a partir do Quadro 4, que regista a tipologia dos itens das provas entre 2006 e 2011. As características 4 e 5 resultaram da análise dos documentos “Informação Exame” e dos critérios de correção das provas.

Quadro 4.

Tipologia de itens das provas de Biologia e Geologia (1.ª fase), entre 2006 e 2011.

Tipos de Item	Exames Nacionais 1.ª Fase – Biologia e Geologia					
	2006	2007	2008	2009	2010	2011
Resposta Alternativa (V/F - Sim/Não)	3	4	4	----	----	----
Resposta Curta	1	----	----	----	----	----
Escolha Múltipla	14	14	22	21	22	24
Resposta restrita	5	5	4	5	4	4
Ordenação	----	3	----	1	1	2
Correspondência Associação	3	----	----	1	2	1
N.º TOTAL DE ITENS DA PROVA	26	26	30	28	29	31

Da análise do Quadro 4 constata-se que:

- relativamente à tipologia dos itens,
 - os itens do tipo alternativa, verdadeiro/falso e/ou sim/não, presentes nas provas de exame de 2006 a 2008, deixaram de integrar as provas de 2009 a 2011;
 - os itens de resposta curta só surgiram na prova de exame de 2006;
 - os itens de escolha múltipla constam em todas as provas de exame;
 - os itens de resposta restrita constam em todas as provas de exame;
 - os itens de ordenação nem sempre têm estado presentes nas provas de exame, tendo surgido só em 2007, 2009, 2010 e 2011;

- os itens de associação nem sempre têm estado presentes nas provas de exame, tendo surgido em 2006 e só voltando a constar nas provas de 2009, 2010 e 2011;
 - mantem-se a mesma desde 2009.
2. relativamente ao número total de itens em cada prova verifica-se que foi variável, com um mínimo de 26 itens nas provas relativas aos exames de 2006 e 2007, e um máximo de 31 itens na prova de 2011. Refere-se que o tempo de duração de cada prova, em todos os anos considerados, foi de 120 minutos, começando a ter tolerância de 30 minutos a partir do ano de 2008.
3. relativamente ao número de itens do mesmo tipo por prova, verifica-se que:
- o número de itens de escolha múltipla nas provas aumentou a partir de 2008, comparativamente com os dois anos anteriores;
 - o tipo de item de escolha múltipla é predominante em qualquer das provas em análise;

Com base nas “Informação Exame” e nos critérios de correção gerais das provas, verifica-se que:

4. Os itens que entre 2006 e 2010 eram designados por itens de resposta aberta passaram a ser designados por itens de construção, em 2011, incluindo os itens de resposta curta e os itens de resposta restrita. Os itens de resposta fechada passaram a ser designados por itens de seleção, incluindo os itens de resposta múltipla, de ordenação e de associação/correspondência.
5. Em relação aos critérios gerais de classificação

Após uma análise comparativa dos critérios gerais de classificação constantes nas provas, verifica-se que:

- (I) nos itens de resposta fechada de:
- escolha múltipla – o critério geral manteve-se idêntico entre 2006 e 2011. À resposta correta é atribuída a cotação total e as respostas incorretas são cotadas com zero pontos, não havendo lugar a classificações intermédias.
 - verdadeiro / falso e sim/não (alternativos) – o critério geral de classificação foi sempre o mesmo. As classificações tiveram em conta os níveis de desempenho revelados na resposta.

- ordenação – o critério geral de classificação tem sido o mesmo. É atribuída a cotação total do item quando a sequência está integralmente correta, não havendo lugar à atribuição de cotações intermédias.
 - resposta curta (atualmente integrado nos itens de construção) – neste tipo de item, pelo facto de só ter constado no exame nacional no ano de 2006, não é possível fazer comparações dos critérios gerais de classificação relativos a outras provas de exame. No entanto sendo itens que têm integrado os testes intermédios do ensino secundário, no período temporal em análise, fez-se uma analogia com os mesmos. Verifica-se, então, que, no exame nacional de 2006, no caso em que a resposta continha elementos que excediam o solicitado, só eram considerados, para efeito da classificação, os elementos que satisfizessem o que era pedido, segundo a ordem pela qual eram apresentados na resposta. Porém, se os elementos referidos revelassem uma contradição entre si, a cotação a atribuir seria zero pontos. Relativamente aos testes intermédios de 2008, a classificação a atribuir traduzia a avaliação das competências específicas da disciplina e era atribuída de acordo com níveis de desempenho; em 2011, constatou-se que as respostas corretas são classificadas com a cotação total do item. As respostas incorretas são classificadas com zero pontos. Não há lugar a classificações intermédias. Caso a resposta contenha elementos que excedam o solicitado, deverá ser classificada com zero pontos. Verifica-se, assim, que os critérios gerais de classificação neste tipo de item sofreram alterações entre 2006 e 2011.
 - associação – a classificação deste tipo de item tem sido feita com base no nível de desempenho revelado na resposta.
- (II) nos itens de resposta aberta/resposta restrita, os critérios gerais de classificação têm apresentado sempre níveis de desempenho. O enquadramento das respostas num determinado nível de desempenho contempla aspetos relativos aos conteúdos, à organização lógico-temática e à utilização da linguagem científica. São sempre avaliadas as competências de comunicação em Língua Portuguesa.

3.1.10. O formato dos exames nacionais entre 2006 e 2011

Com base na análise dos exames suprarreferidos, verifica-se que estes apresentam, de acordo com Nova (1997), um perfil do tipo misto.

Os exames nacionais, no período em apreço, integram um conjunto de itens de categorias diferentes: os itens de resposta fechada/de seleção (objetivos) e os itens de resposta aberta/construção (subjetivos), o que vai ao encontro de Tenbrink (2010), quando refere que um teste pode comportar itens de diferentes tipos. Nos itens de resposta fechada/seleção dos exames incluem-se os itens de verdadeiro/falso, de sim/não, escolha múltipla, associação/correspondência e de ordenação. A partir do ano de 2008 os itens de verdadeiro/falso e de sim/não deixaram de integrar as provas de exame, o que poderá dever-se ao facto de esta tipologia de item dar ao aluno 50% de probabilidades de adivinhação e poder pôr em causa a fiabilidade do exame, tal como já referido anteriormente, no ponto 3.1.3.1.. Nos itens de resposta aberta/construção, incluem-se os itens de resposta curta e de resposta restrita.

Os itens de resposta fechada, os atuais itens de seleção, predominam sobre os itens de resposta aberta/construção, nomeadamente os itens de escolha múltipla. Esta diferença poderá ser explicada considerando Ribeiro e Ribeiro (1990) e Gronlund (1977), quando referem que os itens de resposta restrita, pelas características de que se revestem só devem ser utilizados quando é necessário medir aprendizagens que só esta tipologia permite, o que poderá explicar o menor número de itens deste tipo nos exames nacionais. Por outro lado, as características dos itens de seleção de escolha múltipla, também poderão explicar a razão de estes serem os mais representativos nos exames, uma vez que permitem a medição de aprendizagens desde as mais simples às mais complexas. Como refere Goring (1981), estes itens prestam-se “à análise estatística cujo manejo tem mais implicações em relação com a confiabilidade nos resultados” (p. 94).

Considerando Tenbrink (2010), de acordo com o que se pretende medir e o tempo de duração da prova, assim serão escolhidos os tipos de itens a constar na prova. A diversidade da tipologia de itens integrada nas provas de exame, em virtude das características dos itens, permite medir diferentes níveis de aprendizagem dos níveis mais simples aos mais complexos das taxonomias.

No que concerne à ordenação dos itens na generalidade das provas de exame, em cada grupo, surgem primeiro os itens de seleção, seguidos dos itens de construção, o que vai

ao encontro de Tenbrink (2010), quando refere que os itens devem estar ordenados de forma lógica.

A diversificação da tipologia dos itens nos exames, como diz Fernandes (2008), tem manifestado uma tentativa de equilibrar as exigências da validade, da fiabilidade, da equidade com as exigências do ensino no nosso sistema educativo.

CAPÍTULO 4 – METODOLOGIA DA INVESTIGAÇÃO

4.1. Opções metodológicas e sua justificação

O presente capítulo apresenta as opções metodológicas que estiveram subjacentes ao estudo empírico realizado.

Esta investigação teve como objetivo geral verificar a influência da tipologia dos itens nos resultados dos exames nacionais do ensino secundário – Prova de Biologia e Geologia – código 702, no ano de 2011.

4.1.1. Natureza da investigação

Considerando que a investigadora pretendia conhecer a opinião dos alunos relativamente ao grau de dificuldade dos itens da prova de exame de Biologia e Geologia, código 702, 1.^a fase do ano de 2011, e que pretendia questioná-los para compreender as razões das suas dificuldades, julgou ser o tipo estudo de caso o procedimento mais apropriado. Esta decisão parece ir ao encontro de Ponte (1994), quando refere que o estudo de caso é um tipo de pesquisa com uma forte dimensão descritiva, em que o investigador pretende compreender uma situação tal como ela é. Pode ter igualmente uma outra dimensão analítica, na medida em que pode interrogar a situação, confrontando-a com outras situações já conhecidas e com as teorias existentes. Deste modo, pode contribuir para gerar novas teorias e propor questões para estudos de investigação futuros. Segundo Sousa (2005), o estudo de caso caracteriza-se por uma investigação que visa a compreensão de um grupo de sujeitos perante um dado acontecimento, numa dada situação contextual específica, e irá permitir uma exploração mais aprofundada do problema a investigar no local a determinar.

No presente estudo, desenvolvido em quatro escolas do Alentejo, distrito de Évora, a opção para a recolha e análise da informação recaiu sobre procedimentos de natureza quantitativa, reforçados com elementos qualitativos, visando uma interpretação mais

sustentada, contextualizada e holística, conduzindo desta forma a uma triangulação que, segundo Yin (citado em Meirinhos & Osório, 2010) é

uma estratégia de validação, na medida em que torna possível a combinação de metodologias para estudo do mesmo fenómeno. Portanto, a triangulação permite obter de duas ou mais fontes de informação, dados referentes ao mesmo acontecimento, a fim de aumentar a fiabilidade da informação (p. 60).

Para a recolha de dados quantitativos, recorreu-se ao inquérito por questionário, tendo este sido aplicado nas quatro escolas, a todos os alunos que realizaram o exame nacional de Biologia e Geologia, código 702, na 1.^a fase de 2011. A vertente qualitativa do estudo foi suportada na análise da seguinte documentação:

- (I) informações-exame, emitidas pelo GAVE entre 2006 e 2011;
- (II) provas dos exames nacionais de 2006 a 2011;
- (III) documentos GAVE;
- (IV) relatórios finais dos exames nacionais realizados entre 2006 e 2011;
- (V) informações sobre os resultados obtidos pelos alunos internos na prova de exame nacional de Biologia e Geologia – Código 702 (1.^a fase de 2011), as quais, sendo da autoria do GAVE, foram disponibilizadas pelas escolas participantes no estudo.

4.1.2. Fases da investigação

Para a operacionalização do estudo empírico, foi necessário planear e desenvolver as seguintes fases:

- (I) Levantamento bibliográfico, investigando o “estado da arte” para conhecer e aprofundar os estudos realizados sobre a problemática dos exames ao longo do tempo e para aprofundar o estudo da tipologia dos itens e das suas características, imprescindível para sustentar teoricamente o conteúdo das partes 2 e 3 do questionário;
- (II) Leitura e análise dos documentos de suporte ao desenvolvimento do estudo;

- (III) Construção, validação e aplicação de um instrumento de recolha de informação, assente em um questionário;
- (IV) Análise e reflexão crítica sobre os resultados obtidos.

4.2. Critérios da escolha das escolas em estudo

As quatro escolas envolvidas no estudo situam-se no Alto Alentejo, no distrito de Évora. Por opção da investigadora, as escolas são identificadas neste estudo por Escola A, Escola B, Escola C e Escola D, de modo a garantir o anonimato das mesmas e dos participantes.

Para desenvolver o estudo empírico, optou-se, num primeiro momento, por escolas do distrito de Évora, por se encontrarem relativamente próximas da Universidade de Évora, local onde decorre o curso de mestrado da investigadora; num segundo momento, a opção teve por base a posição, em função das classificações obtidas nos exames nacionais de Biologia e Geologia (Código 702) de 2011, de cada uma das escolas no *ranking* do distrito de Évora. Inicialmente, pretendeu-se desenvolver o estudo nas duas escolas com melhores classificações e nas duas com classificações mais baixas. No entanto, por uma das escolas se ter mostrado indisponível para aplicar o questionário, decidiu-se alterar os critérios iniciais, passando a opção por selecionar a escola que apresentou a segunda melhor classificação, uma das que apresentou pior classificação e duas das escolas com classificações intermédias.

O motivo pelo qual o estudo contemplou escolas com resultados de classificação de exame diferentes prendeu-se com o facto de considerar opiniões de alunos, relativamente ao grau de dificuldade dos itens, de escolas que obtiveram classificações de exame diferentes e compará-las com os resultados obtidos. No entanto, ao longo do estudo, considerou-se pertinente fazer uma comparação desta análise entre as várias escolas. Deste modo, esta análise não se encontra prevista nos objetivos do estudo.

4.3. Instrumento de recolha de dados: Inquérito por questionário

Tendo em conta a abordagem de natureza quantitativa escolhida, utilizou-se, como instrumento de recolha de dados, o inquérito por questionário, dado o número elevado de participantes no estudo e por permitir a obtenção da informação necessária relativa às suas opiniões.

Um questionário é, segundo Ghiglione e Matalon (1997), “um instrumento rigorosamente estandardizado tanto no texto das questões, como na sua ordem” (p. 110). Ressalvam os mesmos autores que, de modo a que as respostas da totalidade dos indivíduos possam ser comparadas, é imprescindível que não haja variações na forma como as questões são colocadas a cada um deles.

Na perspetiva de Tuckman (citado em Afonso, 2005) o questionário pode estar orientado:

- (I) para a recolha de dados sobre o que o representante sabe (conhecimento ou informação);
- (II) para o que o respondente quer ou prefere (valores ou preferências);
- (III) para seleccionar o que o respondente pensa ou crê (atitudes e convicções).

A construção de um questionário, desde a sua fase de conceção até ao seu desenvolvimento, obriga a alguns requisitos imprescindíveis. Com base em Hill e Hill (2012), Iarossi (2011), Moreira (2009) e em Vicente, Reis e Ferrão (2001) foram considerados, para a construção do questionário, os seguintes aspetos:

- a) Definição do conteúdo e da forma do questionário - a formulação dos itens deve fornecer a informação necessária à resposta do(s) objetivo(s) do estudo; a linguagem deve ser cuidada e precisa; deve definir-se o tipo das questões (abertas ou fechadas), a ordem dos temas e das questões e o *layout*;
- b) Estudo piloto ou preliminar do questionário – a aplicação definitiva de um questionário deve ser antecedida de uma aplicação preliminar, com o objetivo de identificar situações pouco claras ao nível lexical, semântico e sintático, formato e número de itens, entre outros aspetos;
- c) Elaboração da versão definitiva do questionário – a versão definitiva deve ser elaborada atendendo as indicações e melhoramentos sugeridos a partir do estudo piloto ou preliminar e do “exame de conteúdos” (Moreira, 2009, p. 361).

4.3.1. A definição da forma e do conteúdo do questionário

Atendendo ao objetivo geral e aos objetivos específicos definidos para o estudo, bem como à revisão bibliográfica efetuada, desenhou-se um questionário, tendo-se tido em conta aspetos como a introdução, o *layout*, o tipo de resposta, o tipo de escala e o método a aplicar. No que diz respeito à introdução, houve a preocupação de, dar conhecimento, de forma muito breve, do objetivo do questionário, sensibilizando-se o respondente para a importância da sua opinião e para a natureza confidencial da mesma. Sendo a primeira informação que o respondente vai ler, a introdução assume-se de grande importância, na medida em que as primeiras impressões podem ser determinantes para uma boa cooperação (Hill & Hill, 2012).

A investigadora preocupou-se, igualmente, com o aspeto estético do questionário, pois considera que esteticamente o questionário deve mostrar-se apelativo de forma a estimular ao seu preenchimento. Como referem Hill e Hill (2012), o tamanho e a existência de espaços adequados entre os vários elementos do questionário, isto é, o seu *layout*, estimulam a boa vontade do respondente. No sentido de organizar as perguntas de forma lógica, optou-se por agrupá-las em temas homogéneos.

Deste modo, o questionário constitui-se de três partes, com a forma de questões fechadas. Nestas, de acordo com Ghiglione e Matalon (1997), é colocada uma questão ao indivíduo e de seguida é-lhe pedido, de entre uma lista preestabelecida de respostas possíveis, que indique a que melhor corresponde à que deseja dar. Neste sentido, e considerando o conceito de questionário apresentado por Ghiglione e Matalon (1997) (referido no ponto anterior), a standardização abrange não só o texto das questões como também as respostas.

Na primeira parte do questionário, pretende-se obter uma breve caracterização do aluno, contemplando-se a idade, o sexo e a fase em que realizou o exame nacional de Biologia e Geologia, código 702, em 2011.

Na segunda parte, apresenta-se os quatro tipos de itens constantes da prova suprarreferida, pedindo-se ao aluno que indique o grau de dificuldade de cada tipo de item, com base numa escala ordinal do tipo de Likert de quatro alternativas: 1 - Nada difícil; 2 - Pouco difícil; 3 - Difícil e 4 - Muito difícil.

Na terceira parte, elenca-se possíveis razões associadas ao grau de dificuldade dos diferentes tipos de itens, de entre as quais os alunos têm de selecionar as três que consideram principais.

No que diz respeito à escolha das escalas, na fase que antecedeu a escolha do número de respostas alternativas, foram analisadas exaustivamente as vantagens e desvantagens da utilização de um número par (quatro) ou ímpar (cinco) de respostas alternativas; a opção por uma delas constituiu-se uma decisão delicada.

Em virtude de a escala com número ímpar de alternativas poder conduzir a respostas neutras e havendo esta tendência por parte dos respondentes (Hill & Hill, 2012), e considerando a investigadora que havia todo o interesse em que o respondente manifestasse a sua verdadeira opinião, optou-se por uma escala com quatro intervalos. Um número par de respostas alternativas obriga a que todos os inquiridos deem uma opinião, definitivamente positiva ou negativa, não sendo possível dar uma resposta neutra (Hill & Hill, 2012).

Procurou-se que o questionário assentasse na clareza e na objetividade, recorrendo a afirmações curtas, precisas, diretas e limitadas a um só problema e que, simultaneamente, garantissem a abrangência e a pertinência no conteúdo a que se destinavam. No que concerne à redação de cada pergunta (item), Larossi (2011) considera que tem de ser “*breve, objectiva, simples e específica* (ou BOSE)” (p. 62).

Na parte dois do questionário, houve a preocupação de garantir que os respondentes tivessem presente em que consiste cada tipo de item, sendo, por isso, esta informação colocada junto de cada um dos tipos de itens, de forma simples, clara e objetiva.

Na parte três do questionário, procurou utilizar-se afirmações claras e objetivas, com uma linguagem clara e adequada aos respondentes (neste caso, alunos do 12.º ano).

Houve uma atenção muito dirigida ao conteúdo de cada afirmação, no sentido de este transmitir exatamente aquilo para que foi feito. Neste sentido, e para garantir a maior validade possível de cada afirmação, considerou-se a literatura específica na área da construção dos itens e das suas características, bem como os documentos do GAVE, nomeadamente os relatórios finais dos exames nacionais, nos quais são referidos os níveis de desempenho dos alunos nos itens e identificadas as dificuldades por si reveladas. Com base em toda a informação desta forma recolhida, houve grande preocupação em contemplar as diversas razões subjacentes ao grau de dificuldade dos itens.

Assim, tendo em consideração os pressupostos e procedimentos anteriores, construiu-se uma primeira versão do questionário.

4.3.2. Validação do questionário

Após a elaboração da versão inicial, o questionário deve ser submetido à análise e à crítica de especialistas (Moreira, 2009), capazes de detetar quaisquer problemas, como erros técnicos ou gramaticais, que obriguem a revisões dos itens. No sentido de pôr em prática estas orientações, solicitou-se a validação deste instrumento de recolha de informação a três especialistas do Departamento de Pedagogia e Educação da Universidade de Évora e a outros dois especialistas em avaliação. O contributo prestado por todos estes especialistas foi essencial para melhorar a primeira versão. Assim, com base nas sugestões de melhoria, foram feitas as seguintes alterações da 1ª para 2ª versão:

- (I) A escala ordinal do tipo Likert – 1- Muito reduzido, 2 - Reduzido, 3 - Elevado e 4 - Muito elevado – foi substituída por: 1 - Nada difícil, 2 - Pouco difícil, 3 - Difícil e 4 -Muito difícil; com esta substituição pretendeu-se homogeneizar os intervalos, dado que o intervalo entre os pontos 2 e 3 da escala inicial se apresentava maior comparativamente aos outros intervalos.
- (II) Alteração de alguns itens ao nível sintático e lexical, com vista a melhorar a clareza e a objetividade das afirmações.

Após a introdução das alterações referidas e de uma nova análise exaustiva, foi criada a versão definitiva do questionário (Apêndice A).

4.3.3. Estudo piloto ou preliminar do questionário

Conforme referem Pardal e Correia (1995), antes de se administrar um questionário, este deve ser aplicado a uma amostra reduzida, que em termos de estrutura deve ser homóloga à da amostra destinada ao estudo. Esta pré-testagem é necessária para o investigador se assegurar da qualidade das perguntas e da razoabilidade da sua ordenação.

De modo a respeitar as orientações anteriores, foram aplicados seis questionários a seis alunos do 12.º ano de uma escola secundária do distrito de Setúbal, sendo que todos eles tinham realizado exame nacional de Biologia e Geologia, código 702, no ano de 2011. Esta aplicação visou não só a identificação de eventuais dúvidas ou obstáculos sentidos pelos respondentes, assim como a obtenção da sua opinião sobre a pertinência das questões para a problemática em estudo, as dimensões omissas e o tempo de preenchimento do questionário. O questionário foi preenchido na presença da investigadora, em dia e hora previamente combinados com os alunos e respetivos encarregados de educação, altura em que foi dada a conhecer a finalidade do questionário. A aplicação foi feita fora do ambiente escolar, numa biblioteca municipal, para garantir um ambiente favorável à concentração dos alunos.

A investigadora, após a distribuição dos questionários pelos alunos, explicou-lhes, num primeiro momento, os objetivos deste instrumento de recolha de dados e a sua organização, tendo-os sensibilizado para a importância da sua opinião relativamente a aspetos relacionados com a estrutura, os conteúdos e a compreensão do questionário. Num segundo momento, os alunos preencheram o questionário; o tempo utilizado foi de sensivelmente 12 minutos, sem terem revelado qualquer observação ou dúvida.

Num terceiro momento, foi feita a apreciação do questionário pelos alunos, a qual se revelou favorável no que diz respeito à compreensão quer das orientações de preenchimento quer das afirmações. No que aludiu à segunda parte do questionário, relativa às razões da justificação das suas dificuldades em cada tipo de item, os alunos referiram que as razões elencadas no questionário abrangiam todas as razões possíveis, não tendo apresentado qualquer sugestão de alteração ou de melhoria.

4.3.4. Análise e aprovação do questionário pela DGIDC

Após concluídas as fases anteriormente descritas, submeteu-se o questionário à DGIDC, para que esta autorizasse a sua aplicação (Apêndice B), tal como estipulado no Despacho n.º 15847/2007, de 23 de Julho, que estabelece os procedimentos a cumprir nos pedidos de autorização para aplicação de inquéritos / realização de estudos de investigação em meio escolar.

A sua aprovação foi concedida (Apêndice C), da qual foi dado conhecimento à direção das escolas.

4.4. Aplicação do questionário

Depois de alterado o questionário e não havendo sugestões resultantes do estudo piloto, foi pedida autorização aos diretores das quatro escolas, via telefone, para a aplicação do questionário, tendo a investigadora obtido o aval de todos para o efeito (Apêndices D a G). Assim sendo, a versão final do questionário foi distribuída às escolas no dia dezasseis de maio. Não tendo sido possível realizar, por incompatibilidade de horários, uma pequena reunião para explicar os objetivos e o conteúdo do questionário aos professores que o iriam aplicar, esta tarefa foi feita, pela investigadora, de forma exaustiva, junto de cada diretor de escola, tendo sido ainda o diretor sensibilizado para a importância do retorno dos questionários. Ficou acordado, que a recolha destes seria feita pela investigadora dois dias depois, o que se veio a confirmar.

4.5. Recolha e tratamento dos dados

Os questionários foram recolhidos no mês de maio de 2012, em quatro escolas do distrito de Évora. A Tabela 1 mostra o número de questionários distribuídos e recolhidos, em cada uma das quatro escolas envolvidas no estudo:

Tabela 1

Número de questionários distribuídos em cada escola e respetiva taxa de retorno

Questionários	Escola A		Escola B		Escola C		Escola D		Total	
	N	Taxa retorno	N	Taxa retorno						
Distribuídos	69		16		19		29		133	
		76,8		62,5		78,9		96,6		79,7
Recolhidos	53		10		15		28		106	

Como se pode observar, a taxa de retorno foi de 79,7%, correspondente à entrega de 106 dos 133 questionários inicialmente distribuídos. A Escola A, com 53 alunos, é a mais representada na amostra (50%), seguida da Escola D, com 28 alunos (26,4%), e das escolas C e B com 15 (14,2%) e 10 (9,4%) alunos, respetivamente. A Escola D foi a que apresentou uma taxa de retorno mais elevada (96,6%).

Os dados obtidos nos questionários foram submetidos a tratamento estatístico pelo programa *Statistical Package for the Social Sciences* (SPSS), versão 20, e os resultados das provas de exames de Biologia e Geologia, Código 702, de 2011, gentilmente cedidos pelos diretores das respetivas escolas, foram tratados estatisticamente e graficamente no MS-Excel 2007 e MS-Word 2007.

Para proceder à análise quantitativa dos resultados, as duas escolas com classificações intermédias no *ranking* do distrito de Évora (Escola B e Escola C) foram agregadas de forma a equilibrar o número de alunos com as outras duas escolas (Escola com melhor resultado, Escola A, com N= 53 alunos, Escola com pior resultado, Escola D, com N= 28 alunos). As escolas com resultados intermédios, escolas B e C, com N=10 e N=15, respetivamente, ficaram assim, agrupadas, com designação B+C, com 25 alunos.

4.5.1. Caracterização dos alunos

Neste ponto apresentar-se-á a caracterização dos participantes quanto ao sexo, à idade e à fase de realização da prova de Biologia e Geologia no ano de 2011.

4.5.1.1 Caracterização dos alunos por sexo

Como se pode observar na Figura 1, dos 106 alunos, 56,6% são do sexo feminino e 43,4% são do sexo masculino.

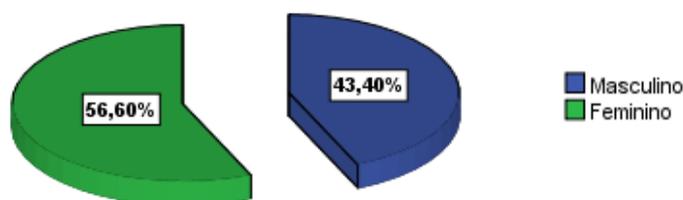


Figura 1. Distribuição dos alunos participantes por sexo.

A Tabela 2 apresenta os valores desta variável por escola

Tabela 2
Distribuição do sexo dos alunos participantes por escola

Escola	Sexo	F	%
A	Masculino	25	47,2
	Feminino	28	52,8
	Total	53	100,0
B	Masculino	5	50,0
	Feminino	5	50,0
	Total	10	100,0
C	Masculino	5	33,3
	Feminino	10	66,7
	Total	15	100,0
D	Masculino	11	39,3
	Feminino	17	60,7
	Total	28	100,0

Nas escolas C e D a percentagem de raparigas é mais elevada, respetivamente de 66,7% e 60,7%. Na Escola B não existem diferenças e, na Escola A, o número de raparigas é de 28 (52,8%), mais 3 que os rapazes.

4.5.1.2 Caracterização dos alunos por idade

Os alunos inquiridos frequentavam o 12.º ano de escolaridade, tendo realizado a prova de exame no ano letivo de 2010/2011, no final do 11.º ano. Assim, a idade indicada no questionário não corresponde à idade com que realizaram a prova, uma vez que esta foi realizada em junho de 2011 e o questionário foi aplicado em maio de 2012.

As idades encontram-se distribuídas no intervalo entre 17 anos e 20 anos, sendo a idade média de 17,55. O valor modal e mediano encontrado foi de 17 anos. A Figura 2 possibilita uma visualização mais facilitada da distribuição destes alunos pelas várias categorias de idades.

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

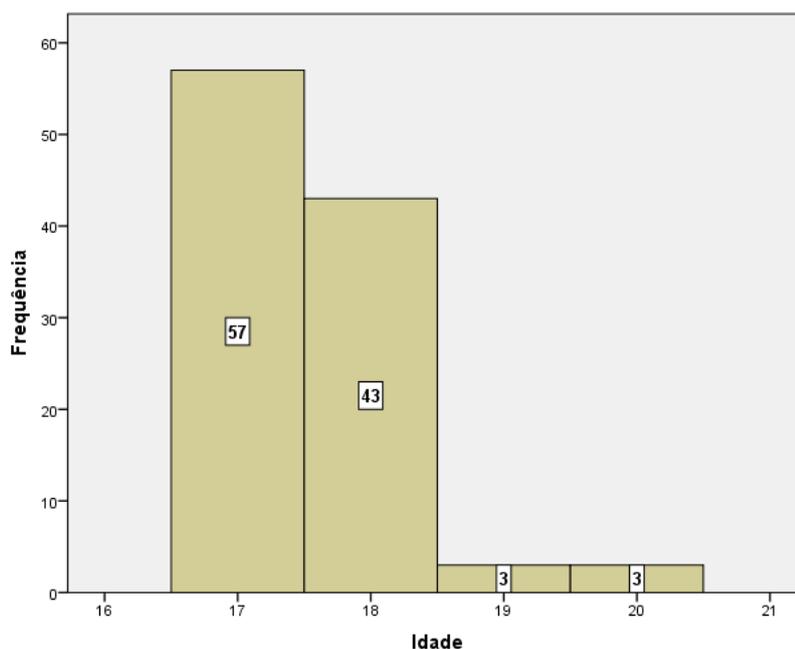


Figura 2. Distribuição dos alunos participantes por idade.

A maioria dos alunos (57%) tinha 17 anos, seguindo-se a classe dos 18 anos com 43 alunos (40,6%). Havia 3 alunos com 19 anos (2,8%) e outros tantos com 20 anos (2,8%), o que parece indicar que se está na presença de um grupo de alunos com um percurso escolar regular.

A Tabela 3 apresenta a distribuição das idades por escola.

Tabela 3

Distribuição das idades dos alunos participantes por escola

Escola	Idade	F	%
A	17	28	52,8
	18	23	43,4
	20	2	3,8
	Total	53	100,0
B	17	4	40,0
	18	6	60,0
	Total	10	100,0
C	17	11	73,3
	18	4	26,7
	Total	15	100,0
D	17	14	50,0
	18	10	35,7
	19	3	10,7
	20	1	3,6
	Total	28	100,0

Verifica-se, que, em todas as escolas, à exceção da B, predominam os participantes com 17 anos. Nas escolas B e C, todos os alunos tinham 17 ou 18 anos, enquanto na Escola A há a registar 2 alunos com 20 anos (3,8%) e, na Escola D, 3 alunos com 19 anos (10,7%) e um com 20 (3,6%).

4.5.1.3. Caracterização dos alunos por fase de realização da prova de Biologia e Geologia

Na questão 3 da 1.^a parte do questionário, «No ano letivo passado (2011), em que fase realizaste a prova de Biologia e Geologia (Código 702)?», os alunos indicaram a fase em que realizaram a prova de Biologia e Geologia, sendo que esta podia ter sido realizada na 1.^a ou na 2.^a fase ou em ambas. A Tabela 4 apresenta a distribuição dos alunos participantes por fase de realização da prova de Biologia e Geologia.

Tabela 4

Distribuição dos alunos participantes por fase de realização da prova de Biologia e Geologia

Fases	F	%	% Válida
1 ^a	75	70,8	71,4
2 ^a	7	6,6	6,7
1 ^a e 2 ^a	23	21,7	21,9
Total	105	99,1	100,0
<i>Missing</i>	1	,9	
	106	100,0	

Com uma percentagem de 70,8%, a maioria dos alunos (75) realizou a prova apenas na primeira fase, sendo que 7 alunos (6,6%) optaram pela segunda fase, e 23 alunos prestaram prova nas duas fases. Um dos alunos (0,9%) não respondeu à questão.

CAPÍTULO 5 – APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Tendo em conta os objetivos do presente estudo: (I) saber quais os resultados obtidos no exame nacional de Biologia e Geologia (Código 702), realizado no ano de 2011, em diferentes tipos de itens; (II) conhecer as perceções que alunos do 12.º ano de escolaridade têm quanto ao grau de dificuldade de cada tipo de item e compará-las com os resultados obtidos nas provas de exame; e (III) identificar as razões que contribuem para o grau de dificuldade das perguntas (itens) dos Exames Nacionais de Biologia e Geologia e recorrendo às técnicas estatísticas referidas no ponto 4.5, passa-se agora a descrever e interpretar os resultados obtidos.

Assim, começa-se por analisar os resultados obtidos pelos alunos na referida prova, por escola, para cada um dos tipos de itens. Posteriormente, faz-se a análise relativa à perceção dos alunos quanto ao grau de dificuldade dos itens seguida de uma análise comparativa com os resultados obtidos. Por último, será feita uma análise da perceção dos alunos quanto às razões que contribuem para o grau de dificuldade dos itens.

5.1. Resultados obtidos, pelos alunos internos, na prova de exame nacional de Biologia e Geologia (Código 702), em 2011

Uma vez que o presente estudo se debruçou sobre os resultados do exame nacional de Biologia e Geologia, código 702, 1ª fase de 2011, das quatro escolas em estudo, era imprescindível conhecer os resultados dos alunos por item constante da prova.

Neste sentido, a investigadora solicitou junto dos diretores de cada uma das escolas os dados suprarreferidos, fornecidos a cada escola pelo GAVE. Constatou-se que os dados

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

se referem aos resultados médios por item, da prova, e às respetivas cotações máximas e nulas.

Estes dados foram tratados com recurso ao MS-Excel 2007. Passamos agora a descrever e interpretar esses resultados, apresentados por escola, para os quatro tipos de item (Tabelas 5, 6 e 7) e, para facilitar a comparação, por tipologia de item para as quatro escolas (Tabelas 8 a 12). Os resultados obtidos corroboram a classificação destas escolas relativamente à sua classificação no *ranking* (a nível do distrito de Évora), uma vez que o valor médio dos resultados médios obtidos é superior em A ($M = 59,150$), intermédio para B+C ($M = 54,125$) e inferior para D ($M = 39,250$).

Tabela 5

Resultados médios e cotação nula por tipo de item: Escola A

Tipos de Item	Resultados médios	Cotação nula
	(em %)	(% alunos)
Escolha Múltipla	65,7	34,3
Ordenação	64,4	35,6
Resposta Restrita	29,0	46,9
Correspondência / Associação	77,5	7,5
Média	59,150	31,075

Na Escola A, a melhor do *ranking*, os melhores resultados dizem respeito aos itens de tipo correspondência / associação (77,5%), o mesmo se verificando nas restantes escolas, embora com valores mais baixos (B+C = 75,8%; D = 60%). Os itens de resposta restrita foram os que apresentaram piores resultados (29%), com uma taxa de cotação nula de 46,9%. Os itens de Escolha Múltipla e de Ordenação apresentaram valores médios superiores a 50%, respetivamente 65,7% e 64,4%.

Tabela 6

Resultados médios e cotação nula por tipo de item: Escolas B+C

Tipos de Item	Resultados médios	Cotação nula
	(em %)	(% alunos)
Escolha Múltipla	55,3	31,0
Ordenação	59,9	40,1
Resposta Restrita	25,5	49,0
Correspondência / Associação	75,8	8,8
Média	54,125	32,225

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

Nas escolas B+C há um decréscimo dos valores médios obtidos para os quatro tipos de itens ($M^A - M^{B+C}=5,025\%$). Nestas escolas, contrariando o que acontece em A (Tabela 5) e D (Tabela 7), os itens de ordenação apresentaram melhores resultados médios que os itens de escolha múltipla.

Tabela 7

Resultados médios e cotação nula por tipo de item: Escola D

Tipos de Item	Resultados médios	Cotação nula
	(em %)	(% alunos)
Escolha Múltipla	42,0	38,0
Ordenação	33,3	66,7
Resposta Restrita	21,7	55,0
Correspondência / Associação	60,0	20,0
Média	39,250	44,925

A Escola D apresentou, claramente, os piores resultados, coincidindo, assim, com a sua posição no *ranking* do distrito. Os itens de correspondência foram os únicos em que se registaram resultados médios acima dos 50% (60%). Seguiram-se os itens de escolha múltipla, com 42%, e de ordenação, com 33,3%, ambos com taxas de cotação nula superiores a 50%: 66,7% e 55%, respetivamente. Em termos de comparação com a melhor escola, a Escola D sofre uma redução de 19,9% nos resultados médios.

Em suma, constata-se que a ordenação dos resultados obtidos na prova de Biologia e Geologia coincidiu com a posição das escolas no *ranking* do distrito de Évora.

De seguida, apresenta-se os resultados médios obtidos para os quatro tipos de itens (Tabelas 8 a 11), das quais resulta a Tabela 12, que permite uma visualização mais facilitada e comparada dos valores apresentados:

Tabela 8

Resultados médios e cotação nula para os itens de escolha múltipla

GRUPOS - ITENS	Resultados médios	Cotação nula	ESCOLAS
	(em %)	(% alunos)	
I – 1,2,3,4,5,6	65,7	34,3	A
II – 1,2,3,4,5,6,7	55,3	31,0	B + C
III – 1,2,3,4,5	42,0	38,0	D
	Média – Resultados médios	53,7	
	Média - Cotação nula	34,4	

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

A prova de Biologia e Geologia apresentava 24 itens de escolha múltipla, distribuídos por quatro grupos. Os resultados médios variaram entre 42% na Escola D e 65,7% na Escola A, originando um valor médio de 53,7 % para o total de escolas. A percentagem média de cotação nula foi de 34,4%.

Tabela 9

Resultados médios e cotação nula para os itens de ordenação

GRUPOS - ITENS	Resultados médios	Cotação nula	ESCOLAS
	(em %)	(% alunos)	
III - 6	64,4	35,6	A
IV - 7	59,9	40,1	B + C
	33,3	66,7	D
	Média – Resultados médios		52,5
	Média - Cotação nula		47,5

A prova de Biologia e Geologia apresentava 2 itens de ordenação, distribuídos por dois grupos. Os resultados médios variaram entre 33,3% na Escola D e 64,4% na Escola A, originando um valor médio de 52,5 % para o total de escolas. A percentagem média de cotação nula foi de 47,5%.

Tabela 10

Resultados médios e cotação nula para os itens de resposta restrita

GRUPOS - ITENS	Resultados médios	Cotação nula	Escolas
	(em %)	(% alunos)	
I - 8	29,0	46,9	A
II - 8	25,5	49,0	B + C
III - 7	21,7	55,0	D
IV - 8			
	Média – Resultados médios		25,4
	Média - Cotação nula		50,3

A prova de Biologia e Geologia apresentava 4 itens de resposta restrita, distribuídos por quatro grupos. Os resultados médios variaram entre 21,7% na Escola D e 29,0% na Escola A, originando um valor médio de 25,4 % para o total de escolas. A percentagem média de cotação nula foi de 50,3%.

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

Tabela 11

Resultados médios e cotação nula para os itens de correspondência / associação

GRUPOS - ITENS	Resultados médios	Cotação nula	Escolas
	(em %)	(% alunos)	
I - 7	77,5	7,5	A
	75,8	8,8	B + C
	60,0	20,0	D
Média – Resultados médios		71,1	
Média - Cotação nula		12,1	

A prova de Biologia e Geologia apresentava apenas um item de correspondência / associação, no grupo I. Os resultados médios variaram entre 60% na Escola D e 77,5% na Escola A, originando um valor médio de 71,1 % para o total de escolas. A percentagem média de cotação nula foi de 12,1%.

De forma a facilitar a comparação dos valores médios para cada item, construiu-se a Tabela 12.

Tabela 12

Valores médios e cotação nula, para cada tipo de item

Tipo de itens	Resultados médios	Cotação nula
	(em %)	(% alunos)
Itens de Escolha Múltipla	53,7	34,4
Itens de Ordenação	52,5	47,5
Itens de Resposta Restrita	25,4	50,3
Itens de Correspondência /Associação	71,1	12,1

Há uma clara evidência de que os itens de correspondência / associação são os que apresentam melhores resultados médios (71,1%), em contraposição com os itens de resposta restrita (25,4%). Os itens de escolha múltipla e de ordenação parecem apresentar níveis de dificuldade semelhante (mas com resultados médios inferiores aos itens de correspondência). Há que salientar que os itens de escolha múltipla se encontram na prova em maior número (N = 24), quando comparados com os itens de

ordenação (N = 2), o que pode estar a influenciar os resultados, até porque a percentagem de cotação nula é superior para os itens de ordenação. Os itens de resposta restrita são os únicos que apresentaram uma percentagem de cotação nula superior a 50% (50,3%), ou seja, cerca de metade dos alunos não obtêm qualquer pontuação nestes itens.

Com base na experiência profissional da investigadora, a ordem relativa dos tipos de itens no que aos resultados médios respeita era expectável, parecendo estar em consonância com os resultados médios dos alunos nas provas de avaliação realizadas a nível de escola. Nestas, à semelhança do que se constatou no exame nacional em estudo, os alunos tendem a ter melhor desempenho nos itens de correspondência / associação, sendo os itens de resposta restrita aqueles em que obtêm piores resultados.

Os itens de associação apresentam características apropriadas para obter informação sobre o conhecimento do indivíduo, podendo avaliar resultados simples da aprendizagem, (Tenbrink, 2010), o que se pode constatar no item de associação da prova de exame em estudo e, podendo, deste modo, levar a um bom desempenho dos alunos, o que vai ao encontro dos estudos realizados pelo GAVE, em 2002, que integraram a disciplina de Biologia. Estes estudos concluíram que quando o que está em causa são saberes que exigem pouca elaboração cognitiva, tais como o conhecimento de conceitos, de procedimentos, de factos ou a mera reprodução de informação, o desempenho dos alunos é médio ou mesmo bom (Fernandes, 2008a).

Relativamente aos itens de escolha múltipla, ordenação e de resposta restrita, pelas características que apresentam, normalmente, testam níveis de aprendizagem mais elevados (Tenbrink, 2010, Goring, 1981, Ribeiro, 1990), o que pode levar a níveis de desempenho mais baixos. No caso dos itens de resposta restrita, que apresentaram os valores médios mais baixos de entre os vários tipos de itens, os alunos tendem a revelar dificuldades ao nível da compreensão das indicações fornecidas, pelo que, muitas vezes, as suas respostas não se coadunam com o que lhes é solicitado, a avaliar pela percentagem de cotação nula verificada nas quatro escolas em estudo.

Os resultados médios constatados, e nomeadamente os que dizem respeito às respostas restritas, vão, uma vez mais ao encontro dos estudos do GAVE, realizados em 2002, nos quais se concluiu que quando é pedido aos alunos que apliquem conhecimentos que integram a mobilização de saberes em novas situações ou que elaborem raciocínios mais complexos, o seu desempenho é fraco ou muito fraco. Para além disso e considerando o relatório relativo aos exames nacionais de 2011, elaborado por aquele organismo,

verifica-se que os itens em que os alunos revelaram pior desempenho foram os de resposta restrita e alguns de escolha múltipla, em que nos primeiros se avaliava a explicação de contextos em análise, a interpretação de dados de um suporte, a aplicação a novas situações e a relação entre conceitos, enquanto que nos itens de escolha múltipla, se pretendia avaliar o conhecimento e a compreensão de dados, conceitos, modelos e teorias.

5.2. Perceção dos alunos quanto ao grau de dificuldade dos itens

Para se saber a opinião dos alunos, quanto ao grau de dificuldade dos diferentes tipos de itens, apresentam-se em tabelas as frequências absolutas de resposta e respetivas percentagens, por tipo de item e para cada escola, bem como os valores referentes à média e à moda para as diferentes variáveis (tipo de item).

No decorrer da investigação e apesar de não fazer parte dos objetivos do presente estudo, considerou-se que seria pertinente saber se o grau médio de dificuldade sentido pelos estudantes, para cada uma das quatro variáveis, é ou não idêntico nas escolas em estudo, pelo que houve a necessidade de proceder à testagem da Normalidade e à averiguação da homogeneidade das variâncias.

5.2.1. Resultado das variáveis «Item de escolha múltipla», «Item de ordenação», «Item de resposta restrita (aberta)» e «Item de associação / correspondência»

Na segunda parte do questionário, respeitante à opinião dos alunos relativamente ao grau de dificuldade dos diferentes tipos de itens, é pedido aos alunos que indiquem o grau de dificuldade para cada tipo de item, considerando a seguinte escala: 1 - Nada difícil; 2 – Pouco difícil; 3 – Difícil; e 4 – Muito difícil. A Tabela 13 apresenta as frequências absolutas de resposta e respetivas percentagens, por tipo de item, para cada escola:

**Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de
Biologia e Geologia.**

Tabela 13

Respostas e respetivas percentagens, por tipo de item, para cada escola

Escola	Grau de dificuldade	Escolha Múltipla		Ordenação		Resposta restrita		Associação	
		N	%	N	%	N	%	N	%
A	Nada difícil	3	5,7	3	5,7	3	5,7	16	30,2
	Pouco difícil	32	60,4	23	43,4	13	24,5	31	58,5
	Difícil	17	32,1	19	35,8	28	52,8	4	7,5
	Muito difícil	1	1,9	8	15,1	9	17,0	2	3,8
	Total	53	100,0	53	100,0	53	100,0	53	100,0
B+C	Nada difícil	0	0	1	4,0	1	4,0	3	12,0
	Pouco difícil	9	36,0	3	12,0	4	16,0	7	28,0
	Difícil	15	60,0	14	56,0	7	28,0	15	60,0
	Muito difícil	1	4,0	7	28,0	13	52,0	0	0
	Total	25	100,0	25	100,0	25	100,0	25	100,0
D	Nada difícil	0	0	0	0	0	0	0	0
	Pouco difícil	7	25,0	8	28,6	5	17,9	14	50,0
	Difícil	19	67,9	13	46,4	10	35,7	10	35,7
	Muito difícil	2	7,1	7	25,0	13	46,4	4	14,3
	Total	28	100,0	28	100,0	28	100,0	28	100,0
Total da amostra	Nada difícil	3	2,8	4	3,8	4	3,8	19	17,9
	Pouco difícil	48	45,3	34	32,1	22	20,8	52	49,1
	Difícil	51	48,1	46	43,4	45	42,5	29	27,4
	Muito difícil	4	3,8	22	20,8	35	33,0	6	5,7
	Total	106	100,0	106	100,0	106	100,0	106	100,0

Pela análise das frequências relativas a cada um dos itens, podemos verificar que 60,4% dos alunos da Escola A considera os itens de escolha múltipla pouco difíceis, enquanto os alunos das escolas B+C (60%) e D (67,9%) os consideram difíceis. Em termos de total da amostra, 93,4% dos alunos considera este tipo de item difícil (48,1%) ou pouco difícil (45,3%).

Relativamente ao item de ordenação, os alunos da Escola A voltam a apresentar uma opinião diferente da dos seus colegas, considerando este item pouco difícil (43,4%) contrariamente aos 56% e 46,4% dos seus colegas das escolas B+C e D, respetivamente, que consideram o item de ordenação como difícil.

O item de resposta restrita foi considerado muito difícil nas escolas B+C (52,0%) e na Escola D (46,4%); por sua vez, os alunos da Escola A voltam a considerar um grau de dificuldade imediatamente abaixo dos seus colegas, «Difícil», com um valor de 52,8%.

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

Estes alunos consideram como «pouco difícil» o item de associação, com 58,5%, o mesmo acontecendo com os alunos da Escola D (50,0%). Apenas os alunos das escolas B+C referiram ser «difícil» (60,0%) o item de associação.

A Tabela 14 apresenta os valores para a média e moda para as variáveis «Item de escolha múltipla», «Item de ordenação», «Item de resposta restrita» e «Item de associação / correspondência», permitindo, assim, uma mais fácil comparação dos valores médios obtidos para cada tipo de item, para cada escola, bem como a classe modal para cada tipo de item por escola:

Tabela 14

Média e moda por tipo de item e por escola

	A	B+C	D	Total	A	B+C	D	Total
	Média				Moda			
Item escolha múltipla	2,30	2,68	2,82	2,53	Pouco Difícil (2)	Difícil (3)	Difícil (3)	Difícil (3)
Item de ordenação	2,60	3,08	2,96	2,81	Pouco Difícil (2)	Difícil (3)	Difícil (3)	Difícil (3)
Item de resposta restrita	2,81	3,28	3,29	3,05	Difícil (3)	Muito Difícil (4)	Muito Difícil (4)	Difícil (3)
Item de associação	1,85	2,48	2,64	2,21	Pouco Difícil (2)	Difícil (3)	Pouco Difícil (2)	Pouco Difícil (2)
Valor médio para os 4 tipos de itens por escola	2,39	2,88	2,93	2,65				

A análise dos resultados indica, que os itens de resposta restrita foram considerados, pelos alunos, como mais difíceis ($M^A = 2,81$; $M^{B+C} = 3,28$; $M^D = 3,29$; $M^{Total} = 3,05$), seguidos dos itens de ordenação ($M^A = 2,60$; $M^{B+C} = 3,08$; $M^D = 2,96$; $M^{Total} = 2,81$) e dos itens de escolha múltipla ($M^A = 2,30$; $M^{B+C} = 2,68$; $M^D = 2,82$; $M^{Total} = 2,53$). Os itens de associação foram considerados os menos difíceis ($M^{Total} = 2,21$), principalmente pelos alunos da Escola A ($M^A = 1,85$). No global, os alunos da escola melhor classificada referiram sentir menos dificuldade ($M = 2,39$) do que os seus colegas das escolas B+C ($M = 2,88$) e D ($M = 2,93$) nos tipos de itens referidos.

Apesar de não fazer parte dos objetivos do estudo saber se o grau médio de dificuldade, sentido pelos estudantes, para cada uma das quatro variáveis, é ou não idêntico nas

**Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de
Biologia e Geologia.**

escolas em estudo, considerou-se ser pertinente proceder à testagem da Normalidade e à averiguação da homogeneidade das variâncias, com a finalidade de perceber se se poderia aplicar testes paramétricos, mais robustos, em que a probabilidade de rejeitar corretamente H_0 é maior. Assim, apresentam-se, agora, os resultados do Teste à normalidade Kolmogorov-Smirnov (K-S) - Tabela 15, e do Teste de Levene à homogeneidade das variâncias - Tabela 16:

Tabela 15

Teste à normalidade da amostra

Teste à Normalidade							
Tipo de Item	Escola	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Estatística	Gl	<i>p</i>	Estatística	Gl	<i>p</i>
II a - Item de escolha múltipla	A	,351	53	,000	,771	53	,000
	B+C	,357	25	,000	,721	25	,000
	D	,378	28	,000	,720	28	,000
II b - Item de ordenação	A	,261	53	,000	,854	53	,000
	B+C	,298	25	,000	,812	25	,000
	D	,233	28	,000	,813	28	,000
II c - Item de resposta restrita	A	,293	53	,000	,847	53	,000
	B+C	,311	25	,000	,776	25	,000
	D	,290	28	,000	,776	28	,000
II d - Item de associação	A	,304	53	,000	,771	53	,000
	B+C	,367	25	,000	,708	25	,000
	D	,310	28	,000	,761	28	,000

a. Correção de Lilliefors

A tabela apresenta o Teste Kolmogorov-Smirnov com correção de Lilliefors para as três amostras da variável «Escola» nas variáveis em estudo. Com uma probabilidade de erro de 5%, pode concluir-se que a distribuição das variáveis nas três amostras não segue a Normal (*p-value* = .000 com $\alpha = 0,05$). Note-se, ainda que no caso das escolas B+C e D, o valor de significância deve ser lido na parte direita da tabela, referente ao teste de Shapiro-Wilk, uma vez que este teste é mais indicado para amostras com menos de 50 observações (Maroco, 2011).

Tabela 16

Teste à homogeneidade das variâncias

Teste de Homogeneidade das Variâncias					
		Teste de Levene	GI 1	GI 2	p
II a - Item escolha múltipla	<i>Com base na Média</i>	,971	2	103	,382
	<i>Com base na Mediana</i>	,318	2	103	,728
	<i>Com base na Mediana corrigida</i>	,318	2	102,020	,728
	<i>Com base na Média aparada</i>	,919	2	103	,402
II b - Item de ordenação	<i>Com base na Média</i>	1,867	2	103	,160
	<i>Com base na Mediana</i>	1,566	2	103	,214
	<i>Com base na Mediana corrigida</i>	1,566	2	101,808	,214
	<i>Com base na Média aparada</i>	1,625	2	103	,202
II c - Item de resposta restrita	<i>Com base na Média</i>	,883	2	103	,417
	<i>Com base na Mediana</i>	,787	2	103	,458
	<i>Com base na Mediana corrigida</i>	,787	2	84,109	,459
	<i>Com base na Média aparada</i>	,988	2	103	,376
II d - Item de associação	<i>Com base na Média</i>	1,115	2	103	,332
	<i>Com base na Mediana</i>	1,039	2	103	,358
	<i>Com base na Mediana corrigida</i>	1,039	2	87,313	,358
	<i>Com base na Média aparada</i>	,706	2	103	,496

Uma vez que a variável dependente não segue a distribuição Normal, usou-se o Teste de Levene baseado na mediana: $F(2,103) = 0.318$; $p = 0.728$ para a variável «Item de associação»; $F(2,103) = 1,566$; $p = 0.214$ para a variável «Item de ordenação»; $F(2,103) = 0.789$; $p = 0.458$ para a variável «Item de resposta restrita»; e $F(2,103) = 1.039$; $p = 0.358$ para a variável «Item de associação». Pode, assim, concluir-se, que as variâncias estimadas são homogéneas, uma vez que todos os *p-values* calculados são superiores a $\alpha = 0,05$.

Com estes resultados, optou-se por realizar o Teste de Kruskal-Wallis, para proceder à realização da comparação múltipla das médias das ordens, uma vez que a versão 20 do SPSS já possui este módulo de testes não paramétricos, para amostras independentes. A Figura 3, relativa ao *output* obtido, indica que podemos rejeitar que a distribuição das 4 variáveis é a mesma entre as categorias da Escola.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of II a - Item escolha múltipla is the same across categories of 1 - Escola BC.	Independent-Samples Kruskal-Wallis Test	,001	Reject the null hypothesis.
2	The distribution of II b - Item de ordenação is the same across categories of 1 - Escola BC.	Independent-Samples Kruskal-Wallis Test	,021	Reject the null hypothesis.
3	The distribution of II c - Item de resposta restrita is the same across categories of 1 - Escola BC.	Independent-Samples Kruskal-Wallis Test	,009	Reject the null hypothesis.
4	The distribution of II d - Item de associação is the same across categories of 1 - Escola BC.	Independent-Samples Kruskal-Wallis Test	,000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Figura 3. Output do teste de Kruskal-Wallis

Com a análise do *output* das comparações múltiplas (Figura 4), disponível no *Model Viewer*, pode concluir-se que as diferenças significativas ocorrem entre a Escola A e B+C ($p_{aj} = .011$), e entre A e D ($p_{aj} = .000$), com $\alpha = 0,05$:

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
A-B+C	-16,893	6,654	-2,539	,011	,033
A-D	-23,202	6,407	-3,621	,000	,001
B+C-D	-6,309	7,546	-,836	,403	1,000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.

Figura 4. Output das comparações múltiplas

Com um valor de $p_{aj} = 0.403$, podemos rejeitar que existem diferenças significativas entre as escolas B+C e D, enquanto entre as escolas A e B+C e A e D se registam diferenças.

5.3. Análise comparativa dos resultados obtidos no exame com as percepções dos alunos quanto ao grau de dificuldade dos itens

A Tabela 17 compara os resultados médios obtidos em cada um dos tipos de item com as opiniões dos alunos, relativamente ao grau de dificuldade.

Tabela 17

Comparação dos resultados médios reais e o grau de dificuldade percebido pelos alunos

Tipo de itens	Resultados médios (em %)	Grau de dificuldade (percebido pelos alunos)	
		Média	Grau de dificuldade (Moda)
Itens de Correspondência /Associação	71,1	2,21	2 (Pouco difícil)
Itens de Escolha Múltipla	53,7	2,53	3 (Difícil)
Itens de Ordenação	52,5	2,81	3 (Difícil)
Itens de Resposta Restrita	25,4	3,05	3 (Difícil)

Como se pode observar na Tabela 17, parece existir uma correspondência entre a opinião dos alunos sobre a dificuldade de um tipo de item e os resultados obtidos, isto é, o item em que foi declarada maior dificuldade, «resposta restrita» ($M = 3,05$), foi também aquele em que os alunos obtiveram piores resultados (25,4%). No pólo oposto surge o item de correspondência, com 71,1% de resultado médio, e um valor de média de 2,21 (Pouco difícil).

5.4. Perceção dos alunos quanto às razões que contribuem para o grau de dificuldade das perguntas (itens)

Neste ponto apresentam-se os dados obtidos na 3.^a parte do questionário, fazendo a análise para cada uma das variáveis: itens de escolha múltipla, itens de ordenação, itens de resposta restrita e itens de associação.

Para cada uma destas variáveis serão apresentadas para a globalidade das escolas do estudo, a frequência das razões que, na opinião dos alunos, contribuem para o grau de dificuldade dos itens e o cruzamento das razões assinaladas pelos alunos em função do

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

grau de dificuldade que por eles foi atribuído em cada um dos itens. Este último procedimento será efetuado para cada uma das escolas.

A questão relativa às razões que podem contribuir para o grau de dificuldade dos itens tem várias opções de resposta, das quais três teriam de ser selecionadas. Assim, optou-se por tratar cada questão como múltipla. Como cada variável só pode assumir um valor válido, cada opção de resposta foi assumida como uma variável, dicotómica, definida por «Selecioneada» com valor 1, e «Não Selecionada» com valor 2. Realizado este procedimento no programa SPSS (v.20), foi posteriormente selecionado o menu *Analyse*, escolhido o comando *Multiple response* e marcada a opção *Define set*, definindo, assim, o conjunto de variáveis múltiplas (*set*), com o qual se passou a trabalhar, como nova variável. Foi também definido o valor 1 como valor de contagem.

5.4.1. – Resultado das variáveis para «Item de escolha múltipla»

A questão relativa às razões que podem contribuir para o grau de dificuldade dos itens de escolha múltipla tem 9 opções de resposta, das quais três teriam de ser selecionadas.

A Tabela 18 indica a frequência das razões que, na opinião dos alunos, contribuem para o grau de dificuldade dos itens de escolha múltipla.

Tabela 18

Razões que contribuem para o grau de dificuldade dos itens de escolha múltipla

Itens de Escolha múltipla Razões que contribuem para o grau de dificuldade	Respostas		% de casos
	N	%	
1 - Compreender os textos fornecidos	30	9,5	28,3
2 - Interpretar tabelas / gráficos / esquemas fornecidos	20	6,3	18,9
3 - Compreender o que é pedido	32	10,1	30,2
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)	50	15,8	47,2
5 - Relacionar as diversas informações fornecidas (textos / tabelas / gráficos / esquemas).	27	8,5	25,5
6 - Complexidade da matéria.	37	11,7	34,9
7 - As opções de resposta são muito semelhantes entre si.	65	20,5	61,3
8 - As opções de resposta requerem conhecimento aprofundado da matéria.	34	10,7	32,1
9 - Selecionar a informação necessária para a resposta.	22	6,9	20,8
Total	317	100,0	299,1

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

Relativamente aos itens de escolha múltipla, as três principais razões apontadas pelos alunos para o grau de dificuldade sentido são «As opções de resposta são muito semelhantes entre si», com 61,3% de escolha, seguindo-se «Relacionar os conhecimentos com as informações fornecidas (textos/tabelas/gráficos e esquemas)», com 47,2%, e, por último, «a complexidade da matéria», com 34,9%. A razão menos escolhida como responsável pelo grau de dificuldade sentido pelos alunos neste tipo de item é «Interpretar tabelas/ gráficos/ esquemas fornecidos», com 18,9%.

A Tabela 19 cruza as razões assinaladas pelos alunos em função do grau de dificuldade que por eles foi atribuído aos itens de escolha múltipla.

Tabela 19

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de escolha múltipla

Itens de Escolha múltipla		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
Razões que contribuem para o grau de dificuldade						
1 - Compreender os textos fornecidos	N	1	16	13	0	30
	% Total	0,3	5,0	4,1	0,0	9,5
2 - Interpretar tabelas / gráficos / esquemas fornecidos	N	1	11	6	2	20
	% Total	0,3	3,5	1,9	0,6	6,3
3 - Compreender o que é pedido	N	1	15	15	1	32
	% Total	0,3	4,7	4,7	0,3	10,1
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)	N	2	23	23	2	50
	% Total	0,6	7,3	7,3	0,6	15,8
5 - Relacionar as diversas informações fornecidas (textos / tabelas / gráficos / esquemas).	N	2	10	14	1	27
	% Total	0,6	3,2	4,4	0,3	8,5
6 - Complexidade da matéria.	N	1	15	19	2	37
	% Total	0,3	4,7	6,0	0,6	11,7
7 - As opções de resposta são muito semelhantes entre si.	N	1	29	34	1	65
	% Total	0,3	9,1	10,7	0,3	20,5
8 - As opções de resposta requerem conhecimento aprofundado da matéria.	N	0	11	22	1	34
	% Total	0,0	3,5	6,9	0,3	10,7
9 - Selecionar a informação necessária para a resposta.	N	0	14	6	2	22
	% Total	0,0	4,4	1,9	0,6	6,9
Total	N	9	144	152	12	317
	% Total	2,8	45,4	47,9	3,8	100,0

A análise dos resultados indica que, nos itens de escolha múltipla, os alunos atribuíram, maioritariamente, um grau de dificuldade partilhado entre «Difícil» (47,9%) e «Pouco Difícil» (45,4%).

As três principais justificações para estes graus de dificuldade foram «As opções de resposta são muito semelhantes entre si», com 34 escolhas para «Difícil» e 29 escolhas para «Pouco difícil», num total de 65 escolhas, seguida da razão «Relacionar os conhecimentos com as informações fornecidas (textos/tabelas/gráficos e esquemas)», com 23 escolhas para o grau de dificuldade «Pouco difícil» e 23 escolhas para o grau de dificuldade «Difícil», num total de 50, e, em terceiro lugar, a «Complexidade da matéria», com 19 escolhas para «Difícil» e 15 para o «Pouco difícil» num total de 37 escolhas.

A razão menos escolhida para justificar o grau de dificuldade deste tipo de item foi «Interpretar tabelas/gráficos/esquemas fornecidos», apresentando 11 escolhas para os alunos que consideraram o item «Pouco difícil» e 6 para quem considerou o item «Difícil», num total de 20 escolhas.

Relembra-se que esta tipologia de item apresentou uma percentagem de cotação média de 53,7% e de 34,4% de cotação nula.

Salienta-se o caso da Escola D, que apresenta percentagens médias abaixo dos 50% neste tipo de item.

Considerando a experiência profissional da investigadora, esta tipologia de item apresenta algumas flutuações em termos do desempenho dos alunos. Os itens de escolha múltipla são muito versáteis, podendo avaliar aprendizagens mais ou menos complexas, o que conduz a maiores ou menores dificuldades dos alunos e conseqüentemente a melhores ou piores desempenhos. Por outro lado, são itens com critérios de correção que só admitem a cotação total da resposta ou então zero.

Tendo em conta as opiniões dadas pelos alunos, estes consideram que, no caso dos itens de escolha múltipla, as suas dificuldades se prendem com “as opções de resposta serem muito semelhantes entre si”, o que é uma característica do próprio tipo de item. De acordo com Pinto (2001), “um distrator (alternativa incorreta) deve assemelhar-se a uma resposta correta e gerar dúvidas relevantes naqueles estudantes que não possuem um conhecimento específico” (p. 34). Consideram ainda que um dos motivos das suas dificuldades se prende com a relação dos conhecimentos com as informações fornecidas (textos/tabelas/gráficos/esquemas), o que se poderá explicar com o facto de “relacionar” corresponder a um nível de aprendizagem mais complexo e poder ter subjacente dificuldades de níveis de aprendizagem mais baixos, designadamente «Compreender o que é pedido», como acontece nas escolas B+C ou «As opções de resposta requerem conhecimento aprofundado da matéria», como é o caso da Escola D. O outro motivo

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

apresentado pelos alunos está ligado à complexidade da matéria e, portanto, é uma dificuldade inerente aos próprios conteúdos.

Relativamente ao relatório do GAVE (GAVE, 2012) sobre os exames realizados em 2011, verifica-se que os itens de escolha múltipla coincidem com os itens de melhor desempenho, apesar de um destes tipos de itens constar dos itens com pior desempenho. Como já foi referido, estes itens foram os mais representativos da prova de exame (24 itens), o que aumenta a probabilidade de corresponderem aos itens com melhor desempenho.

A Tabela 20 cruza as razões assinaladas pelos alunos da Escola A em função do grau de dificuldade que por eles foi atribuído aos itens de escolha múltipla.

Tabela 20

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de escolha múltipla – Escola A

Escola A						
Itens de escolha múltipla		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Compreender os textos fornecidos	N	1	8	3	0	12
	% Total	0,6	5,1	1,9	0,0	7,6
2 - Interpretar tabelas / gráficos / esquemas fornecidos	N	1	7	0	1	9
	% Total	0,6	4,4	0,0	0,6	5,7
3 - Compreender o que é pedido	N	1	6	7	1	15
	% Total	0,6	3,8	4,4	0,6	9,5
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)	N	2	18	10	0	30
	% Total	1,3	11,4	6,3	0,0	19,0
5 - Relacionar as diversas informações fornecidas (textos / tabelas / gráficos / esquemas).	N	2	9	5	0	16
	% Total	1,3	5,7	3,2	0,0	10,1
6 - Complexidade da matéria.	N	1	13	8	0	22
	% Total	0,6	8,2	5,1	0,0	13,9
7 - As opções de resposta são muito semelhantes entre si.	N	1	18	7	0	26
	% Total	0,6	11,4	4,4	0,0	16,5
8 - As opções de resposta requerem conhecimento aprofundado da matéria.	N	0	9	8	0	17
	% Total	0,0	5,7	5,1	0,0	10,8
9 - Selecionar a informação necessária para a resposta.	N	0	8	2	1	11
	% Total	0,0	5,1	1,3	0,6	7,0
Total	N	9	96	50	3	158
	% Total	5,7	60,8	31,6	1,9	100,0

Na Escola A verifica-se que o grau de dificuldade atribuído aos itens de escolha múltipla seguiu a tendência verificada na generalidade das escolas, estando distribuído pelos graus «Difícil» (31,6%) e «Pouco difícil» (60,8%), levando a concluir que mais de 50% dos alunos considerou este tipo de item pouco difícil.

A razão mais escolhida, nesta escola, para justificar os graus de dificuldade considerados foi «Relacionar os conhecimentos com as informações fornecidas (textos/tabelas/gráficos/esquemas)», com 18 escolhas para o grau «Pouco difícil» e 10 para o «Difícil», num total de 30 escolhas.

Em segundo lugar, os alunos consideraram como justificação para a sua dificuldade «As opções de resposta são muito semelhantes entre si», com 18 e 7 escolhas para os graus de dificuldade «Pouco difícil» e «Difícil», respetivamente, num total de 26 escolhas.

A terceira razão mais apontada foi «Complexidade da matéria», com 13 e 8 escolhas para os graus «Pouco difícil» e «Difícil», respetivamente.

No que diz respeito à razão menos apontada para as suas dificuldades «Interpretar tabelas/gráficos e esquemas», a mesma seguiu a tendência verificada na generalidade das escolas, com sete escolhas para «Pouco difícil», num total de nove.

Verifica-se, assim, que as razões apresentadas pelos alunos da Escola A são as mesmas que as apresentadas no geral das escolas, apesar das duas primeiras razões se encontrarem numa ordem diferente. Há a registar a coincidência na terceira razão apresentada e na razão menos assinalada pelos alunos.

A Tabela 21 que a seguir se apresenta cruza as razões assinaladas pelos alunos das Escolas B+C em função do grau de dificuldade que por eles foi atribuído aos itens de escolha múltipla.

**Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de
Biologia e Geologia.**

Tabela 21

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de escolha múltipla – Escolas B+C

Escolas B+C						
Itens de escolha múltipla		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Compreender os textos fornecidos	N	0	5	4	0	9
	% Total	0,0	6,7	5,3	0,0	12,0
2 - Interpretar tabelas / gráficos / esquemas fornecidos	N	0	0	2	0	2
	% Total	0,0	0,0	2,7	0,0	2,7
3 - Compreender o que é pedido	N	0	6	3	0	9
	% Total	0,0	8,0	4,0	0,0	12,0
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)	N	0	4	4	1	9
	% Total	0,0	5,3	5,3	1,3	12,0
5 - Relacionar as diversas informações fornecidas (textos / tabelas / gráficos / esquemas).	N	0	1	4	1	6
	% Total	0,0	1,3	5,3	1,3	8,0
6 - Complexidade da matéria.	N	0	0	6	0	6
	% Total	0,0	0,0	8,0	0,0	8,0
7 - As opções de resposta são muito semelhantes entre si.	N	0	7	13	0	20
	% Total	0,0	9,3	17,3	0,0	26,7
8 - As opções de resposta requerem conhecimento aprofundado da matéria.	N	0	0	7	1	8
	% Total	0,0	0,0	9,3	1,3	10,7
9 - Selecionar a informação necessária para a resposta.	N	0	4	2	0	6
	% Total	0,0	5,3	2,7	0,0	8,0
Total	N	0	27	45	3	75
	% Total	0,0	36,0	60,0	4,0	100,0

A análise da Tabela 21 permite constatar que nas escolas B+C, relativamente aos itens de escolha múltipla, os alunos não consideraram o grau de dificuldade «Nada Difícil», sendo as suas opiniões distribuídas pelos restantes graus. O grau de dificuldade do item está maioritariamente distribuído pelos graus «Pouco Difícil» (36%) e «Difícil» (60%), tal como na Escola A, e seguindo, portanto, a tendência da globalidade das escolas. Verifica-se, portanto, que mais de metade dos alunos considera este tipo de item difícil.

A principal razão apresentada por estas escolas para justificar a dificuldade do item foi «As opções de resposta são muito semelhantes», com 13 escolhas para «Difícil» e 7 para «Pouco difícil», num total de 20 escolhas.

Como segunda razão estas escolas atribuem a dificuldade dos itens a vários motivos: «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)», com 4 escolhas para Difícil e 4 para Pouco difícil, «Compreender o que é pedido», com 6 escolhas para «Pouco difícil» e 3 para «Difícil», e «Compreender os

**Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de
Biologia e Geologia.**

textos fornecidos», com 5 escolhas para «Pouco difícil» e 4 para «Difícil», em todos os casos num total de 9 escolhas.

Nestas escolas verifica-se, que a razão apontada em primeiro lugar segue a tendência geral da globalidade das escolas, o mesmo sucedendo com uma das razões que os alunos apontam em segundo lugar – «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)».

A razão menos escolhida para justificar o grau de dificuldade deste tipo de item – «Interpretar tabelas/gráficos/esquemas fornecidos» – coincide com as escolhas verificadas na Escola A e na globalidade das escolas.

A Tabela 22 cruza as razões assinaladas pelos alunos da Escola D em função do grau de dificuldade que por eles foi atribuído aos itens de escolha múltipla.

Tabela 22

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de escolha múltipla – Escola D

Escola D						
Itens de escolha múltipla		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Compreender os textos fornecidos	N	0	3	6	0	9
	% Total	0,0	3,6	7,1	0,0	10,7
2 - Interpretar tabelas / gráficos / esquemas fornecidos	N	0	4	4	1	9
	% Total	0,0	4,8	4,8	1,2	10,7
3 - Compreender o que é pedido	N	0	3	5	0	8
	% Total	0,0	3,6	6,0	0,0	9,5
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)	N	0	1	9	1	11
	% Total	0,0	1,2	10,7	1,2	13,1
5 - Relacionar as diversas informações fornecidas (textos / tabelas / gráficos / esquemas).	N	0	0	5	0	5
	% Total	0,0	0,0	6,0	0,0	6,0
6 - Complexidade da matéria.	N	0	2	5	2	9
	% Total	0,0	2,4	6,0	2,4	10,7
7 - As opções de resposta são muito semelhantes entre si.	N	0	4	14	1	19
	% Total	0,0	4,8	16,7	1,2	22,6
8 - As opções de resposta requerem conhecimento aprofundado da matéria.	N	0	2	7	0	9
	% Total	0,0	2,4	8,3	0,0	10,7
9 - Selecionar a informação necessária para a resposta.	N	0	2	2	1	5
	% Total	0,0	2,4	2,4	1,2	6,0
Total	N	0	21	57	6	84
	% Total	0,0	25,0	67,9	7,1	100,0

Pela análise da Tabela 22, constata-se que, à semelhança das escolas B+C, também na Escola D, os alunos não consideraram o grau de dificuldade «Nada Difícil», no item de escolha múltipla.

O grau de dificuldade do item está distribuído pelos graus «Pouco Difícil» (25%) e «Difícil» (67,9%), tal como nas escolas A e B+C, e seguindo igualmente a tendência da globalidade das escolas. Mais uma vez se verifica que mais de metade dos alunos considera este tipo de item difícil.

Na opinião destes alunos, a principal razão justificativa das suas dificuldades foi «As opções de resposta são muito semelhantes entre si». As escolhas recaíram no grau «Difícil» em número de 14 e 4 para «Pouco difícil», num total de 19 escolhas.

A segunda razão atribuída às suas dificuldades foi «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)», com 9 escolhas para «Difícil» num total de 11 escolhas.

Na terceira e última opção estes alunos encontraram como razão para as suas dificuldades «As opções de resposta requerem conhecimento aprofundado da matéria» com 7 escolhas para «Difícil», num total de 9, e «Complexidade da matéria», com 5 escolhas para «Difícil» num total de 9 e «Interpretar tabelas / gráficos / esquemas fornecidos» com 4 escolhas para «Difícil» e 4 para «Pouco difícil», num total de 9 escolhas e «Compreender os textos fornecidos», com 6 escolhas para «Difícil» num total de 9. Constata-se que as primeira e segunda razões apresentadas seguem a tendência dos resultados quer da globalidade das escolas quer das escolas B+C. Comparativamente à Escola A as razões apresentadas são as mesmas, embora numa ordem diferente.

Quanto à razão menos escolhida na Escola D, não é a mesma das escolas anteriores, recaindo a escolha sobre «Relacionar as diversas informações fornecidas (textos / tabelas / gráficos / esquemas)», com 5 escolhas para «Difícil», num total de 5 escolhas, e «Selecionar a informação necessária para a resposta», com 2 escolhas para «Difícil» e 2 para «Pouco difícil», num total de 5 escolhas.

5.4.2. – Resultado das variáveis para «Item de ordenação»

A questão relativa às razões que podem contribuir para o grau de dificuldade dos itens de ordenação tem 6 opções de resposta, das quais três teriam de ser selecionadas.

A Tabela 23 indica as frequências das razões que contribuem para o grau de dificuldade dos itens de ordenação.

Tabela 23

Razões que contribuem para o grau de dificuldade dos itens de ordenação

Ordenação	Respostas		% de casos
	N	%	
1 - Compreender os textos fornecidos.	51	16,3	48,1
2 - Interpretar tabelas / gráficos / esquemas fornecidos.	32	10,3	30,2
3 - Compreender o que é pedido.	34	10,9	32,1
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	67	21,5	63,2
5 - Complexidade da matéria.	61	19,6	57,5
6 - A ordenação dos elementos requer um conhecimento aprofundado da matéria.	67	21,5	63,2
Total	312	100,0	294,3

Relativamente aos itens de ordenação, as três principais razões apontadas pelos alunos, para o grau de dificuldade sentido, foram «A ordenação dos elementos requer um conhecimento aprofundado da matéria», com 63,2%, em igualdade com a razão «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)». Como terceira razão para as suas dificuldades, os alunos escolheram a opção «Complexidade da matéria», com 57,5%. A razão menos escolhida pelos alunos foi «Interpretar tabelas / gráficos / esquemas fornecidos», com 30,2%.

A Tabela 24 cruza as razões assinaladas pelos alunos em função do grau de dificuldade que por eles foi atribuído aos itens de ordenação.

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

Tabela 24

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de ordenação

Item de ordenação		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Compreender os textos fornecidos.	N	2	22	18	9	51
	% Total	0,6	7,1	5,8	2,9	16,3
2 - Interpretar tabelas / gráficos / esquemas fornecidos.	N	2	12	14	4	32
	% Total	0,6	3,8	4,5	1,3	10,3
3 - Compreender o que é pedido.	N	3	12	11	8	34
	% Total	1,0	3,8	3,5	2,6	10,9
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	N	1	22	27	17	67
	% Total	0,3	7,1	8,7	5,4	21,5
5 - Complexidade da matéria.	N	3	14	33	11	61
	% Total	1,0	4,5	10,6	3,5	19,6
6 - A ordenação dos elementos requer um conhecimento aprofundado da matéria.	N	1	17	33	16	67
	% Total	0,3	5,4	10,6	5,1	21,5
Total	N	12	99	136	65	312
	% Total	3,8	31,7	43,6	20,8	100,0

A análise dos resultados indica que, nos itens de ordenação, os alunos atribuíram, claramente, um grau de dificuldade partilhado entre «Pouco Difícil», «Difícil» e «Muito difícil», com 31,7%, 43,6% e 20,8%, respetivamente. Verifica-se que o grau «Nada difícil» foi considerado muito menos vezes que os restantes graus de dificuldade, com 3,8%.

As três principais justificações para estes graus de dificuldade foram «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)» e «A ordenação dos elementos requer um conhecimento aprofundado da matéria», ambos com 67 escolhas. Seguiu-se, na terceira opção, a «Complexidade da matéria», com 61 escolhas. Verifica-se que nas 3 razões apresentadas, o grau «Difícil» foi o que obteve maior número de escolhas. A razão menos apontada foi “Interpretar tabelas / gráficos / esquemas fornecidos”, com 32 escolhas. Como já foi referido anteriormente, no que diz respeito aos resultados da cotação média e da cotação nula dos itens de ordenação, estes situam-se numa posição intermédia em que mais de metade dos alunos obteve classificação média de 52,5% e uma cotação nula de 47,5%. Salienta-se o caso da Escola D, que apresenta percentagens médias abaixo dos 50% para este tipo de item.

Com base na experiência profissional da investigadora, esta tipologia de item apresenta algumas flutuações em termos do desempenho dos alunos. Os itens de ordenação ao

**Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de
Biologia e Geologia.**

solicitar ao aluno uma ordenação numa sequência lógica de determinados fenómenos, acontecimentos ou processos, podem exigir do aluno operações mentais com diferentes níveis de complexidade o que, por sua vez, conduz a melhor ou pior desempenho por parte dos alunos. Por outro lado são itens com critérios de classificação que só admitem a cotação total da resposta ou então zero pontos, tal como acontece com os critérios relativos aos itens de escolha múltipla.

De acordo com as respostas dos alunos, estes referiram que as razões das suas dificuldades se prendem com o facto de a ordenação dos elementos requerer um conhecimento aprofundado da matéria, bem como com a complexidade da mesma e com a relação dos conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas), esta última já referida em qualquer um dos tipos de itens anteriores.

De acordo com o relatório da GAVE (GAVE, 2012), relativo a 2011, os itens de ordenação (2 itens) não estão referidos nem nos itens que apresentaram melhor desempenho, nem nos itens que apresentaram pior desempenho.

A Tabela 25 cruza as razões assinaladas pelos alunos da Escola A em função do grau de dificuldade que por eles foi atribuído aos itens de ordenação.

Tabela 25

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de ordenação – Escola A

Escola A						
Item de ordenação		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Compreender os textos fornecidos.	N	2	14	7	3	26
	% Total	1,3	9,0	4,5	1,9	16,7
2 - Interpretar tabelas / gráficos / esquemas fornecidos.	N	2	10	5	1	18
	% Total	1,3	6,4	3,2	0,6	11,5
3 - Compreender o que é pedido.	N	2	9	6	3	20
	% Total	1,3	5,8	3,8	1,9	12,8
4 - Relacionar os conhecimentos com as informações fornecidas (textos/tabelas/gráficos/esquemas)	N	1	13	8	8	30
	% Total	0,6	8,3	5,1	5,1	19,2
5 - Complexidade da matéria.	N	2	11	16	4	33
	% Total	1,3	7,1	10,3	2,6	21,2
6 - A ordenação dos elementos requer um conhecimento aprofundado da matéria.	N	0	11	13	5	29
	% Total	0,0	7,1	8,3	3,2	18,6
Total	N	9	68	55	24	156
	% Total	5,8	43,6	35,3	15,4	100,0

**Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de
Biologia e Geologia.**

Da análise da Tabela 25, verifica-se que, na Escola A, a maior parte das escolhas recaem sobre os graus «Pouco difícil», com 43,6%, «Difícil», com 35,3% e «Muito difícil» com 15,4%, com maior incidência no «Pouco difícil» e no «Difícil».

As três razões mais escolhidas pelos alunos para justificar as suas dificuldades foram a «Complexidade da matéria», com 16 escolhas para o grau «Difícil», 11 para o «Pouco difícil» e 4 para «Muito difícil», num total de 33 escolhas. A segunda razão apresentada pelos alunos foi «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)», com 13 escolhas para «Pouco difícil» e 8 para «Difícil» e «Muito difícil», num total de 30 escolhas. A terceira razão foi «A ordenação dos elementos requer um conhecimento aprofundado da matéria», com 13 escolhas para «Difícil», 11 para «Pouco difícil» e 5 para «Muito difícil», num total de 29 escolhas.

A razão menos escolhida foi «Interpretar tabelas / gráficos / esquemas fornecidos», com 10 escolhas para «Pouco difícil» e 5 para «Difícil», num total de 18 escolhas.

Na Escola A, as razões apresentadas são as mesmas que na globalidade das escolas, embora a sua ordem seja inversa. A razão menos escolhida coincide com a da globalidade das escolas.

A Tabela 26 cruza as razões assinaladas pelos alunos das Escolas B+C em função do grau de dificuldade que por eles foi atribuído aos itens de ordenação.

Tabela 26

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de ordenação – Escolas B+C

Escolas B+C						
II b - Item de ordenação		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Compreender os textos fornecidos.	N	0	2	4	4	10
	% Total	0,0	2,7	5,4	5,4	13,5
2 - Interpretar tabelas / gráficos / esquemas fornecidos.	N	0	0	4	3	7
	% Total	0,0	0,0	5,4	4,1	9,5
3 - Compreender o que é pedido.	N	1	1	3	3	8
	% Total	1,4	1,4	4,1	4,1	10,8
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	N	0	2	9	4	15
	% Total	0,0	2,7	12,2	5,4	20,3
5 - Complexidade da matéria.	N	1	2	10	1	14
	% Total	1,4	2,7	13,5	1,4	18,9
6 - A ordenação dos elementos requer um conhecimento aprofundado da matéria.	N	1	2	12	5	20
	% Total	1,4	2,7	16,2	6,8	27,0
Total	N	3	9	42	20	74
	% Total	4,1	12,2	56,8	27,0	100,0

**Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de
Biologia e Geologia.**

Nas escolas B+C, a análise da Tabela 26 permite constatar que o grau de dificuldade mais apontado pelos alunos foi «Difícil», com 56,8%, seguido de «Muito difícil», com 27,0%. As principais razões apontadas para as dificuldades sentidas foram «A ordenação dos elementos requer um conhecimento aprofundado da matéria», com 12 escolhas para o grau «Difícil» e 5 para o «Muito difícil» num total de 20 escolhas. Como segunda opção os alunos consideraram a razão «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)», com 9 escolhas para «Difícil», seguido de 4 escolhas para «Muito difícil», num total de 15 escolhas. No que diz respeito à terceira razão foi considerada a «Complexidade da matéria», com 10 escolhas, para «Difícil», num total de 14. Perante estes resultados, pode constatar-se que os alunos das escolas B+C, para justificar as suas dificuldades, atribuem as mesmas razões que as escolas na globalidade. Relativamente à Escola A, as razões apresentadas são as mesmas, embora numa ordem diferente.

A opção menos escolhida foi «Interpretar tabelas/gráficos/esquemas fornecidos», com 4 escolhas para «Difícil» e 3 para «Muito difícil», num total de 7, constatando-se que, os alunos das escolas B+C partilham da mesma opinião da generalidade das escolas.

A Tabela 27 cruza as razões assinaladas pelos alunos da Escola D em função do grau de dificuldade que por eles foi atribuído aos itens de ordenação.

Tabela 27

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de ordenação – Escola D

Escola D						
Item de ordenação		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Compreender os textos fornecidos.	N	0	6	7	2	15
	% Total	0,0	7,3	8,5	2,4	18,3
2 - Interpretar tabelas / gráficos / esquemas fornecidos.	N	0	2	5	0	7
	% Total	0,0	2,4	6,1	0,0	8,5
3 - Compreender o que é pedido.	N	0	2	2	2	6
	% Total	0,0	2,4	2,4	2,4	7,3
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	N	0	7	10	5	22
	% Total	0,0	8,5	12,2	6,1	26,8
5 - Complexidade da matéria.	N	0	1	7	6	14
	% Total	0,0	1,2	8,5	7,3	17,1
6 - A ordenação dos elementos requer um conhecimento aprofundado da matéria.	N	0	4	8	6	18
	% Total	0,0	4,9	9,8	7,3	22,0
Total	N	0	22	39	21	82
	% Total	0,0	26,8	47,6	25,6	100,0

Da análise da Tabela 27, constata-se que, na Escola D, os graus de dificuldade considerados pelos alunos em relação aos itens de ordenação recaíram no «Difícil», com 47,6%, no «Pouco difícil», com 26,8% e no «Muito difícil», com 25,6%, estando portanto mais centrados no «Difícil». A opção «Nada difícil» não foi selecionada por qualquer aluno.

As três razões mais apresentadas para justificar as dificuldades sentidas foram, primeiro, «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)», com 10 escolhas, para «Difícil», 7 para «Pouco difícil» e 5 para «Muito difícil», num total de 22 escolhas. A razão imediatamente a seguir foi a «Ordenação dos elementos requer um conhecimento aprofundado da matéria», com 4 escolhas para «Pouco difícil», 8 para «Difícil» e 6 para «Muito difícil», num total de 18 escolhas.

A terceira opção escolhida por esta escola foi «Compreender os textos fornecidos», com 6 escolhas para «Pouco difícil», 7 para «Difícil» e 2 para «Muito difícil», num total de 15.

Das três razões apontadas pela Escola D, duas coincidem com as razões apontadas pelas escolas A e B+C, apesar de em ordens diferentes, o mesmo acontecendo quando comparados estes resultados com os obtidos na globalidade das escolas. No que se refere à opção menos escolhida é diferente da apresentada pelos alunos das outras escolas - «Compreender os textos fornecidos».

5.4.3. – Resultado das variáveis para «Item de resposta restrita»

A questão relativa às razões que podem contribuir para o grau de dificuldade dos itens de resposta restrita tem 11 opções de resposta, das quais três teriam de ser selecionadas.

A Tabela 28 indica a frequência das razões que contribuem para o grau de dificuldade dos itens de resposta restrita.

Tabela 28

Razões que contribuem para o grau de dificuldade dos itens de resposta restrita

Resposta Restrita	Respostas		% de casos
	N	%	
1 - Compreender os textos fornecidos.	16	5,0	15,1
2 - Interpretar tabelas / gráficos / esquemas fornecidos.	8	2,5	7,5
3 - Compreender o que é pedido.	40	12,5	37,7
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	27	8,5	25,5
5 - Relacionar as diversas informações fornecidas (textos / tabelas / gráficos / esquemas).	15	4,7	14,2
6 - Complexidade da matéria.	34	10,7	32,1
7 - Organizar mentalmente as ideias / os conhecimentos necessários para a resposta.	36	11,3	34,0
8 - Expor as ideias por escrito de forma organizada e estruturada.	43	13,5	40,6
9 - Necessidade de conhecimento aprofundado da matéria.	21	6,6	19,8
10 - Utilização de linguagem escrita adequada e rigorosa.	29	9,1	27,4
11 - Selecionar os conhecimentos essenciais para a resposta.	50	15,7	47,2
Total	319	100,0	300,9

Com base na análise da Tabela 28, observa-se que na globalidade das escolas, as três principais razões escolhidas para as dificuldades sentidas pelos alunos foram, em primeiro lugar, «Selecionar os conhecimentos essenciais para a resposta», com 47,2%, em segundo lugar, «Expor as ideias por escrito de forma organizada e estruturada», com 40,6%, e, por último, «Compreender o que é pedido», com 37,7%.

A razão menos apontada pelos alunos prende-se com «Interpretar tabelas / gráficos / esquemas fornecidos», com 7,5%.

A Tabela 29 cruza as razões assinaladas pelos alunos em função do grau de dificuldade que por eles foi atribuído aos itens de resposta restrita.

**Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de
Biologia e Geologia.**

Tabela 29

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de resposta restrita

Item de resposta restrita		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Compreender os textos fornecidos.	N	1	6	6	3	16
	% Total	0,3	1,9	1,9	0,9	5,0
2 - Interpretar tabelas / gráficos / esquemas fornecidos.	N	1	1	3	3	8
	% Total	0,3	0,3	0,9	0,9	2,5
3 - Compreender o que é pedido.	N	0	10	17	13	40
	% Total	0,0	3,1	5,3	4,1	12,5
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	N	1	4	12	10	27
	% Total	0,3	1,3	3,8	3,1	8,5
5 - Relacionar as diversas informações fornecidas (textos / tabelas / gráficos / esquemas).	N	1	4	5	5	15
	% Total	0,3	1,3	1,6	1,6	4,7
6 - Complexidade da matéria.	N	2	6	16	10	34
	% Total	0,6	1,9	5,0	3,1	10,7
7 - Organizar mentalmente as ideias / os conhecimentos necessários para a resposta.	N	1	13	14	8	36
	% Total	0,3	4,1	4,4	2,5	11,3
8 - Expor as ideias por escrito de forma organizada e estruturada.	N	0	8	20	15	43
	% Total	0,0	2,5	6,3	4,7	13,5
9 - Necessidade de conhecimento aprofundado da matéria.	N	0	4	12	5	21
	% Total	0,0	1,3	3,8	1,6	6,6
10 - Utilização de linguagem escrita adequada e rigorosa.	N	2	5	10	12	29
	% Total	0,6	1,6	3,1	3,8	9,1
11 - Selecionar os conhecimentos essenciais para a resposta.	N	3	5	21	21	50
	% Total	0,9	1,6	6,6	6,6	15,7
Total	N	12	66	136	105	319
	% Total	3,8	20,7	42,6	32,9	100,0

Com base na análise da Tabela 29, verifica-se que na globalidade das escolas os alunos consideraram o grau de dificuldade desta tipologia de item distribuído maioritariamente por «Difícil», com 42,6%, «Muito difícil», com 32,9%, e «Pouco difícil», com 20,7%.

Para justificar as suas dificuldades, as três opções mais escolhidas foram «Selecionar os conhecimentos essenciais para a resposta», escolhida 21 vezes para quem considerou o item «Difícil» e mais 21 vezes para os que o consideraram «Muito difícil», num total de 50 escolhas. Como segunda razão surge «Expor as ideias por escrito de forma organizada e estruturada», com 20 escolhas para «Difícil» e 15 para «Muito difícil», num total de 43 escolhas. Por último, surge «Compreender o que é pedido», com 17 escolhas para «Difícil», 13 para «Muito difícil» e 10 para «Pouco difícil», num total de 40.

A razão menos escolhida foi «Interpretar tabelas / gráficos / esquemas fornecidos», distribuída equitativamente entre o «Difícil» e o «Muito difícil», com 3 escolhas, num total de 8.

Da análise global dos dados, considerando os resultados obtidos pelos alunos na prova de exame, verifica-se que a tipologia de item que apresenta resultados mais baixos é a resposta restrita, com 25,4% de resultado médio, sendo que cerca de metade dos alunos não obteve cotação neste tipo de item (média de cotação nula de 50,3%). Salienta-se, que os resultados da Escola A e das escolas B+C coincidem com os resultados gerais, mas, na Escola D, todos os itens apresentaram resultados médios negativos, com exceção da tipologia de associação / correspondência.

Estes resultados não se apresentam surpreendentes, para a investigadora, uma vez que, a nível da escola, os alunos, na generalidade, apresentam, nesta tipologia de item, fragilidades que se repercutem no seu fraco desempenho. Essas fragilidades parecem prender-se com a tipologia do item, uma vez que esta requer, normalmente, a interpretação de suportes fornecidos e a utilização e a mobilização de dados, conceitos e teorias e a explicitação de contextos exigindo, desta forma, operações mentais mais ou menos complexas. Para além disso, é um tipo de item em que o aluno tem de apresentar uma resposta escrita adequada e estruturada e em que a liberdade de resposta do aluno tem limites inerentes à própria tipologia do item (resposta restrita), tal como referem Gronlund (1977) e Ribeiro e Ribeiro (1990).

Nesta investigação, ao se solicitar aos alunos as razões das suas dificuldades, pretendia conhecer-se a opinião que estes tinham relativamente às razões das mesmas. Constatou-se, que, na generalidade das escolas, os alunos apresentam diversas dificuldades, essencialmente ao nível da compreensão do que lhes é pedido, em selecionar os conhecimentos essenciais para a resposta e em expor as suas ideias por escrito de forma organizada e estruturada.

Ao recorrer-se ao relatório nacional da prova de exame de Biologia e Geologia, código 702, do ano de 2011 (GAVE, 2012), no sentido de ver se as dificuldades dos alunos destas escolas tendiam ou não para as prováveis dificuldades sentidas a nível nacional, constatou-se, que as justificações encontradas para as dificuldades dos alunos nesta tipologia de item parecem prender-se com a explicação de contextos em análise e com fragilidades de carácter transversal, nomeadamente o estabelecimento de relações entre conceitos, a interpretação de dados de suportes (texto) e a aplicação a novas situações.

Considerando estes pareceres, parece haver uma certa convergência entre os mesmos, na medida em que as dificuldades de compreensão apresentadas pelos alunos e as suas dificuldades em perceber os conhecimentos que têm de utilizar para a resposta, correspondendo estas a níveis de aprendizagem baixos das taxonomias cognitivas

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

(Tenbrink, 2010), poderão estar subjacentes às dificuldades encontradas pela investigadora e pelo GAVE, uma vez que o que se exigiu a nível desta tipologia de item, no exame de 2011, foram níveis de aprendizagem considerados mais elevados e mais ou menos complexos.

Relembrando Cheng (2006), nesta tipologia de item, os examinandos têm de, geralmente, percorrer várias etapas para construírem a sua resposta, o que torna os itens mais difíceis e, portanto, itens que avaliem capacidades complexas, como por exemplo a aplicação, são em geral mais difíceis do que os que avaliam capacidades mais simples, como a replicação. Para além disso, estes itens podem ainda passar por solicitar um estabelecimento de relações mais simples ou mais complexas tornando-se, portanto, mais ou menos difíceis.

A Tabela 30 cruza as razões assinaladas pelos alunos da Escola A em função do grau de dificuldade que por eles foi atribuído aos itens de resposta restrita.

Tabela 30

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de resposta restrita – Escola A

Escola A						
Item de resposta restrita		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Compreender os textos fornecidos.	N	0	4	6	0	10
	% Total	0,0	2,5	3,8	0,0	6,2
2 - Interpretar tabelas / gráficos / esquemas fornecidos.	N	1	1	1	1	4
	% Total	0,6	0,6	0,6	0,6	2,5
3 - Compreender o que é pedido.	N	0	7	11	4	22
	% Total	0,0	4,4	6,9	2,5	13,8
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	N	1	2	8	3	14
	% Total	0,6	1,2	5,0	1,9	8,8
5 - Relacionar as diversas informações fornecidas (textos / tabelas / gráficos / esquemas).	N	0	1	2	2	5
	% Total	0,0	0,6	1,2	1,2	3,1
6 - Complexidade da matéria.	N	2	4	11	5	22
	% Total	1,2	2,5	6,9	3,1	13,8
7 - Organizar mentalmente as ideias / os conhecimentos necessários para a resposta.	N	0	8	8	1	17
	% Total	0,0	5,0	5,0	0,6	10,6
8 - Expor as ideias por escrito de forma organizada e estruturada.	N	0	6	11	3	20
	% Total	0,0	3,8	6,9	1,9	12,5
9 - Necessidade de conhecimento aprofundado da matéria.	N	0	1	7	2	10
	% Total	0,0	0,6	4,4	1,2	6,2
10 - Utilização de linguagem escrita adequada e rigorosa.	N	2	2	5	3	12
	% Total	1,2	1,2	3,1	1,9	7,5
11 - Selecionar os conhecimentos essenciais para a resposta.	N	3	3	15	3	24
	% Total	1,9	1,9	9,4	1,9	15,0
Total	N	9	39	85	27	160
	% Total	5,6	24,4	53,1	16,9	100,0

Com base na análise da Tabela 30, constata-se que quanto ao nível de dificuldade deste tipo de item, os alunos distribuíram-se essencialmente por «Difícil», com 53,1%, o «Pouco difícil», com 24,4%, e o «Muito difícil» com 16,9%, permitindo concluir que mais de metade dos alunos considera este tipo de item difícil.

As três razões que os alunos apresentaram como sendo justificativas das suas dificuldades foram «Selecionar os conhecimentos essenciais para a resposta», com 15 escolhas correspondentes ao grau «Difícil», 3 ao grau «Pouco difícil» e 3 ao grau «Muito difícil», num total de 24 escolhas, seguida das razões «Compreender o que é pedido», com 11 escolhas para o grau «Difícil», 7 para «Pouco difícil» e 4 para «Muito difícil», num total de 22 escolhas e «Complexidade da matéria», com 11 escolhas para o «Difícil», cinco para «Muito difícil» e 4 para «Pouco difícil», num total de 22 escolhas.

A razão menos escolhida pelos alunos foi «Interpretar tabelas / gráficos / esquemas fornecidos», com uma escolha para cada um dos graus de dificuldade, num total de 4 escolhas.

Duas das razões apontadas pela Escola A vão ao encontro das apresentadas na globalidade das escolas, embora numa ordem diferente. Uma das razões referidas em segundo lugar - «Complexidade da matéria» - não é referenciada na globalidade das escolas. Relativamente à razão menos escolhida pelos alunos na Escola A, esta coincide com a assinalada na globalidade das escolas.

A Tabela 31 cruza as razões assinaladas pelos alunos das Escolas B+C em função do grau de dificuldade que por eles foi atribuído aos itens de resposta restrita.

**Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de
Biologia e Geologia.**

Tabela 31

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de resposta restrita – Escolas B+C

Escolas B+C						
Item de resposta restrita		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Compreender os textos fornecidos.	N	1	1	0	1	3
	% Total	1,3	1,3	0,0	1,3	4,0
2 - Interpretar tabelas / gráficos / esquemas fornecidos.	N	0	0	0	1	1
	% Total	0,0	0,0	0,0	1,3	1,3
3 - Compreender o que é pedido.	N	0	1	2	6	9
	% Total	0,0	1,3	2,7	8,0	12,0
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	N	0	0	2	5	7
	% Total	0,0	0,0	2,7	6,7	9,3
5 - Relacionar as diversas informações fornecidas (textos / tabelas / gráficos / esquemas).	N	1	2	1	2	6
	% Total	1,3	2,7	1,3	2,7	8,0
6 - Complexidade da matéria.	N	0	0	3	4	7
	% Total	0,0	0,0	4,0	5,3	9,3
7 - Organizar mentalmente as ideias / os conhecimentos necessários para a resposta.	N	1	4	2	3	10
	% Total	1,3	5,3	2,7	4,0	13,3
8 - Expor as ideias por escrito de forma organizada e estruturada.	N	0	1	3	4	8
	% Total	0,0	1,3	4,0	5,3	10,7
9 - Necessidade de conhecimento aprofundado da matéria.	N	0	1	4	2	7
	% Total	0,0	1,3	5,3	2,7	9,3
10 - Utilização de linguagem escrita adequada e rigorosa.	N	0	1	1	1	3
	% Total	0,0	1,3	1,3	1,3	4,0
11 - Selecionar os conhecimentos essenciais para a resposta.	N	0	1	3	10	14
	% Total	0,0	1,3	4,0	13,3	18,7
Total	N	3	12	21	39	75
	% Total	4,0	16,0	28,0	52,0	100,0

Nas escolas B+C, a análise da tabela permite constatar que 52,0%, isto é, mais de metade dos alunos consideraram este tipo de item «Muito difícil», 28% «Difícil» e 16 % «Pouco difícil».

As três razões mais consideradas pelos alunos para justificar as suas dificuldades foram «Selecionar os conhecimentos essenciais para a resposta», com 10 escolhas para os que consideram o item «Muito difícil,» num total de 14, e 3 e 1 escolhas para os que, respetivamente, consideraram «Difícil» e «Pouco difícil».

A segunda razão apresentada foi «Organizar mentalmente as ideias / os conhecimentos necessários para a resposta», com 4 escolhas para «Pouco difícil», 3 para «Muito difícil» e 2 para «Difícil», num total de 10 escolhas.

Em terceiro lugar os alunos selecionaram a razão «Compreender o que é pedido» com 6 escolhas para os alunos que consideraram o item «Muito difícil», 2 para «Difícil» e 1 para «Pouco difícil», num total de 9 escolhas.

**Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de
Biologia e Geologia.**

A razão menos escolhida é a que se prende com «Interpretar tabelas / gráficos / esquemas fornecidos», com uma única escolha e a corresponder ao grau «Muito difícil».

Nas escolas B+C, verifica-se, então, que duas das razões mais apontadas pelos alunos para justificar as suas dificuldades coincidem com as apresentadas na Escola A e com as apresentadas na globalidade das escolas. Há a registar a apresentação da razão «Organizar mentalmente as ideias / os conhecimentos necessários para a resposta», a qual não consta das 3 razões mais escolhidas na globalidade das escolas.

A opção menos escolhida pelas escolas B+C, «Interpretar tabelas / gráficos / esquemas fornecidos», com uma única escolha e a corresponder ao grau «Muito difícil», coincide com a escolha da Escola A e com a da globalidade das escolas.

A Tabela 32 cruza as razões assinaladas pelos alunos da Escola D em função do grau de dificuldade que por eles foi atribuído aos itens de resposta restrita.

Tabela 32

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de resposta restrita – Escola D

Escola D						
Item de resposta restrita		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Compreender os textos fornecidos.	N	0	1	0	2	3
	% Total	0,0	1,2	0,0	2,4	3,6
2 - Interpretar tabelas / gráficos / esquemas fornecidos.	N	0	0	2	1	3
	% Total	0,0	0,0	2,4	1,2	3,6
3 - Compreender o que é pedido.	N	0	2	4	3	9
	% Total	0,0	2,4	4,8	3,6	10,7
4 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	N	0	2	2	2	6
	% Total	0,0	2,4	2,4	2,4	7,1
5 - Relacionar as diversas informações fornecidas (textos / tabelas / gráficos / esquemas).	N	0	1	2	1	4
	% Total	0,0	1,2	2,4	1,2	4,8
6 - Complexidade da matéria.	N	0	2	2	1	5
	% Total	0,0	2,4	2,4	1,2	6,0
7 - Organizar mentalmente as ideias / os conhecimentos necessários para a resposta.	N	0	1	4	4	9
	% Total	0,0	1,2	4,8	4,8	10,7
8 - Expor as ideias por escrito de forma organizada e estruturada.	N	0	1	6	8	15
	% Total	0,0	1,2	7,1	9,5	17,9
9 - Necessidade de conhecimento aprofundado da matéria.	N	0	2	1	1	4
	% Total	0,0	2,4	1,2	1,2	4,8
10 - Utilização de linguagem escrita adequada e rigorosa.	N	0	2	4	8	14
	% Total	0,0	2,4	4,8	9,5	16,7
11 - Selecionar os conhecimentos essenciais para a resposta.	N	0	1	3	8	12
	% Total	0,0	1,2	3,6	9,5	14,3
Total	N	0	15	30	39	84
	% Total	0,0	17,9	35,7	46,4	100,0

No que concerne à Escola D, ressalta desde logo o facto de os alunos não terem considerado o grau de dificuldade «Nada difícil». A maioria dos alunos, 46,4%, considera esta tipologia de itens «Muito difícil», sendo que 35,7% a considera «Difícil» e 17,9% «Pouco difícil».

As 3 razões atribuídas pelos alunos para justificar as dificuldades sentidas foram «Expor as ideias por escrito de forma organizada», com 8 escolhas para o grau «Muito difícil» e 6 para o «Difícil», num total de 15 escolhas.

Em segunda opção foi seleccionada a razão «Utilização de linguagem escrita adequada e rigorosa», com 8 escolhas para o item «Muito difícil» e 4 para «Difícil», num total de 14 escolhas.

Por último, como terceira opção, foi escolhida a razão «Selecionar os conhecimentos essenciais para a resposta», com 8 escolhas para «Muito difícil» e três para «Difícil», num total de 12 escolhas.

Deste modo, a Escola D partilha de uma opinião com a Escola A e com as escolas B+C - «Selecionar os conhecimentos essenciais para a resposta» - e de duas com a globalidade das escolas, em qualquer dos casos em ordens diferentes.

Relativamente à razão menos escolhida, esta é coincidente com a da Escola A e com a das escolas B+C, com 2 escolhas para «Difícil», num total de 3 escolhas.

5.4.4. – Resultado das variáveis para «Item de associação»

A questão relativa às razões que podem contribuir para o grau de dificuldade dos itens de associação tem 6 opções de resposta, das quais três teriam de ser seleccionadas.

A Tabela 33 indica a frequência das razões que contribuem para o grau de dificuldade dos itens de associação.

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.

Tabela 33

Razões que contribuem para o grau de dificuldade dos itens de associação

Item de Associação	Respostas		% de casos
	N	%	
1 - Interpretar tabelas / gráficos / esquemas fornecidos.	45	14,3	42,5
2 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	62	19,7	58,5
3 - Complexidade da matéria.	52	16,6	49,1
4 - Necessidade de conhecimento aprofundado da matéria.	41	13,1	38,7
5 - Os elementos apresentados, em cada uma das colunas, são muito semelhantes.	63	20,1	59,4
6 - Os elementos apresentados requerem um conhecimento aprofundado da matéria.	51	16,2	48,1
Total	314	100,0	296,2

Relativamente aos itens de associação, as três razões mais escolhidas para justificar as dificuldades sentidas pelos alunos foram «Os elementos apresentados, em cada uma das colunas, são muito semelhantes», com 59,4%, seguido de «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)», com 58,5% e por último «Complexidade da matéria», com 49,1%. A razão menos escolhida pelos alunos foi «Interpretar tabelas / gráficos / esquemas fornecidos» com 42,5%.

A Tabela 34 cruza as razões assinaladas pelos alunos em função do grau de dificuldade que por eles foi atribuído ao item de associação.

Tabela 34

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de associação

Item de associação		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Interpretar tabelas / gráficos / esquemas fornecidos.	N	9	22	13	1	45
	% Total	2,9	7,0	4,1	0,3	14,3
2 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	N	10	30	19	3	62
	% Total	3,2	9,6	6,1	1,0	19,7
3 - Complexidade da matéria.	N	12	26	12	2	52
	% Total	3,8	8,3	3,8	0,6	16,6
4 - Necessidade de conhecimento aprofundado da matéria.	N	6	22	9	4	41
	% Total	1,9	7,0	2,9	1,3	13,1
5 - Os elementos apresentados, em cada uma das colunas, são muito semelhantes.	N	10	30	20	3	63
	% Total	3,2	9,6	6,4	1,0	20,1
6 - Os elementos apresentados requerem um conhecimento aprofundado da matéria.	N	10	24	14	3	51
	% Total	3,2	7,6	4,5	1,0	16,2
Total	N	57	154	87	16	314
	% Total	18,2	49,0	27,7	5,1	100,0

Com base na análise da Tabela 34, verifica-se que no que se refere aos itens de associação, a maior parte dos alunos, 49,0%, considera esta tipologia de item «Pouco difícil», 27,7% consideram-na «Difícil» e 18,2 % consideram este tipo de itens «Nada difícil».

As razões mais apontadas para justificar as dificuldades dos alunos foram «Os elementos apresentados, em cada uma das colunas, são muito semelhantes», com 30 escolhas para «Pouco difícil» e 20 para «Difícil», num total de 63 escolhas; como segunda escolha está «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)», com 30 escolhas para «Pouco difícil», e dezanove para «Difícil», num total de 62 escolhas. Por último, a razão escolhida foi a «Complexidade da matéria», com 26 escolhas para «Pouco difícil», e doze para «Difícil» e para «Nada difícil». A opção menos escolhida foi «Necessidade de conhecimento aprofundado da matéria», com 22 escolhas para «Pouco difícil» e 9 para «Difícil», num total de 41 escolhas.

No que se refere aos resultados médios mais elevados obtidos, 71,1%, e à cotação nula, 12,1%, estes recaíram na tipologia de itens de associação/correspondência, o que era exetável pela investigadora, uma vez que a nível de escola os alunos apresentam, normalmente, um bom desempenho neste tipo de item.

Relativamente à opinião dos alunos, as razões apresentadas para as suas dificuldades prendem-se com a semelhança dos elementos apresentados em cada uma das colunas, com a relação dos conhecimentos com as informações fornecidas e com a complexidade da matéria. Tecnicamente este tipo de item passa por apresentar elementos parecidos para "forçar o aluno a discriminar entre as diversas premissas e respostas" (Tenbrink, 2010, p.329). As dificuldades, no que se refere à relação dos conhecimentos com as informações fornecidas, como já foi referenciado para os itens de resposta restrita, poderão ter subjacentes dificuldades ao nível da compreensão que impeçam o desenvolvimento dos níveis mais complexos de aprendizagem.

Quando consultado o relatório nacional do GAVE (2012), os itens que revelaram melhor desempenho não foram os de associação / correspondência, mas os de escolha múltipla. Importa referir que este gabinete não faz a análise dos itens pela sua tipologia mas por item. Uma vez que a prova é essencialmente constituída por itens de escolha múltipla, 24 itens, e tendo estes critérios que garantem ou a sua cotação total ou zero pontos, têm fortes probabilidades de serem os itens que apresentam melhor desempenho. No caso dos itens de correspondência ou associação, para além da prova só ter um item desta

**Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de
Biologia e Geologia.**

tipologia, estão ainda sujeitos a critérios baseados em níveis de desempenho, podendo, portanto, haver lugar a classificações intermédias.

A Tabela 35 cruza as razões assinaladas pelos alunos da Escola A em função do grau de dificuldade que por eles foi atribuído aos itens de associação.

Tabela 35

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de associação – Escola A

Escola A						
Item de associação		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Interpretar tabelas / gráficos / esquemas fornecidos.	N	8	11	1	1	21
	% Total	5,1	7,0	0,6	0,6	13,4
2 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	N	9	20	1	2	32
	% Total	5,7	12,7	0,6	1,3	20,4
3 - Complexidade da matéria.	N	9	17	2	1	29
	% Total	5,7	10,8	1,3	0,6	18,5
4 - Necessidade de conhecimento aprofundado da matéria.	N	5	12	2	1	20
	% Total	3,2	7,6	1,3	0,6	12,7
5 - Os elementos apresentados, em cada uma das colunas, são muito semelhantes.	N	8	14	3	0	25
	% Total	5,1	8,9	1,9	0,0	15,9
6 - Os elementos apresentados requerem um conhecimento aprofundado da matéria.	N	9	17	3	1	30
	% Total	5,7	10,8	1,9	0,6	19,1
Total	N	48	91	12	6	157
	% Total	30,6	58,0	7,6	3,8	100,0

Na análise da Escola A, mais de metade dos alunos, 58,0%, consideram este item «Pouco difícil», seguindo-se 30,6% de alunos a considerarem esta tipologia de item «Nada difícil».

As três razões apontadas pelos alunos como sendo a justificação para as suas dificuldades são «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)», com 20 escolhas para «Pouco difícil» e 9 para «Nada difícil», num total de 32 escolhas. Como segunda justificação foi seleccionada «Os elementos apresentados requerem um conhecimento aprofundado da matéria», com 17 escolhas para «Pouco Difícil» e 9 para «Nada difícil» em 30 escolhas. Em terceiro lugar foi escolhida a razão «Complexidade da matéria», com 17 escolhas para «Pouco difícil», e 9 para «Nada difícil», num total de 29 escolhas. A razão menos escolhida foi «Necessidade de conhecimento aprofundado da matéria», com 12 escolhas para «Pouco difícil» e 5 para «Nada difícil», num total de 20 escolhas.

**Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de
Biologia e Geologia.**

Verifica-se, deste modo, que, na Escola A, todas as razões apresentadas pelos alunos coincidem com as razões apresentadas na globalidade das escolas, incluindo a razão menos escolhida.

A Tabela 36 cruza as razões assinaladas pelos alunos das escolas B+C em função do grau de dificuldade que por eles foi atribuído aos itens de associação.

Tabela 36

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de associação – Escola B+C

Escolas B+C						
Item de associação		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Interpretar tabelas / gráficos / esquemas fornecidos.	N	1	2	5	0	8
	% Total	1,3	2,7	6,7	0,0	10,7
2 - Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	N	1	3	10	0	14
	% Total	1,3	4,0	13,3	0,0	18,7
3 - Complexidade da matéria.	N	3	4	7	0	14
	% Total	4,0	5,3	9,3	0,0	18,7
4 - Necessidade de conhecimento aprofundado da matéria.	N	1	3	4	0	8
	% Total	1,3	4,0	5,3	0,0	10,7
5 - Os elementos apresentados, em cada uma das colunas, são muito semelhantes.	N	2	7	10	0	19
	% Total	2,7	9,3	13,3	0,0	25,3
6 - Os elementos apresentados requerem um conhecimento aprofundado da matéria.	N	1	2	9	0	12
	% Total	1,3	2,7	12,0	0,0	16,0
Total	N	9	21	45	0	75
	% Total	12,0	28,0	60,0	0,0	100,0

Nas escolas B+C, verifica-se que mais de metade dos alunos, 60,0%, consideraram os itens de associação «Difícil», e 28% consideraram-nos «Pouco difícil». Regista-se que nenhum aluno considerou este tipo de item «Muito difícil».

As três razões apontadas pelos alunos para justificar as suas dificuldades foram «Os elementos apresentados, em cada uma das colunas, são muito semelhantes», com 10 escolhas para «Difícil» e 7 para «Pouco difícil», num total de 19 escolhas; seguido de «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)», com 10 escolhas para «Difícil» e a «Complexidade da matéria» com 7 escolhas para «Difícil», 4 para «Pouco difícil» e 3 para «Nada difícil», ambas com um total de 14 escolhas.

**Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de
Biologia e Geologia.**

Nas escolas B+C, as razões escolhidas são todas coincidentes com as razões apresentadas pela globalidade das escolas e pela Escola A.

As razões menos escolhidas foram «Interpretar tabelas / gráficos / esquemas fornecidos» e «Necessidade de conhecimento aprofundado da matéria», cada uma com um total de 8 escolhas. A segunda razão mencionada coincide com a apresentada pela globalidade das escolas e a Escola A.

A Tabela 37 cruza as razões assinaladas pelos alunos da Escola D em função do grau de dificuldade que por eles foi atribuído aos itens de associação.

Tabela 37

Razões assinaladas pelos alunos em função do grau de dificuldade atribuído aos itens de associação – Escola D

Escola D						
Item de associação		Nada difícil	Pouco difícil	Difícil	Muito difícil	Total
1 - Interpretar tabelas / gráficos / esquemas fornecidos.	N	0	9	7	0	16
	% Total	0,0	11,0	8,5	0,0	19,5
2 - Relacionar os conhecimentos com as informações fornecidas (textos/tabelas/gráficos/esquemas).	N	0	7	8	1	16
	% Total	0,0	8,5	9,8	1,2	19,5
3 - Complexidade da matéria.	N	0	5	3	1	9
	% Total	0,0	6,1	3,7	1,2	11,0
4 - Necessidade de conhecimento aprofundado da matéria.	N	0	7	3	3	13
	% Total	0,0	8,5	3,7	3,7	15,9
5 - Os elementos apresentados, em cada uma das colunas, são muito semelhantes.	N	0	9	7	3	19
	% Total	0,0	11,0	8,5	3,7	23,2
6 - Os elementos apresentados requerem um conhecimento aprofundado da matéria.	N	0	5	2	2	9
	% Total	0,0	6,1	2,4	2,4	11,0
Total	N	0	42	30	10	82
	% Total	0,0	51,2	36,6	12,2	100,0

Pela análise da tabela, constata-se, que, na Escola D, 51,2% dos alunos consideram este tipo de item «Pouco difícil» e 36,6%, consideram-no «Difícil». Salienta-se que nenhum aluno considerou o grau de dificuldade «Nada difícil».

As três razões mais apontadas por estes alunos foram «Os elementos apresentados, em cada uma das colunas, são muito semelhantes», com 9 escolhas para «Pouco difícil» e 7 para «Difícil». Segue-se, com 16 escolhas, as razões, «Interpretar tabelas / gráficos / esquemas fornecidos» com 9 e 7 escolhas para «Pouco difícil» e «Difícil» e «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)», com 7 e 8 escolhas para os graus «Pouco difícil» e «Difícil», respetivamente.

Verifica-se, que, comparando as razões escolhidas pela Escola D, com as escolas A e B+C, constata-se que há duas opiniões coincidentes. Quando comparados os resultados da Escola D com os da globalidade das escolas, há duas opiniões em comum.

A razão menos escolhida pela Escola D foi a «Complexidade da matéria» com 5 escolhas para «Pouco difícil», e 3 para «Difícil», num total de 9 escolhas, não coincidindo com nenhuma das escolhas das restantes escolas.

CONSIDERAÇÕES FINAIS

Chegados ao termo deste trabalho é, agora, pertinente tecer algumas considerações sobre os condicionalismos e os contributos de que se revestiu este estudo.

Proceder-se-á, assim, a uma síntese reflexiva sobre as implicações dos dados recolhidos, subjacentes ao objetivo geral desta investigação: conhecer “*A influência da tipologia dos itens nos resultados das provas de Biologia e Geologia*”, procurando

- (I) saber quais os resultados obtidos nos anos de 2011, em diferentes tipos de itens;
- (II) conhecer as perceções que alunos do 12.º ano de escolaridade têm quanto ao grau de dificuldade de cada tipo de item e compará-las com os resultados obtidos nas provas de exame;
- (III) identificar as razões que contribuem para o grau de dificuldade das perguntas (itens) dos Exames Nacionais.

Esta investigação não teve a pretensão de generalizar de forma efetiva os resultados encontrados, mas contribuir para uma melhor compreensão do desempenho dos alunos submetidos ao estudo nos exames nacionais. Face ao exposto, apresenta-se a seguir, de forma resumida, os principais resultados alcançados com a investigação.

Objetivo (I) - saber quais os resultados obtidos no ano de 2011, em diferentes tipos de itens.

No sentido de dar cumprimento ao objetivo, foram solicitados, junto dos diretores das escolas, os resultados dos exames nacionais de Biologia e Geologia, código 702, 1ª fase, no ano de 2011. Foi gentilmente cedida por todos os diretores a informação que o Ministério da Educação envia a todas as escolas. Nesta constam os resultados médios por item da prova e as respetivas cotações máximas e nulas dos alunos que realizaram o exame nacional suprarreferido. Com base nestes dados encontraram-se, para cada

escola, os resultados médios por tipo de item e a percentagem de alunos com cotação nula, tendo-se estabelecido, posteriormente, uma comparação entre as quatro escolas.

Tal como era esperado, foi na escola melhor posicionada no *ranking* (Escola A) que surgiram resultados médios mais elevados para todos os tipos de itens. Considerando a globalidade das quatro escolas, os itens de correspondência / associação foram os que melhores resultados proporcionaram, 71,1%, com apenas 12,1% de cotação nula. Os itens de ordenação e de escolha múltipla apresentaram valores de resultados médios semelhantes, 52,5% e 53,7% respetivamente. Os itens de resposta restrita foram os que apresentaram valores mais baixos, provavelmente, devido ao facto de testarem níveis de aprendizagem mais elevados. Com base nestas constatações, estabeleceu-se uma base de análise para os resultados do segundo objetivo desta investigação, que vem a ser:

Objetivo (II) - conhecer as perceções que alunos do 12.º ano de escolaridade têm quanto ao grau de dificuldade de cada item e compará-las com os resultados obtidos nas provas de exame

Pelos resultados obtidos, a partir da aplicação do questionário, foi possível inferir que a globalidade dos alunos das quatro escolas perceciona os itens de resposta restrita como sendo os mais difíceis e os itens de associação / correspondência os mais fáceis. Os itens de escolha múltipla e de ordenação revestem-se de um grau de dificuldade intermédio, o que coincide com os resultados realmente obtidos na prova de Biologia e Geologia (objetivo anterior). Pelo exposto, poder-se-á concluir que os alunos estão conscientes dos diferentes níveis de dificuldade dos itens, sentindo como mais difíceis os itens em que realmente obtiveram piores resultados. Comparando por escola, foram os alunos das escolas B+C (2,88) e D (2,93) os que declararam sentir níveis de dificuldade mais elevados e, recorrendo a comparações múltiplas, com o teste de Kruskal-Wallis, foi concluído que as diferenças significativas se encontram entre a Escola A, melhor classificada, e as escolas B+C ($p\text{-value} = ,033$) e D ($p\text{-value} = ,001$), escolas pior classificadas. As diferenças encontradas entre B+C e D não foram significativas.

Objetivo (III) - identificar as razões que contribuem para o grau de dificuldade das perguntas (itens) dos Exames Nacionais de Biologia e Geologia

Relativamente aos itens de escolha múltipla, as principais razões apresentadas para o grau de dificuldade considerado pelos alunos foram «As opções de resposta são muito semelhantes entre si», «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)» e a «Complexidade da matéria».

Para os itens de ordenação, as razões indicadas foram «A ordenação dos elementos requer um conhecimento aprofundado da matéria», «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)» e «Complexidade da matéria».

As razões mais apontadas para os itens de resposta restrita foram «Selecionar os conhecimentos essenciais para a resposta», «Expôr as ideias por escrito de forma organizada e estruturada» e «Compreender o que é pedido».

Nos itens de associação foram mais referidas as opções «Os elementos apresentados, em cada uma das colunas, são muito semelhantes», «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)» e «Complexidade da matéria».

Verifica-se, que as razões que são apresentadas pelos alunos como justificação para os graus de dificuldade sentidos nos itens de escolha múltipla e associação / correspondência são as mesmas, o que pode estar relacionado com a tipologia dos itens. Em qualquer destas tipologias são apresentados elementos similares que fazem parte das características dos próprios itens, que naturalmente exigem um conhecimento aprofundado da matéria, podendo conduzir à utilização de níveis de aprendizagem mais elevados.

Constata-se, que as razões «Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas)» e «Complexidade da matéria» são as mais apontadas, aparecendo em itens de escolha múltipla, itens de ordenação e itens de associação. Isto parece estar relacionado com o facto de os alunos terem de utilizar níveis de aprendizagem com uma certa complexidade, como a mobilização de conhecimentos ou conceitos, aplicação de conhecimentos em novas situações, entre outros, nos quais parece manifestarem dificuldades.

No que diz respeito aos itens de resposta restrita, as razões apresentadas para as dificuldades sentidas são muito particulares, decorrendo da natureza da própria tipologia do item (já referenciadas na discussão dos resultados), uma vez que solicitam ao aluno uma resposta organizada e estruturada e avaliam níveis de aprendizagem, normalmente complexos.

No que concerne aos itens de ordenação, a justificação mais apontada para as suas dificuldades foi «a ordenação dos elementos requer um conhecimento aprofundado da matéria». São itens que solicitam ao aluno uma ordenação numa sequência lógica de determinados fenómenos, acontecimentos ou processos, exigindo, normalmente, do aluno operações mentais com diferentes níveis de complexidade.

Considerando o objetivo geral desta investigação - *A influência da tipologia dos itens nos resultados das provas de Biologia e Geologia* – e tendo em conta os principais resultados deste estudo que deram resposta aos três objetivos definidos, verifica-se, por um lado, que os itens de associação/correspondência foram percecionados pelos alunos como sendo os mais fáceis, correspondendo à tipologia de item que proporcionou os melhores resultados médios da prova. Por outro lado, os itens de resposta restrita percecionados pelos alunos como sendo os mais difíceis correspondem à tipologia de item em que os resultados médios foram mais baixos. No que diz respeito aos itens de escolha múltipla e de ordenação, apura-se que foram as tipologias que os alunos percecionaram como tendo um grau de dificuldade intermédio, coincidindo com as tipologias que, na prova de exame, apresentaram valores de resultados intermédios.

Como se pode constatar neste estudo, as tipologias de itens apresentam características diferentes que podem influenciar o desempenho dos alunos. Os itens de associação/correspondência são, geralmente, utilizados para testar níveis de aprendizagem baixos das taxonomias, podendo proporcionar um melhor desempenho. Os itens de resposta restrita são os mais apropriados para avaliar níveis de aprendizagem complexos, como a organização e síntese de ideias, análise crítica de suportes diversos e a capacidade de expressão escrita, podendo conduzir a desempenhos mais fracos dos alunos. Quanto às tipologias de escolha múltipla e ordenação podem solicitar operações mentais complexas. A versatilidade e quantidade dos itens de escolha múltipla na prova podem conduzir a diferentes níveis de desempenho.

Assim, e atendendo aos resultados obtidos na presente investigação, parece confirmar-se que a tipologia dos itens influencia os resultados da prova de Biologia e Geologia, uma vez que os resultados médios por item e a cotação nula estão diretamente relacionados com o tipo de item e o grau de dificuldade percebido pelos alunos.

Limitações do estudo

Neste ponto apresenta-se algumas limitações que se considerou existir no desenvolvimento do trabalho de investigação.

O estudo poderia incluir mais fontes de informação, tais como entrevistas aos alunos, no sentido de aprofundar as opiniões destes relativamente às dificuldades sentidas por si nos exames nacionais, e confrontá-los com as ilações retiradas da análise do questionário.

Por motivos profissionais, o tempo disponível para o desenvolvimento deste trabalho foi difícil de gerir, impedindo que o calendário inicialmente previsto fosse cumprido. Não obstante, pensa-se que o trabalho foi conseguido, tendo-se revelado esclarecedor, quanto ao objetivo principal da investigação. Este estudo poderá contribuir para uma reflexão sobre a problemática dos exames nacionais e servir de suporte a futuros trabalhos.

Sugestões para futuras investigações

Este estudo foi centrado em aspetos particulares dos exames nacionais, neste caso a influência da tipologia dos itens nos resultados das provas de exame. Seria pertinente determinar o impacto de outros fatores nos resultados dos exames e o impacto dos exames nas práticas pedagógicas dos professores. Seria importante considerar nestes estudos as opiniões de outros intervenientes no processo de exames, nomeadamente a opinião dos professores (sobre a dificuldade dos diferentes tipos de itens e de que forma lidam com estes aspetos nas suas práticas letivas) e dos profissionais envolvidos na elaboração dos itens.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abrantes, P. (2010). Políticas de avaliação e avaliação de políticas: o caso português no contexto ibero-americano. *Revista ibero-americana de educación, n.º 53*, abril 2010, pp. 25-42. Consultado em 10 de agosto de 2012, através de <http://www.rieoei.org/rie53a01.pdf>
- Adão, A. (1997). Os primeiros anos de ensino secundario [sic] liceal, em Portugal: realidades, necessidades [sic]. *Educació i Història: Revista d'Història de l'Educació*, 3, 29-44. Versão eletrónica. Consultado em 18 de agosto de 2012, através de <http://publicacions.iec.cat/repository/pdf/00000157%5C00000009.pdf>
- Afonso, A. (2009). Políticas avaliativas e *accountability* em educação — subsídios para um debate iberoamericano. *Sísifo. Revista de Ciências da Educação*, 09, pp 57-70. Consultado em 26 de julho de 2012, através de <http://sisifo.fpce.ul.pt/pdfs/Revista%209%20AAFONSO.pdf> <http://sisifo.fpce.ul.pt>
- Afonso, A. (2011). Questões polémicas no debate sobre políticas educativas contemporâneas: o caso da *accoutability* baseada em testes standardizados e rankings escolares. In M. Alves & J. De Ketele (Orgs.), *Do currículo à avaliação, da avaliação ao currículo* (pp. 84-10). Porto: Porto Editora.
- Afonso, N. (2005). *Investigação Naturalista em Educação – Um guia prático e crítico*. Porto: ASA Edições.
- Azevedo, M. (2009). *Teses, relatórios e trabalhos escolares: sugestões para estruturação da escrita*. Lisboa: Universidade Católica Portuguesa.
- Beato, C. (2011). *Os liceus e as ciências (1836-1860). Um estudo sobre o processo de criação das disciplinas de ciências físicas e naturais nos liceus portugueses*. Tese de doutoramento inédita. Lisboa: Universidade de Lisboa, Instituto de Educação.
- Bonboir, A. (1976). *Como avaliar os alunos?* Lisboa: Seara Nova.
- Castilho, S. (2011). *O ensino passado a limpo*. Porto: Porto Editora.

Cheng, L. (2006, Maio). On varying the difficulty of test items. A paper presented at the 32nd Annual Conference of the International Association for Educational Assessment, Singapore. Consultado em 5 junho de 2012 através de http://www.iaea.info/documents/paper_1162a1d9f3.pdf

Escudero Escorza., T. (2003). Desde los tests hasta la investigación evaluativa actual. Un siglo, el XX, de intenso desarrollo de la evaluación en educación. RELIEVE, v. 9, n. 1, p. 11-43. Consultado em 3 de agosto, através de http://www.uv.es/RELIEVE/v9n1/RELIEVEv9n1_1.htm

Estanqueiro, A. (2010). *Boas práticas na educação. O papel dos professores*. Queluz: Editorial Presença.

Fernandes, D. (2007a). A avaliação das aprendizagens no Sistema Educativo Português. *Educação e Pesquisa*, v. 33, n.º 3, 581-600. São Paulo. Consultado em 2 de agosto de 2012, através de <http://www.scielo.br/pdf/ep/v33n3/a13v33n3.pdf>

Fernandes, D. (2007b). A avaliação tem limites. *Jornal a página da educação*, n.º 170, 2007, p. 35. Consultado em 12 de agosto de 2012, através de <http://www.apagina.pt/?aba=7&cat=170&doc=12553&mid=2>

Fernandes, D. (2008a). Algumas reflexões acerca dos saberes dos alunos em Portugal. *Revista Educação & Sociedade*. Campinas, vol. 29, n.º 102, 275-292, jan/abr 2008. Consultado em 12 de agosto 2012, através de <http://www.scielo.br/pdf/es/v29n102/a1429102.pdf>

Fernandes, D. (2008b). *Avaliação das aprendizagens: desafios às teorias, práticas e políticas*. Lisboa: Texto Editora.

Fernandes, D. (2009). Avaliação das aprendizagens em Portugal - Investigação e teoria da actividade. *Revista de Ciências da Educação*, n.º 9, 87-100. Consultado em 2 de agosto de 2012, através de <http://sisifo.fpce.ul.pt/pdfs/Revista%209%20DFernandes%20PTG.pdf>

- Fernandes, D. (2011). Avaliar para melhorar as aprendizagens: análise e discussão de algumas questões essenciais. In I. Fialho & H. Salgueiro (Orgs.) *Turma Mais e sucesso escolar: contributos teóricos e práticos*, pp. 81-107. Évora: Centro de Investigação em Educação e Psicologia da Universidade de Évora.
- Fernandes, J. & Leiria, I (2002). *O Ranking Revela Uma Cartografia Interessantíssima do País*. Entrevista com David Justino. Jornal O Público (outubro). Consultado em 14 de agosto de 2012, através de http://pascal.iseg.utl.pt/~ncrato/Recortes/Justino1_Publico_20021006.htm
- Ferraz, M, Carvalho, A. Dantas, C. Cavaco, H. Barbosa, J. Tourais, L. & Neves, N. (1994). Avaliação criterial /Avaliação normativa. In *Pensar avaliação, melhorar a aprendizagem*. Lisboa: Ministério da Educação - IIE.
- Froemel, J. E. (2009). La Efectividad Y la Eficacia de las Mediciones Estandarizadas y de Las Evaluaciones. *Educación. Revista Iberoamericana de Evaluación Educativa*, Volume 2, Número 1, 10-28. Consultado em 15 de Agosto de 2012, através de <http://dialnet.unirioja.es/servlet/articulo?codigo=3038074>
- Gabinete de Avaliação Educacional [GAVE] (s.d.a). Supervisores da Classificação - Provas de Aferição e Provas Finais, consultado em 30 de agosto de 2012 através de <http://www.gave.min-edu.pt/np3/formacao>
- Gabinete de Avaliação Educacional [GAVE] (s.d.b). Supervisão da Classificação de Provas, consultado em 30 de agosto de 2012 através de <http://www.gave.min-edu.pt/np3/161.html>
- Gabinete de Avaliação Educacional [GAVE] (s.d.c). Documentos GAVE, consultado em 30 de agosto de 2012 através de <http://www.gave.min-edu.pt/np3/455.html>
- Gabinete de Avaliação Educacional [GAVE] (s.d.d). Terminologia adoptada na classificação de itens de instrumentos de avaliação externa http://www.gave.min-edu.pt/np3content/?newsId=393&fileName=Terminologia_Itens.pdf
- Gabinete de Avaliação Educacional [GAVE] (2005). Informação n.º 03.05, janeiro de 2005. Prova de Exame Nacional de Biologia e Geologia. Lisboa: Ministério da Educação.

Gabinete de Avaliação Educacional [GAVE] (2006). Informação n.º 124.06, dezembro 2006. Prova de Exame Nacional de Biologia e Geologia. Lisboa: Ministério da Educação.

Gabinete de Avaliação Educacional [GAVE] (2008a). Informação n.º 06.08, fevereiro 2008. Prova de Exame Nacional de Biologia e Geologia. Lisboa: Ministério da Educação.

Gabinete de Avaliação Educacional [GAVE] (2008b). Informação n.º 06.09, novembro 2008. Prova de Exame Nacional de Biologia e Geologia. Lisboa: Ministério da Educação.

Gabinete de Avaliação Educacional [GAVE] (2009). Informação n.º 04.10, dezembro 2009. Prova de Exame Nacional de Biologia e Geologia. Lisboa: Ministério da Educação.

Gabinete de Avaliação Educacional [GAVE] (2010a). Informação n.º 07.11, novembro 2010. Prova de Exame Nacional de Biologia e Geologia. Lisboa: Ministério da Educação.

Gabinete de Avaliação Educacional [GAVE] (2010b). Relatório: Um olhar sobre os resultados dos exames nacionais. Lisboa: Ministério da Educação.

Gabinete de Avaliação Educacional [GAVE] (2011). Exames nacionais: Relatório 2010. Lisboa: Ministério da Educação.

Gabinete de Avaliação Educacional [GAVE] (2012). Exames nacionais: Relatório 2011. Lisboa: Ministério da Educação.

Ghiglione, R. & Matalon, B. (1997). *O Inquérito: teoria e prática*. Oeiras: Celta Editora (trabalho original em francês publicado em 1985).

Goring, P. (1981). *Manual de medições e avaliação do rendimento escolar*. Coimbra: Livraria Almedina.

Grilo, M. (2003). Ranking / expresso do ensino secundário. *Opinião – Os rankings*. Jornal o Expresso, 27 de setembro. Consultado em 14 de agosto, através de http://pascal.iseg.utl.pt/~ncrato/Recortes/MarcalGrilo_Expresso_20030927.htm

- Gronlund, N. (1977). *Constructing achievement tests*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc..
- Hadji, C. (1994). *A Avaliação, regras do jogo. Das intenções aos instrumentos*. Porto: Porto Editora.
- Hargreaves, A., Earl, L. & Ryan, J. (2001). *Educação para a mudança - Reinventar a escola para os jovens adolescentes*. Porto: Porto Editora.
- Hill, M. & Hill, A. (2012). *Investigação por questionário*. Lisboa: Edições Sílabo.
- Iarossi, G. (2011). *O poder da concepção em inquéritos por questionário*. Lisboa: Fundação Calouste Gulbenkian.
- Inspeção-Geral da Educação e Ciência [IGEC], (2012). Provas de Aferição do Ensino Básico e Exames Nacionais dos Ensinos Básico e Secundário. Consultado em 10 de agosto de 2012 através de http://www.ige.min-edu.pt/content_01.asp?BtreeID=03/01&treeID=03/01/01&auxID=&newsID=315#content
- Justino, D. (2010). *Diffícil é educa-los*. Lisboa: Fundação Francisco Manuel dos Santos.
- Lansheere, G. (1979). *A avaliação contínua e exames*. Coimbra: Almedina Coimbra.
- Leal, L. (1997). Exames: uma via a prosseguir? *Educação e Matemática*, (43).Lisboa: Universidade de Lisboa. Consultado em 2 agosto de 2012, através de <http://www.educ.fc.ul.pt/docentes/msantos/exa.pdf>
- Maroco, J. (2011). *Análise estatística com SPSS statistics* (5.^a edição). Pêro Pinheiro: Reports Number, Análise e gestão de informação Lda.
- Matos, M., Lopes, C.T., Nunes, S. & Venâncio, I. (2006). Reflexões sobre os *rankings* do Secundário. *Associação Portuguesa de Investigação Operacional*, 26, 1-21. Consultado em 4 de agosto de 2012 através de <http://apdio.pt/documents/10180/15548/n1.pdf>

- Meirinhos, M. & Osório, A. (2010). O estudo de caso como estratégia de investigação em educação. *EDUSER: revista de educação*, Vol 2(2), 49-65. Consultado em 30 de agosto, através de <https://bibliotecadigital.ipb.pt/bitstream/10198/3961/1/O%20estudo%20de%20caso%20como%20estrat%C3%A9gia%20de%20investiga%C3%A7%C3%A3o%20em%20educa%C3%A7%C3%A3o.pdf>
- Mendes, A. N., Costa, J. A. & Ventura, A. (2003). Ranking de Escolas em Portugal: um estudo exploratório. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educacion*, vol.1, n.º 1, 1-13. Consultado em 4 de agosto de 2012 através de <http://www.ice.deusto.es/RINACE/reice/vol1n1/NCV.pdf>
- Miranda, M. J. (1982) A docimologia em perspectiva. *Revista Faculdade de Educação*, São Paulo; 8 (1); 39-69. Consultado em 10 de agosto de 2012, através de <http://educa.fcc.org.br/pdf/rfe/v8n1/v8n1a04.pdf>
- Moreira, J. (2009). *Questionários: Teoria e prática*. Coimbra: Edições Almedina.
- Nova, E. (1997). *Avaliação dos alunos*. Lisboa: Texto Editora.
- Nóvoa, A. (2005). *Evidentemente. Histórias da Educação*. Consultado em 18 de agosto de 2012, através de <http://repositorio.ul.pt/bitstream/10451/4810/1/9789724142142.pdf>
- Ó, J. R. (s.d.). *História da Educação II. Relatório da unidade curricular*. Consultado em 18 de agosto de 2012, através de <https://grupos.ufrgs.br/pipermail/edp-53-l/attachments/20091112/c3a1cb84/attachment-0001.doc>
- Ó, J. R. (2009). *Ensino liceal (1836-1975)*. Consultado em 18 de agosto de 2012, através de <http://repositorio.ul.pt/bitstream/10451/6296/1/Ensino%20Liceal%20.pdf>
- Pacheco, J. (1994). *A avaliação dos alunos na perspectiva da reforma*. Porto: Porto Editora.
- Pardal, L. & Correia, E. (1995). *Métodos e técnicas de investigação social*. Porto: Areal Editores.

- Perrenoud, P. (1992). Não mexam na minha avaliação! Para uma abordagem sistémica da mudança pedagógica. In A. Estrela & A. Nóvoa (Org.), *Avaliações em educação: novas perspectivas* (pp.155-173). Lisboa: Educa.
- Piéron, H. (1974). *Ciência e técnica dos exames*. Lisboa: Moraes Editores.
- Pinto, A. (2001). Factores relevantes na avaliação escolar por perguntas de escolha múltipla. *Psicologia, Educação e Cultura*, 5(1), 23-44.
- Pinto, J. & Santos, J. (2006). *Modelos de avaliações das aprendizagens*. Lisboa: Universidade Aberta.
- Pires, E. L. (1997). *Lei de Bases do Sistema Educativo. Apresentação e comentários*. Porto: Edições Asa.
- Ponte, J. P. (1994). *O estudo de caso na investigação em educação matemática. Quadrante*, 3(1), 3-18. Consultado em 30 de agosto, através de [http://www.educ.fc.ul.pt/docentes/jponte/docs-pt%5C94-Ponte\(Quadrante-Estudo%20caso\).pdf](http://www.educ.fc.ul.pt/docentes/jponte/docs-pt%5C94-Ponte(Quadrante-Estudo%20caso).pdf)
- Ramalho, G. (2002). O que nos podem dizer e dizem os resultados das provas realizadas a nível nacional. In CNE. *Qualidade e avaliação da educação*, pp. 47-59. Lisboa: Editorial do Ministério da Educação.
- Reis, J. L. F. V. (2009). *Nacionalismo e europeísmo. A incidência na questão escolar em Portugal na revista O Professor (1973-2004)*. Tese de mestrado inédita. Coimbra: Universidade de Coimbra, Faculdade de Letras. Consultado em 21 de agosto, através de 2012, de https://estudogeral.sib.uc.pt/bitstream/10316/13550/1/Tese_mestrado_Jo%C3%A3o%20Lu%C3%ADs%20Reis.pdf
- Ribeiro, L. (1990). *Avaliação da Aprendizagem*. Lisboa: Texto Editora.
- Ribeiro, A. & Ribeiro, L. (1990). *Planificação e avaliação do ensino-aprendizagem*. Lisboa: Universidade Aberta.
- Rodrigues, P. (1992). A Avaliação Curricular. In A. Estrela & A. Nóvoa, *Avaliações em Educação. Novas Perspectivas* (pp.15-72).Lisboa: Educa.

- Rodrigues, P. M. (2002). *Avaliação da formação pelos participantes em entrevista de investigação*. Porto: Fundação Calouste Gulbenkian - Fundação para a ciência e a tecnologia.
- Rosário, M. A. (2007). *Influência do Exame Nacional do 9.º ano de escolaridade nas práticas de ensino e de avaliação em matemática*. Tese de mestrado inédita. Universidade do Minho, Instituto de Educação e Psicologia.
- Santiago, R. Correia, M., Tavares, O. & Pimenta, C. (2004). *Um Olhar sobre os Rankings*. Coimbra: Fundação das universidades portuguesas.
- Santos, L. (Org), Pinto, J., Rio, F., Leite, F., Varandas, J., Moreirinha, O., Dias, P., Dias, S. & Bondoso, T. (2010). *Avaliar para Aprender- Relatos de experiências de sala de aula do pré-escolar ao ensino secundário*. Porto: Porto Editora.
- Seixas, A. M. (2005). Aprender a democracia: Jovens e protesto no ensino secundário em Portugal. *Revista Crítica de Ciências Sociais*, 72, 187-209.
- Silva, A. (2000). Os "Rankings" Destroem a Avaliação, *Jornal O Público* (12 de outubro). Consultado em 14 de agosto de 2012, através de http://pascal.iseg.utl.pt/~ncrato/Recortes/ASantosSilva_Publico_20021012.htm
- Sousa, A. B. (2005). *Investigação em Educação*. Lisboa: Livros Horizonte.
- Tenbrink, T. (2010) *Evaluacion Guia practica para profesores* (11.ª ed.) Madrid: Narcea, S.A. de Ediciones.
- Vallejo. P. (1979). *Manual de avaliação escolar*. Coimbra: Livraria Almedina.
- Vicente, P., Reis, E. & Ferrão, F. (2001). *Sondagens: a amostragem como fator decisivo de qualidade*. Lisboa: Edições Sílabo.

LEGISLAÇÃO

Decreto 36508, de 17 de setembro de 1947.

Decreto-Lei n.º 491/77, de 23 de novembro.

Decreto-Lei n.º 240/80, de 19 de julho.

Decreto-Lei n.º 354/88, de 12 de outubro.

Decreto-Lei n.º 33/90, de 24 de janeiro.

Decreto-Lei n.º 189/92, de 3 de setembro.

Decreto-Lei n.º 318/95, de 28 de novembro.

Decreto-Lei n.º 28-B/96, de 4 de abril.

Decreto-Lei n.º 75/97, de 3 de abril.

Decreto-Lei n.º 296-A/98, de 25 de setembro.

Decreto-Lei n.º 74/2004, de 26 de março.

Decreto-Lei n.º 24/2006, de 6 de fevereiro.

Decreto-Lei n.º 213/2006, de 27 de outubro.

Decreto-Lei n.º 125/2011, de 29 de dezembro.

Decreto Regulamentar Regional n.º 9/2008/M, de 30 de abril.

Decreto Regulamentar Regional n.º 25/2011/A, de 25 de novembro.

Despacho Normativo n.º 338/93, de 21 de outubro.

Despacho Normativo n.º 55/95, de 19 de setembro.

Despacho Normativo n.º 45/96, de 31 de outubro.

Despacho Normativo n.º 6/2012, de 10 de abril.

Lei n.º 46/86, de 14 de Outubro.

Portaria n.º 210/78, de 15 de abril.

Portaria n.º 505-A/99, de 15 de julho.

APÊNDICES



Apêndice A – Questionário (Versão final)



QUESTIONÁRIO de OPINIÃO

A aplicação deste questionário, inserido numa pesquisa a realizar no âmbito do curso de Mestrado em Ciências da Educação – especialização em Avaliação Educacional, tem como objetivo saber qual a **influência da tipologia dos itens nos resultados dos Exames Nacionais de Biologia e Geologia** (código 702).

Por conseguinte, a tua opinião é fundamental para que se possa saber qual a perceção dos alunos quanto ao grau de dificuldade das perguntas (itens) dos exames de Biologia e Geologia que realizaste no ano passado (2011).

Não há respostas certas ou erradas. Pretende-se que dês a tua opinião pessoal e sincera. Agradeço que respondas a todas as questões do questionário.

Este questionário é **anónimo** e de natureza **confidencial**. Não escrevas o teu nome em nenhum local. A informação recolhida destina-se exclusivamente ao presente estudo.

Obrigada pela tua colaboração.

1.ª Parte – Caracterização.

1	Idade	_____ anos
Assinala com X a opção de resposta		
2	Sexo	<input type="checkbox"/> Masculino <input type="checkbox"/> Feminino
3	No ano letivo passado (2011) em que Fase realizaste a prova de Biologia e Geologia (código702)?	<input type="checkbox"/> - 1.ª Fase <input type="checkbox"/> - 2.ª Fase

v.s.f.f.

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.



2.^a Parte - Percepção dos alunos quanto ao grau de dificuldade das perguntas (itens) dos Exames Nacionais de Biologia e Geologia (código 702) realizados no ano letivo 2010-2011.

As provas de exame integram uma grande diversidade de itens que podem ser classificados de diferentes maneiras, consoante o contexto, o enquadramento teórico e a finalidade visada pela classificação.

A partir do ano escolar de 2010-2011, a referência a *itens de resposta fechada* e a *itens de resposta aberta* é substituída por uma nomenclatura que prevê itens de seleção e itens de construção.

Nos itens de seleção, o aluno/examinando escolhe a resposta a partir de várias hipóteses dadas no item. Para responder, regista os elementos que identificam a sua seleção (a forma de registar as respostas é explicitada nas instruções dadas nos enunciados das provas). Constituem exemplo os itens de escolha múltipla, de associação/correspondência e de ordenação.

Nos itens de construção, o aluno/examinando produz a resposta. São exemplo os itens de resposta curta, restrita e extensa.

Baseado em: GAVE (2010). Terminologia adotada na classificação de itens de instrumentos de avaliação externa

Indica com **X** o grau de dificuldade para cada tipo de item, considerando a seguinte escala:

1 = Nada difícil; 2 = Pouco difícil; 3 = Difícil; 4 = Muito difícil

ITEM DE ESCOLHA MÚLTIPLA	1	2	3	4
Este tipo de item consiste em selecionar a opção que está correta de entre as quatro opções que são apresentadas.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

ITEM DE ORDENAÇÃO	1	2	3	4
Este tipo de item consiste em apresentar uma sequência de letras que identificam elementos (termos, expressões, fases, etapas,...) a ordenar.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

ITEM DE RESPOSTA RESTRITA (ABERTA)	1	2	3	4
Este tipo de item consiste em produzir uma resposta.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

ITEM DE ASSOCIAÇÃO / CORRESPONDÊNCIA	1	2	3	4
Este tipo de item consiste em associar um elemento (termos, expressões, fases, etapas,...) de uma coluna a um elemento de outra coluna.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.



3.^a Parte - Perceção dos alunos quanto às razões que contribuem para o grau de dificuldade das perguntas (itens) dos Exames Nacionais de Biologia e Geologia (código 702) realizados no ano letivo 2010-2011.

Nesta parte do questionário pretende-se encontrar razões que contribuam para o grau de dificuldade de cada um dos tipos de itens.

Assim, são apresentadas para cada tipo de item, várias razões que podem ser, ou não, comuns a vários tipos de item.

Assinala com **X** as **3 (três) principais** razões que, **na tua opinião**, consideras serem justificativas do grau de dificuldade **de cada tipo de item**.

ITEM DE ESCOLHA MÚLTIPLA

Este tipo de item consiste em selecionar a opção que está correta de entre as quatro opções que são apresentadas.

1	Compreender os textos fornecidos.	<input type="checkbox"/>
2	Interpretar tabelas / gráficos / esquemas fornecidos.	<input type="checkbox"/>
3	Compreender o que é pedido.	<input type="checkbox"/>
4	Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	<input type="checkbox"/>
5	Relacionar as diversas informações fornecidas (textos / tabelas / gráficos / esquemas).	<input type="checkbox"/>
6	Complexidade da matéria.	<input type="checkbox"/>
7	As opções de resposta são muito semelhantes entre si.	<input type="checkbox"/>
8	As opções de resposta requerem conhecimento aprofundado da matéria.	<input type="checkbox"/>
9	Selecionar a informação necessária para a resposta.	<input type="checkbox"/>

ITEM DE ORDENAÇÃO

Este tipo de item consiste em apresentar uma sequência de letras que identificam elementos (termos, expressões, fases, etapas,...) a ordenar.

1	Compreender os textos fornecidos.	<input type="checkbox"/>
2	Interpretar tabelas / gráficos / esquemas fornecidos.	<input type="checkbox"/>
3	Compreender o que é pedido.	<input type="checkbox"/>
4	Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	<input type="checkbox"/>
5	Complexidade da matéria.	<input type="checkbox"/>
6	A ordenação dos elementos requer um conhecimento aprofundado da matéria.	<input type="checkbox"/>

Exames Nacionais: a influência da tipologia dos itens nos resultados das provas de Biologia e Geologia.



Mestrado em Ciências da Educação – especialização em Avaliação Educacional

Assinala com **X** as **3 (três) principais** razões que, **na tua opinião**, consideras serem justificativas do grau de dificuldade **de cada tipo de item**.

ITEM DE RESPOSTA RESTRITA (ABERTA)

Este tipo de item consiste em produzir uma resposta.

1	Compreender os textos fornecidos.	<input type="checkbox"/>
2	Interpretar tabelas / gráficos / esquemas fornecidos.	<input type="checkbox"/>
3	Compreender o que é pedido.	<input type="checkbox"/>
4	Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	<input type="checkbox"/>
5	Relacionar as diversas informações fornecidas (textos / tabelas / gráficos / esquemas).	<input type="checkbox"/>
6	Complexidade da matéria.	<input type="checkbox"/>
7	Organizar mentalmente as ideias / os conhecimentos necessários para a resposta.	<input type="checkbox"/>
8	Expor as ideias por escrito de forma organizada e estruturada.	<input type="checkbox"/>
9	Necessidade de conhecimento aprofundado da matéria.	<input type="checkbox"/>
10	Utilização de linguagem escrita adequada e rigorosa.	<input type="checkbox"/>
11	Selecionar os conhecimentos essenciais para a resposta.	<input type="checkbox"/>

ITEM DE ASSOCIAÇÃO / CORRESPONDÊNCIA

Este tipo de item consiste em associar um elemento (termos, expressões, fases, etapas,...) de uma coluna a um elemento de outra coluna.

1	Interpretar tabelas / gráficos / esquemas fornecidos.	<input type="checkbox"/>
2	Relacionar os conhecimentos com as informações fornecidas (textos / tabelas / gráficos / esquemas).	<input type="checkbox"/>
3	Complexidade da matéria.	<input type="checkbox"/>
4	Necessidade de conhecimento aprofundado da matéria.	<input type="checkbox"/>
5	Os elementos apresentados, em cada uma das colunas, são muito semelhantes.	<input type="checkbox"/>
6	Os elementos apresentados requerem um conhecimento aprofundado da matéria.	<input type="checkbox"/>

O questionário terminou!

Verifica, por favor, se não te esqueceste de assinalar alguma resposta!

Muito obrigada pela tua colaboração.

Apêndice B – Pedido de autorização à DGICD para aplicação do questionário



Apêndice C – Autorização da DGIC para aplicação do questionário



Apêndice D – Autorização do Diretor da Escola A para aplicação do
questionário

*Autorização a
solicitar a
questionário*

Exmo(a). Sr.(a) Diretor(a)

Escola :

4/5/2012
Assunto: Pedido de autorização para aplicação de um *Questionário de opinião*.

Maria Florentina Alves Gomes Lopes de Andrade, mestranda da Universidade de Évora, encontra-se a desenvolver uma dissertação de mestrado em Ciências da Educação – Avaliação Educacional, sob orientação da Professora Doutora Isabel José Botas Bruno Fialho. O tema de dissertação abordará a temática dos Exames Nacionais de Biologia e Geologia (código 702), em concreto **A influência da tipologia dos itens nos resultados das provas de Biologia e Geologia**. Neste sentido procuraremos:

- (i) conhecer a opinião dos alunos quanto ao grau de dificuldade que têm para cada tipo item;
- (ii) perceber se existe uma relação entre os resultados dos tipos de itens com menor e maior desempenho e aqueles que os respondentes referem ter mais dificuldade.

Para o efeito, como método de recolha de dados, recorrer-se-á ao inquérito por questionário, por constituir um procedimento mais adequado e que permitirá à investigadora a obtenção da informação necessária relativa às perceções dos participantes no estudo. O questionário será aplicado a alunos do 12.º ano de escolaridade (e do 11.º ano que tenham ficado retidos) de quatro escolas do distrito de Évora onde foram obtidos os melhores e os piores resultados no exame de Biologia e Geologia no ano de 2011. Os métodos e as técnicas a utilizar são exclusivamente quantitativos.

O instrumento foi validado por um painel de três especialistas em avaliação em Educação e por seis alunos do 12.º ano (que realizaram exame de Biologia e Geologia em 2011), num estudo exploratório.

A aplicação é individual e de natureza confidencial, pelo que será preservado o anonimato dos respondentes. Pretende-se que sejam dadas opiniões sinceras e não exista qualquer intervenção por parte dos aplicadores do questionário. A informação recolhida destina-se exclusivamente ao presente estudo. O tempo necessário para a aplicação do questionário (apresentação e respetivo preenchimento), tendo por base o estudo exploratório realizado, não durará mais do que 20 minutos.

O questionário é composto por itens de seleção, tendo como objetivo auscultar a opinião dos alunos quanto ao grau de dificuldade que estes têm dos diferentes tipos de itens e procurar saber quais as razões que justificam o respetivo grau de dificuldade.

A aplicação e a recolha do questionário será acordada entre a investigadora e os(as) Diretores(as) das escolas. No âmbito da realização de estudos de investigação em meio escolar foi solicitada a autorização de aplicação do referido questionário à DGIDC.

Face ao exposto, venho por este meio solicitar a V.ª Ex.ª que se digne autorizar a aplicação do questionário aos alunos das turmas de 12.º ano (e do 11.º ano que tenham ficado retidos) que realizaram o exame de Biologia e Geologia (código 702), em 2011, de forma a garantir a amostra definida para a consecução do referido estudo.

A signatária disponibiliza-se a efetuar uma pequena explicação aos professores que irão aplicar o referido questionário, que agendará com V.ª Ex.ª, dentro das disponibilidades e conveniências de serviço.

Aguardo com a maior brevidade possível a V/ resposta,

Grata antecipadamente pela atenção dispensada.

Com os melhores cumprimentos,

M.ª Florentina A. G. Lopes Andrade

Apêndice E – Autorização do Diretor da Escola B para aplicação do questionário

Exmo(a). Sr.(a) Diretor(a) / Presidente da CAP

Assunto: Pedido de autorização para aplicação de um *Questionário de opinião*.

Maria Florentina Alves Gomes Lopes de Andrade, mestranda da Universidade de Évora, encontra-se a desenvolver uma dissertação de mestrado em Ciências da Educação – Avaliação Educacional, sob orientação da Professora Doutora Isabel José Botas Bruno Fialho. O tema de dissertação abordará a temática dos Exames Nacionais de Biologia e Geologia (código 702), em concreto *A influência da tipologia dos itens nos resultados das provas de Biologia e Geologia*. Neste sentido procuraremos:

- (i) conhecer a opinião dos alunos quanto ao grau de dificuldade que têm para cada tipo item;
- (ii) perceber se existe uma relação entre os resultados dos tipos de itens com menor e maior desempenho e aqueles que os respondentes referem ter mais dificuldade.

Para o efeito, como método de recolha de dados, recorrer-se-á ao inquérito por questionário, por constituir um procedimento mais adequado e que permitirá à investigadora a obtenção da informação necessária relativa às perceções dos participantes no estudo. O questionário será aplicado a alunos do 12.º ano de escolaridade (e do 11.º ano que tenham ficado retidos) de quatro escolas do distrito de Évora onde foram obtidos os melhores e os piores resultados no exame de Biologia e Geologia no ano de 2011. Os métodos e as técnicas a utilizar são exclusivamente quantitativos.

O instrumento foi validado por um painel de três especialistas em avaliação em Educação e por seis alunos do 12.º ano (que realizaram exame de Biologia e Geologia em 2011), num estudo exploratório.

A aplicação é individual e de natureza confidencial, pelo que será preservado o anonimato dos respondentes. Pretende-se que sejam dadas opiniões sinceras e não exista qualquer intervenção por parte dos aplicadores do questionário. A informação recolhida destina-se exclusivamente ao presente estudo. O tempo necessário para a aplicação do questionário (apresentação e respetivo preenchimento), tendo por base o estudo exploratório realizado, não durará mais do que 20 minutos.

O questionário é composto por itens de seleção, tendo como objetivo auscultar a opinião dos alunos quanto ao grau de dificuldade que estes têm dos diferentes tipos de itens e procurar saber quais as razões que justificam o respetivo grau de dificuldade.

A aplicação e a recolha do questionário será acordada entre a investigadora e os(as) Diretores(as) das escolas. No âmbito da realização de estudos de investigação em meio escolar foi solicitada a autorização de aplicação do referido questionário à DGIDC.

Face ao exposto, venho por este meio solicitar a V.ª Ex.ª que se digne autorizar a aplicação do questionário aos alunos das turmas de 12.º ano (e do 11.º ano que tenham ficado retidos) que realizaram o exame de Biologia e Geologia (código 702), em 2011, de forma a garantir a amostra definida para a consecução do referido estudo.

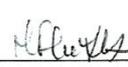
A signatária disponibiliza-se a efetuar uma pequena explicação aos professores que irão aplicar o referido questionário, que agendará com V.ª Ex.ª, dentro das disponibilidades e conveniências de serviço.

Aguardo com a maior brevidade possível a V/ resposta,

Grata antecipadamente pela atenção dispensada.

Com os melhores cumprimentos,


16 de Maio 12.


M.ª Florentina A. G. Lopes Andrade

Contacto: Telemóvel: 919 191 824 // <mailto:florentina.andrade@sapo.pt?subject=Aplicação Questionário>

Apêndice F – Autorização do Diretor da Escola C para aplicação do questionário



Exmo(a). Sr.(a) Diretor(a)

Assunto: Pedido de autorização para aplicação de um *Questionário de opinião*.

Maria Florentina Alves Gomes Lopes de Andrade, mestranda da Universidade de Évora, encontra-se a desenvolver uma dissertação de mestrado em Ciências da Educação – Avaliação Educacional, sob orientação da Professora Doutora Isabel José Botas Bruno Fialho. O tema de dissertação abordará a temática dos Exames Nacionais de Biologia e Geologia (código 702), em concreto *A influência da tipologia dos itens nos resultados das provas de Biologia e Geologia*. Neste sentido procuraremos:

- (i) conhecer a opinião dos alunos quanto ao grau de dificuldade que têm para cada tipo item;
- (ii) perceber se existe uma relação entre os resultados dos tipos de itens com menor e maior desempenho e aqueles que os respondentes referem ter mais dificuldade.

Para o efeito, como método de recolha de dados, recorrer-se-á ao inquérito por questionário, por constituir um procedimento mais adequado e que permitirá à investigadora a obtenção da informação necessária relativa às percepções dos participantes no estudo. O questionário será aplicado a alunos do 12.º ano de escolaridade (e do 11.º ano que tenham ficado retidos) de quatro escolas do distrito de Évora onde foram obtidos os melhores e os piores resultados no exame de Biologia e Geologia no ano de 2011. Os métodos e as técnicas a utilizar são exclusivamente quantitativos.

O instrumento foi validado por um painel de três especialistas em avaliação em Educação e por seis alunos do 12.º ano (que realizaram exame de Biologia e Geologia em 2011), num estudo exploratório.

A aplicação é individual e de natureza confidencial, pelo que será preservado o anonimato dos respondentes. Pretende-se que sejam dadas opiniões sinceras e não exista qualquer intervenção por parte dos aplicadores do questionário. A informação recolhida destina-se exclusivamente ao presente estudo. O tempo necessário para a aplicação do questionário (apresentação e respetivo preenchimento), tendo por base o estudo exploratório realizado, não durará mais do que 20 minutos.

O questionário é composto por itens de seleção, tendo como objetivo auscultar a opinião dos alunos quanto ao grau de dificuldade que estes têm dos diferentes tipos de itens e procurar saber quais as razões que justificam o respetivo grau de dificuldade.

A aplicação e a recolha do questionário será acordada entre a investigadora e os(as) Diretores(as) das escolas. No âmbito da realização de estudos de investigação em meio escolar foi solicitada a autorização de aplicação do referido questionário à DGIDC.

Face ao exposto, venho por este meio solicitar a V.ª Ex.ª que se digne autorizar a aplicação do questionário aos alunos das turmas de 12.º ano (e do 11.º ano que tenham ficado retidos) que realizaram o exame de Biologia e Geologia (código 702), em 2011, de forma a garantir a amostra definida para a consecução do referido estudo.

A signatária disponibiliza-se a efetuar uma pequena explicação aos professores que irão aplicar o referido questionário, que agendará com V.ª Ex.ª, dentro das disponibilidades e conveniências de serviço.

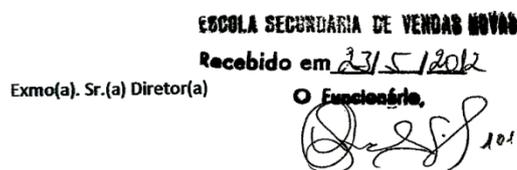
Aguardo com a maior brevidade possível a V/ resposta,

Grata antecipadamente pela atenção dispensada.

Com os melhores cumprimentos,

M.ª Florentina A. G. Lopes Andrade

Apêndice G – Autorização do Diretor da Escola D para aplicação do questionário



Assunto: Pedido de autorização para aplicação de um *Questionário de opinião*.

Maria Florentina Alves Gomes Lopes de Andrade, mestranda da Universidade de Évora, encontra-se a desenvolver uma dissertação de mestrado em Ciências da Educação – Avaliação Educacional, sob orientação da Professora Doutora Isabel José Botas Bruno Fialho. O tema de dissertação abordará a temática dos Exames Nacionais de Biologia e Geologia (código 702), em concreto *A influência da tipologia dos itens nos resultados das provas de Biologia e Geologia*. Neste sentido procuraremos:

- (i) conhecer a opinião dos alunos quanto ao grau de dificuldade que têm para cada tipo item;
- (ii) perceber se existe uma relação entre os resultados dos tipos de itens com menor e maior desempenho e aqueles que os respondentes referem ter mais dificuldade.

Para o efeito, como método de recolha de dados, recorrer-se-á ao inquérito por questionário, por constituir um procedimento mais adequado e que permitirá à investigadora a obtenção da informação necessária relativa às perceções dos participantes no estudo. O questionário será aplicado a alunos do 12.º ano de escolaridade (e do 11.º ano que tenham ficado retidos) de quatro escolas do distrito de Évora onde foram obtidos os melhores e os piores resultados no exame de Biologia e Geologia no ano de 2011. Os métodos e as técnicas a utilizar são exclusivamente quantitativos.

O instrumento foi validado por um painel de três especialistas em avaliação em Educação e por seis alunos do 12.º ano (que realizaram exame de Biologia e Geologia em 2011), num estudo exploratório.

A aplicação é individual e de natureza confidencial, pelo que será preservado o anonimato dos respondentes. Pretende-se que sejam dadas opiniões sinceras e não exista qualquer intervenção por parte dos aplicadores do questionário. A informação recolhida destina-se exclusivamente ao presente estudo. O tempo necessário para a aplicação do questionário (apresentação e respetivo preenchimento), tendo por base o estudo exploratório realizado, não durará mais do que 20 minutos.

O questionário é composto por itens de seleção, tendo como objetivo auscultar a opinião dos alunos quanto ao grau de dificuldade que estes têm dos diferentes tipos de itens e procurar saber quais as razões que justificam o respetivo grau de dificuldade.

A aplicação e a recolha do questionário será acordada entre a investigadora e os(as) Diretores(as) das escolas. No âmbito da realização de estudos de investigação em meio escolar foi solicitada a autorização de aplicação do referido questionário à DGIDC.

Face ao exposto, venho por este meio solicitar a V.ª Ex.ª que se digne autorizar a aplicação do questionário aos alunos das turmas de 12.º ano (e do 11.º ano que tenham ficado retidos) que realizaram o exame de Biologia e Geologia (código 702), em 2011, de forma a garantir a amostra definida para a consecução do referido estudo. A signatária disponibiliza-se a efetuar uma pequena explicação aos professores que irão aplicar o referido questionário, que agendará com V.ª Ex.ª, dentro das disponibilidades e conveniências de serviço.

Aguardo com a maior brevidade possível a V/ resposta,
Grata antecipadamente pela atenção dispensada.
Com os melhores cumprimentos,


M.ª Florentina A. G. Lopes Andrade