

Vinte anos de avaliação das aprendizagens: Uma síntese interpretativa de artigos publicados em Portugal¹

Domingos Fernandes²

Neste trabalho elabora-se uma síntese interpretativa do conteúdo dos artigos publicados entre 1985 e 2005 em revistas portuguesas de índole científica na área da avaliação das aprendizagens. Foram identificadas 19 revistas e seleccionados 59 artigos que foram organizados em categorias tais como *Avaliação Interna*, *Concepções e Práticas de Professores/Alunos* e *Métodos e Técnicas*. A análise e a síntese interpretativa permitiram que se produzissem conclusões tais como: a) apenas 35% dos artigos se baseiam em investigação empírica indicando a falta de uma agenda de investigação em avaliação das aprendizagens; b) a maioria dos artigos é de natureza descritiva e prescritiva, mais baseados *no que há* ou *no que está feito* do que na reflexão teórica; c) são muito poucos os artigos em que há reflexão acerca de métodos alternativos de recolha de informação; e d) em geral, a avaliação formativa, a avaliação sumativa e as relações entre si são discutidas com pouca profundidade e abrangência.

Apesar dos progressos alcançados, é necessário reconhecer as fragilidades que caracterizam a investigação e a produção teórica no domínio da avaliação das aprendizagens. Um domínio reconhecidamente imprescindível para transformar e melhorar as actuais realidades educativas.

Introdução

A investigação que se apresenta e discute neste artigo faz parte de um projecto mais amplo iniciado em Maio de 2003 que se desenvolveu em dois momentos. No primeiro elaboraram-se sínteses relativas a cada um dos seguintes conjuntos de literatura na área da avaliação das aprendizagens: a) teses de doutoramento realizadas nas universidades com assento no Conselho de Reitores das Universidades Portuguesas (CRUP); b) dissertações concluídas entre 1994 e 2005 (inclusive) no âmbito de programas de

1 O projecto em que se enquadra a investigação que aqui se apresenta e discute tem sido financiado pela Unidade de Investigação & Desenvolvimento em Ciências da Educação da Faculdade de Psicologia e de Ciências da Educação da Universidade de Lisboa. Agradeço na pessoa do seu Coordenador, Professor Doutor Rui Canário, todo o apoio concedido.

2 Universidade de Lisboa

mestrado realizados em universidades com assento no CRUP; c) livros de autores portugueses publicados entre 1980 e 2005 (inclusive); e d) artigos publicados entre 1985 e 2005 (inclusive) em revistas portuguesas de índole científica. No segundo momento fez-se uma meta-síntese e uma meta-interpretação (Weed, 2005) das sínteses elaboradas na primeira fase. Até ao presente foram consultadas cerca de 350 publicações e directamente analisadas mais de 200.

O presente texto é uma parte da síntese baseada em artigos que foi elaborada na primeira fase do projecto.

Para além desta introdução o presente trabalho está organizado nas seguintes partes: a) Problema e Questões da Investigação; b) Enquadramento da Investigação; c) Método; d) Uma Caracterização Geral Dos Artigos; e) Uma Descrição E Análise Dos Artigos; f) Uma Síntese Interpretativa; e g) Considerações e Reflexões Finais.

Problema e questões da investigação

Nos últimos 25/30 anos foram publicados, por autores portugueses, algumas centenas de trabalhos de natureza diversa no domínio da avaliação das aprendizagens dos alunos. Este facto, em si mesmo, torna clara a necessidade de se desenvolver investigação que sintetize o que é que, em Portugal, tem vindo a ser produzido no domínio da avaliação das aprendizagens e que seja capaz de: a) caracterizar globalmente o que tem sido produzido pelos investigadores e estudiosos portugueses; b) identificar e discutir os problemas e os temas que mais têm merecido a sua atenção; c) analisar como é que esses problemas têm sido abordados; e d) contribuir para o desenvolvimento teórico e prático de um domínio científico e pedagógico de reconhecida importância (Black & Wiliam, 1998; Stake, 2006).

A investigação aqui discutida teve em conta as seguintes questões orientadoras:

1. Como é que se poderão caracterizar globalmente os artigos analisados?
2. Que questões teóricas e práticas da área da avaliação das aprendizagens são mais e menos privilegiadas pelos autores?
3. Que sínteses e reflexões se poderão elaborar relativamente ao conteúdo dos artigos analisados?

Enquadramento da investigação

Esta parte do texto está organizada em duas secções. Na primeira apresentam-se e discutem-se os principais contornos do debate que se tem vindo a desenvolver nos últimos anos no domínio da avaliação das aprendizagens. Na segunda discutem-se brevemente questões inerentes à concepção e utilização de sínteses de literatura.

Acerca da avaliação das aprendizagens

A investigação realizada nos últimos anos evidencia com clareza o papel que a avaliação das aprendizagens poderá ter na melhoria do ensino e das aprendizagens (e.g., Figari & Achouche, 2001; Grégoire, 1996; Shepard, 2000, 2001; Stiggins, 2004). Em particular, Black & Wiliam (1998a; 1998b), num artigo de revisão de literatura sobre práticas de avaliação formativa, sublinham três resultados de grande alcance e significado: a) a prática sistemática de avaliação formativa melhora substancialmente as aprendizagens dos alunos; b) os alunos que mais beneficiam das práticas de avaliação formativa são os que têm mais dificuldades de aprendizagem; e c) os alunos que frequentam aulas em que a avaliação formativa é predominante obtêm melhores resultados em provas de avaliação externa (e.g., exames) do que os alunos que frequentam aulas em que a avaliação que predomina é de natureza sumativa.

No seguimento deste seminal trabalho de Paul Black e Dylan Wiliam houve uma espécie de revigoramento da discussão em torno da avaliação das aprendizagens, após uma certa euforia vivida nos anos 80 e, muito especialmente, no início dos anos 90, que se traduziu na publicação de artigos e livros de referência insubstituível (e.g., Abrecht, 1991; Allal, 1986; Allal, Cardinet & Perrenoud, 1979; Berlak, 1992a, 1992b; Cardinet, 1986, 1991; Crooks, 1988; Gipps, 1994; Natriello, 1987; Nuttall, 1986; Stiggins & Conklin, 1992).

De facto, após aquela publicação de Black & Wiliam, não só se assistiu a uma certa proliferação de reacções ao artigo propriamente dito (e.g., Biggs, 1998; Perrenoud, 1998) como, sobretudo, se relançaram as investigações empíricas e a construção teórica com base em renovadas visões epistemológicas, em novos desenvolvimentos das teorias das aprendizagens e curriculares e numa variedade de contributos tais como os que são provenientes da sociologia, das ciências da cognição, da antropologia e das teorias da comunicação (e.g., Black & Wiliam, 2006a, 2006b, 2006c; Earl, 2003; Gardner, 2006a, 2006b; Gipps & Stobart, 2003; Harlen, 2005, 2006; Kellaghan & Madaus, 2003; Stiggins & Chappuis, 2005; Stobart, 2006).

Mas o que se poderá dizer de fundamental, com base na investigação empírica e na reflexão teórica que se tem produzido nestes últimos anos, que possa ter interesse no contexto deste trabalho? Ou seja, quais são as questões de interesse a estudar e a debater? No essencial podem destacar-se três linhas de preocupação e/ou de desenvolvimento: a) A necessidade de consolidar uma teoria da avaliação formativa que possa apoiar e sustentar as práticas realizadas nas salas de aula; b) A necessidade de credibilizar as avaliações internas, nomeadamente através da melhoria das práticas a todos os níveis e da caracterização fina e exaustiva dos processos utilizados; e c) A necessidade de credibilizar as avaliações externas através da melhoria da elaboração e selecção de itens, dos processos de correcção e, naturalmente, através de uma melhor definição da natureza das provas tornando-as mais adequadas às populações a que se destinam.

A avaliação formativa ocupa um lugar de destaque dadas as suas comprovadas potencialidades na melhoria do ensino e das aprendizagens e, simultaneamente, a grande e persistente dificuldade em concretizá-la nas salas de aula. Por isso a construção teórica parece fundamental para a clarificação conceptual que sustente a melhoria das práticas. Neste aspecto há um considerável esforço a fazer que passa, nomeadamente, pela clarificação do próprio conceito de avaliação formativa, pela integração e depuração terminológica, por uma mais adequada caracterização das tarefas e métodos de avaliação e por uma definição mais clara dos papéis de professores e alunos no processo de avaliação formativa. As tarefas, os alunos e os professores constituem pilares fundamentais na construção de uma teoria da avaliação formativa. Consequentemente, há um complexo sistema de relações entre aqueles *elementos* que é necessário descrever, analisar e interpretar (Fernandes, 2006). Parece também ser relevante estudar como é que as perspectivas sociocognitivas e socioculturais da avaliação das aprendizagens se poderão articular para que possamos ter uma visão mais profunda e abrangente de questões tais como o papel e a natureza do *feedback*, o papel e a natureza da auto-regulação e do auto-controlo e o papel e a natureza de processos tais como a co-avaliação, a auto-avaliação e a hetero-avaliação (Fernandes, 2005).

Assim, no domínio da avaliação interna, em que avulta evidentemente o papel da avaliação formativa, parece importante que a investigação e a reflexão teórica e prática aborde questões tais como:

1. O estudo das relações entre a avaliação formativa e a avaliação sumativa.
2. O estudo das relações entre as práticas de avaliação formativa e as aprendizagens efectivamente realizadas pelos alunos.
3. Os papéis de alunos e de professores no processo de avaliação formativa.

4. Os métodos de recolha de evidências de aprendizagem.
5. As utilizações da informação gerada pelas práticas de avaliação formativa.
6. Os processos de auto-avaliação, de auto-regulação e de auto- controlo.
7. As relações entre o *feedback*, a regulação e a avaliação formativa.
8. As relações entre conhecimentos, concepções e práticas de avaliação formativa dos professores.
9. As relações entre o desenvolvimento curricular, a selecção de tarefas e avaliação formativa (integração da avaliação formativa nos processos de ensino e aprendizagem).
10. As relações entre a avaliação formativa e as teorias da aprendizagem.
11. A qualidade das avaliações formativas (validade e fiabilidade).
12. O estudo das relações entre métodos de avaliação sumativa interna (e.g., provas realizadas ao nível da escola) e a avaliação formativa realizada ao nível das salas de aula (pense-se, por exemplo, no papel moderador e regulador que tais provas internas poderão ter).

Ao nível da avaliação externa as áreas problemáticas mais relevantes são as seguintes:

- a) efeitos dos exames e das provas aferidas nas escolas, nos professores e nos alunos;
- b) natureza e qualidade psicométrica das provas (validade, fiabilidade, discriminação e dificuldade) e dos critérios e processos de correcção; c) análise dos resultados dos exames e das provas aferidas e medidas associadas a essas análises tomadas pela administração e pelas escolas.

Esta lista de áreas problemáticas e de investigação no domínio da avaliação interna e da avaliação formativa, assim como da avaliação externa decorre da literatura da especialidade e contribuiu significativamente para analisar e sintetizar o conjunto de artigos que aqui se apresentam e discutem. Não cabia no âmbito deste artigo elaborar sobre cada uma das áreas problemáticas enunciadas; no entanto, serão abordados os seus principais contornos à medida que se vão discutindo os artigos seleccionados para análise.

Acerca das sínteses de literatura

As sínteses de literatura têm vindo a ganhar uma importância crescente como actividades primárias de investigação (Weed, 2005). É possível identificar uma variedade de abordagens relativas à concepção e à elaboração de sínteses que revelam o reconhecimento crescente desta forma de investigação (ver, por exemplo, Cooper, 1988, 1998,

2003; Glass, 1976; Noblit & Hare, 1988; Slavin, 1984, 1986, 2004; Suri, 1999a, 1999b, 2002; Weed, 2005). As sínteses constituem reflexões sistematizadas que integram o trabalho teórico ou o trabalho de investigação empírica realizado por outros e que, no final, deverão traduzir mais do que *a soma das partes* analisadas.

Uma das vantagens dos trabalhos de síntese da literatura tem a ver com o facto de permitirem que os leitores e investigadores possam ter uma visão abrangente e profunda de desenvolvimentos teóricos e práticos numa dada área do conhecimento ao longo de um dado período de tempo. Interessa, porém, diferenciar entre sínteses ou revisões de literatura que são desenvolvidas pelos investigadores com o propósito de enquadrarem conceptualmente uma dada investigação primária e as sínteses da literatura cujo principal objectivo é o de contribuir para a compreensão de um dado problema ou domínio de investigação a partir da análise dos resultados e conclusões de um conjunto mais ou menos alargado de investigações primárias. A este propósito Suri (1999a, 1999b) considera que enquanto no primeiro caso a síntese é um meio para atingir um fim – contextualizar teórica e conceptualmente uma investigação -- no segundo caso ela é um fim em si mesmo – produzir compreensão sobre um dado fenómeno. (Na presente investigação a síntese é encarada nesta última perspectiva.)

Vários autores (e.g., Cooper & Lindsay, 1998; Suri, 1999b, 2002; Suri & Clarke, 1999) referem que as sínteses são um desafio à imaginação para que se encontrem formas de tornar visível, de forma integrada e simples, a complexidade, a diversidade, os propósitos, métodos, perspectivas e conclusões constantes num número mais ou menos elevado de investigações ou outros trabalhos. Neste sentido, contribuem para identificar e delinear novos projectos de investigação.

Há várias sínteses no domínio da avaliação das aprendizagens que se tornaram referências incontornáveis nesta área do conhecimento. Algumas das mais recentes foram elaboradas no âmbito da avaliação formativa por Allal & Lopez (2005), que fazem uma revisão da literatura teórica e empírica de trabalhos publicados em língua francesa, e por Koller (2005) que faz uma revisão da literatura empírica alemã. Bem conhecida é a revisão da literatura de Black & Wiliam (1998), também no domínio da avaliação formativa. São também incontornáveis os trabalhos de síntese já clássicos de Crooks (1988) e de Natriello (1987), ambos referentes a impactos dos processos de avaliação nos estudantes. Sousa (1995) fez uma síntese de literatura de investigação na área da avaliação das aprendizagens publicada no Brasil entre 1930 e 1980.

Em Portugal foram recentemente elaboradas duas sínteses centradas quase exclusivamente em trabalhos de investigação empírica (Barreira & Pinto, 2006; Neves, Jordão & Santos, 2004).

Suri & Clarke (1999a) identificam os seguintes tipos de sínteses de investigação existentes na literatura: a) as chamadas revisões narrativas de investigação; b) as meta-análises; c) as sínteses realizadas a partir das “melhores evidências” (*best-evidence synthesis*); e d) as sínteses de investigações de natureza qualitativa. É uma categorização que parece ter sido determinada pelos métodos utilizados em cada um dos tipos de síntese. Porém, em trabalhos mais recentes, Suri (1999b, 2002) faz referência a sínteses cuja categorização parece ter sido mais orientada pela natureza dos objectos de análise e/ou do estado em que se encontra a investigação no domínio de interesse. Assim são consideradas: a) *sínteses agregadoras*, quando estamos perante um conjunto mais ou menos homogêneo de evidências; b) *sínteses interpretativas*, quando estamos a lidar com um conjunto heterogêneo de resultados decorrentes de diferentes metodologias e a examinar diferentes conceitos; c) *sínteses exploratórias*, utilizadas quando se pretende identificar linhas de investigação que de algum modo são promissoras ou quando se está perante um domínio pouco investigado; e d) *sínteses confirmatórias*, que serão preferíveis quando estamos a analisar estudos de uma área já bastante investigada e consolidada.

Naturalmente que, tal como acontece em estudos de avaliação e em investigações, as abordagens a utilizar nas sínteses de literatura devem estar mais dependentes dos seus contextos e propósitos do que de qualquer ortodoxia teórica ou metodológica.

Método

O processo de identificação, selecção e recolha dos artigos incluídos nesta investigação foi inicialmente desenvolvido a partir das revistas que foi possível identificar através das Bibliotecas de Universidades, de Faculdades e/ou dos respectivos Departamentos e das Escolas Superiores de Educação, quase todas disponíveis *on-line*, do Centro de Documentação e Informação da Direcção Geral de Inovação e de Desenvolvimento Curricular (DGIDC), que inclui o acervo da Biblioteca do extinto Instituto de Inovação Educacional. Foram ainda pesquisadas as bases de dados ColCat, PORBASE e SIRIUS que também permitiram a identificação de revistas e artigos.

Após este processo de pesquisa organizou-se uma base de dados integrando as referências essenciais de cada um dos artigos que se considerava ter condições para integrar a investigação. Para seleccionar um artigo foram considerados os seguintes critérios:

1. A avaliação das aprendizagens dos alunos era um dos principais temas organizadores do artigo ou de alguma, ou algumas, das suas secções.
2. O artigo tinha de ter sido publicado entre 1985 e 2005 (inclusive).

3. O primeiro autor do artigo tinha de ser de nacionalidade portuguesa.
4. O artigo tinha de estar publicado numa revista portuguesa de educação de índole científica.

Tem de reconhecer-se que este último critério poderá suscitar algumas dúvidas no leitor. No entanto, se fossem apenas consideradas as revistas indexadas em bases de dados internacionalmente conceituadas e com sistemas de arbitragem reconhecidos, critérios normalmente seguidos a nível internacional em trabalhos de investigação desta natureza, não seria possível analisar a grande maioria dos artigos publicados pela comunidade científica portuguesa que tem desenvolvido trabalho na área da avaliação das aprendizagens. Assim, para efeitos desta investigação, por *Revista Portuguesa de Educação de Índole Científica* entende-se uma revista que obedece pelo menos a um dos seguintes critérios:

1. É editada por uma instituição, faculdade, escola, departamento ou centro de investigação do ensino superior na área da educação.
2. É editada por um qualquer organismo público ou privado e tem um conselho científico ou editorial maioritariamente constituído por investigadores da área da educação.
3. É editada por uma sociedade científica e tem um conselho científico ou editorial que inclui investigadores da área da educação.
4. É editada por uma associação de professores ou qualquer outro tipo de associação de profissionais da educação e tem um conselho científico ou editorial que inclui investigadores da área da educação.

Assim, não foram seleccionados artigos constantes numa variedade de revistas e boletins de grande divulgação e de interesse geral, assim como artigos constantes em livros ou em actas de congressos de natureza científica e/ou profissional.

A pesquisa das revistas e a correspondente identificação dos artigos acabou por determinar o intervalo de tempo em que incidiu o processo final de selecção. Na verdade, rapidamente se constatou que a grande maioria das revistas da área da educação, sobretudo as que estão associadas a instituições do ensino superior, publicaram os seus primeiros números a partir de meados dos anos 80. Nestas condições, pareceu mais adequado centrar a pesquisa entre 1985 e 2005 (inclusive) e não entre 1980 e 2005 como estava inicialmente previsto.

A aplicação dos critérios acima enunciados resultou na consulta de 86 artigos e na selecção, para análise, de 59. O Anexo A é um quadro em que se identificam os 59

artigos seleccionados e no qual se indicam: a data de publicação, o nome do autor ou dos autores, o título e a revista onde foi publicado.

Tendo em vista a organização, a análise e a síntese dos artigos seleccionados foram tidas em conta as seguintes seis categorias gerais: *Avaliação Interna*, *Avaliação Externa*, *Concepções e Práticas de Professores/Alunos*, *Métodos e Técnicas*, *Reacção a Normativos da Administração* e *Reflexão/Análise de Literatura*. Estas categorias decorreram da literatura e da análise dos artigos e foram consideradas suficientemente amplas para *acomodar* uma grande variedade de trabalhos. No Quadro 1 apresenta-se a distribuição dos artigos pelas categorias consideradas.

Quadro 1. Distribuição dos artigos pelas categorias seleccionadas e de acordo com os quatro lustros abrangidos pela investigação.

	Avaliação Interna	Avaliação Externa	Concepções e Práticas	Métodos e Técnicas	Reacções a Normativos	Reflexão/ Análise de Literatura
85-89	5	1	-	-	1	1
90-94	6	-	2	5	3	2
95-99	-	3	2	3	2	1
00-05	4	4	5	5	1	3
Total	15	8	9	13	7	7

Foram utilizados procedimentos de recolha de informação que passaram pela leitura holística e focada dos artigos ou das suas secções de interesse para os propósitos desta investigação. Procedeu-se então ao registo escrito, através de comentários breves de natureza descritiva, da informação referente a cada categoria. Ou seja, cada categoria considerada deu origem a uma análise horizontal *através* de todos os artigos seleccionados. Além disso, relativamente a cada artigo, procedeu-se a uma análise vertical *através* de todas as categorias consideradas. Assim, elaboraram-se sínteses para cada artigo (verticais) e sínteses para cada categoria (horizontais).

A análise cruzada daqueles dois conjuntos de sínteses revelou-se importante para que se produzisse uma *síntese global*, identificando os aspectos que deveriam ser destacados. Além disso, permitiu fundamentar a interpretação, a discussão e as reflexões conclusivas acerca dos dados recolhidos a partir do material seleccionado para a investigação.

Uma caracterização geral dos artigos

Como acima se referiu foram identificados e analisados 59 artigos publicados entre 1985 e 2005 (inclusive): oito foram publicados nos anos 80, 29 nos anos 90 e 22 na primeira metade da presente década. Estes dados parecem indiciar que tem havido um aumento significativo de artigos publicados na área da avaliação das aprendizagens. Repare-se que só nos primeiros cinco anos desta década já foi publicado um número de artigos que corresponde a cerca de 75% dos artigos publicados em toda a década anterior.

O Quadro 2 mostra a distribuição dos artigos pelas dezanove revistas em que foram publicados, identificando cada revista e o ano em que foi publicado o primeiro número. Doze revistas estão directamente associadas a instituições do ensino superior, quatro a sociedades científicas, duas a associações de natureza profissional e/ou científica e uma ao extinto Instituto de Inovação Educacional do Ministério da Educação. Como acima se afirmou a grande maioria das revistas (14) surge a partir de meados da década de 80 o que parece poder explicar-se pelo facto de a educação como domínio de investigação ter tido um forte incremento entre nós a partir desse período, nomeadamente com a entrada em funcionamento das Escolas Superiores de Educação (ESE) e com a consolidação e o desenvolvimento dos departamentos da área nas universidades. Repare-se que das seis revistas que iniciaram a sua publicação nos anos 80 duas são da iniciativa de departamentos universitários (*Revista de Educação* e *Revista Portuguesa de Educação*) e duas de ESE (*Aprender* e *Revista da ESE de Santarém*). Além disso, é também a partir desta altura que começam a surgir projectos de investigação associados a programas de mestrado e de doutoramento assim como financiamentos a projectos na área da educação. Consequentemente, gerou-se uma dinâmica que pode explicar o surgimento de 14 revistas em 18 anos.

A análise do Quadro 2 permite verificar que cerca de 70% dos artigos analisados foram publicados em apenas sete revistas: a *Inovação*, com onze artigos, a *Revista Portuguesa de Pedagogia*, com nove, a *Quadrante*, com seis, o *Boletim da Sociedade Portuguesa de Química* com quatro, e a *Aprender*, o *Boletim da Sociedade Portuguesa de Educação Física* e a *Revista Portuguesa de Educação*, com três artigos cada uma.

Quadro 2. Identificação das revistas com artigos seleccionados: Primeiro ano de publicação e número de artigos publicados a partir de 1985.

Revista	Primeiro ano de publicação	Número de artigos publicados
Gazeta da Matemática Sociedade Portuguesa de Matemática.	1940	2
Gazeta da Física Sociedade Portuguesa de Física.	1946	2
Revista Portuguesa de Pedagogia Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra.	1960	9
Ludens Faculdade de Motricidade Humana da Universidade Técnica de Lisboa.	1976	1
Boletim da Sociedade Portuguesa de Química Sociedade Portuguesa de Química.	1977	4
Boletim da Sociedade Portuguesa de Educação Física Sociedade Portuguesa de Educação Física.	1986	3
Revista de Educação Faculdade de Ciências da Universidade de Lisboa (Departamento de Educação).	1986	2
Aprender Escola Superior de Educação de Portalegre.	1987	3
Inovação Instituto de Inovação Educacional do Ministério da Educação, entre-tanto extinto.	1988	11
Revista Portuguesa de Educação Instituto de Educação e Psicologia da Universidade do Minho.	1988	3
Revista da Escola Superior de Educação de Santarém Escola Superior de Educação de Santarém.	1989	1
Quadrante Associação de Professores de Matemática.	1992	6
Educação, Sociedade & Culturas Faculdade de Psicologia e de Ciências da Educação da Universidade do Porto.	1994	2
Educare-Educere Escola Superior de Educação de Castelo Branco.	1995	2
Revista da Escola Superior de Educação de Viana do Castelo Escola Superior de Educação de Viana do Castelo.	1996	1
Educação & Comunicação Escola Superior de Educação de Leiria.	1999	1
Da Investigação às Práticas-Estudos de Natureza Educacional Escola Superior de Educação de Lisboa.	2001	2
Revista Portuguesa de Investigação Educacional Instituto de Educação da Universidade Católica Portuguesa.	2002	2
Revista de Estudos Curriculares Associação Portuguesa de Estudos Curriculares	2003	2

Cerca de 80% dos artigos analisados têm como primeiro autor, ou como um dos autores, docentes e investigadores do ensino superior. Os restantes artigos são da autoria de docentes dos ensinos básico ou secundário que, em vários casos, apresentam e discutem trabalhos de investigação realizados no âmbito de programas de mestrado. Este facto revela que as investigações realizadas no âmbito destes programas têm ainda uma reduzida divulgação através de artigos publicados em revistas de índole científica. Na verdade, tendo em conta que no período abrangido por esta investigação se pode estimar com segurança que foram concluídas cerca de 60 dissertações de mestrado na área da avaliação das aprendizagens, são ainda uma pequena minoria os investigadores que procedem à sua divulgação sob a forma de artigo.

Há apenas vinte e um artigos, cerca de 35%, que são claramente baseados na recolha de dados empíricos ou num processo de natureza investigativa. Ou seja, cerca de 65% dos artigos analisados centra-se essencialmente na discussão e reflexão teórica, mais ou menos fundamentada, acerca de um ou mais aspectos relacionados com a avaliação das aprendizagens. Em certos casos há alguma referência a dados obtidos pelo autor ou algum tipo de análise de dados obtidos por outrém. Vários artigos constituem reacções analíticas a medidas tomadas pelo Ministério da Educação no domínio da avaliação das aprendizagens (e.g., exames, despachos regulamentadores). Em suma, na maioria dos casos, os artigos não descrevem ou discutem investigação empírica realizada pelos seus autores. São reflexões, análises teóricas ou revisões de literatura através das quais, por vezes, os autores retiram *lições* ou *ilações* para as práticas.

A análise dos dados mostra que a grande maioria dos autores surge apenas uma vez como autor ou co-autor dos artigos seleccionados. As excepções são João Boavida e Carlos Barreira que aparecem como autores ou co-autores em cinco artigos cada um (Barreira, 2001, 2002, 2005; Boavida, 1985; Boavida, Lopes & Vaz, 1982; Boavida & Vaz, 1987; Boavida & Barreira, 1992; Boavida & Barreira, 1993); e ainda Ana Veiga Simão, António Rosado, Clara Vasconcelos, Leonor Santos e Paulo Abrantes que são autores, ou co-autores, de dois artigos cada um (Abrantes, 1989; Leal (actualmente Santos) & Abrantes, 1990; Rosado & Dias, 2002; Santos, 2003; Simão, 1992, 2005; Silva, Rosado & Dias, 2003; Vasconcelos, 2002; Vasconcelos, Praia & Almeida, 2004).

Globalmente, considerando como referência apenas os primeiros autores nos casos em que há dois ou mais, há um equilíbrio entre géneros no que se refere à autoria dos 59 artigos seleccionados para análise. De facto, 30 artigos foram escritos por autores do género feminino, 25 das quais docentes do ensino superior, e 29 por autores do género masculino, 22 dos quais docentes do ensino superior. Trinta e nove dos 59 artigos ana-

lisados foram escritos por um único autor: vinte e um do género feminino e dezoito do género masculino.

Em suma, ao longo dos 20 anos abrangidos por esta investigação são evidentes alguns factos que merecem referência.

Os dados evidenciam que, ao longo dos anos, há um aumento consistente do número de artigos publicados. Nos últimos dez anos verifica-se que em 2000-2005 foram publicados 22 artigos, o dobro dos que se publicaram em 1995-1999 (Anexo A).

Há um claro desequilíbrio entre o número de artigos que resultam de processos de investigação empírica e o número de artigos que são elaborados a partir de análises de índole mais teórica, que constituem a grande maioria. Ou seja, existe alguma dificuldade em divulgar as investigações realizadas no âmbito de programas de pós-graduação através de artigos publicados em revistas. Por outro lado, esta situação pode também indiciar a inexistência, ou a fragilidade, de programas de investigação empírica suficientemente consolidados capazes de gerar produção e divulgação de conhecimento no domínio da avaliação das aprendizagens.

São poucos os artigos que abordam um e um só assunto; normalmente, a propósito da avaliação das aprendizagens dos alunos, os autores acabam por se referir a outros temas tais como as teorias da aprendizagem, os paradigmas de investigação, as questões específicas de avaliação numa dada disciplina, a questão da formulação de objectivos educacionais ou da recolha de evidências de aprendizagem. Dos 59 artigos analisados, 22 não estão contextualizados num ano ou nível de escolaridade específico ou num dado curso ou ainda numa dada disciplina (e.g., Benavente, 1990; Barreira, 2005; Boavida *et al.*, 1987) e nove fazem claras incursões no domínio das políticas educativas propondo recomendações no domínio da formação de professores, criticando as condições existentes no sistema educativo ou referindo supostas ineficiências ou incongruências dos normativos que regulam o seu funcionamento (e.g., Barreira, 2002; Boavida & Barreira, 1992; Roldão, 1989).

Doze artigos abordam a avaliação das aprendizagens no contexto da disciplina de Matemática (e.g. Abrantes, 1989; Oliveira & Pereira, 1993; Graça, 2003), sete no contexto de disciplinas de Ciências (e.g., Castro & Pereira, 1994; Martins & Cachapuz, 1988; Vasconcelos *et al.*, 2004), cinco no contexto da Língua Portuguesa (e.g., Abreu *et al.*, 2004; Rocha, 1990), três no contexto da Educação Física (e.g., Carvalho, 1994; Rosado & Dias, 2002), três no contexto do Pré-Escolar (e.g., Nabuco, 2000; Pereira, 1997) e dois no domínio das artes visuais (Eça, 2004; Paulo & Santos, 1998). Há apenas um artigo no contexto da disciplina de Inglês (Carvalho, 1997) e outro no contexto do

1.º ciclo (Peixoto, 1998). Estes números evidenciam a existência de uma certa dinâmica no que se refere à reflexão e/ou investigação nos contextos das disciplinas de Matemática, Ciências e Língua Portuguesa que, no entanto, talvez fique aquém do que seria desejável ou necessário. Por outro lado, evidenciam lacunas ou mesmo ausências muito significativas em domínios disciplinares e não disciplinares e, muito particularmente, ao nível do 1.º ciclo do ensino básico.

Uma descrição e análise dos artigos

Nesta secção apresenta-se uma descrição e análise dos artigos agrupando-os segundo as décadas em que foram publicados e segundo as seis categorias já enunciadas: *Avaliação Interna*, *Avaliação Externa*, *Concepções e Práticas de Professores/Alunos*, *Métodos e Técnicas*, *Reacções a Normativos da Administração* e *Reflexão/Análise de Literatura*.

Os artigos da segunda metade dos anos 80

Dos oito artigos publicados neste período cinco estão mais centrados nas questões relativas à *Avaliação Interna* (Boavida & Vaz, 1987; Abrantes, 1989; Martins & Cachapuz, 1988; Traguil, 1988; Valadares & Teodoro, 1988) e os outros três distribuem-se por cada uma das seguintes categorias: *Avaliação Externa* (Boavida, 1985), *Reflexão/Análise de Literatura* (Boavida et al., 1986) e *Reacção a Normativos da Administração* (Roldão, 1989). Em três dos artigos apresentam-se e discutem-se dados de natureza empírica (Boavida, 1985; Boavida & Vaz, 1987; Martins & Cachapuz, 1988).

Boavida & Vaz (1987) discutem a objectividade e a subjectividade da avaliação referindo que no primeiro caso se valoriza a classificação (de natureza analítica) e no segundo a avaliação propriamente dita (de natureza sintética). Há cerca de 20 anos atrás aqueles autores referiam que “A partir de uma avaliação pedagógica feita com base em critérios essencialmente subjectivos ter-se-á evoluído no sentido de reduzir e controlar esses factores de subjectividade, como condição para uma avaliação menos aleatória”(p. 465). A partir daqui consideravam que “a subjectividade parece ser um elemento em perda” (p. 465). Trata-se de uma perspectiva que parece estar baseada na docimologia experimental ou na chamada fase positiva da docimologia (Boavida, 1985). Ou seja, a procura da objectividade através do desenvolvimento e construção de métodos e técnicas adequados em busca de uma disciplina exacta. Porém, questionando “se a subjectividade na avaliação é necessariamente um factor limitativo” (p. 467), os autores perguntam: “(...) concorrerá a subjectividade necessariamente para a desvalorização

da avaliação?" (p. 467). São questões epistemológicas que continuam a ser discutidas. Alguns autores questionam se deveremos continuar a procurar a objectividade ou se deveremos olhar para a avaliação com outras lógicas ou racionalidades menos técnicas e mais críticas ou sociocríticas (e.g., Afonso, 1994; Gipps, 1994).

O trabalho de Boavida & Vaz (1987) apresenta um estudo empírico em que se compararam as classificações atribuídas por futuros professores a trabalhos por si realizados, que os autores designam como auto-avaliação, com as classificações atribuídas a trabalhos realizados por outros, que os autores designam como hetero-avaliação. Concluiu-se que os participantes atribuíram melhores classificações aos seus próprios trabalhos e piores aos que foram realizados por outros. Para explicar este facto os autores referem que "(...) a auto-avaliação é um processo que envolve mais parâmetros, alguns deles habitualmente não considerados na hetero-avaliação" (p. 473) e, por isso, "(...) a hetero-avaliação é limitada e limitativa" (p. 474). Por conseguinte defendem as virtualidades pedagógicas da auto-avaliação afirmando "(...) que por certo uma função importante lhe estará reservada num estágio mais avançado da avaliação pedagógica (...)" (p. 475). Interessante esta premonição dos autores elaborada há cerca de 20 anos atrás já que a auto-avaliação é hoje um conceito mais abrangente e mais denso e um processo indispensável no desenvolvimento das aprendizagens dos alunos (Fernandes, 2005).

No artigo de Martins & Cachapuz (1988) descreve-se uma recolha de dados feita através de um inquérito por questionário a alunos do 8.º ano de escolaridade destinada a analisar a natureza das suas dificuldades na aprendizagem do tema Energia e Reacções Químicas. Os autores concluem que as duas questões utilizadas podem ser "adaptadas como instrumento diagnóstico" (p.10) e fazem recomendações de natureza pedagógica mas não chegam a discutir questões de avaliação propriamente ditas.

Abrantes (1989) enuncia cinco princípios considerados fundamentais para o desenvolvimento de uma avaliação pedagógica que responda às exigências decorrentes de "Alterações significativas ao nível dos objectivos, dos conteúdos e dos métodos (...)" (p.10): a) Integração no ensino e na aprendizagem; b) Positividade, valorizando mais o que os alunos sabem e são capazes de fazer; c) Generalidade, realçando objectivos a atingir a longo prazo; d) Variedade, incidindo em todos os domínios da aprendizagem; e e) Consistência, com os objectivos e com as actividades realizadas. Aquele autor discute ainda a necessidade de se diversificarem os métodos de recolha de dados através de ensaios escritos, discussões orais, relatórios e observações porque "Avaliar novas capacidades com as actuais técnicas seria com certeza um desastre" (p.11). O artigo de Traguil (1988) também enuncia um conjunto de princípios e orientações semelhantes

dando alguma ênfase à participação dos alunos, dos pais e de outros professores. Esta autora remete-nos para o processo de meta-avaliação quando pergunta: “E não será que também o próprio processo de avaliação está sujeito a avaliação?” (p. 61).

O artigo de Valadares & Teodoro (1988) é de natureza algo técnica e prescritiva, fazendo uma apresentação fundamentada de uma variedade de assuntos no domínio da avaliação que vai das características psicométricas de uma prova de avaliação (validade, fiabilidade, discriminação e dificuldade) aos tipos de itens e suas características, passando pela formulação de juízos de valor e pela processo de recolha e interpretação sistemática de informação.

Boavida (1985) descreve uma investigação destinada a analisar como é que alunos do 4.º ano de licenciaturas em ensino, atribuem classificações a trabalhos dos seus colegas. Os alunos atribuíam as classificações em grupo e individualmente, tendo-se variado o número de níveis das escalas utilizadas (três, cinco e sete) e também o número de critérios de avaliação dos trabalhos (três, quatro e cinco). Entre outras conclusões o autor refere que o aumento do número de critérios de avaliação não está associado à atribuição de um espectro mais alargado de níveis de classificação e que não há diferenças significativas entre as classificações atribuídas individualmente e as que são atribuídas em grupo. Neste artigo o autor refere as discrepâncias nas classificações que são atribuídas em contexto de provas de avaliação externa (e.g., exames) apoiando-se em investigação realizada nos anos 20 do século XX e mesmo em investigação realizada no século XIX. A este propósito refere que os exames são pobres na orientação e selecção dos alunos pois são “muito aleatórios na atribuição de classificações, ou na distribuição justa dos candidatos” (Boavida, 1985, p. 165). Esta discussão é enquadrada tendo em conta três períodos no desenvolvimento da docimologia: a) a docimologia negativa, resultante em boa medida do descrédito provocado pela aleatoriedade das classificações atribuídas nos exames; b) a docimologia positiva, resultante do esforço realizado no sentido de melhorar os exames a partir de uma concepção de avaliação pedagógica que “(...) não deve ser vista como um processo exterior e essencialmente punitivo, mas sim como um elemento indissociável do processo, e donde se podem eliminar quase completamente os elementos aversivos” (p.168).; e c) a docimologia experimental resultante das abordagens que se desenvolveram para aperfeiçoar os exames a partir de melhorias na construção dos itens, na fiabilidade inter-classificadores ou nos processos de standardização.

Boavida *et al.* (1986) chamam a atenção para problemas decorrentes da *contaminação* entre a avaliação formativa e a avaliação sumativa, afirmando que “Reduzir esta contaminação é um dos grandes objectivos no domínio da avaliação (...)” (p. 269), e para

a necessidade de se evitar que a avaliação formativa se transforme num somatório de avaliações sumativas. A posição daqueles autores acerca do papel da avaliação formativa fica bem ilustrada na seguinte afirmação "(...) está em jogo uma mentalidade nova, a desenvolver entre professores e alunos, que encare as tarefas de avaliação como tarefas explicitamente pedagógicas. E sê-lo-á na medida em que a avaliação for formativa (...)" (p. 274). Além disso, associam a avaliação contínua à avaliação formativa que, por isso mesmo, não deverá ter funções classificativas. Traguil (1988) chega a afirmar que a avaliação contínua é "(...) uma das mais inovadoras propostas da Pedagogia dos nossos dias" (p. 59). Por seu lado, Abrantes (1989) refere que a avaliação formativa não se pode reduzir a uma emulação dos testes sumativos, devendo fazer parte integrante do ensino e da aprendizagem e responder às inadequações dos testes para avaliar um alargado espectro de capacidades, processos e estratégias de resolução de problemas.

Boavida *et al.* (1986) fazem ainda uma discussão acerca do papel dos exames e da sua relação com a avaliação pedagógica. No fundo estes autores referem que os exames têm sido identificados com a avaliação o que tem contribuído para a desvalorização desta nos sistemas educativos. Preocupados com esta questão afirmam que

(...) a desvalorização ou supressão da avaliação no sistema educativo em nada serve uma correcta e eficaz promoção social dos alunos. (...) Porque a ausência generalizada duma avaliação correcta, isto sim, é uma das grandes causas do abaixamento geral do ensino, e com isto ninguém verdadeiramente ganha (p. 265).

Roldão (1989), a propósito de uma decisão do Ministério da Educação destinada a simplificar uma ficha de registo de avaliação dos alunos do ensino básico abolindo a descrição e deixando cair "o carácter individualizado da apreciação dos alunos e a sua apreciação global (...)" (p. 9), faz uma reflexão acerca do funcionamento do sistema educativo com especial incidência nas questões relativas à avaliação das aprendizagens. A abordagem às questões do registo das informações avaliativas e das escalas de classificação utilizadas é muito rara entre nós e daí a relevância e oportunidade deste artigo. A autora chama a atenção para a distinção entre avaliação como processo de acompanhamento do ensino e da aprendizagem e avaliação como registo ou expressão de resultados e faz um conjunto de apreciações críticas quanto à forma como, na prática, se tem lidado com os registos e com as escalas de classificação. Consequentemente, critica políticas educativas "(...) em que se muda primeiro e se pensa depois. Em que se mexe muito nas formas, nos nomes, no que se vê - e se deixa quase intacto o conteúdo, indiscutido o significado, não analisados os porquês" (p. 9). Prosseguindo, a autora chama a atenção para a necessidade de se "(...) repensar a estrutura e o fun-

cionamento da Escola no sentido de tornar pedagogicamente eficazes os conselhos de turma (...)” (p. 9).

Em síntese, através dos artigos analisados verifica-se que foi dado especial realce à clarificação do conceito de avaliação formativa e à sua efectiva integração nas práticas escolares (e.g., Abrantes, 1989; Boavida *et al.*, 1986); às questões da objectividade e da subjectividade da avaliação e, implicitamente, às relações entre as funções classificativa e formativa da avaliação pedagógica (e.g., Boavida & Vaz, 1987); aos problemas de credibilidade da avaliação pedagógica, motivados pelas questões relacionadas com a falta de fiabilidade inter-correctores nos exames externos, e aos processos de atribuição de classificações (e.g., Boavida, 1985; Boavida & Vaz, 1987; Boavida *et al.*, 1986); e aos princípios a que deve obedecer uma avaliação pedagógica que possa responder aos desafios curriculares e aos desafios da aprendizagem (e.g., Abrantes, 1989; Traguil, 1988).

Os artigos dos anos 90

Os anos 90 ficaram marcados pelos esforços de reforma que se desenvolveram na sequência da publicação da Lei de Bases do Sistema Educativo em 1986 (Lei n.º 46/86). Efectivamente, após a publicação, em 1989, do Decreto-Lei n.º 286/89, que definiu a nova organização curricular dos ensinos básico e secundário, o sistema de avaliação dos alunos destes níveis de ensino foi regulado através, respectivamente, do Despacho Normativo n.º 98-A/92 e do Despacho Normativo n.º 338/93. Entre outras medidas igualmente importantes ambos os despachos consagravam orientações tais como: a) a avaliação formativa deve prevalecer nas salas de aula sobre qualquer outra modalidade de avaliação, nomeadamente a avaliação sumativa; b) a avaliação pedagógica deve fazer parte integrante do ensino e da aprendizagem; c) o processo de avaliação deve ser partilhado com os alunos, com os encarregados de educação e com outros professores e intervenientes no processo educativo; e d) a progressão dos alunos deve ser a regra e a retenção uma excepção. Ambos os despachos introduziam modalidades de avaliação externa: provas aferidas e exames nacionais no caso do ensino secundário.

Foi também nos anos 90 que se assistiu a uma expansão significativa de cursos de mestrado na área da educação tendo-se iniciado, em 1992/1993, na Universidade Católica Portuguesa, os primeiros cursos com especialização na área da avaliação em educação. Estes factos parecem ter contribuído para que a avaliação das aprendizagens tivesse uma visibilidade que até então nunca tivera. Organizaram-se múltiplos debates, surgiram artigos de toda a natureza na imprensa especializada e não especializada, desen-

volveram-se investigações e produziram-se materiais. Gerou-se uma dinâmica tão pouco habitual entre nós em torno das questões avaliativas que alguns intervenientes chegaram a considerar que se vivia o *Tempo da Avaliação* (Fernandes, 1992, 2005).

Dos 29 artigos publicados nos anos 90 apenas sete são de natureza claramente empírica. A grande maioria são reflexões teóricas, mais ou menos fundamentadas, mais ou menos profundas, havendo ainda alguns artigos que se podem considerar sínteses de literatura e outros que são análises de dados recolhidos no contexto de investigações realizadas por outros.

Na categoria *Avaliação Interna* incluíram-se seis artigos que abordam questões relacionadas com as práticas de avaliação nas salas de aula. A avaliação formativa e as práticas inerentes à sua utilização estão presentes em todos os artigos mas de forma bastante diferenciada. Nos trabalhos de Vieira (1992) e Carvalho (1992) as práticas de avaliação formativa consistem na resolução de testes cujas perguntas são geradas aleatoriamente a partir de uma base de dados acessível por computador; ou seja, a avaliação formativa parece ser vista como uma espécie de preparação para os testes sumativos feita através da resolução de baterias de exercícios por parte dos alunos num terminal de computador, mais ou menos assistidos pelos professores. Aqueles autores descrevem uma “experiência” que desenvolveram mas não discutem quaisquer concepções de avaliação formativa nem o seu papel no desenvolvimento das aprendizagens dos alunos e na melhoria do ensino. Carvalho (1992) descreve a construção da base de dados de itens e de testes referindo as suas principais características e Vieira (1992), após apresentar o projecto e discutir propriedades psicométricas dos testes, acaba por concluir que “O interesse despertado por este tipo de trabalho foi enorme, mas os resultados em termos de avaliação formativa foram diminutos” (p. 152).

Os artigos de Rocha (1990) e de Castro & Pereira (1994) também referem expressamente a avaliação formativa. No primeiro caso a autora concebe a avaliação formativa como um processo de regulação que permite a tomada de decisões a vários níveis estabelecendo a sua relação com a construção do conhecimento. Depois de elaborar sobre o “(...) que significa o ensino/aprendizagem de uma língua (...)” (p. 79), a autora refere que a avaliação formativa é fundamental para compreender como os alunos aprendem pois “(...) preocupa-se em determinar os estados de organização ou níveis de representação que os alunos construíram a partir dos quais dão um sentido a qualquer novo problema” (Rocha, 1990, p. 83). Rocha faz ainda alusão à necessidade de se entender a correcção dos erros cometidos pelos alunos como um *ajustamento interactivo* ou *controlo pragmático* e um meio para se perceber como se constrói o conhecimento.

O artigo de Castro & Pereira (1994) descreve e discute um estudo empírico em que se analisaram relatórios produzidos por alunos relativos às demonstrações experimentais realizadas por seis professores da disciplina de Física e Química. Os relatórios foram elaborados a partir de cinco demonstrações experimentais, duas no 9.º ano e três no 8.º ano, e foram objecto de análise de conteúdo. Apesar de reconhecerem que os relatórios são meios úteis para recolher informação acerca da compreensão dos alunos sobre o que observam, os professores participantes na investigação raramente os utilizaram para efeitos de avaliação. Para além de se referirem a várias relações que se podem estabelecer entre o conteúdo dos relatórios e o que se passa nas salas de aula durante os períodos de demonstrações experimentais, as autoras não discutem a avaliação formativa nem o seu papel no ensino/aprendizagem das ciências. Porém, concluem que “O relatório parece, portanto, ser útil como meio de avaliação formativa” (p. 174).

Leal & Abrantes (1990) ao descreverem e discutirem as características pedagógicas de um projecto de inovação curricular no contexto da disciplina de Matemática no 3.º ciclo do ensino básico, chamam a atenção para a necessidade de a avaliação se articular com o desenvolvimento do currículo e da aprendizagem. Trata-se de um artigo em que é possível discernir uma concepção de avaliação eminentemente formativa, muito associada à didáctica. Os autores afirmam ser “(...) indispensável conceber e praticar formas de avaliação do trabalho dos alunos compatíveis com os objectivos, as orientações metodológicas e a natureza das actividades do currículo” (p. 67). Por isso, enunciam alguns princípios tais como a consistência da avaliação com os objectivos e as metodologias seguidas, a ênfase que deve ser dada à avaliação dos processos ou a natureza positiva da avaliação. Além disso, descrevem formas alternativas de recolha de informação tais como o teste em duas fases, os ensaios e relatórios escritos e o trabalho de projecto.

Num artigo de natureza reflexiva e, de certo modo, prescritiva, Carvalho (1994) conceptualiza a *avaliação inicial* prevista no currículo de Educação Física e a sua integração num projecto pedagógico consistente. A autora diferencia a avaliação formativa contínua, interactiva e informal da avaliação formativa pontual e formal que é uma distinção pouco discutida na literatura. Além disso, apresenta e discute as diferentes fases da avaliação formativa, os momentos em que se poderão desenvolver avaliações formativas formais, os objectos de avaliação, os procedimentos a utilizar e quem deverá participar no processo de avaliação. A autora faz referência à questão dilemática da objectividade e da subjectividade da avaliação: “Assumir a subjectividade não significa renunciar à objectividade e rigor da avaliação. Não equivale a deixar de preparar e planificar cuidadosamente a avaliação” (Carvalho, 1994, p. 157). Mais uma vez se aborda

esta questão epistemológica que tem a maior relevância em termos da organização do processo de avaliação.

Na categoria *Avaliação Externa* foram analisados três artigos. Um deles faz a análise dos resultados dos alunos dos distritos de Braga e de Viana do Castelo na prova específica de acesso ao ensino superior de Física realizada em 1994 (Ferreira, Almeida, Cunha, Rolo e Vaz, 1995). Trata-se de um trabalho em que, para além da apresentação e discussão dos resultados dos alunos em cada uma das seis questões da prova, se fazem algumas apreciações e recomendações decorrentes dos fracos resultados obtidos pelos alunos. Um pouco na mesma linha, o artigo de Ramalho (1995) apresenta os resultados de alunos portugueses de nove e de 13 anos de idade na componente de Matemática do *Second International Assessment of Educational Progress*, que se realizou entre 1989 e 1992, e que envolveu 20 países. Ainda na mesma linha, Amaro (1997) descreve um projecto centrado na avaliação externa das aprendizagens dos alunos do ensino básico (4.º, 6.º e 9.º anos de escolaridade), em Matemática e em Língua Portuguesa, que decorreu da necessidade de se desenvolverem as provas aferidas previstas nos normativos legais em vigor desde 1992. O artigo centra-se essencialmente na apresentação dos procedimentos adoptados na recolha da informação e dos resultados obtidos pelos alunos.

No que se refere à avaliação externa não se identificaram artigos em que se discutam questões de avaliação propriamente ditas tais como a questão da fiabilidade inter-correctores, as características psicométricas das provas, a sua validade pedagógica, a sua fiabilidade, a sua natureza, o *poder* avaliativo dos itens, os critérios de correcção ou a relação entre a avaliação externa e a avaliação interna.

Na categoria *Métodos e Técnicas* foram considerados oito artigos que, no essencial, se centram na construção e/ou validação de instrumentos ou abordam questões técnicas de avaliação. Curiosamente é esta a categoria com um maior número de artigos o que poderá indiciar a preocupação dos autores em encontrar procedimentos e instrumentos que permitam *medir* (?) ou *avaliar* (?) as aprendizagens dos alunos. Metade dos artigos refere-se à apresentação, construção e/ou validação de testes de diversa natureza (Carvalho, 1990; Henriques, 1997; Leitão, 1993; Oliveira & Pereira, 1993).

Um artigo descreve a construção de um instrumento destinado a avaliar a tarefa de resumir (Simão, 1992), outro faz a análise dos exames nacionais de Inglês de 1996 (Carvalho, 1997) e outro descreve e discute a construção de um quadro de referência para a elaboração de provas aferidas de Inglês (Carvalho, 1999). Finalmente, há um artigo que

faz uma discussão acerca dos aspectos de natureza técnica que se devem ter em conta, particularmente na elaboração e utilização de exames e de testes (Pedrosa, 1991).

Oliveira & Pereira (1993) discutem a concepção e desenvolvimento de um teste de avaliação das aprendizagens de alunos do 6.º ano de escolaridade na disciplina de Matemática, num contexto abrangente e integrado de avaliação pedagógica. Para além de referirem princípios tais como o da participação dos alunos no processo de avaliação, o da criteriosa utilização de *feedback* pelo professor ou o da relevância de nos centrarmos no desenvolvimento dos processos de pensamento mais elaborados dos alunos, as autoras discutem os papéis e as relações entre a avaliação formativa e a avaliação sumativa. A perspectiva das autoras parece ser a de considerarem que a avaliação sumativa e a avaliação formativa são dimensões distintas de um mesmo conceito e não uma dicotomia. Na verdade, chamam a atenção para “(...) a necessidade de estabelecer um contínuo entre a avaliação formativa e a sumativa, sendo suposto que esta mede não só conhecimentos e capacidades através de testes escritos mas também saberes práticos, valores e atitudes (...)” (p. 114). Trata-se de uma questão que continua a merecer discussão e que necessita de clarificação (Fernandes, 2006; Harlen, 2005, 2006). Oliveira & Pereira fazem uma detalhada descrição dos processos utilizados para a elaboração do teste discutindo questões técnicas e pedagógicas tais como a natureza dos itens, as grelhas de classificação (analítica e holística focada), a natureza criterial do instrumento e a análise psicométrica e pedagógica dos itens.

O artigo de Henriques (1997), centrado na construção de um teste para o 1.º ciclo do ensino básico no domínio da Matemática, também faz um conjunto de considerações do domínio da avaliação pedagógica na linha do que se constatou no artigo de Oliveira & Pereira (1993). Leitão (1993) limita-se praticamente a apresentar um teste de Química do 12.º ano e Carvalho (1990) a descrever procedimentos e a analisar os dados obtidos na sequência do processo de validação de um teste de compreensão de leitura junto de crianças de 6 e 7 anos de idade.

No artigo de Simão (1992) relaciona-se a conceptualização teórica da tarefa de resumir com o que deverão ser os procedimentos a seguir na sua avaliação. Partindo da ideia de que “A tarefa de resumir um texto, ao pôr em evidência o uso de estratégias, permite identificar as principais dificuldades que o leitor tem para compreender o que lê” (p. 253), a autora desenvolve uma grelha organizada em três blocos a que correspondem três critérios considerados fundamentais para a avaliação da tarefa em análise: fidelidade, selecção e coerência.

Em que medida é que grelhas relativamente simples, mas não simplistas, bem enquadradas teórica e conceptualmente, poderão constituir bons meios de avaliação de tarefas que exijam a mobilização de processos complexos de pensamento? Trata-se de uma questão que o artigo de Simão suscita e que a investigação na área da avaliação das aprendizagens está ainda longe de ter respondido. Na verdade, continua a ser necessário investigar as potencialidades de avaliação de métodos que vão para além dos testes.

Num artigo baseado numa perspectiva técnica da avaliação, Pedrosa (1991) faz uma analogia entre duas medidas de grandezas físicas, exactidão e precisão, e duas medidas educacionais, validade e fiabilidade. A partir desta comparação elabora sobre os diferentes tipos de validade e sobre a fiabilidade referindo que “Existe sempre um erro inerente ao acto de medir, quer em medidas de grandezas físicas quer nas classificações dos estudantes em testes (...)” (p. 83). Consequentemente, a autora identifica e discute as fontes de erro na classificação dos testes e faz recomendações para melhorar a sua fiabilidade. A avaliação é identificada com a medição e, nesta perspectiva, o artigo centra-se nas questões técnicas relativas aos instrumentos, procurando que eles permitam obter medidas tão exactas quanto possível do que os alunos sabem. Daí a discussão que é feita sobre os índices psicométricos e a necessidade de os melhorar.

No que se refere à categoria *Concepções e Práticas de Professores/Alunos* foram analisados quatro artigos (Caria, 1994a; Duarte, 1994; Paulo & Santos, 1998; Pereira, 1997).

Caria (1994a) organizou a secção *Diálogos Sobre o Vivido*, do primeiro número da revista *Educação, Sociedade & Culturas*, em torno do *Sistema de Avaliação dos Alunos do Ensino Básico* previsto no Despacho Normativo n.º 98-A/92. Trata-se de uma secção que contém *Registos Etnográficos*, baseados num conjunto de diálogos induzidos pelas medidas previstas naquele normativo legal, que permitem discernir concepções e práticas de professores em relação à avaliação das aprendizagens dos alunos. A secção termina com depoimentos reflexivos de Telmo Caria e de Almerindo Afonso (Afonso, 1994; Caria, 1994b).

Julga-se desnecessário dar mais pormenores acerca das orientações constantes naquele despacho para além das que acima já se indicaram pois estão amplamente divulgadas através de várias publicações (e.g., Barreira & Boavida, 1992, 1993).

A complexa relação entre as concepções e as práticas de professores tem sido muito estudada e está caracterizada na literatura (e.g., Boavida, 1996; Campos, 1996; Fernandes & Vale, 1994; Pajares, 1992; Thompson, 1992). O trabalho de Caria (1994a) ilustra

concepções de professores que parecem estar relacionadas com as dificuldades em alterar e melhorar as práticas de avaliação. Vejamos um exemplo de um dos diálogos.

2.º Diálogo

Conversa entre dois professores em meados de Setembro de 1992:

1.ª Prof.ª - Vê lá tu que há bocado disseram-me que já não era preciso fazer testes. Eu fiquei pasmada.

2.ª Prof.ª - Sim, eu também já ouvi dizer que agora como todos passam já não faz sentido haver testes.

1.ª Prof.ª - Mas isso não faz sentido. Como é que depois se faz prova se houver algum protesto? É sempre necessário ter documentos.

2.ª Prof.ª - Talvez isso tenha a ver com o facto de a avaliação agora ser predominantemente formativa.

1.ª Prof.ª - Ah, pois é! Agora chamam formativa àquela que era sumativa e a gente fica sem saber para onde vai a sumativa (Caria, 1994a, pp. 131-132).

Caria (1994b) refere que as concepções que este e outros diálogos indiciam resultam do facto de a interpretação das medidas previstas para a avaliação das aprendizagens "(...) não ser realizada no contexto do seu conteúdo e finalidades, mas sim no contexto da estrutura da interacção social e das estratégias práticas do agir que constituem a cultura do professor" (Caria, 1994b, p. 159). A verdade é que nem os testes acabaram, nem a avaliação passou a ser predominantemente formativa nem os alunos começaram a passar todos. No ano lectivo de 2003/2004, passados onze anos da vigência do *novo sistema de avaliação*, reprovaram e abandonaram a escola cerca de 130 000 alunos nos três ciclos do ensino básico, correspondendo a cerca de 12% do número total de alunos inscritos nos percursos escolares ditos regulares (GIASE, 2005a, 2005b). Portugal continua a ser o país da Europa onde mais alunos reprovam.

Caria (1994b) refere que a *fala* dos professores acaba por revelar uma tensão e uma desarticulação entre o conhecimento cientificamente informado e o conhecimento que decorre da experiência e que é muito baseado nas solidariedades inter-pessoais. Talvez este último acabe por prevalecer e constitua, pelo menos em parte, um obstáculo à mudança e à melhoria das práticas.

As concepções reveladas pelos professores acerca da avaliação formativa, a essência do *novo sistema de avaliação*, talvez ajudem a compreender as persistentes dificuldades em levá-la efectivamente à prática. De facto, Caria refere que, quanto à avaliação formativa, "(...) os professores afirmam optar por valorizar o que é positivo e ignorar, no desempenho dos alunos, o que é negativo" (p. 163) e interroga-se "É esta a avaliação formativa possível nas actuais condições das práticas dos professores?" (p. 163).

No seu comentário Afonso (1994) reconhece virtualidades no sistema de avaliação considerando que “há aspectos que devem mudar e que os professores não podem mudar e aspectos que poderão mudar se os professores quiserem” (p. 176). A propósito desta última possibilidade o autor refere que as concepções dos professores acerca da avaliação formativa e da avaliação sumativa e ainda as práticas burocráticas dos conselhos de turma podem dificultar ou facilitar as necessárias mudanças nas práticas de avaliação. A questão, mais uma vez presente, da objectividade e da subjectividade da avaliação é equacionada por Afonso da seguinte forma:

Há algumas crenças e representações em torno da avaliação que se traduzem, por exemplo, na consideração da avaliação formativa como uma modalidade de avaliação qualitativa, subjectiva, intuitiva e na consideração da avaliação sumativa como uma modalidade de avaliação quantitativa e objectiva. Ora, não há razão (científica ou técnica, pelo menos) que justifique os epítetos que frequentemente se associam a estas modalidades de avaliação. Por um lado, não há nenhuma modalidade de avaliação totalmente objectiva, e, por outro, é possível (e os professores sabem isso por experiência) praticar de forma extremamente rigorosa a avaliação formativa e de uma forma muito pouco rigorosa a avaliação sumativa (Afonso, 1994, p. 176).

Afonso discute a tensão entre as visões objectivas e subjectivas da avaliação das aprendizagens, chamando a atenção para o papel que os conselhos de turma podem ter na validação inter-subjectiva dos juízos avaliativos dos professores.

Também Duarte (1994), com professores de várias disciplinas do 2.º ciclo do ensino básico, Pereira (1997), com duas educadoras de infância e Paulo & Santos (1998), com três futuros professores de Educação Visual e Tecnológica, desenvolveram estudos empíricos em que se investigaram concepções e/ou práticas dos docentes participantes. Da análise destes artigos verifica-se que os professores tendem a ver as suas práticas de avaliação e as dos outros de forma bastante diferente, referindo, por exemplo, que enquanto os outros dão ênfase à aquisição de conhecimentos, eles privilegiam a progressão na aprendizagem (Duarte, 1994). Outro resultado dos estudos de Duarte (1994) e de Pereira (1997) refere-se ao facto de os professores desenvolverem uma avaliação intuitiva que tem em conta finalidades mas que não se baseia na definição de objectivos e de critérios precisos.

A investigação de Paulo & Santos (1998) desenvolveu-se num contexto em que a investigação-acção, com frequente recurso à reflexão e ao trabalho cooperativo, foi assumida como estratégia de formação inicial. As investigadoras concluíram que houve uma notória evolução nas concepções e práticas de avaliação dos futuros professores nomeadamente na clarificação dos critérios e finalidades de avaliação e no envolvi-

mento dos alunos nas tarefas e na discussão dos resultados. Os participantes privilegiaram a função formativa da avaliação embora aquelas investigadoras considerem que há um campo comum partilhado pela avaliação formativa e pela avaliação sumativa no sentido da melhoria das aprendizagens dos alunos. Trata-se de uma alusão às relações entre aquelas duas modalidades de avaliação que, como já se referiu acima continua a suscitar reflexão e a necessitar de clarificação (e.g., Harlen, 2006; Harlen & James, 1996; Fernandes, 2006).

Na categoria *Reflexão/Análise de Literatura* foram analisados três artigos (Benavente, 1990; Fernandes, 1999; Gonçalves, 1994). Apesar de nenhum deles se centrar especialmente no domínio da avaliação das aprendizagens, a sua selecção resulta do facto de poderem contribuir para a construção dos seus fundamentos a partir do debate paradigmático (Fernandes, 1999), para o desenvolvimento de uma visão mais alargada e sistémica da avaliação como processo social e pedagógico associado à inovação (Benavente, 1990) e para estabelecer relações entre os processos de avaliação e os processos de investigação educacional tendo em conta a sua evolução nos últimos 40 anos (Gonçalves, 1994).

Finalmente, na categoria *Reacção a Normativos da Administração*, foram seleccionados e analisados cinco artigos. Repare-se que é uma das categorias com maior número de artigos depois das *Métodos e Técnicas* (com oito) e da *Avaliação Interna* (com seis) o que se pode explicar pelo grande impacto que os normativos legais de avaliação acima referidos tiveram na sociedade portuguesa e, em particular, nos meios pedagógicos e educacionais. Repare-se ainda que o trabalho de Caria (1994a) também se poderia incluir nesta categoria pois todo ele está baseado nas reacções de professores ao Despacho Normativo n.º 98-A/92. Apesar disso, considerou-se que a recolha empírica de dados junto de professores lhe confere uma natureza diferente dos outros artigos permitindo formular reflexões acerca das concepções e práticas dos professores no domínio da avaliação das aprendizagens.

Boavida & Barreira (1992) analisam o Despacho n.º 162/ME/91, que na prática não chegou a ser aplicado, e sobretudo o Despacho Normativo n.º 98-A/92. Trata-se de um artigo que comenta e enquadra teoricamente este normativo legal, em que basicamente se consideram positivas as medidas propostas, muito particularmente no que se refere à avaliação formativa e ao papel que esta poderá desempenhar na melhoria das aprendizagens de todos os alunos da escolaridade obrigatória. Contudo, formulam-se objecções que, no essencial, têm a ver com as condições existentes no sistema educativo pois “(...) legisla-se e manda-se executar sem cuidar verdadeiramente das condições para que a lei se cumpra com proveito geral” (p. 361). Por outro lado, teme-se a

eventualidade de se caminhar para a degradação pedagógica se as medidas preconizadas não apoiarem os alunos a vencer as suas dificuldades. E teme-se ainda a existência prática de *dois níveis de sucesso*: o das escolas com bons professores e o das escolas com professores que não são bons e que baixam os níveis de exigência. Apesar de tudo, os autores reafirmam a necessidade de se mudar as práticas de avaliação mas também as atitudes dos principais intervenientes no processo pedagógico: "Vai pôr-se em prática um tipo de avaliação que é de algum modo revolucionário e que implica comportamentos e atitudes diferentes tanto da parte dos alunos como dos professores" (p. 362).

Num outro artigo, Boavida & Barreira (1993) voltam a analisar o referido despacho elaborando, a propósito, uma reflexão teórica sobre questões pertinentes da avaliação das aprendizagens. Em particular, produzem uma discussão em torno da avaliação formativa e da avaliação sumativa focando as suas funções e as suas relações. Partindo do princípio de que "A verdadeira função da avaliação é a de informar o aluno sobre a sua evolução, as suas deficiências, o que é necessário corrigir; a real função da avaliação é ajudar cada um a progredir efectivamente" (p. 101), os autores consideram que "(...) todas as formas de avaliação podem (e devem) ser formativas (...)" (p. 101). Consequentemente, "(...) a avaliação sumativa pode também desempenhar um papel crucial no aperfeiçoamento da prática educativa" (p. 100). Estamos assim perante a discussão do valor pedagógico da avaliação sumativa e, mais geralmente, de uma concepção de avaliação como processo que serve para melhorar, associado ao progresso e ao desenvolvimento de todos e de cada um dos alunos.

Pacheco (1993) também faz uma análise comentada daquele normativo acabando por se referir a um "decálogo curricular da avaliação formativa" (p. 11) a partir do qual se questionam "(...) pressupostos, orientações, concepções, processos e práticas curriculares (...)" (p. 11). Ou seja, Pacheco defende que se a avaliação formativa se tornar a modalidade predominante no sistema educativo devem daí retirar-se as devidas consequências curriculares. Relativamente às relações entre a avaliação formativa e a avaliação sumativa a posição deste autor é clara "(...) a avaliação sumativa é uma modalidade baseada na avaliação formativa que lhe serve de referência e que a precede" (p.18).

O artigo de Marques (1997) também faz uma análise comentada do Despacho 98-A/92 referindo-se à relevância das provas globais no 9.º ano de escolaridade nos seguintes termos: "A introdução das provas globais (...) pode acrescentar credibilidade ao sistema e à escola e pode ser um importante factor extrínseco de motivação para o estudo. (...) podem constituir um instrumento de promoção da qualidade da escolaridade básica (...)" (p. 13). A questão das provas globais é pertinente e tem sido muito pouco estudada entre nós. Trata-se de uma avaliação interna que pode ter um papel relevante na

moderação da avaliação feita pelos diferentes professores e na forma como o currículo é desenvolvido. É uma área a necessitar de investigação empírica que clarifique o real papel que desempenha na vida escolar de professores e alunos.

Peixoto (1998) também parte do Despacho 98-A/92 para fazer uma reflexão acerca do papel da avaliação no sistema educativo, particularmente no 1.º ciclo, e na sociedade. A autora faz sistematicamente o contraste entre a “avaliação tradicional” e a avaliação formativa procurando enquadrar e contextualizar as oportunidades e os obstáculos às mudanças necessárias. Nas sistemáticas referências à avaliação formativa faz alusão à sua natureza privada em contraste com a natureza pública da avaliação sumativa “É sobretudo através da avaliação formativa (...) que se detectam os problemas e se tenta ultrapassá-los. E tudo deve acontecer na intimidade da relação educativa (...)” (p. 202). É uma questão que não tem merecido atenção especial dos investigadores e que será interessante analisar à luz do que se consideram ser as funções da avaliação formativa e da avaliação sumativa.

Os artigos da primeira metade da presente década

Numa primeira análise aos artigos publicados no primeiro lustro desta década verificam-se alguns factos que, pelo seu significado, se destacam desde já:

1. Foram identificados 22 artigos o que representa o dobro dos que se identificaram no lustro anterior.
2. Metade dos artigos apresenta e discute investigação de natureza empírica. Este facto pode constituir uma indicação positiva no que se refere ao desenvolvimento de investigação no domínio da avaliação das aprendizagens e contraria o que se observou nos 15 anos precedentes.
3. Em relação à década anterior mantém-se bastante semelhante a distribuição dos artigos pelas categorias genéricas definidas. No entanto, parece haver uma tendência consistente na diminuição do número de artigos enquadráveis na categoria *Reacção a Normativos da Administração*.

Os quatro artigos incluídos na categoria *Avaliação Interna* abrangem temas tais como a discussão teórica acerca da natureza das avaliações de natureza formativa e as suas implicações para as práticas de avaliação pedagógica (Barreira, 2001, 2005; Simão, 2005) e a análise dos processos utilizados por professores quando classificam os seus alunos (Pais, 2000).

Barreira (2005) identificou dificuldades dos professores com as práticas sugeridas no sistema de avaliação do ensino básico, referindo que a avaliação sumativa continua a predominar em relação à avaliação formativa e que esta é de natureza pontual e retro-activa e muito pouco utilizada. É ainda mencionada a dificuldade dos professores em discernirem as duas modalidades de avaliação e as relações entre elas. Por estas razões Barreira sublinha que é "(...) imprescindível que os professores aprofundem a prática de avaliação formativa contínua e interactiva (...) pois só interagindo de perto com os alunos poderão ajudá-los a colmatar as suas dificuldades e a desenvolver as suas potencialidades" (p. 139). O autor discute eventuais medidas que poderão contribuir para melhorar as práticas de avaliação formativa tais como as adaptações curriculares, a formação de professores e a utilização de estratégias de trabalho com pequenos grupos de alunos como alternativa ao trabalho exclusivo com o grupo turma.

Num outro artigo de marcada índole pedagógica Barreira (2001) discute a avaliação formadora e a avaliação autêntica como duas estratégias complementares de avaliação das aprendizagens. A primeira implica uma regulação das aprendizagens feita pelo próprio aluno em vez de pelo professor e, por isso, a auto-avaliação, a auto-correcção, a auto-regulação e o auto-controlo são processos que lhe são inerentes. A avaliação autêntica é associada à avaliação do desempenho dos alunos quando estão envolvidos na resolução de tarefas *autênticas* tais como "(...) a resolução de problemas, o desenvolvimento de uma investigação, de uma discussão ou de uma composição" (Barreira, 2001, p. 9). Depois de discutir a importância e as características da observação no contexto da avaliação autêntica, este autor destaca o papel da auto-avaliação como processo integrado no ensino e aprendizagem, formativo, que está para além da mera auto-correcção e que é "(...) uma componente essencial na regulação do processo de aprendizagem" (p. 18).

Simão (2005) parte da aprendizagem auto-regulada para considerar que a avaliação das aprendizagens tem um duplo valor regulador: formativo no que se refere ao ensino e formador no que se refere à aprendizagem. Tal como Barreira (2001) esta autora destaca a relevância do processo de auto-avaliação e também sugere o portfolio como "(...) instrumento de aprendizagem e avaliação" (p. 266).

Os artigos de Barreira (2001) e Simão (2005) suscitam questões de interesse relacionadas com a construção teórica no domínio da avaliação formativa tais como: a) De que estamos realmente a falar quando falamos de avaliação formativa?; b) Em que medida é que se poderá dizer que a avaliação formadora é formativa?; c) E a avaliação formativa será formadora? E autêntica?; d) Será possível encontrar um conceito suficientemente robusto que dê significado e sentido a todo o tipo de avaliações formativas

que vão surgindo na literatura?; e e) Que papéis desempenham professores, alunos e tarefas no processo de avaliação formativa? Estas e outras questões devem ser objecto de análise e reflexão pois o que está em causa é, como se referiu, a necessidade de se construir uma teoria sólida da avaliação formativa (Black & Wiliam, 2006; Fernandes, 2006; Gipps, 1994).

O artigo de Pais (2000) é pioneiro entre nós pois descreve e discute uma investigação empírica em que se analisaram processos de classificação utilizados por professores do ensino secundário. É um trabalho que também suscita várias questões de interesse para a investigação em avaliação das aprendizagens. Na verdade, o autor mostra-nos que a classificação não é uma mera questão técnica, mas sim um processo imbuído de valores, uma vez que os professores não deixam de ter em conta os significados e as consequências das classificações que atribuem. Consequentemente as classificações não estão livres dos conhecimentos, das concepções e dos valores sustentados pelos professores. Será interessante continuar a aprofundar as conclusões da investigação de Pais, nomeadamente no sentido de procurar caracterizar detalhadamente o real significado das classificações que se atribuem. Evidentemente que a questão da *construção* da classificação, ou seja, a forma como os professores determinam as classificações permanece largamente ignorada e, por isso, é um desafio a enfrentar.

No que se refere à *Avaliação Externa* Martins & Silva (2000) enunciam alguns dos problemas inerentes aos exames nacionais tais como a fiabilidade das classificações, a sua influência, por vezes nefasta, no desenvolvimento do ensino e das aprendizagens dos alunos, os seus índices de dificuldade, os erros de elaboração, a falta de critérios nas correcções. Fazem ainda referência aos problemas de articulação entre o ensino secundário e o superior. O que é interessante é que estes autores ilustram a sua discussão com base em documentos produzidos nos anos 40 e nos anos 60 do século passado pela Sociedade Portuguesa de Matemática e por dois eminentes matemáticos e pedagogos como foram Bento de Jesus Caraça e Sebastião e Silva. A propósito dos exames e do sistema educativo português dizia Sebastião e Silva em 1968:

(...) reduziram o ensino à preparação em massa para o exame, e, portanto, à degradação e à mecanização dos processos. (...) além de ficarem pelo caminho cerca de 80% dos alunos, consta que há numerosos casos de esgotamento! (...) estamos em presença de um sistema educacional que não ensina a observar nem a experimentar, nem a reflectir, nem a raciocinar, nem a escrever, nem a falar: ensina apenas a repetir mecanicamente, a imitar e, por conseguinte, a não ter personalidade. (...) um sistema feito à medida da mediocridade obediência, que acerta o passo enquadrada em legiões de explicadores. É, portanto, um ensino em regime de desdobramento: professor-explicador (e o mais grave

é que o professor já conta com o explicador). (...) Depois na Universidade, o drama atinge o ápice. (...) Em certas cadeiras, a percentagem de reprovações atinge 90%. (...) verifica-se um divórcio quase total entre o ensino secundário e o ensino universitário (...) (Sebastião e Silva, citado em Martins & Silva, 2000, p. 53).

É a visão de Sebastião e Silva há cerca de 40 anos atrás que Martins & Silva evocam para nos referirem que há problemas que persistem e que não são simples de resolver. É certo que, para efeitos da certificação no ensino secundário, há hoje um maior peso da avaliação interna (70%) em relação à avaliação externa (30%). Em 1968 os exames tinham um peso de 100% e a avaliação interna só servia como filtro para acesso ao exame e não tinha qualquer peso na classificação certificativa do aluno. Mas hoje os mesmos problemas persistem, são da mesma natureza e, por isso, como dizem os autores, é necessário melhorar a qualidade técnica dos exames, articular o que neles se pergunta com o que os alunos efectivamente tiveram oportunidade para aprender e "(...) por fim é indispensável a intensificação do diálogo entre os professores do ensino secundário e os professores do ensino superior." (Martins & Silva, 2000, p. 54).

Martins & Silva discutem ainda o dilema sempre presente entre o papel moderador e regulador que os exames poderão ter e a sua abolição pura e simples. Há argumentos a favor e contra cada uma destas visões. Um deles, favorável à abolição, baseia-se nos nefastos efeitos de *backwash* dos exames que, por exemplo, levam os professores a ignorar as metodologias mais adequadas para a aprendizagem dos alunos, limitando-se a treiná-los e mecanizá-los nos procedimentos de resolução do tipo de questões que sai no exame, como referia Sebastião e Silva. Trata-se do *estreitamento* ou *afunilamento* do currículo. O outro, favorável à manutenção, baseia-se em evidências que mostram a tendência dos professores para dedicarem mais tempo aos assuntos de que mais gostam ou que conhecem melhor, menorizando outros igualmente importantes, ficando assim o *currículo ensinado* aquém do que seria expectável e desejável. É uma outra forma de *estreitamento* do currículo, dizem os defensores da manutenção.

Os autores formulam ainda uma outra questão relevante: Qual deverá ser a natureza dos exames? Trata-se de uma questão ainda não resolvida mesmo em termos internacionais. Os exames devem ter como referente o que está prescrito nos programas? Devem avaliar as aplicações práticas que cada professor faz do programa? Devem basear-se num programa essencial a cumprir rigorosamente por todas as escolas? Devem basear-se em provas modelo? Que tipo de equilíbrio deverá existir entre os diferentes tipos de itens possíveis?

O artigo de Martins & Silva (2000) acaba por constituir uma síntese de alguns dos problemas de investigação que existem no domínio da avaliação externa, área que, no nosso país, tem estado bastante abandonada e sem produção de conhecimento.

Eça (2004) baseia-se num quadro conceptual para avaliação de sistemas de avaliação externa que inclui elementos tais como a validade e a fiabilidade das provas, os instrumentos de avaliação e os processos de correcção e de classificação utilizados. O seu objectivo é comparar os sistemas de avaliação externa no domínio das artes visuais existentes em Portugal e em Inglaterra. Com base na sua investigação a autora tece um conjunto de críticas fundamentadas aos exames de artes visuais portugueses questionando a sua validade de conteúdo e a sua validade pedagógica, assim como a sua fiabilidade e a consistência das correcções (fiabilidade inter-correctores), a falta de definição clara dos domínios de aprendizagem e a ambiguidade dos critérios de correcção. São problemas pertinentes que têm que ser encarados e que a autora sistematiza a partir de um inquérito realizado junto de 40 professores e 100 ex-examinandos.

O processo de avaliação externa das aprendizagens de artes visuais que tem lugar em Inglaterra, incluindo uma moderação interna das classificações e, posteriormente, uma moderação externa por parte de um ou mais correctores externos, abre um plano de discussão e de investigação que não se deverá desprezar. Em questões de avaliação externa precisamos de inovar e de criar condições para que ela se possa vir a tornar um processo com real valor pedagógico e educativo.

O artigo da Gazeta da Matemática (2004) baseia-se nas opiniões de alguns professores e, no essencial, acaba por enunciar vantagens e desvantagens dos exames. Lobo (2001) faz uma análise de uma prova modelo de Química e produz um conjunto de recomendações específicas no sentido de se aproximar a prova da realidade vivida pelos alunos e criticando a utilização pouco criteriosa dos itens de escolha múltipla. No fundo a discussão desenvolvida por este autor tem a ver com a natureza dos exames, uma questão que, como acima se referiu, é da maior relevância e tem de ser objecto de investigação e de reflexão devidamente fundamentada.

No domínio das *Métodos e Técnicas* Nabuco (2000), no seu artigo, considera a observação como uma técnica a privilegiar no contexto da educação pré-escolar porque é através dela que "(...) os educadores avaliam as suas necessidades, dão consistência às suas experiências, facilitam a sua aprendizagem, mudam as práticas educativas e reestruturam os currículos" (p. 84). Ou seja, Nabuco parece reconhecer um papel regulador e formativo à observação. Depois de fazer considerações e recomendações acerca da melhor forma de observar (e.g., método da narrativa, método da amostra), a

autora discute dois instrumentos de avaliação do desenvolvimento e envolvimento das crianças.

O artigo de Vasconcelos, Praia & Almeida (2004) centra-se essencialmente no processo de construção e de validação de uma escala de 28 itens destinada a avaliar métodos de estudo utilizados por alunos do ensino básico no contexto da área curricular não disciplinar do *Estudo Acompanhado*. A escala foi construída tendo em vista o estudo das Ciências Naturais por parte de alunos dos 7.º e 8.º anos de escolaridade e incluía três dimensões: a) Organização e compreensão da informação em Ciências Naturais; b) Esforço e persistência nas tarefas de Ciências Naturais; e c) Ansiedade em situação de teste de Ciências Naturais. Face aos resultados encontrados quanto à consistência da escala os autores consideram que esta "(...) serve aos professores como instrumento de diagnóstico das estratégias de aprendizagem dos alunos (...)" (p. 45). Na mesma linha o artigo de Vasconcelos (2002) discute a validação de uma escala com os mesmos fins.

Estes artigos suscitam algumas questões de interesse que seria importante investigar e estudar. De que formas é que a avaliação dos métodos de estudo poderá contribuir para o seu desenvolvimento e para a melhoria das aprendizagens dos alunos? Como é que professores e alunos utilizam as escalas desenvolvidas pelos investigadores? Ou ainda: Como é que se poderá descrever a utilidade prática deste tipo de instrumentos? Que questões se poderão formular relativamente à exequibilidade da sua utilização?

Abreu, Nunes & Rosa (2004) descrevem os procedimentos utilizados na validação de um teste de avaliação da escrita destinado a alunos do 2.º ao 4.º ano de escolaridade. Com base nos resultados obtidos através da administração do teste a cerca de 600 crianças, as autoras concluem que a fonologia e a morfologia são duas fontes relevantes para o conhecimento da escrita e que o teste reflecte bem as competências dos alunos nos domínios para que foi construído.

Do artigo de Menino (2005), que descreve e discute uma investigação empírica em que se analisou a utilização do portfolio por parte de alunos do 6.º ano de escolaridade na aula de Matemática, importa salientar dois aspectos que parecem relevantes e que devem continuar a merecer a atenção dos investigadores. Um deles prende-se com o facto de parecer importante clarificar o conceito de portfolio e, em função desse trabalho, retirar as necessárias ilações pedagógicas em termos da organização e do funcionamento do trabalho na sala de aula. Esta questão é relevante e também passa por clarificarmos se o portfolio é um *instrumento*, um *método*, uma *estratégia* ou uma *tarefa* de avaliação. Trata-se de um problema teórico e epistemológico e não meramente

semântico (Fernandes, 2005, 2006). O outro aspecto tem a ver com o facto de Menino ter identificado um conjunto de dificuldades reveladas pelas crianças do 6.º ano quando utilizaram os portefólios. Ora parece relevante continuar a caracterizar e a categorizar exaustivamente as dificuldades dos alunos e a sua natureza mas também me parece que deveremos procurar fazer o mesmo relativamente aos professores. Só desta forma poderemos construir em bases sólidas formas alternativas de avaliação, mais consistentes com o que sabemos acerca do desenvolvimento do currículo, da aprendizagem e da avaliação.

Relativamente à categoria *Concepções e Práticas de Professores/Alunos* foram identificados e analisados cinco artigos. No artigo de Rosado & Dias (2002) descreve-se uma investigação cujo principal objectivo era o de investigar as concepções e práticas de professores de Educação Física acerca da avaliação formativa. Os dados foram obtidos através da administração de um questionário a 94 professores. Os autores referem que os professores avaliam os seus alunos nos domínios motor, sócio-afectivo e cognitivo e as conclusões parecem evidenciar que a avaliação realizada é mais baseada em concepções que estarão mais próximas de uma visão certificativa ou sumativa do que de uma visão formativa da avaliação. Porém, no que se refere à avaliação formativa, o artigo não chega a explicitar e a elaborar sobre as concepções dos professores.

Graça (2003) descreve e discute uma investigação destinada a analisar as relações entre as concepções e as práticas de avaliação de três professores de Matemática em contextos de resolução de problemas. A autora refere que verificou existir consistência entre as concepções e práticas dos professores relativamente à resolução de problemas mas inconsistência no que se refere às concepções e práticas dos professores relativamente à avaliação da resolução de problemas. Quanto às práticas de avaliação dos professores refere que são pouco diversificadas e sobretudo de natureza informal.

Um pouco na mesma linha dos artigos anteriores Oliveira & Gaspar (2004) também descrevem uma investigação empírica destinada a analisar as concepções e práticas de 96 educadores de infância através de um inquérito por questionário. Foram consideradas e relacionadas as seguintes variáveis: a opinião dos educadores sobre avaliação, práticas de avaliação, interesse na mudança das práticas de avaliação, anos de serviço, escola de formação, instituição onde exercem a sua actividade profissional e mobilidade docente. Das conclusões elaboradas pelas autoras destacam-se as seguintes: a) "(...) as educadoras que constituem a amostra têm opiniões positivas sobre a avaliação na educação pré-escolar" (p. 470).; b) "(...) a mobilidade docente poderá influenciar a opinião das educadoras relativamente à avaliação" (p. 471).; c) "(...) são as educadoras que se formaram nas ESE's, com menos tempo de serviço e a exercerem funções em

instituições privadas que têm opiniões mais positivas sobre as funções da avaliação" (p. 472); d) "(...) todas as dimensões consideradas são avaliadas pelas educadoras (...)" (p. 474); e e) "(...) a observação é o procedimento mais referido (...)" (p. 474).

Fernandes, Alves & Arantes (2004), descrevem e discutem uma investigação que analisou a relação entre a formação inicial nas concepções e práticas de avaliação de dois professores estagiários de Matemática. Os autores referem que, após o processo de formação (estágio), as concepções dos professores estagiários sofreram uma alteração bastante sensível "(...) afastando-se do paradigma tradicional e aproximando-se do paradigma cognitivista e, em alguns aspectos, dos paradigma pensamento do professor" (p. 292).

A leitura e análise destes quatro artigos (Rosado & Dias, 2002; Graça, 2003; Oliveira & Gaspar, 2004; Fernandes *et al.*, 2004) suscita várias questões de interesse relacionadas com o desenvolvimento de projectos de investigação na área da avaliação das aprendizagens e, particularmente, no domínio das concepções e práticas de avaliação de professores ou educadores. Deixam-se aqui apenas três dessas questões ou áreas de interesse. Em primeiro lugar, tratando-se de uma das áreas mais investigadas nos anos 90, pois existem algumas dezenas de trabalhos de investigação produzidos no âmbito de programas de mestrado, em que medida é que as investigações mais recentes não ganhariam se houvesse algum esforço de integração do que tem sido desenvolvido em Portugal. Por outro lado, parece haver alguma dificuldade em "entrar" na área das práticas; ou seja, normalmente as investigações ficam-se mais pelas concepções que são estudadas através de inquéritos por questionário ou por entrevista mas raramente através das interacções dos professores com outros professores (veja-se o artigo de Caria que mais acima se discute) ou com os seus alunos. Ora essa análise parece ser fundamental para que se possam estudar mais profundamente as concepções dos professores. Finalmente, sente-se a falta de descrições detalhadas do que realmente se passa dentro das salas de aula, dentro da "caixa negra", conforme nos dizem Black & Wiliam (1998). Não será fácil progredir-se em termos de consolidação teórica que possa informar e sustentar as práticas sem que se descrevam, analisem e interpretem essas mesmas práticas.

O artigo de Silva, Rosado & Freitas (2003) é o único em que se discute uma investigação que estudou concepções de alunos acerca da avaliação das aprendizagens. A investigação incidiu sobre uma amostra de 147 alunos do 9.º ano de escolaridade todos inquiridos por questionário e alguns ainda por entrevista. Os autores concluem que quer os alunos de diferentes géneros quer os alunos com níveis de sucesso diferenciados, percebem de forma diferente o que devem ser objectos desejáveis de avaliação e o

que entendem que é efectivamente avaliado (e.g., esforço, participação nas aulas, aquisição de conhecimentos, interesse, estudo). Os alunos tendem a valorizar a avaliação de variáveis sócio-afectivas (e.g., comportamento, esforço, interesse) e as variáveis de relação interpessoal e não as consideram devidamente valorizadas pelos professores. Identificam sobretudo os testes mas gostariam de ver utilizados outros métodos e de participar mais no processo de avaliação. O que se parece concluir dos dados obtidos junto dos alunos é que estes têm uma visão da avaliação como um processo centrado no professor, que ocorre essencialmente no fim do período, baseado em testes e sem a sua participação activa. Para além disso os melhores alunos parecem estar melhor informados acerca da avaliação. Uma das conclusões mais relevantes terá a ver com o facto de que, para os alunos, não é visível qualquer cultura de avaliação formativa. Para estes jovens do 9.º ano avaliar é, ainda, sinónimo de classificar.

Temos aqui uma outra *frente* de investigação que merece ser explorada na medida em que pode contribuir para se perceber a relação que os alunos estabelecem com as tarefas de avaliação ou a forma como percebem o *feedback* dos professores ou ainda como desenvolvem processos complexos tais como a auto-avaliação ou a autorregulação das suas aprendizagens. Para tal são precisos estudos realizados dentro das salas de aula que procurem investigar *in loco* o que os alunos sabem, fazem e pensam.

Na categoria *Reflexão/Análise de Literatura* foram incluídos três artigos, dois dos quais se poderão considerar análises de literatura (Neves, Jordão & Santos, 2004; Santos, 2003) e o outro uma reflexão de natureza histórica (Freitas, 2001).

No seu artigo Neves *et al.* (2004) fazem uma análise de literatura com base numa selecção de "(...) 46 estudos/investigações cujo objecto se enquadra na área da avaliação das aprendizagens. Destas, 37 são dissertações de mestrado, 1 é dissertação de doutoramento e 6 são investigações (...)" (p. 54). A maioria das investigações analisadas por estes autores são de natureza qualitativa ou mista (80%) tendo a recolha de dados sido essencialmente baseada em inquéritos por questionário ou por entrevista e em observações (82%). Uma grande parte das investigações analisadas centraram-se no estudo das concepções de professores, das reacções aos normativos legais sobre avaliação e das estratégias, métodos e técnicas de avaliação.

A análise deste artigo permite referir que ainda há muitos professores com concepções de avaliação associadas a racionalidades de natureza técnica e que, ao nível do estudo das concepções, parece verificar-se a necessidade de clarificação uma vez que há uma certa ambiguidade e até contradição nos resultados. Este facto pode, naturalmente, estar relacionado com as metodologias utilizadas pelos investigadores. Constatam-se

também as persistentes dificuldades dos professores com a avaliação formativa, apesar de a considerarem um processo relevante na regulação do ensino e da aprendizagem. A maioria das investigações revistas por aqueles autores evidencia que os testes são os instrumentos mais utilizados, havendo pouca diversificação de métodos.

Nas suas considerações finais Neves *et al.* (2004) referem nomeadamente: a) (...) nas investigações analisadas regista-se uma clara presença dos normativos legais na contextualização do estudo (...)” (p. 62-63); e b) “São poucos os estudos em profundidade que permitam um conhecimento das práticas reais de avaliação e que usem a observação de situações reais” (p. 64).

Santos (2003) analisa e discute um conjunto de oito documentos de natureza programática no domínio da Matemática em que a avaliação das aprendizagens está presente, de forma mais ou menos profunda e mais ou menos específica. Os documentos, dois portugueses e seis de países anglo-saxónicos, foram publicados entre 1975 e 2000 (inclusive). A partir da sua análise comparativa a autora fez, entre outras, as seguintes constatações: a) a avaliação formativa foi ganhando cada vez mais predominância ao longo dos anos; b) os testes escritos foram cedendo lugar a outras formas de recolha de informação mais informais, associadas às acções, orientadas para apoiar os alunos; c) o paradigma da avaliação como medida foi dando lugar ao paradigma da avaliação como interacção social; e d) a avaliação interna foi tendo uma relevância crescente conduzindo a um maior equilíbrio com a avaliação externa.

Embora reconheça a evolução que se tem registado ao longo do tempo, Santos refere que nem sempre há coerência entre os princípios enunciados e as orientações de natureza prática. No entanto, afirma que a avaliação surge cada vez mais articulada com as teorias da aprendizagem e com o ensino e como um processo mais próximo dos actores directos da acção pedagógica. Além disso, mais recentemente, os documentos atribuem um papel central aos alunos no processo de avaliação. A este propósito refere a autora: “Apenas em 2000 a auto-avaliação e a co-avaliação, que atribuem ao aluno um papel central, são visivelmente reconhecidas como partes essenciais e insubstituíveis de uma avaliação ao serviço da aprendizagem do aluno” (Santos, 2004, p.18).

Apesar do artigo de Freitas (2001) não ser propriamente um artigo sobre avaliação das aprendizagens tem real interesse pois contém um conjunto de informações e de reflexões que são relevantes para os investigadores da área. Com base em vários exemplos Freitas afirma que “O conceito de avaliação não era praticamente utilizado entre nós antes dos anos 70” (p. 98) e que tal só se verificou após o que designa como a

internacionalização da educação em Portugal e que passou pela integração do país em organizações internacionais.

É interessante constatar que nos finais de 1971 se realizou um curso breve sobre *Avaliação do Rendimento do Ensino* em que se abordaram questões tais como a formulação de objectivos, a taxonomia de Bloom e a construção de instrumentos objectivos de avaliação. Também em 1971, no Congresso do Ensino Liceal, os professores faziam questão de propor que os exames fossem progressivamente substituídos por processos de avaliação contínua e objectiva.

O autor considera que 1976 foi um ano chave no desenvolvimento da avaliação em Portugal talvez porque foi criado um grupo de trabalho no âmbito do Gabinete de Estudos e Planeamento do Ministério da Educação destinado a estudar a questão da *avaliação do rendimento escolar dos alunos*. Surge então a diferenciação clara entre avaliação e classificação, a relevância da avaliação formativa e dos processos de avaliação conducentes à progressão na aprendizagem e um conjunto de brochuras sobre avaliação das aprendizagens dos alunos editadas pelo Instituto de Tecnologia Educativa.

Não deixa também de ser interessante a informação de que em 1972/1973 circulavam entre professores estagiários cópias de traduções de capítulos do livro de Bloom, Hastings e Madaus (1971)! Na mesma linha o autor refere que em 1976/1977 havia vários documentos de apoio aos professores baseados nos trabalhos de Bloom e de Tyler.

Finalmente, Freitas (2001) faz alusão ao papel que as chamadas universidades novas tiveram na disseminação das novas ideias sobre a avaliação referindo que os professores nos anos 80 já não pensavam a avaliação como nos anos 70 devido à referida influência dos cursos de formação de professores das universidades novas, às influências internacionais e ao papel de organismos do Ministério da Educação.

Na categoria *Reacção a Normativos da Administração* foi apenas identificado e analisado um artigo neste primeiro lustro da década. Trata-se de um artigo de Barreira (2005) em que se relaciona a reorganização curricular do ensino básico com a avaliação das aprendizagens tendo como referência a legislação em vigor, nomeadamente o Decreto-Lei n.º 6/2001 e o Despacho Normativo n.º 30/2001. Na verdade, o autor quer contribuir para se possam "(...) perceber as implicações que a gestão flexível do currículo poderá ter ao nível da avaliação das aprendizagens (...) " (p. 13). Assim, faz uma comparação entre o Despacho Normativo n.º 98-A/92 e o Despacho Normativo n.º 30/2001, concluindo que apesar de este último não introduzir mudanças significativas na concepção de avaliação que está subjacente no primeiro, "(...) aperfeiçoa a terminologia usada ao nível das modalidades de avaliação, nomeadamente no que se refere à avaliação sumativa,

dando, também, maior importância à avaliação contínua e à auto-avaliação por parte do aluno” (Barreira, 2005, p. 103).

Considerando que eventualmente haverá produção legislativa a mais e acção concreta e prática a menos Carlos Barreira afirma “Chega de mudanças legislativas sobre a avaliação das aprendizagens dos alunos. É preciso agora investir na formação adequada dos professores, tendo em conta as suas verdadeiras necessidades. Torna-se, pois, urgente dedicar tempo aos professores” (p. 103).

Uma síntese interpretativa

Esta síntese interpretativa está organizada de acordo com quatro das seis categorias em que os artigos seleccionados nesta investigação foram agrupados para descrição e análise. Dada a extensão do artigo decidiu-se não apresentar as sínteses das categorias *Reflexão/Análise de Literatura* e *Reacção a Normativos da Administração*. Como já se referiu as categorias são suficientemente amplas para organizar as questões teóricas e práticas de interesse constantes nos artigos. Também é necessário ter em conta que muito dificilmente um artigo não poderá ser considerado em mais do que uma categoria. Nesses casos, a opção baseou-se no conteúdo do artigo que se se considerou mais relevante e/ou mais elaborado teórica e conceptualmente.

Avaliação interna

Esta é a categoria com o maior número de artigos seleccionados. O seu conteúdo, organização e estrutura têm essencialmente a ver com a avaliação que é da integral responsabilidade dos professores e das escolas. Isto significa um alargado espectro de temas que vão desde as modalidades, funções e princípios da avaliação, até às questões da integração da avaliação no ensino e na aprendizagem ou à organização das respectivas práticas.

A grande maioria dos 15 artigos analisados nesta categoria parte de uma reflexão e de uma base teórica mais ou menos profunda para prescrever ou sugerir modos de actuação prática (e.g., Abrantes, 1989; Carvalho, 1994; Valadares & Teodoro, 1988; Barreira, 2001). São poucos os artigos que se escreveram a partir de dados empíricos (e.g., Barreira, 2005; Pais, 2000).

Apesar de ser notória uma certa sofisticação e elaboração teórica nos artigos da presente década (e.g., Barreira, 2005; Pais, 2000), também é possível identificar tal sofis-

ticação e elaboração em alguns dos artigos das décadas precedentes (e.g., Abrantes, 1989; Boavida & Vaz, 1987; Carvalho, 1994).

Alguns artigos traduzem o debate em torno de duas concepções bem distintas de avaliação das aprendizagens (e.g., Barreira, 2005; Boavida & Vaz, 1987; Carvalho, 1994; Pais, 2000). Num caso temos uma visão mais científica, marcada por preocupações com a objectividade, com os aspectos técnicos e com a neutralidade dos avaliadores e em que a qualidade da avaliação é garantida através da utilização de instrumentos *cientificamente* construídos. Noutro caso temos uma visão mais crítica, ou sociocrítica, em que a avaliação é assumida como um processo interactivo e subjectivo e em que a sua qualidade é garantida pela intersubjectividade dos participantes. No fundo, trata-se do debate entre uma visão da avaliação como disciplina exacta, livre de valores, e uma visão da avaliação como uma construção social complexa imbuída de valores e significados que lhe são atribuídos pelos participantes e/ou principais interessados no processo (Gipps, 1994; Guba & Lincoln, 1989).

Na mesma linha, surge também em alguns artigos a clarificação dos conceitos de avaliação formativa e de avaliação sumativa e das relações entre si (e.g., Abrantes, 1989; Barreira, 2005; Carvalho, 1994; Leal & Abrantes, 1990). De facto, verifica-se que a distinção entre uma avaliação formativa de natureza pontual, retroactiva e formal e uma avaliação formativa contínua, interactiva e informal (Allal, 1986) é feita por alguns autores, apesar de nem todos associarem esta distinção às perspectivas teóricas - behaviorismo e cognitivismo - que lhe são subjacentes.

Verificou-se que, na maioria dos casos, os autores não discutem questões relevantes para o desenvolvimento da avaliação interna tais como a natureza da avaliação formativa e da avaliação sumativa, as tarefas de avaliação, o *feedback* e os processos de auto-avaliação e de auto-regulação das aprendizagens.

O que se poderá dizer é que são ainda em número reduzido os artigos que abordam questões relevantes no domínio da avaliação interna e que não foram encontrados quaisquer artigos em áreas tais como o lugar da avaliação nos projectos educativos e curriculares das escolas, o papel que as avaliações sumativas de escola poderão ter no desenvolvimento do currículo e na moderação das classificações internas, a caracterização de práticas de avaliação formativa em contexto de sala de aula e a relação entre essas práticas e as aprendizagens dos alunos. Na verdade, fica-se com a sensação que, neste domínio, muito estará ainda por fazer para que os processos relacionados com a avaliação interna possam ser compreendidos em benefício de práticas mais consis-

tentes com os desafios pedagógicos e didácticos impostos pelos desenvolvimentos nas teorias do currículo e das aprendizagens.

Avaliação externa

Foram incluídos oito artigos nesta categoria que, de algum modo, se centravam em processos de avaliação das aprendizagens que são da responsabilidade de uma entidade exterior às escolas tais como a elaboração, administração e correcção de provas de exame, a análise de procedimentos utilizados em avaliações externas e efeitos da avaliação externa nas escolas, nos alunos e nos professores.

Dos oito artigos analisados apenas três discutem com algum desenvolvimento questões de avaliação externa das aprendizagens (Boavida, 1985; Eça, 2004; Martins & Silva, 2000). Os restantes cinco apresentam essencialmente resultados de avaliações externas nacionais ou internacionais (Amaro, 1997; Ferreira *et al.*, 1995; Ramalho, 1995), veiculam opiniões de professores sobre os exames (Gazeta da Matemática, 2004) ou analisam uma prova de exame fazendo recomendações (Lobo, 2001).

A avaliação externa das aprendizagens dos alunos pode ser analisada ou estudada sob pontos de vista tão diferentes como o pedagógico, o técnico-científico ou o político. De facto, podem estudar-se questões relacionadas com a natureza global das provas, com a sua articulação com o currículo e com a avaliação interna, ou com a natureza dos itens, assim como questões que têm a ver com os seus índices psicométricos (validade, fiabilidade, discriminação, dificuldade) e com a análise e interpretação dos resultados e ainda questões relacionadas com as vantagens, desvantagens e limitações dos exames. Além de tudo o que está inerente a estas áreas de estudo e de investigação poderemos considerar os efeitos dos exames sobre as formas como os professores ensinam e avaliam, como os alunos estudam e se preparam para os exames ou como as escolas organizam as turmas e o trabalho pedagógico em geral; trata-se do efeito de *backwash* dos exames para se referir o termo utilizado na literatura anglo-saxónica.

Com base nos três artigos que efectivamente abordam temas de avaliação externa das aprendizagens dos alunos, pode dizer-se que todos enunciam e discutem áreas tão relevantes como a correcção das provas (e.g., fiabilidade inter-correctores, critérios de correcção), a natureza das provas ou os seus efeitos nos alunos, professores e escolas. Mas há necessidade de investigação empírica em todas as áreas acima identificadas. No domínio da avaliação externa o nosso desconhecimento acerca do que efectivamente se passa no sistema educativo é praticamente total. Desde a análise de séries de dados baseados nos exames nacionais do ensino secundário, que se realizam há mais

de dez anos, às formas como as escolas reagem e lidam com os resultados dos seus alunos nos exames ou ainda à natureza das questões e à sua relação com o currículo, há um amplo espectro de matérias que urge estudar e investigar.

As provas aferidas dos 4.º e 6.º anos de escolaridade suscitam uma variedade de questões de investigação que, por exemplo, permitam compreender as suas relações com as formas como o currículo está a ser desenvolvido nas escolas ou com o tipo de reflexões e acções que os professores têm empreendido.

Métodos e técnicas

A partir da análise dos 13 artigos incluídos nesta categoria parece claro que, em geral, os autores questionam pouco os métodos e técnicas que sugerem para a recolha de informação avaliativa. E a verdade é que muitas das sugestões que abundam na literatura acabam por constituir uma espécie de emulações das avaliações em grande escala como os exames. Curiosamente, na grande maioria dos artigos aqui revistos apresenta-se ou discute-se testes de avaliação das aprendizagens dos alunos. Apenas foi encontrado um artigo em que se descreve e analisa um método alternativo de avaliação (Menino, 2005). Barreira (2005), e Simão (2005) fazem-no nos seus artigos mas estão incluídos noutras categorias porque o seu foco não foi considerado nos métodos e técnicas de avaliação.

A ênfase nos *instrumentos* de avaliação e nos testes está normalmente associada a uma lógica de ensino, de aprendizagem e de avaliação que está muito baseada na formulação de objectivos e/ou na definição de critérios, na utilização de taxonomias de objectivos educacionais e na elaboração e utilização das chamadas matrizes de objectivos/conteúdos ou matrizes de competências/domínios do currículo. Trata-se de uma abordagem que tem uma marca epistemológica relacionada com a quantificação, a objectividade ou a neutralidade dos avaliadores. Por isso há autores que consideram que a designação *instrumento* está associada a uma concepção da avaliação como medida, preferindo, por isso, outras designações tais como *tarefa* de avaliação ou *método* de avaliação por parecerem mais coerentes com a natureza das racionalidades que estão na base das concepções socioculturais e sociocognitivas da aprendizagem e da avaliação (Fernandes, 2005, 2006).

É fundamental que a recolha de informação seja credível, se faça de forma sistemática e com base em métodos diversificados mais ou menos estruturados. Porém, pode perguntar-se: Será que a avaliação das aprendizagens dos alunos, realizada no âmbito das salas de aula, é sobretudo uma questão de concepção e administração de instru-

mentos? Ou ela será antes uma questão essencialmente pedagógico-didáctica? Ou um processo que não pode ignorar questões de natureza ético-política? Ou ainda um processo em que é necessário compreender sua natureza sociocultural e sociocognitiva? Ou um pouco de todas estas coisas? São questões que devem ser debatidas numa abordagem ao problema dos métodos e técnicas de avaliação a utilizar nas salas de aula e, muito particularmente, em contextos em que a avaliação formativa é tendencialmente predominante.

Outro plano de discussão, não menos relevante, tem a ver com a qualidade dos ditos instrumentos ou, se quisermos, com a qualidade ou a credibilidade da informação recolhida. Também aqui se verifica que existe essa preocupação por parte de vários autores ao discutirem conceitos tais como a validade, a fiabilidade, a objectividade, o índice de discriminação ou o índice de dificuldade dos chamados instrumentos utilizados, particularmente dos testes (e.g., Oliveira & Pereira, 1993; Pedrosa, 1991). Repare-se que esta preocupação permanece desde os artigos dos anos 80 até aos nossos dias e, curiosamente, apesar da evolução teórica nos domínios da avaliação, da aprendizagem e do ensino, a abordagem à questão da qualidade das avaliações continua a ser feita numa perspectiva psicométrica ou de racionalidade técnica. Na verdade, incide sobretudo nas avaliações de natureza sumativa e baseia-se em conceitos criados para serem utilizados no âmbito de abordagens psicométricas. Existe nesta questão um desafio que passa, em parte, por uma valorização da avaliação formativa e por novas concepções acerca da sua validade e acerca da sua fiabilidade que, por exemplo, tenham em conta as consequências das avaliações (e.g., Black & Wiliam, 2006c; Stobart, 2006).

Concepções e práticas de professores/alunos

Todos os artigos constantes nesta categoria são baseados em estudos de natureza empírica destinados a caracterizar as concepções e as práticas de avaliação de professores (em oito artigos) e as concepções de alunos (no artigo de Silva *et al.*, 2003). Em todos os casos os dados foram obtidos através de inquéritos por questionário e entrevista e só em dois deles se fizeram algumas observações das práticas de professores em contexto de sala de aula (Graça, 2003; Paulo & Santos, 1998).

O que se pode dizer com base na análise daqueles oito artigos é que se torna bastante evidente que não há praticamente trabalhos que descrevam, analisem e interpretem as práticas de avaliação de professores em contextos reais de sala de aula. Os dados obtidos exclusivamente através de questionários e/ou entrevistas dizem-nos o que os participantes dizem que fazem ou uma sua interpretação acerca do que fazem.

Consequentemente, nalguns casos, os dados obtidos são algo estranhos e até pouco congruentes podendo a sua credibilidade ser questionada quando temos elevadíssimas percentagens de professores, ou de educadores a referirem que, por exemplo, desenvolvem predominantemente avaliação formativa e que utilizam uma variedade de métodos de recolha de informação, contrariando o que nos diz a literatura empírica internacional e a pouca que se tem realizado em Portugal.

É evidente que é importante ouvir os professores através de entrevistas, pois são poderosos meios de obtenção de informação acerca do que os professores sabem, pensam e fazem, mas a triangulação metodológica parece essencial para credibilizar qualquer caracterização que se pretenda fazer das práticas avaliativas dos professores.

Parece ainda ser de referir nesta altura que a investigação das práticas e das concepções de professores terá necessariamente de partir de uma sólida base conceptual que permita tornar claro o que se entende por avaliação formativa, por avaliação sumativa ou por *feedback*. Caso contrário pode correr-se o risco de se estar a caracterizar de uma forma aquilo que realmente deveria ser caracterizado de uma outra maneira. Além disso, sentiu-se dificuldade em poder identificar relações significativas entre as práticas de avaliação dos professores, as suas concepções e outros elementos de interesse como é o caso das tarefas ou o conhecimento sobre avaliação. Ou porque tais relações não são tornadas evidentes ou porque é difícil identificar elementos suficientemente claros e credíveis que permitam a sua compreensão.

Finalmente, a investigação das concepções dos alunos deve ser rodeada do mesmo tipo de cuidados metodológicos pois constitui uma área da maior relevância para que se possa compreender melhor as práticas de avaliação que ocorrem nas salas de aula.

Considerações e reflexões finais

Após a descrição, a análise e a síntese dos 59 artigos revistos nesta investigação é desejável elaborar um conjunto de reflexões que possa traduzir a realidade de 20 anos de trabalhos publicados no domínio da avaliação das aprendizagens em revistas portuguesas de índole científica. Seleccionaram-se apenas as que se consideraram mais relevantes.

Ausência de uma agenda de investigação

A maioria dos autores dos artigos analisados são docentes e investigadores do ensino superior. Este facto poderia indiciar que os artigos pudessem traduzir, implícita ou explicitamente, uma agenda de investigação no domínio da avaliação das aprendizagens. Dito de outra forma, poderia esperar-se que uma boa parte dos artigos emanasse

do trabalho empírico desenvolvido pelos autores. Na verdade, apenas em 21 artigos, cerca de 35% do total, se pode dizer que há uma apresentação e discussão de dados de natureza empírica. Esta situação pode talvez explicar-se pelo facto da comunidade científica nesta área ser relativamente recente e muito pouco numerosa e ainda pelo facto de a oferta de cursos de pós-graduação ser ainda reduzida.

Talvez estas razões expliquem o que parece ser a ausência de consistência entre os conteúdos dos artigos analisados; ou seja, os temas abordados resultantes de investigação empírica abordam questões de investigação que não se enquadram num todo coerente que permita desenvolvimentos teóricos significativos. Mesmo no domínio das concepções e práticas de professores, que é o que está mais estudado empiricamente, quer as questões investigadas quer as metodologias utilizadas parecem não contribuir para a consistência e coerência em que a construção de conhecimento se possa basear.

Tal como se tem vindo a afirmar (Fernandes, 1994; 2005), parece ser necessário definir uma agenda de investigação que oriente os esforços dos investigadores, tornando o seu trabalho mais útil, pertinente e significativo.

Ênfase na descrição e na prescrição

A grande maioria dos artigos analisados é de natureza descritiva e prescritiva. Ou seja, há sobretudo uma apresentação e descrição de um conjunto de factos, definições e, por vezes, de perspectivas teóricas que aparecem como um corpo acabado e completo de conhecimentos na área da avaliação das aprendizagens; uma apresentação *do que há* ou *do que está feito* sem que, verdadeiramente, se produza uma reflexão que mobilize e integre um conjunto de contribuições provenientes de domínios tais como a sociologia, a psicologia social, a antropologia, as teorias da comunicação, as teorias curriculares ou as teorias da aprendizagem.

Este problema relaciona-se com o que acima se discutiu pois é essencialmente através da investigação empírica que se podem suscitar reflexões, enunciar problemas, formular questões e relacionar fenómenos, factos e teorias.

Métodos e técnicas de recolha de informação

A questão dos métodos de recolha de informação destinada a avaliar o que os alunos sabem e são capazes de fazer tem de ser mais investigada, discutida e reflectida. São muito poucos os artigos em que se pode dizer que há uma reflexão acerca de um domínio tão relevante. É notória a tendência para fazer incidir o trabalho neste domí-

nio sobre os testes sendo muitíssimo poucos os artigos em que se discutem métodos alternativos de recolher informação avaliativa junto dos alunos.

Há necessidade de se compreender como é que essas formas alternativas de avaliação funcionam nas salas de aula.

As perguntas que a seguir se formulam traduzem apenas algumas das preocupações que decorrem da discussão que acima se desenvolveu.

1. Será possível desenvolver uma avaliação menos dependente de verdadeiras baterias de instrumentos e que garanta a qualidade da informação que se pretende obter? Como?
2. Como se poderão seleccionar e utilizar tarefas que possam suscitar actividades de ensino, de aprendizagem e de avaliação?
3. De que formas é que a partilha do processo de avaliação com os alunos e com outros intervenientes poderá permitir a obtenção de informação avaliativa de boa qualidade? Que estratégias parecem mais adequadas para concretizar essa partilha?
4. Como é que os processos de auto-avaliação, de co-avaliação e de hetero-avaliação poderão contribuir para melhorar e credibilizar o processo de avaliação?

Estas e, com certeza, outras questões podem contribuir para que se procurem métodos de avaliação das aprendizagens dos alunos mais consistentes com as concepções e conhecimentos contemporâneos nos domínios do ensino, do currículo e da aprendizagem. Mas também mais simples, mais realistas, mais significativas para professores e alunos e mais aptas para promoverem e melhorarem as aprendizagens.

Avaliação formativa e a sua teoria

A avaliação formativa é referida na grande maioria dos artigos. No entanto, são escasos os artigos em que se discute em profundidade e com a necessária abrangência, quer a avaliação formativa, quer a avaliação sumativa, quer ainda as relações entre estes dois processos de avaliação. Provavelmente muitos autores partem do princípio de que quando se referem à avaliação formativa todas as pessoas terão um entendimento semelhante. A verdade é que existem vários entendimentos possíveis, alguns dos quais pouco terão a ver com uma avaliação formativa contínua, cuja função primordial é a melhoria das aprendizagens dos alunos, interactiva, integrada nos processos de ensino e de aprendizagem e que cria condições para proporcionar *feedback* de qualidade aos alunos. Além disso, persistem concepções erróneas acerca da natureza da avaliação

formativa e da natureza da avaliação sumativa. A primeira normalmente associada à falta de rigor, à subjectividade sem controlo e aos processos informais e a segunda à garantia da qualidade, à objectividade e aos processos formais. Nestas condições, quer no plano da discussão teórica, quer no plano da investigação empírica, é necessário empreender esforços que contribuam para clarificar estas questões.

O estudo das relações entre a avaliação formativa e a avaliação sumativa não pode depender exclusivamente do facto de uma estar integrada nos processos de ensino e de aprendizagem e servir para os melhorar e a outra ser realizada após um dado período para que se possa fazer um balanço ou uma súmula do que os alunos aprenderam tendo em vista a sua classificação. É necessário clarificar se estas duas avaliações constituem uma dicotomia, se são dimensões diferentes de um mesmo conceito ou se são complementares. É necessário caracterizar um conjunto de práticas de avaliação mais ou menos indiferenciadas que ocorrem nas salas de aula e que poderão *estar* entre a avaliação formativa e a avaliação sumativa. E será neste terreno que as relações entre as duas terão de ser estudadas e compreendidas.

Estas e outras questões têm de estar presentes na construção de uma teoria da avaliação formativa que parece ser essencial para o desenvolvimento informado e sustentado de práticas que, comprovadamente, melhoram substancialmente as aprendizagens dos alunos (Black & Wiliam, 1998).

Outras questões passarão necessariamente pelos papéis que professores e alunos podem e devem desempenhar no processo de avaliação e que tão pouco discutidos e investigados têm sido de acordo com a literatura que aqui se analisou e sintetizou. Assim, uma teoria da avaliação formativa poderá basear-se em três elementos fundamentais e nas múltiplas relações entre eles: a) compreender os processos de desenvolvimento do currículo nas salas de aula e a sua relação com os processos de avaliação; b) compreender os papéis dos alunos e professores nos processos de ensino, aprendizagem e avaliação; e c) compreender os contextos, dinâmicas e ambientes de ensino, aprendizagem e avaliação nas salas de aula (Fernandes, 2006).

A construção teórica estará sempre muito dependente de investigação que seja capaz de nos proporcionar retratos tão nítidos quanto possível das realidades e dos fenómenos avaliativos que ocorrem nas salas de aula. É necessário e urgente produzir investigação e conhecimento acerca das práticas avaliativas que ocorrem em contextos reais. É necessário e urgente que a investigação se debruce sobre a natureza das interacções sociais que se desenvolvem entre professores e alunos. E tantas outras coisas são necessárias! Por isso é importante referir que os 59 artigos publicados nos últimos 20

anos revelam que a investigação empírica no domínio da avaliação das aprendizagens ainda está longe de ter a presença e o significado que permitam o desenvolvimento desta área científica.

Independentemente de se reconhecer o caminho que se foi desbravando ao longo de duas décadas e tudo o que de positivo já se conseguiu, é necessário ter consciência das fragilidades que caracterizam a investigação e a produção teórica num domínio que é reconhecidamente imprescindível para transformar e melhorar as actuais realidades educativas.

Bibliografia

- Abrantes, P. (1989). Avaliação em Matemática: A necessidade de mudar de óptica. *Aprender*, 7, pp. 9-14.
- Abrecht, R. (1991). *L'évaluation formative: Une analyse critique*. Bruxelles: De Boeck.
- Afonso, A. (1994). Comentário. A prática avaliativa não mudou. *Educação, Sociedade & Culturas* (Secção Diálogos Sobre o Vivido), 1, pp. 174-179.
- Allal, L. (1986). Estratégias de avaliação formativa: Concepções psicopedagógicas e modalidades de aplicação. In L. Allal, J. Cardinet e Ph. Perrenoud (Orgs.), *A avaliação formativa num ensino diferenciado*, pp. 175-209. Coimbra: Almedina.
- Allal, L., Cardinet, J. & Perrenoud, Ph. (Eds.) (1979). *L'évaluation formative dans un enseignement différencié*. (Actes du colloque à l' Université de Genève, Mars 1978.) Berne: Peter Lang.
- Amaro, G. (1997). Qualidade em educação: A avaliação externa das aprendizagens dos alunos em Portugal. *Inovação*, 10, 2-3, pp. 259-275.
- Barreira, C. (2001). Duas estratégias complementares para a avaliação das aprendizagens: A avaliação formadora e a avaliação autêntica. *Revista Portuguesa de Pedagogia*, Ano 35, 3, pp. 3-33.
- Barreira, C. (2002). Reorganização dos currículos no ensino básico e suas implicações na avaliação das aprendizagens. *Revista Portuguesa de Pedagogia*, Ano 36, 1-3, pp. 89-106.
- Barreira, C. (2005). Soluções para a prática da avaliação formativa. *Revista Portuguesa de Pedagogia*, Ano 39, 2, pp. 137-149.
- Barreira, C. & Pinto, J. (2006). *A investigação em Portugal sobre avaliação das aprendizagens dos alunos (1990-2005)*. Retirado em 6 de Novembro, 2006, de <http://www.educ.fc.ul.pt/docentes/msantos/spce.pdf> (versão html).
- Benavente, A. (1990). Avaliação e inovação educacional: Notas e reflexões. *Inovação*, 3, 4, pp. 33-46.
- Berlak, H. (1992a). The need for a new science of assessment. In H. Berlak, F. Newmann, E. Adams, D. Archbald, T. Burgess, J. Raven e T. Romberg (Eds.), *Toward a new science of educational testing and assessment*, pp. 1-22. Albany, NY: State University of New York Press.

- Berlak, H. (1992b). Toward the development of a new science of educational testing and assessment. In H. Berlak, F. Newmann, E. Adams, D. Archbald, T. Burgess, J. Raven e T. Romberg (Eds.), *Toward a new science of educational testing and assessment*, pp. 181-206. Albany, NY: State University of New York Press.
- Biggs, J. (1998). Assessment and classroom learning: A role for summative assessment? *Assessment in Education: Principles, Policy & Practice*, 5, 1, pp. 103-110.
- Black, P. & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5, 1, pp. 7-74.
- Black, P. & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. Retirado em 22 de Outubro de 2004 de www.pdkintl.org/kappan/kbla9810.htm
- Black, P. & Wiliam, D. (2006a). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and learning*, pp. 9-25. London: Sage.
- Black, P. & Wiliam, D. (2006b). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning*, pp. 81-100. London: Sage.
- Black, P. & Wiliam, D. (2006c). The reliability of assessments. In J. Gardner (Ed.), *Assessment and learning*, pp. 119-131. London: Sage.
- Boavida, J. (1985). Análise das avaliações realizadas pelos alunos do ramo educacional da Faculdade de Ciências e Tecnologia da Universidade de Coimbra. *Revista Portuguesa de Pedagogia*, Ano 29, pp. 165-180.
- Boavida, J. L. (1996). *Concepções e práticas de avaliação das aprendizagens de professores do 1.º ciclo do ensino básico: Três estudos de caso*. Tese de mestrado em ciências da educação (avaliação em educação) não publicada. Universidade Católica Portuguesa.
- Boavida, J. & Barreira, C. (1992). Nova avaliação no ensino básico: Análise dos despachos 162/ME/91 e 98-A/92. *Revista Portuguesa de Pedagogia*, Ano 19, pp. 344-366.
- Boavida, J. & Barreira, C. (1993). Nova avaliação: Novas exigências. *Inovação*, 6, pp. 97-105.
- Boavida, J. & Vaz, P. (1987). Auto-avaliação: Contributo para a sua revalorização no processo pedagógico. *Revista Portuguesa de Pedagogia*, Ano 21, pp. 463-477.
- Boavida, J., Lopes, M. C. & Vaz, P. (1986). Avaliação: Tópicos para uma mudança de atitudes. *Revista Portuguesa de Pedagogia*, 20 (2.ª Série, Ano 20), pp. 261-277.
- Campos, C. (1996). *Concepções e práticas de professores sobre avaliação das aprendizagens: Dois estudos de caso*. Tese de mestrado em ciências da educação (avaliação em educação) não publicada. Universidade Católica Portuguesa.
- Cardinet, J. (1986). *Évaluation scolaire et mesure*. Bruxelas: De Boeck.
- Cardinet, J. (1991). L'apport sociocognitif à la régulation interactive. In J. Weiss (Ed.), *L'évaluation: Problème de communication*, pp. 199-213. Cousset (Fribourg): Delval.
- Caria, T. (Org.) (1994a). Sistema de avaliação dos alunos do ensino básico. *Educação, Sociedade & Culturas* (Secção Diálogos Sobre o Vivido), 1, pp. 129-179.
- Caria, T. (1994b). Comentário. A reforma da avaliação dos alunos do ensino básico analisada no contexto da(s) cultura(s) do(s) professor(es). *Educação, Sociedade & Culturas* (Secção Diálogos Sobre o Vivido), 1, pp. 159-173.
- Carvalho, A. (1990). Validação da tradução portuguesa do teste de compreensão de leitura do Institut Supérieur de Pédagogie de Hainaut. *Inovação*, 3, 4, pp. 87-96.
- Carvalho, J. (1997). Uma análise dos exames de Inglês de 1995/1996 na perspectiva da quali-

- dade do ensino e da aprendizagem. *Inovação*, 10, 2-3, pp. 295-313.
- Carvalho, J. (1999). Provas aferidas de Inglês: Construção de um quadro de referência. *Inovação*, 12, 1, pp. 145-156.
- Carvalho, L. (1994). Avaliação das aprendizagens em Educação Física. *Boletim da Sociedade Portuguesa de Educação Física*, 10/11, pp. 135-151.
- Carvalho, M. P. (1992). Avaliação formativa por computador. *Quadrante*, 1, pp. 139-147.
- Castro, M.L. & Pereira, M. (1994). O relatório como um meio de avaliação formativa no 3.º ciclo. *Revista Portuguesa de Educação*, 7, 1-2, pp. 165-176.
- Cooper, H. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, 1, pp. 104-126.
- Cooper, H. (1998). *Synthesising research: A guide for literature reviews* (3rd ed.). London: Sage.
- Cooper, H. (2003). Editorial. *Psychological Bulletin*, 129(1), pp. 3-9.
- Cooper, H. & Lindsay, J. (1998). Research synthesis and meta-analysis. In L. Bickman & D. Rog (Eds.), *Handbook of social research methods*, pp. 315-337. London: Sage.
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, pp. 438-481.
- Duarte, R.S. (1994). Alguns aspectos das concepções e práticas avaliativas dos professores de uma escola do 2.º ciclo do ensino básico. *Boletim da Sociedade Portuguesa de Educação Física*, 10/11, pp. 79-109.
- Earl, L. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.
- Eça, T. (2004). Avaliação externa nas artes visuais em Portugal e Inglaterra. *Educare-Educere*, 2, pp. 35-48.
- Fernandes, D. (1992). O tempo da avaliação. *Noesis*, 23, pp. 18-21.
- Fernandes, D. (1994). Avaliação das aprendizagens: Das prioridades de investigação e de formação às práticas na sala de aula. *Revista de Educação*, 8, pp. 15-20.
- Fernandes, D. (2005). *Avaliação das aprendizagens: Desafios às teorias, práticas e políticas*. Cacém: Texto Editores.
- Fernandes, D. (2006). Para uma teoria da avaliação formativa. *Revista Portuguesa de Educação*, 19(2). (em publicação).
- Fernandes, D. e Vale, I. (1994). Two young teachers' conceptions and practices about problem solving. In *Proceedings of the Eighteenth International Conference for the Psychology of Mathematics Education* (Vol. II, pp. 328-335). Lisboa: Program Committee of the 18th PME Conference.
- Fernandes, J., Alves, M. P. & Arantes, M. J. (2004). Impacto de uma experiência de formação nas concepções e práticas de avaliação de professores estagiários. *Revista de Estudos Curriculares*, 2, 2, pp. 263-298.
- Fernandes, M. (1998). A mudança de paradigma na avaliação educacional. *Educação, Sociedade & Culturas*, 9, pp. 7-32.
- Ferreira, J., Almeida, A., Cunha, L., Rolo, A. & Vaz, J. (1995). Uma análise dos resultados da prova específica de Física de 1994: Distritos de Braga e Viana do Castelo. *Gazeta de Física*, 18, 1, pp. 9-13.
- Figari, G. & Achouche, M. (Eds.) (2001). *L'activité évaluative réinterrogée*. Bruxelles: De Boeck.
- Freitas, C. V. (2001). Contributo para a história da avaliação educacional em Portugal: Os anos

70. *Revista Portuguesa de Educação*, 14, 1, pp. 95-130.
- Gardner, J. (2006a). Assessment and learning: An Introduction. In J. Gardner (Ed.), *Assessment and learning*, pp. 1-5. London: Sage.
- Gardner, J. (2006b). Assessment for learning: A compelling conceptualization. In J. Gardner (Ed.), *Assessment and learning*, pp. 197-204. London: Sage.
- Gazeta de Matemática. (2004). Inquérito: Mais exames no sistema de ensino?! *Gazeta de Matemática*, Ano LXV, 147, pp. 35-41.
- GIASE (2005a). *Estatísticas da educação 03/04*. Lisboa: GIASE.
- GIASE (2005b). *Recenseamento escolar 05/06*. Lisboa: GIASE.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. Londres: Falmer.
- Gipps, C. e Stobart, G. (2003). Alternative assessment. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 549-576. Dordrecht: Kluwer.
- Glass, G. (1976). Primary, secondary and meta-analysis research. *Educational Researcher*, 5, pp. 3-8.
- Gonçalves, C. (1994). Avaliação do processo de ensino-aprendizagem em Educação Física. *Boletim da Sociedade Portuguesa de Educação Física*, 10/11, pp. 111-134.
- Graça, M. (2003). Avaliação da resolução de problemas: Que relação entre as concepções e as práticas lectivas dos professores? *Quadrante*, 12, 1, pp. 53-73.
- Grégoire, J. (Ed.) (1996). *Évaluer des apprentissages: Les apports de la psychologie cognitive*. Bruxelles: De Boeck.
- Guba, E. & Lincoln, Y. (1989). *Fourth generation evaluation*. London: Sage.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning: Tensions and synergies. *Curriculum Journal*, 16 (2), pp.207-223.
- Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning*, pp. 103-118. London: Sage.
- Harlen, W. e James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in education: Principles, policy and practice*, 4, 3, pp. 365-379.
- Henriques, H. (1997). Avaliação aferida: Concepção e construção de um instrumento de avaliação da aprendizagem em Matemática - 1.º ciclo. *Inovação*, 10, 2-3, pp. 267-294.
- Kellaghan, T. & Madaus, G. (2003). External (public) examinations. In T. Kellaghan & D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 577-602. Dordrecht: Kluwer.
- Leal, L. & Abrantes, P. (1990). Avaliação da aprendizagem/Avaliação na aprendizagem. *Inovação*, 3, 4, pp. 65-75.
- Leitão, J. (1993). Um teste diferente para o 12.º Ano de Química. *Boletim da Sociedade Portuguesa de Química*, 48, pp. 34-35.
- Lobo, V. (2001). A nefasta artificialidade das provas de "exames" do 12.º ano. *Boletim da Sociedade Portuguesa de Química*, 83, pp. 73-75.
- Marques, R. (1997). Avaliação do rendimento escolar no ensino básico. *Revista da Escola Superior de Educação de Santarém*, 8, pp. 9-14.
- Martins, A. & Silva, J. C. (2000). Exames nacionais do ensino secundário: Algumas notas. *Gazeta de Matemática*, Ano LXI, 138, pp. 51-58.
- Martins, I. & Cachapuz, A. (1988). O uso de testes diagnósticos e a identificação de dificuldades

- na aprendizagem do tema Energia e Reações Químicas (8.º ano de escolaridade). *Boletim da Sociedade Portuguesa de Química*, 34, pp. 9-10.
- Menino, H. (2005). Avaliação das aprendizagens em Matemática: Uma experiência de utilização do portefólio no 2.º ciclo. *Educação & Comunicação*, 8, pp. 109-119.
- Nabuco, M. E. (2000). A avaliação como processo central da intencionalidade educativa. A avaliação das crianças dos 0 aos 6 anos de idade. *Da Investigação às Práticas-Estudos de Natureza Educacional*, 1, 1, pp. 81-90.
- Noblit, G. & Hare, R. (1988). *Meta-ethnography: Synthesising qualitative studies*. London: Sage.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22, pp. 155-175.
- Nuttall, D. (Ed.) (1986). *Assessing educational achievement*. London: Falmer.
- Neves, A., Jordão, A. & Santos, L. (2004). Investigação publicada em Portugal, entre 1971 e 2003, acerca da avaliação das aprendizagens: Seu levantamento e análise. *Revista Portuguesa de Investigação Educacional*, 3, pp. 47-71.
- Oliveira, F. & Gaspar, M.F. (2004). Olhares sobre a avaliação em educação pré-escolar: As opiniões e as práticas dos educadores de infância. *Revista Portuguesa de Pedagogia*, 38, 1-2-3, pp. 451-484.
- Oliveira, I. & Pereira, J. (1993). Avaliação diagnóstico: Desenvolvimento de um teste de desempenho em Matemática. *Quadrante*, 2, 1, pp. 113-126.
- Pacheco, J. (1993). O novo sistema de avaliação dos alunos do ensino básico: Do contexto europeu ao contexto de experimentação dos programas e das mudanças curriculares. *Revista Portuguesa de Educação*, 6, 2, pp. 1-22.
- Pais, P. (2000). Práticas classificativas de professores do ensino secundário: Significados e valores. *Inovação*, 13, 2-3, pp. 139-159.
- Pajares, M. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62, 3, pp. 307-372.
- Paulo, A. & Santos, F. (1998). A formação inicial de professores em contexto de investigação-ação: Concepções e práticas de avaliação pedagógica. *Revista de Educação*, 7, 2, pp. 97-113.
- Pedrosa, M. A. (1991). Avaliação e classificação da aprendizagem em Química: Alguns aspectos técnicos. *Boletim da Sociedade Portuguesa de Química*, 46, pp. 81-85.
- Perrenoud, Ph. (1998). From formative evaluation to a controlled regulation of learning processes: Towards a wider conceptual field. *Assessment in Education: Principles, Policy & Practice*, 5, 1, 85-102.
- Peixoto, M. T. (1998). A avaliação no 1.º ciclo do ensino básico. *Revista da Escola Superior de Educação de Viana do Castelo*, 3, 193-208.
- Pereira, M. J. (1997). Opiniões de educadores de infância de Castelo Branco face à avaliação na educação pré-escolar. *Educare-Educere*, 3, pp. 23-33.
- Ramalho, G. (1995). Participação dos estudantes portugueses de 9 e de 13 anos de idade no *Second International Assessment of Educational Progress - Matemática*. *Quadrante*, 4, 1, pp. 43-65.
- Rocha, N. (1990). Avaliação formativa no ensino-aprendizagem de uma língua. *Inovação*, 3, 4, pp. 77-86.
- Roldão, M. C. (1989). O que faz mudar a avaliação? *Aprender*, 9, pp. 5-9.

- Rosado, A. & Dias, L. (2002). Caracterização do pensamento dos professores de Educação Física relativamente à avaliação das aprendizagens. *Ludens*, 17, 1, pp. 19-27.
- Santos, L. (2003). A avaliação em documentos orientadores para o ensino da Matemática: Uma análise sucinta. *Quadrante*, 12, 1, pp. 7-20.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29, 7, pp. 4-14.
- Shepard, L. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th Edition). American Educational Research Association. New York: Macmillan.
- Silva, C., Rosado, A. & Freitas, C. V. (2003). Os alunos e a avaliação: Concepções dos alunos do ensino básico relativamente à avaliação das aprendizagens. *Revista Portuguesa de Investigação Educacional*, 2, pp. 69-81.
- Simão, A.M. (1992). Construção de um instrumento para avaliação do desempenho dos alunos na tarefa de resumir. *Revista Portuguesa de Pedagogia*, Ano 26, 2, pp. 249-267.
- Simão, A.M. (2005). Reforçar o valor regulador, formativo e formador da avaliação das aprendizagens. *Revista de Estudos Curriculares*, 3, 2, pp. 265-289.
- Slavin, R. (1984). Meta-analysis in education. How has it been used? *Educational Researcher*, 13, 8, pp. 6-15, 24-27.
- Slavin, R. (1986). Best-evidence synthesis: An alternative to meta-analysis and traditional reviews. *Educational Researcher*, 15(9), pp. 5-11.
- Slavin, R. (2004). Education research can and must address "what works" questions. *Educational Researcher*, 33(1), pp. 27-28.
- Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappa*, 86, 1, pp. 22-27.
- Stiggins, R. & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory into Practice*, 44, 1, pp. 11-18.
- Stiggins, R. & Conklin, N. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, NY: State University of New York Press.
- Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and learning*, pp. 133-146. London: Sage.
- Suri, H. (1999a). *A methodologically inclusive model for research synthesis*. Retirado em 14 de Setembro, 2006, de <http://www.aare.edu.au/99pap/sur99673.htm>
- Suri, H. (1999b). *The process of synthesising qualitative research: A case study*. Retirado em 14 de Setembro, 2006, de <http://www.latrobe.edu.au/aqr/offer/papers/HSuri.htm>
- Suri, H. (2002). *Essential features of methodologically inclusive research synthesis*. Retirado em 11 de Setembro, 2006, da base de dados ERIC.
- Suri, H. & Clarke, D. (1999). *Revisiting methods of literature synthesis*. Retirado em 11 de Setembro, 2006, da base de dados ERIC.
- Thompson, A. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grows (Ed.), *Handbook of Research in Mathematics Teaching and Learning*, pp. 127-146. New York: MacMillan.
- Traguil, L. (1988). A avaliação no ensino primário: Objectivos-instrumentos-métodos. *Aprender*, 11, pp. 59-64.
- Valadares, J. & Teodoro, V. (1988). Avaliação no ensino da Física. *Gazeta de Física*, 11, 3, pp. 103-105.

- Vasconcelos, C. (2002). Avaliação e intervenção nos métodos de estudo em alunos do 3.º ciclo do ensino básico. *Inovação*, 15, 1-2-3, pp. 103-116.
- Vasconcelos, C., Praia, J. & Almeida, L. (2004). Avaliação das estratégias de estudo dos alunos em ciências naturais: Construção e validação de uma escala. *Revista de Educação*, 12, 1, pp. 41-48.
- Vieira, J. D. (1992). Avaliação formativa: Uma experiência no 7º ano. *Quadrante*, 1, pp. 149-161.
- Weed, M. (2005, Janeiro). "Meta-Interpretation": A method for the interpretive synthesis of qualitative research. *Forum Qualitative Sozialforschung/Forum Qualitative Social Research* (On-line Journal), 6(1), Artigo 37. Retirado em 11 de Setembro, 2006, de <http://www.qualitative-research.net/fqs-texte/1-05/05-1-37-e.htm>

Referências Legislativas

- Lei n.º 46/86 de 14 de Outubro. *Diário da República* n.º 237, I Série.
- Decreto-Lei n.º 286/89 de 29 de Agosto. *Diário da República* n.º 198, I Série.
- Decreto-Lei n.º 6/01 de 18 de Janeiro. *Diário da República* n.º 15, I Série A.
- Despacho Normativo n.º 162/91 de 9 de Setembro. *Diário da República* n.º 244, I Série B.
- Despacho Normativo n.º 98-A/92 de 19 de Junho. *Diário da República* n.º 140, I Série B.
- Despacho Normativo n.º 338/93 de 21 de Outubro. *Diário da República* n.º 247, I Série B.
- Despacho Normativo n.º 30/01 de 22 de Junho. *Diário da República* n.º 166, I Série B.

Resumé

Dans ce travail, on élabore une synthèse interprétative du contenu des articles publiés entre 1985 et 2005 dans des revues portugaises à caractère scientifique en ce qui concerne le domaine de l'évaluation des apprentissages. On a été identifiées 19 revues et sélectionnés 59 articles qui ont été organisés dans des catégories telles que Évaluation Interne, Conceptions et Pratiques de Professeurs/Élèves et Méthodes et Techniques. L'analyse et synthèse interprétative ont permis de produire des conclusions telles que: a) seulement 35% des articles ont eu pour base la recherche empirique donnant des indices de manque d'un plan de recherche à propos d'évaluation des apprentissages; b) la majorité des articles est descriptive et prescriptive, plus fondés sur ce qu'il y a ou sur ce qui est fait que sur la réflexion théorique; c) les articles où il y a de la réflexion à propos de méthodes alternatives de recueil d'information sont réduits; et d) en général, l'évaluation formative, l'évaluation sommative et les relations entre elles sont discutées avec peu de profondeur et extension.

Malgré les progrès achevés, il est nécessaire de reconnaître les fragilités qui caractérisent la recherche et la production théorique dans le domaine de l'évaluation des apprentissages. Un domaine qui se reconnaît indispensable pour transformer et améliorer les actuelles réalités éducatives.

Abstract

This paper is an interpretative synthesis based upon the content of the articles on learning assessment that were published in Portuguese academic journals between 1985 and 2005. Nineteen journals have been identified and 59 articles have been selected and organized through the use of categories such as Internal Assessment, Teachers' and Students' Beliefs and Practices, and Methods and Techniques. Both the analysis and the interpretative synthesis yielded conclusions such as: a) since only 35% of the articles were based upon empirical data it seems that there is a need for a research agenda in the area of learning assessment; b) most articles are of a descriptive and prescriptive nature, that is, they seem to rely more on what exists or on what is done rather than on theoretical reflection; c) very few articles deal with some sort of reflection about alternative methods of collecting assessment information; and d) in general, both formative and summative assessment as well as the relationships between them are discussed with little theoretical elaboration.

Regardless the progresses already made, one needs to acknowledge the limitations that characterize both the empirical research and the theoretical construction on learning assessment. And this is a domain of paramount relevance if one wants to transform and to improve the current educational realities.

Anexo – 20 anos de artigos

	Data	Autor(es)	Título	Revista
1	1985	Boavida, J.	Análise das avaliações realizadas pelos alunos do ramo educacional da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.	Revista Portuguesa de Pedagogia
2	1986	Boavida, J., Lopes, M. C., & Vaz, P.	Avaliação: Tópicos para uma mudança de atitudes.	Revista Portuguesa de Pedagogia
3	1987	Boavida, J. & Vaz, P.	Auto-avaliação: Contributo para a sua revalorização no processo pedagógico.	Revista Portuguesa de Pedagogia
4	1988	Martins, I. & Cachapuz, A.	O uso de testes diagnósticos e a identificação de dificuldades na aprendizagem do tema Energia e Reacções Químicas (8.º ano de escolaridade).	Boletim da Sociedade Portuguesa de Química
5	1988	Valadares, J. & Teodoro, V.	Avaliação no ensino da Física.	Gazeta de Física
6	1988	Traguil, L.	A avaliação no ensino primário: Objectivos, instrumentos, métodos.	Aprender
7	1989	Abrantes, P.	Avaliação em Matemática: A necessidade de mudar de óptica.	Aprender
8	1989	Roldão, M. C.	O que faz mudar a avaliação?	Aprender
9	1990	Benavente, A.	Avaliação e inovação educacional: Notas e reflexões.	Inovação
10	1990	Carvalho, A.	Validação da tradução portuguesa do teste de compreensão de leitura do Institut Supérieur de Pédagogie de Hainaut.	Inovação
11	1990	Leal, L. & Abrantes, P.	Avaliação da aprendizagem/Avaliação na aprendizagem.	Inovação
12	1990	Rocha, N.	Avaliação formativa no ensino-aprendizagem de uma língua.	Inovação
13	1991	Pedrosa, M. A.	Avaliação e classificação da aprendizagem em Química: Alguns aspectos técnicos.	Boletim da Sociedade Portuguesa de Química
14	1992	Boavida, J. & Barreira, C.	Nova avaliação no ensino básico: Análise dos despachos 162/ME/91 e 98-A/92.	Revista Portuguesa de Pedagogia

	Data	Autor(es)	Título	Revista
15	1992	Carvalho, M. P.	Avaliação formativa por computador.	Quadrante
16	1992	Simão, A.M.	Construção de um instrumento para avaliação do desempenho dos alunos na tarefa de resumir.	Revista Portuguesa de Pedagogia
17	1992	Vieira, J. D.	Avaliação formativa: Uma experiência no 7º ano.	Quadrante
18	1993	Boavida, J. & Barreira, C.	Nova avaliação: Novas exigências.	Inovação
19	1993	Oliveira, I. & Pereira, J.	Avaliação diagnóstico: Desenvolvimento de um teste de desempenho em Matemática.	Quadrante
20	1993	Pacheco, J.	O novo sistema de avaliação dos alunos do ensino básico: Do contexto europeu ao contexto de experimentação dos programas e das mudanças curriculares.	Revista Portuguesa de Educação
21	1993	Leitão, J.	Um teste diferente para o 12.º Ano de Química.	Boletim da Sociedade Portuguesa de Química
22	1994	Caria, T. (Org.)	Sistema de avaliação dos alunos do ensino básico.	Educação, Sociedade & Culturas
23	1994	Carvalho, L.	Avaliação das aprendizagens em Educação Física.	Boletim da Sociedade Portuguesa de Educação Física
24	1994	Castro, M.L. & Pereira, M.	O relatório como um meio de avaliação formativa no 3.º ciclo.	Revista Portuguesa de Educação
25	1994	Duarte, R.S.	Alguns aspectos das concepções e práticas avaliativas dos professores de uma escola do 2.º ciclo do ensino básico.	Boletim da Sociedade Portuguesa de Educação Física
26	1994	Gonçalves, C.	Avaliação do processo de ensino-aprendizagem em Educação Física.	Boletim da Sociedade Portuguesa de Educação Física

	Data	Autor(es)	Título	Revista
27	1995	Ferreira, J., Almeida, A., Cunha, L., Rolo, A. & Vaz, J.	Uma análise dos resultados da prova específica de Física de 1994: Distritos de Braga e Viana do Castelo.	Gazeta de Física
28	1995	Ramalho, G.	Participação dos estudantes portugueses de 9 e 13 anos de idade no Second International Assessment of Educational Progress – Matemática.	Quadrante
29	1997	Amaro, G.	Qualidade em educação: A avaliação externa das aprendizagens dos alunos em Portugal.	Inovação
30	1997	Carvalho, J.	Uma análise dos exames de Inglês de 1995/1996 na perspectiva da qualidade do ensino e da aprendizagem.	Inovação
31	1997	Henriques, H.	Avaliação aferida: Conceção e construção de um instrumento de avaliação da aprendizagem em Matemática – 1.º ciclo.	Inovação
32	1997	Pereira, M. J.	Opiniões de educadores de infância de Castelo Branco face à avaliação na educação pré-escolar.	Educare-Educere
33	1997	Marques, R.	Avaliação do rendimento escolar no ensino básico.	Revista da Escola Superior de Educação de Santarém
34	1998	Fernandes, M.	A mudança de paradigma na avaliação educacional.	Educação, Sociedade & Culturas
35	1998	Paulo, A. & Santos, F.	A formação inicial de professores em contexto de investigação-acção: Concepções e práticas de avaliação pedagógica.	Revista de Educação
36	1998	Peixoto, M. T. (1998).	A avaliação no 1.º ciclo do ensino básico.	Revista da Escola Superior de Educação de Viana do Castelo
37	1999	Carvalho, J.	Provas aferidas de Inglês: Construção de um quadro de referência.	Inovação
38	2000	Martins, A. & Silva, J. C.	Exames nacionais do ensino secundário: Algumas notas.	Gazeta de Matemática

	Data	Autor(es)	Título	Revista
39	2000	Nabuco, M. E.	A avaliação como processo central da intencionalidade educativa. A avaliação das crianças dos 0 aos 6 anos de idade.	Da Investigação às Práticas-Estudos de Natureza Educacional
40	2000	Pais, P.	Práticas classificativas de professores do ensino secundário: Significados e valores.	Inovação
41	2001	Lobo, V.	A nefasta artificialidade das provas de “exames” do 12.º ano.	Boletim da SPQ
42	2001	Barreira, C.	Duas estratégias complementares para a avaliação das aprendizagens: A avaliação formadora e a avaliação autêntica.	Revista Portuguesa de Pedagogia
43	2001	Freitas, C. V.	Contributo para a história da avaliação educacional em Portugal: Os anos 70.	Revista Portuguesa de Educação
44	2002	Barreira, C	Reorganização dos currículos no ensino básico e suas implicações na avaliação das aprendizagens.	Revista Portuguesa de Pedagogia
45	2002	Rosado, A. & Dias, L.	Caracterização do pensamento dos professores de Educação Física relativamente à avaliação das aprendizagens.	Ludens
46	2002	Vasconcelos, C.	Avaliação e intervenção nos métodos de estudo em alunos do 3.º ciclo do ensino básico.	Inovação
47	2003	Santos, L.	A avaliação em documentos orientadores para o ensino da Matemática: Uma análise sucinta.	Quadrante
48	2003	Silva, C., Rosado, A. & Freitas, C. V.	Os alunos e a avaliação: Concepções dos alunos do ensino básico relativamente à avaliação das aprendizagens.	Revista Portuguesa de Investigação Educacional
49	2003	Graça, M.	Avaliação da resolução de problemas: Que relação entre as concepções e as práticas lectivas dos professores?	Quadrante
50	2004	Abreu, G., Nunes, T. & Rosa, J.	Desenvolvimento da escrita: Criação e validação de um teste de avaliação da escrita do 2.º ao 4.º ano. Da investigação às práticas.	Da Investigação às práticas - Estudos de Natureza Educacional

	Data	Autor(es)	Título	Revista
51	2004	Eça, T.	Avaliação externa nas artes visuais em Portugal e Inglaterra.	Educare-Educere
52	2004	Gazeta de Matemática	Inquérito: Mais exames no sistema de ensino?!	Gazeta de Matemática
53	2004	Fernandes, J., Alves, M.P. & Arantes, M. J.	Impacto de uma experiência de formação nas concepções e práticas de avaliação de professores estagiários.	Revista de Estudos Curriculares
54	2004	Neves, A., Jordão, A. & Santos, L.	Investigação publicada em Portugal, entre 1971 e 2003, acerca da avaliação das aprendizagens: Seu levantamento e análise.	Revista Portuguesa de Investigação Educacional
55	2004	Vasconcelos, C., Praia, J. & Almeida, L.	Avaliação das estratégias de estudo em ciências naturais: Construção e validação de uma escala.	Revista de Educação
56	2004	Oliveira, F. & Gaspar, M.F.	Olhares sobre a avaliação em educação pré-escolar: As opiniões e as práticas dos educadores de infância.	Revista Portuguesa de Pedagogia
57	2005	Barreira, C.	Soluções para a prática da avaliação formativa.	Revista Portuguesa de Pedagogia
58	2005	Menino, H.	Avaliação das aprendizagens em Matemática: Uma experiência de utilização do portefólio no 2.º ciclo.	Educação & Comunicação
59	2005	Simão, A. M.	Reforçar o valor regulador, formativo e formador da avaliação das aprendizagens.	Revista de Estudos Curriculares