

Originalmente publicado em: (Outubro 2008) *Actas do 7.º Encontro Nacional (5.º Internacional) de Investigação em Leitura, Literatura Infantil e Ilustração*. Braga: Universidade do Minho.

Comparação dos níveis de desempenho nas Provas de Aferição

Rosária Maria Jorge dos Santos Rodrigues Correia*

Resumo

Este estudo descreve a forma como as competências de leitura são avaliadas pelo instrumento de avaliação externa de tipo criterial Provas de Aferição de Língua Portuguesa, para o 1.º Ciclo. Para o efeito, realizámos um estudo longitudinal, entre 2000 e 2008. A recolha e o tratamento de dados basearam-se na análise e na categorização dos textos das provas, dos diferentes tipos de itens – *objectivos* e *não objectivos* – e em critérios específicos de classificação. Os resultados mostram que não é possível generalizar/aféir e comparar os resultados obtidos nas diferentes provas, uma vez que não há uniformização entre eles, por terem critérios e níveis de dificuldade diferentes todos os anos. Podemos concluir que estas provas têm validade enquanto instrumento que nos permite avaliar em que posição se situa um aluno face ao resultado do grupo, mas são insuficientes para aferir o conhecimento dos alunos em termos longitudinais.

Abstract

This study describes the way in which reading abilities are evaluated by the criterion/referenced instrument of external evaluation – Tests of Educational Progress in Portuguese, for the 1st Cycle. For the effect, we carried out a longitudinal study between 2000 and 2008. The collection and analysis of data was based on an analysis and categorization of the texts in the tests, of the different types of item – objective and open ended – and on the specific criteria for classification. The results show that it is not possible to generalize/to assess and to compare the results obtained in the different tests, since they are not standardized: they have different criteria and different levels of difficulty every year. We can conclude that these tests have validity while instruments that allow us to evaluate how a student scores in relation to the whole group, but they are inefficient to assess the knowledge of students in longitudinal terms.

1. Introdução

O Decreto-lei n.º 6/2001 implementou no sistema educativo português várias alterações no que concerne à avaliação do desenvolvimento do currículo nacional e das provas nacionais de aferição. Definidas no despacho n.º 5437/2000, as Provas de Aferição (PA) foram testadas no primeiro conjunto de alunos do 4.º ano de escolaridade, no ano lectivo de 1998/1999, tendo-se generalizado ao universo dos alunos em conclusão do 1.º ciclo do Ensino Básico no ano lectivo seguinte. Esta tónica manteve-se até ao ano lectivo de 2003/2004, ano em que se voltou ao método de amostragem, tendo o mesmo vigorado até ao ano lectivo de 2005/2006, altura em que as Provas de Aferição voltaram a ser aplicadas à generalidade da população em conclusão de ciclo. Estas Provas pretendem «medir o grau de cumprimento dos objectivos essenciais, definidos a nível nacional, para

* Instituto Superior de Educação e Ciências (rosariacorreia@gmail.com).

cada ciclo do Ensino Básico, com o propósito de contribuir para a tomada de decisões no sentido de melhorar a qualidade das aprendizagens» (Despacho n.º 5437/2000, de 9 de Março: ponto 1), destacando o Ministério da Educação / Direcção Geral de Inovação e de Desenvolvimento Curricular (2004: 7) «o papel regulador e de instrumento de avaliação da eficácia do sistema» que a avaliação aferida deve ter.

A construção e a realização dos instrumentos de avaliação aferida são da responsabilidade do Gabinete de Avaliação Educacional (GAVE), sendo a selecção dos itens a incluir nas provas muito criteriosa, uma vez que tem de dar garantias de que representam as dimensões/processos/competências que estão a ser avaliadas. As provas são elaboradas para aferir «alguns aspectos do desempenho dos alunos, em determinadas competências» (ME / DGIDC, 2004: 8); como tal, todo o processo de validade tem de estar assegurado.

De acordo com Roldão (2003: 41), a avaliação é «um conjunto organizado de processos que visam: (1) o acompanhamento regulador de qualquer aprendizagem pretendida, e que incorporam, [...] (2) a verificação da sua consecução». Uma vez que os métodos de avaliação do sistema escolar pretendem verificar (Noizet e Caverni, 1985) se os objectivos sociais e pedagógicos foram alcançados, importa distinguir, ainda que sumariamente, dois tipos de avaliação: contínua e final. Se a primeira é, na grande maioria das vezes, interna, uma vez que quem avalia efectuou também a acção pedagógica, a segunda é externa, porque quem avalia é exterior à formação pedagógica do grupo. Estes dois tipos de avaliação distinguem-se também pelas suas finalidades. A avaliação contínua tem uma função mais pedagógica do que a avaliação final, ocorre ao longo do ano lectivo e através dela se pretende recolher dados que permitam compreender de que modo as aprendizagens estão a ser integradas pelos alunos, possibilitando dessa forma corrigir eventuais lacunas ou falhas.

Bloom, Hastings e Madaus (1971) apresentam-nos três tipos ou categorias de avaliação: a) diagnóstica, b) formativa e c) sumativa. Distinguem-se entre si pelos seus objectivos e efeitos no interior do sistema. Uma vez que as Provas de Avaliação se inscrevem no último tipo de avaliação referida, sumativa, poderemos dizer que pretendem «diferenciar» em vez de «homogeneizar» (Noizet e Caverni, 1985). Esta avaliação de produto tem por objectivo ver se o nível de desempenho dos alunos se encontra próximo do nível esperado, isto é, em que medida a aquisição das competências definidas para a conclusão de ciclo foram alcançadas.

No que concerne ao processo de correcção das provas, o mesmo é feito por «equipas específicas de correctores sob a responsabilidade de um supervisor» (ME / DGIDC, 2004: 8), sendo este último acompanhado pelo GAVE, que deve assegurar «as condições diferentes ou iguais de critérios necessários a uma correcta e verídica aferição dos resultados» (ob. cit.: 8), sendo cada «item ou conjunto de itens [...] codificado de acordo com vários níveis de resposta» (ob. cit.: 8). Os instrumentos de avaliação são, na opinião de Noizet e Caverni (1985: 18), «as linguagens utilizadas para codificar e para comunicar os juízos emitidos». Desde 2006/2007, os resultados destas provas, sendo expressos em categorias ordenadas (A, B, C, D, E), são divulgados através de uma escala ordinal (A > B > C > D), e não uma escala de intervalos. Porém, esta «relação de ordem não fornece nenhuma informação sobre a grandeza dos intervalos» (Noizet e Caverni, 1985: 19).

Em relação aos métodos empíricos de moderação das divergências, isto é, ao «conjunto de medidas tomadas para diminuir as divergências na classificação» (Noizet e Caverni, 1985: 47), tornando assim possível a comparação de resultados entre si, eles encontram-se assegurados nas Provas de Aferição, apesar de os métodos de moderação, quer *a priori*, quer *a posteriori*, a que os referidos instrumentos de avaliação são sujeitos não serem passíveis de divulgação.

Das várias técnicas de avaliação existentes, seleccionadas em função dos objectivos da avaliação, as Provas de Aferição inscrevem-se no conjunto dos *testes estandardizados*, uma vez que incluem procedimentos uniformes e consistentes, para a administração, avaliação e interpretação de resultados (Moore, 1983; Wiersma, 1995), o que implica a presença de um conjunto de procedimentos controlados.

Os testes estandardizados podem classificar-se com base num *referente*: a) *referenciados a normas* quando o desempenho de cada aluno no teste é comparado com os de todos os outros que já realizaram a mesma prova, e que constituem o *grupo norma* (ou *normativo*) que serve de comparação para os desempenhos considerados normais (Moore, 1983); b) *referenciados a critério* quando o desempenho do aluno é descrito tendo por base um *critério* previamente definido, a alcançar pelo aluno, sem que haja uma comparação interindividual (Black, 1999). As provas de aferição podem ser consideradas como instrumentos de *avaliação de referência criterial* (Valadares e Graça, 1999), pois cada aluno é avaliado por si, independentemente dos outros alunos.

Assim, e apesar da fiabilidade dos testes estandardizados referenciados a critério e de a sua correcção estar assegurada, será que cada uma das provas de Língua Portuguesa para o 4.º ano de escolaridade é equivalente à do ano anterior, logo, passível de comparação? (Creswell, 2003). Este estudo longitudinal pretende responder a esta questão, tendo por base a avaliação da competência «compreensão da leitura». Para tal, importa clarificar alguns conceitos.

2. A leitura e a sua compreensão

Como é sabido, os conceitos e definições em torno do acto de ler ou da leitura têm proliferado, sendo, no entender de Viana e Teixeira (2002: 11), «mais suposições hipotéticas do que a leitura deve ser, e não descrições neutras do que a leitura é».

Encarada como definição científica, a leitura é o «estabelecimento de uma correspondência entre um padrão visual, composto por uma sequência de letras, na pronúncia que lhe corresponde» (Leite *et al.*, 2006: 131), sendo a sua finalidade a compreensão da mensagem escrita (Morais, 1997) e, em nossa opinião, a interpretação. Todavia, tal como referem Sim-Sim e Micaelo (2006: 12) esta compreensão não é uma realidade dicotómica reduzida «a compreende» ou «não compreende», mas a um «produto variável» no qual intervém o leitor, a sua experiência, o conhecimento que possui sobre o assunto, os seus conhecimentos linguísticos, «a capacidade e rapidez da descodificação e da eficácia de mobilização de estratégias que activam a compreensão», numa interacção em que intervêm três tipos de variáveis, esquematizada na figura 1.

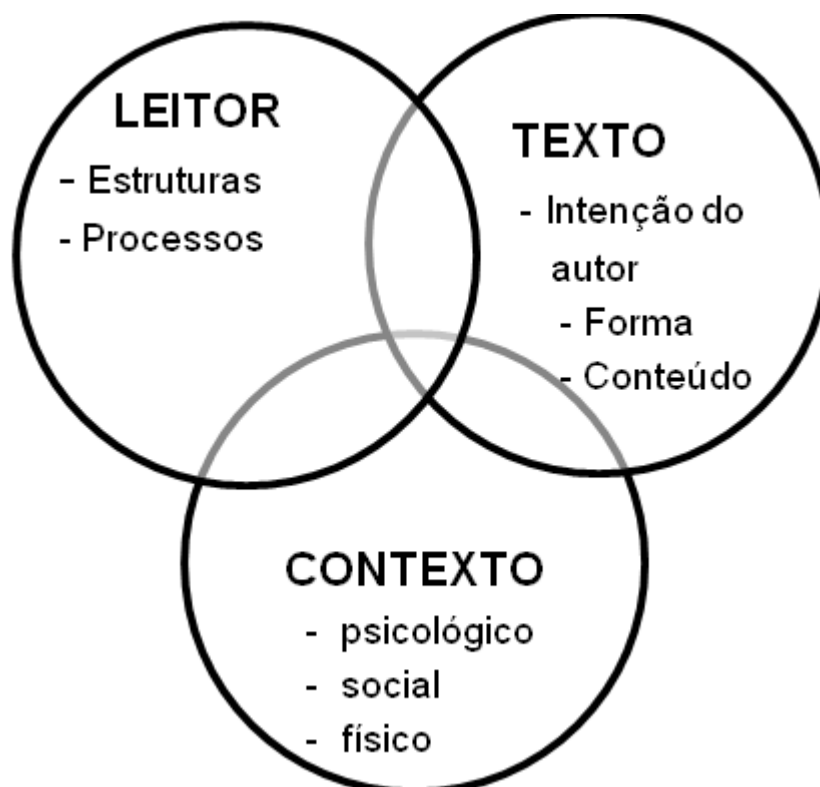


Figura 1 – Modelo contemporâneo da compreensão na leitura (Giasson, 2000: 21).

Tal como acontece quando se tenta definir o acto de ler, o conceito de compreensão da leitura origina diferentes enunciações e perspectivas. De acordo com Santos (2000: 33) «o objectivo de toda a leitura é, precisamente, compreender o que está escrito», situando-se Sim-Sim e Micaelo (2006: 40) nesta linha de convergência de pensamento, pois «a compreensão da leitura é entendida como uma construção activa de significado do texto, em que a informação de um estímulo se associa a informação prévia de que o leitor dispõe», sendo, assim, a compreensão leitora entendida como a assimilação do significado da mensagem; contudo, esta *aprensão* resulta do «nível de compreensão da interacção da criança com o texto» (Sim-Sim, 2007: 9).

Na perspectiva de Venâncio (2004: 53), «o leitor constrói o significado do texto por meio da interpretação e das inferências. Se o leitor, ao ler um texto, não tiver os conceitos subjacentes à sua compreensão, limita-se a decodificar». Esta visão identifica os processos através dos quais a criança é capaz de extrair informação do texto e de inferir os significados implícitos; contudo, admite também que a acção de inferir é bastante influenciada pelo texto e pelos conhecimentos prévios do leitor.

Alguns investigadores, baseando-se no tipo de perguntas do texto, desenvolveram esquemas de classificação que distinguem os diferentes processos de compreensão (Cunningham, 1987; Viana e Teixeira, 2002; Giasson, 2000; Raphael, 1986). Viana e Teixeira (2002), ao efectuarem uma síntese integradora das taxonomias de Barrett e de Smith, apresentam-nos como níveis de compreensão da leitura: a) *compreensão literal*; b) *compreensão interpretativa ou inferencial*; c) *avaliação ou julgamento*; d) *apreciação*; e e) *criação*. Enquanto que a *compreensão literal* se baseia na reprodução fiel do

significado explícito de um texto, a *compreensão interpretativa ou inferencial* ancora na capacidade de reconhecer o significado implícito de um texto. Por sua vez, a *avaliação ou julgamento* resulta da síntese e da integração dos dois níveis de compreensão anteriores, uma vez que a criança tem de questionar, analisar e realizar inferências, a fim de avaliar a veracidade da mensagem; apela, assim, ao sentido crítico do leitor, o qual deve distinguir a realidade da fantasia, o facto da opinião, bem como avaliar a credibilidade das fontes de informação, detectar e avaliar pontos de vista e técnicas de propaganda. No que concerne à *apreciação*, a criança tem de reagir às qualidades estéticas do texto, estando, pois, sempre envolvidos processos afectivos; mesmo que sejam usados critérios objectivos, estes estão apinhados de atitudes, valores e interesses subjectivos da criança. Por fim, a *criação* reside na capacidade do aluno em criar novas ideias ou interpretações, a partir daquilo que leu.

Araújo (2007) apresenta, de uma forma simplificada (figura 2), dois tipos de compreensão: a) *literal* – porque remete para informação que está explícita no texto e não requer um juízo crítico da parte do leitor; b) *inferencial* – porque se apoia na capacidade de dedução lógica do leitor, nos seus conhecimentos e na sua capacidade para integrar e para avaliar a informação, de forma a apreender o sentido global de um texto.

Informação Textual Explícita	Informação Textual Implícita
<p>Respostas baseadas no texto Localizar informação ou ideias que estão expressas de forma implícita no texto.</p> <p style="text-align: center;">↓</p> <p style="text-align: center;">COMPREENSÃO LITERAL</p>	<p>Respostas baseadas no texto e no leitor Fazer deduções lógicas; usar o conhecimento do mundo; considerar a perspectiva do autor e das personagens; avaliar usos de linguagem.</p> <p style="text-align: center;">↓</p> <p style="text-align: center;">COMPREENSÃO INFERENCIAL</p>

Figura 2 – Tipos de compreensão da leitura (Araújo, 2007).

3. As perguntas na avaliação da compreensão da leitura

De acordo com Valadares e Graça (1999: 49), a escolha dos itens a utilizar nos instrumentos de avaliação depende do tipo do referente: um *critério* ou uma *norma*. Enquanto que nos instrumentos de avaliação do tipo normativo se devem eliminar os itens *muito fáceis* e *muito difíceis*, «porque têm baixo poder discriminante», utilizando-se assim itens de *dificuldade média*, nos instrumentos de avaliação do tipo criterial «deve-se recorrer a itens que são directamente relevantes para medir os resultados de aprendizagem pretendidos, sem a preocupação de esses itens discriminarem muito ou pouco os alunos». Face ao exposto anteriormente, e apesar de não ser forçoso, seria interessante conhecer o *índice de dificuldade* (Almeida e Freire, 1997) ou, na acepção de Sim-Sim, o *índice de facilidade* (2006a) dos itens testados, isto é, a percentagem de sujeitos que responderam correctamente a cada um dos itens.

No que concerne às classificações de perguntas, Giasson (2000) apresenta algumas classificações, das quais destacamos a de Smith e Barrett, por ser baseada em processos

cognitivos, e não ter sido estabelecida uma hierarquia dos níveis de dificuldade das perguntas (Arnold e Wilcox, 1982, *in* Giasson, 2000). Assim, esta classificação compreende: a) *Reconhecimento literal ou reconstituição* – o aluno tem de reconhecer ideias, informações, situações ou acontecimentos expressos explicitamente no texto; b) *Inferência* – o aluno tem de articular a compreensão literal com os conhecimentos prévios e com a imaginação; c) *Avaliação* – o aluno tem de emitir um juízo sobre o texto; d) *Apreciação* – o aluno, apoiando-se nos conhecimentos que possui sobre as técnicas literárias utilizadas pelos autores para «suscitar reacções afectivas e intelectuais» (p. 288), tem de reagir à qualidade do texto. Porém, de acordo com a mesma autora, esta taxonomia é muitas vezes reduzida a três níveis: a) *compreensão literal*; b) *compreensão interpretativa*; e c) *compreensão crítica*.

Todavia, Giasson (2000) salienta o facto de esta classificação ter sido substituída pela taxonomia de Pearson e Johnson (1978, *in* Giasson, 2000), conhecida pela classificação *Relação-Questão-Resposta* (R-Q-R), que se centra na relação que ocorre entre as perguntas formuladas e as respostas dadas pelos alunos. Esta taxonomia compreende três níveis: a) a relação é *explícita e textual* (*textually explicit*), isto é, tanto a pergunta como a resposta giram em torno do texto, sendo a relação pergunta/resposta claramente indicada por indícios do próprio texto; b) a relação é *implícita e textual* (*textually implicit*) quando a pergunta/resposta decorre do texto, mas não há nenhum indício gramatical que as ligue, o que implica uma inferência por parte do aluno; c) a relação é *implícita e baseada nos esquemas do leitor* (*scriptally implicit*) quando apenas a pergunta surge do texto, tendo o aluno de utilizar os seus conhecimentos para responder.

A fim de avaliarem a *compreensão da leitura*, as Provas de Aferição têm utilizado diferentes tipos de questões (ver figura 3), que se podem agrupar em três categorias: *itens objectivos*, *itens não objectivos* e um terceiro grupo em que as questões se podem referir a *itens objectivos ou itens não objectivos*, podendo, por vezes, um item situar-se em mais de uma categoria.

3.1. Tipologia de itens

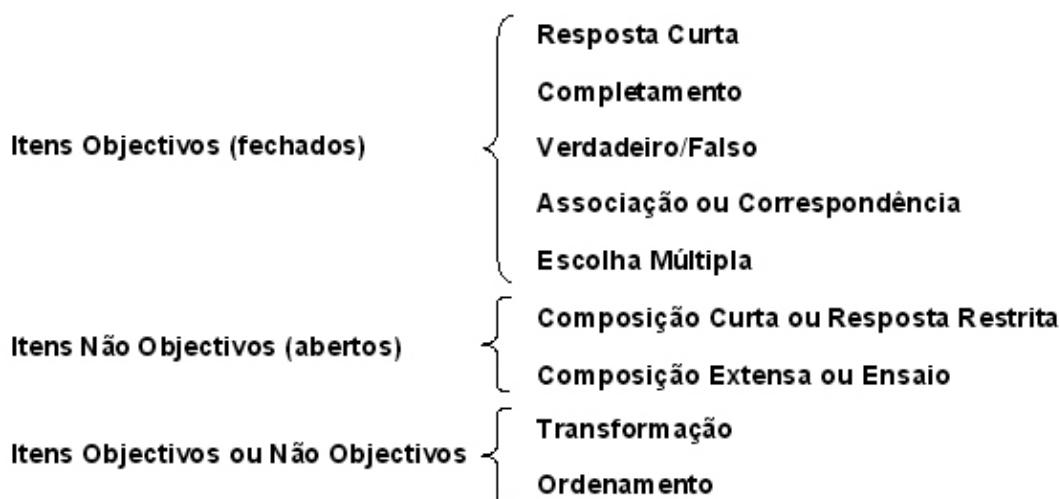


Figura 3 –Tipos de itens. *In* «Avaliação das Aprendizagens» – Formação GAVE.

Caracterizando cada uma das categorias supra referidas, Valadares e Graça (1999) descrevem uma variedade de tipos de itens.

3.1.1. Itens objectivos

Os *itens de resposta curta* têm por base uma pergunta simples, à qual o aluno deve responder de forma breve e sem ambiguidade. Estes itens, tal como os de *complemento*, permitem avaliar definições de conceitos e termos, factos específicos, datas, princípios, procedimentos, etc., enquanto os itens *Verdadeiro/Falso* possibilitam a avaliação de conhecimentos factuais. Os *itens de associação* são indicados para medir relações entre factos, acontecimentos, ideias, etc., sendo os *itens de escolha múltipla* os que nos permitem avaliar objectivos dos mais variados níveis de complexidade; contudo, este tipo de itens não permite avaliar todos os objectivos.

3.1.2. Itens não objectivos

Os *itens de composição* podem ser de dois tipos, dependendo da liberdade que os alunos têm para construir a resposta. Assim, nos *itens de composição curta* ou de *resposta restrita*, o conteúdo da resposta encontra-se bem demarcado, ou por limitações à forma da resposta, ou pela extensão do assunto da questão; enquanto que os *itens de composição extensa* ou de *ensaio* requerem um «modelo de resposta», uma vez que envolvem «altas capacidades cognitivas para serem bem respondidos» (Valadares e Graça, 1999: 87).

3.1.3. Itens objectivos ou não objectivos

Valadares e Graça (1999) não fazem alusão aos *itens de transformação* e de *ordenamento*. Assim, baseando-nos em formação tida na área, promovida pelo GAVE, podemos caracterizar os *itens de transformação* como a transformação do material apresentado, podendo haver condicionamentos à realização da tarefa. No que respeita aos *itens de ordenamento*, eles são considerados como um(uns) conjunto(s) de elemento(s) que devem ser organizados mediante uma ordem bem definida. Concordante com o tipo de itens (não objectivos), poderá haver escolha pessoal na tarefa.

Zabalza (1992) aponta ainda três critérios a considerar aquando da formulação dos itens: a) *representatividade* – as questões devem ser representativas do assunto/conteúdo que se pretende avaliar; b) *significação* – as questões devem verificar se o aluno domina o que é considerado importante acerca do assunto/conteúdo em exame; c) *diferenciação cognitiva* – este critério reporta-se às operações mentais que se pretendem activar; assim, colocam-se questões que activam a memória e o reconhecimento, a interpretação/compreensão, a aplicação, a análise, a síntese e a avaliação, acrescentando ainda Valadares e Graça (1999) a importância da *validade do conteúdo*.

4. Um olhar sobre as Provas de Aferição

Como referimos anteriormente, neste estudo longitudinal (2001/2008) procedeu-se à recolha e à análise de dados das Provas de Aferição. Esta recolha incidiu na análise e categorização dos textos das provas nos diferentes tipos de itens – *objectivos* e *não objectivos* –, bem como nos critérios específicos de classificação.

Ao efectuar-se a leitura comparativa dos enunciados das provas aplicadas ao 4.º ano de escolaridade, na área da Língua Portuguesa, constatámos que a estrutura é comum a todas as provas, mantendo-se inalterada desde 2000. Desta forma, pudemos observar que todas as provas se encontram organizadas em duas partes: na primeira pretende-se avaliar as competências essenciais de *compreensão da leitura e expressão escrita* – ou compreensão da leitura e de *conhecimento explícito da língua* – ou funcionamento da língua; na segunda parte, a intenção é aferir as competências relativas à *expressão escrita compositiva* ou *produção escrita*. No entanto, destacamos a alteração introduzida na nomeação das três competências essenciais aferidas, a partir de 2003, altura em que a competência *compreensão da leitura* passou a designar-se *compreensão da leitura e expressão escrita*, uma vez que é uma competência que exige uma resposta mediante a expressão escrita, tendo passado a ser avaliada não só a interpretação do texto, mas também a forma como os alunos estruturam e formulam a sua resposta.

Neste estudo, centrámos a nossa atenção apenas na primeira parte da prova, isto é, pretendemos apenas saber de que forma as competências essenciais de *leitura* foram testadas.

Importa ainda referir que, na aferição da competência em estudo, foram contempladas a compreensão literal e a compreensão inferencial, havendo, ou não, a presença de outros processos de operacionalização relativos à ordenação sequencial ou cronológica de factos, à identificação de características e localização da narrativa... Ainda de referir a alteração introduzida quanto às designações de alguns dos processos de operacionalização desta competência: até 2003, a compreensão literal era apresentada como a *compreensão e realização de verbatim*, tendo, a partir dessa data, passado a utilizar-se o *reconhecimento e identificação da informação solicitada com as palavras do texto*.

A *compreensão da leitura* foi testada com recurso a um excerto de um texto literário (leitura recreativa), tendo, a partir de 2002, sido introduzido o texto informativo (leitura funcional), visando a identificação de dados e informações para solucionar um problema específico, diversificando, desta forma, as competências e as estratégias mobilizadas pelos alunos na resolução de diferentes tipos de tarefas.

Porém, o tipo de textos utilizado ao longo dos anos, assim como a sua extensão, não tem sido constante. Ao analisarmos a tabela 1 constatamos, à semelhança do que já havia sido referido por Araújo (2007), mas para um período de tempo diferente, que a selecção de textos não obedece aos mesmos critérios todos os anos. Logo, se não há uniformização entre os textos propostos, a comparação entre anos não é possível.

Ano	Leitura Recreativa	Leitura Funcional
2000	Narrativa	
2001	Narrativa	
2002	Narrativa	Informativo
2003	Narrativa	Informativo
2004	Poesia	Informativo
2005	Narrativa	Informativo
2006	Poesia	Instrucional
2007	Dramático	Informativo
2008	Informativo	Poesia

Tabela 1 – Tipo de textos utilizados entre 2000 e 2008.

Contudo, e apesar de os textos apresentados serem, na sua maioria, narrativos, eles não são equivalentes entre si. Uma análise atenta à tabela 2 permite-nos constatar várias divergências.

Ano	Contagens			Média		
	Palavras	Frases / Versos	Parágrafos	Frases por parágrafo	Palavras por frase / verso	Sílabas por palavra
2000	260	35	26	1,5	4,9	4,2
2001	406	40	23	1,6	6,6	4,1
2002	501	47	18	1,8	5,9	4,5
2003	608	49	15	1,6	7,2	4,8
2004	245	50		1,3	3,4	4,5
2005	505	45	13	1,4	7,8	4,2
2006	156	36		1,1	3,5	4,6
2007	480	53		1,5	6,1	4
2008	264	22	7	1,5	8,3	4,9

Tabela 2 – Contagens e médias (leitura recreativa) entre 2000 e 2008.

4.1. Tipos de itens

Os tipos de itens são, na sua maioria, constantes, ao longo dos anos. As provas aferidas recorrem, essencialmente, a itens *objectivos*, logo fechados, predominando as questões de escolha múltipla, resposta curta e verdadeiro/falso. Os itens onde se pede ao aluno para justificar uma resposta ou expressar uma opinião só são testados a partir de 2003.

Baseando-nos na categorização adoptada por Sim-Sim e Ramalho (1993) para a divulgação dos dados do estudo *Reading Literacy* realizado em 1990, adoptámos a mesma categorização seguida pelas autoras para a avaliação das diversas competências no processo de compreensão da leitura, excepção feita ao *reconhecimento global de palavras*, uma vez que a capacidade de identificar o significado de vocábulos fora do contexto frásico não foi aferida, até ao momento, pelo instrumento de avaliação em análise.

Analisando as tabelas 3 e 4, verificamos que as competências testadas oscilam todos os anos, não só no número de itens em que são avaliadas, como também no seu formato, tornando-se assim impossível obter resultados fiáveis, que permitam caracterizar a população em estudo no que concerne à compreensão da leitura.

4.1.1. Itens objectivos – fechados

Anos	Resposta Curta	Completamento	Verdadeiro/ Falso	Associação ou Correspondência	Escolha Múltipla
2000	6		1		4
2001	4	1	1		6
2002	2	1	1	1	4
2003	3	1	1		4
2004	4	2	1		6
2005	1	1	1		8
2006	4	2	1		7
2007	3	1	1		4
2008	3		1	2	5

Tabela 3 – Distribuição dos itens fechados objectivos.

Ano	Composição Curta / Resposta Restrita	Composição Extensa / Ensaio	Transformação	Ordenamento
2000				1
2001				1
2002			1	1
2003	1			1
2004	1			
2005	1			1
2006				
2007	1			
2008	1			

Tabela 4 – Distribuição dos itens abertos não objectivos e itens objectivos/não objectivos.

De acordo com Creswell (2003), a comparação entre provas deve permitir avaliar um determinado desempenho de forma constante e estável ao longo do tempo, devendo esse desempenho ser aferido da mesma forma em provas equivalentes.

4.2. Tipos de compreensão testados

As classificações de níveis de leitura ou de compreensão da leitura apresentados por Sim-Sim e Micaelo (2006) referem a taxonomia proposta por Herber (1978), integradora de três níveis: a) *compreensão literal* – compreensão da informação explícita; b) *compreensão interpretativa* – item relacionado com questões relativas à elaboração de paráfrases, explicações, inferências, conclusões e sínteses; e c) *compreensão aplicada* – articulação da informação expressa no texto com os conhecimentos prévios do leitor, o que se traduz na realização de predições e juízos críticos e na resolução de problemas. Araújo (2007) considera dois tipos de itens: *perguntas de compreensão literal* (as que remetem o aluno para a localização da acção no espaço e no tempo, para a nomeação de personagens e para a identificação da relação causa/efeito) e *perguntas de compreensão inferencial* (que se baseiam na interpretação das acções, intenções e opiniões das personagens e do autor). Apoiados em ambos os estudos, analisámos os níveis de compreensão das provas.

Observando a tabela 5, constatamos que predomina a *compreensão literal*, sendo *verbatim* o tipo de compreensão mais testada ao longo dos anos, seguida da *compreensão inferencial*.

Ano	Compreensão literal		Compreensão inferencial	Extracção da ideia principal	Localização da informação	Total de itens
	Verbatim	Paráfrase				
2000	4	3	2	1	3	12
2001	3	2	5		3	13
2002	3		3	3	2	11
2003	4		3	2	2	11
2004	6		3	1	4	14
2005	6	1	3	1	2	13
2006	3	2	6	3		14
2007	4	1	2	3		10
2008	1	1	5	4	1	12

Tabela 5 – Competências avaliadas no processo de compreensão do texto recreativo e do funcional.

Contudo, além de o número de itens que avalia cada uma das competências no processo de compreensão de leitura não ser constante ao longo dos anos, uma análise conjunta das tabelas 3, 4 e 5 permite verificar que o mesmo se passa em relação ao tipo de perguntas utilizado. O nível de compreensão aplicado anualmente apela a um tipo de raciocínio lógico diferente.

5. Síntese

Como dissemos anteriormente, ao efectuar-se uma análise criteriosa das PA, dos critérios de correcção no âmbito dos seus objectivos e dos descritores de níveis de desempenho, verifica-se que a *compreensão da leitura* (extracção do significado do material escrito) envolve competências distintas de ano para ano.

Desta forma, torna-se difícil concretizar um dos objectivos preconizados por este «instrumento de avaliação da eficácia do sistema», o de «contribuir para a tomada de decisões no sentido de melhorar a qualidade das aprendizagens» (ME/DGIDC, 2006: 7), uma vez que: a) há diferenças entre os enunciados das diferentes provas; b) não há homogeneização dos critérios de classificação; c) há diferenças nas percentagens de itens que avaliam predominantemente os mesmos tipos de competências e as mesmas áreas temáticas; d) algumas aprendizagens foram avaliadas por apenas um item; e) nem em todos os anos todas as competências foram avaliadas; f) as mesmas competências são avaliadas tendo por base tipos de texto diferentes; g) o tipo de textos não é constante em todas as provas; h) o número de palavras dos textos é arbitrário; i) não há uma apresentação psicométrica com os índices de dificuldade dos vários itens. Por isso, corroboramos as conclusões obtidas por Santos, Araújo e Graça (2006): «os resultados das Provas de Aferição não são comparáveis entre si». Contudo, este instrumento de avaliação externa permite comparar os alunos entre si, se nos reportarmos exclusivamente ao ano em que elas são aplicadas.

Referências Bibliográficas

- ▶ ALMEIDA, L. e FREIRE, T. (1997). *Metodologia da Investigação em Psicologia e Educação*. Coimbra: Associação dos Psicólogos Portugueses.
- ▶ ARAÚJO, L. (2007). A compreensão na leitura: Investigação, avaliação e boas práticas. In AZEVEDO, F. (Org.). *Formar Leitores. Das teorias às práticas*, pp. 9-18. Lisboa: Lidel.
- ▶ BLACK, T.R. (1999). *Doing Quantitative Research in the Social Sciences: An integrated approach to research design, measurement and statistics*. London: Sage Publications.
- ▶ BLOOM, B.S.; HASTINGS, J.T. e MADDAUS, G. (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
- ▶ CRESWELL, J. (2003). *Research Design: Qualitative, quantitative and mixed approaches*. London: Sage Publications.
- ▶ CUNNINGHAM, D.J. (1987). Outline of an educational semiotic. In *American Journal of Semiotics*, 5, pp. 201-216.
- ▶ FERREIRA, C.A. (2007). *A Avaliação no Quotidiano da Sala de Aula*. Coleção Currículo, Políticas e Práticas. Porto: Porto Editora.
- ▶ GIASSON, J. (2000). *A Compreensão na Leitura*. Porto: Edições ASA.

- ▶ LEITE, I.; FERNANDES, T.; ARAÚJO, L.; FERNANDES, S.; QUERIDO, L.; CASTRO, S.L.; VENTURA, P. e MORAIS, J. (2006). Dificuldades de aprendizagem da leitura e da escrita. In AZEVEDO, F. (Org.). *Língua Materna e Literatura Infantil*. Lisboa: Lidel.
- ▶ Ministério da Educação e Direcção Geral de Inovação e de Desenvolvimento Curricular (2004). *Provas de Aferição do Ensino Básico 4.º, 6.º e 9.º Anos – Análise comparativa 2001-2003*.
- ▶ Ministério da Educação e Direcção Geral de Inovação e de Desenvolvimento Curricular (2006). *Provas de Aferição do Ensino Básico 4.º, 6.º e 9.º Anos – 2004*.
- ▶ MOORE, G.W. (1983). *Developing and Evaluating Educational Research*. Nova Iorque: Harper Collins Publishers.
- ▶ MORAIS, M. (1997). *A Arte de Ler: Psicologia cognitiva da leitura*. Lisboa: Edições Cosmos.
- ▶ NOIZET, G. e CAVERNI, J.-P. (1985). *Psicologia da Avaliação Escolar*. Coleção Psicopedagogia. Coimbra: Coimbra Editora.
- ▶ RAPHAEL, T.E. (1986). Teaching question-answer relationships, revisited. In *The Reading Teacher*, 39, pp. 516-522.
- ▶ ROLDÃO, M.C. (2003). *Gestão do Currículo e Avaliação de Competências: As questões dos professores*. Lisboa: Editorial Presença.
- ▶ SANTOS, C.P.; ARAÚJO, L. e GRAÇA, M. (2006). Resultados das provas de aferição e dos exames de matemática e de português do 9.º ano: Que conclusões tirar? In *Revista Portuguesa de Pedagogia*, 40, 3, pp. 199-218.
- ▶ SIM-SIM, I. (2007). *O Ensino da Leitura: A compreensão de textos*. Lisboa: Ministério da Educação e Direcção-Geral de Inovação e de Desenvolvimento Curricular.
- ▶ SIM-SIM, I. e MICAÉLO, M. (2006). Determinantes da compreensão da leitura. In SIM-SIM, I. (Org.). *Ler e Ensinar a Ler* pp. 35-62. Lisboa: Edições ASA.
- ▶ SIM-SIM, I. (2006a). *Avaliação da Linguagem Oral* (4.ª ed.). Lisboa: Fundação Calouste Gulbenkian.
- ▶ SIM-SIM, I. e RAMALHO, G. (1993). *Como Lêem as Nossas Crianças? – Caracterização do nível de literacia da população escolar portuguesa*. Lisboa: Ministério da Educação e Gabinete de Estudos e Planeamento.
- ▶ VALADARES, J. e GRAÇA, M. (1999). *Avaliando para Melhorar a Aprendizagem*. Lisboa: Plátano Edições Técnicas.
- ▶ VALATTE, R. (1994). Teaching, testing and assessment. In HANCOOK, C. (org.), *Teaching Testing and Assessment*, pp. 1-39. Chicago: National Textbook Company.
- ▶ VENÂNCIO, L. (2004). *Motivação e Compreensão na Leitura: Estudo exploratório de uma experiência de ensino integrado entre a disciplina de língua portuguesa e ciências naturais*. Covilhã: Universidade da Beira Interior.

- ▶ VIANA, F.L. e TEIXEIRA, M.M. (2002). *Aprender a Ler: Da aprendizagem informal à aprendizagem formal*. Porto: Edições ASA.
- ▶ WIERSMA, W. (1995). *Research Methods in Education: An introduction*. (6.ª ed.). Boston: Allyn and Bacon.
- ▶ ZABALZA, M.A. (1992). *Planificação e Desenvolvimento Curricular na Escola*. Lisboa: Asa.

Documentos Consultados

- ▶ Critérios de Correção (disponível em <http://www.gave.min-edu.pt/np3/7.html>).
- ▶ Despacho n.º 5437/2000 de 09 de Março de 2000.
- ▶ Decreto Lei n.º 6/2001 de 18 de Janeiro de 2001.
- ▶ Provas de Aferição (disponível em <http://www.gave.min-edu.pt/np3/7.html>).