Título

Para um Enquadramento Teórico da Avaliação Formativa e da Avaliação Sumativa das Aprendizagens Escolares

Resumo

A avaliação formativa e a avaliação sumativa continuam a ser conceitos estruturantes no âmbito chamada Avaliação das Aprendizagens cuja conceitualização tem evoluído desde a sua criação, em 1967, por Michael Scriven. Na verdade, a definição destes conceitos há cerca de meio século, pouco ou nada tem a ver com a sua definição atual. A construção teórica tem contribuído para a clarificação e consolidação destes dois conceitos-chave. No entanto, persiste ainda alguma confusão acerca da natureza, dos propósitos e das caraterísticas definidoras de cada uma daquelas modalidades de avaliação. Este texto foi pensado para discutir o enquadramento teórico da avaliação formativa e da avaliação sumativa, clarificar as suas definições e analisar as suas relações. Foram consultados uma diversidade de trabalhos produzidos, sensivelmente, nos últimos 30 anos. Verificou-se que houve assinaláveis progressos no domínio da construção teórica daqueles conceitos, das relações entre si e das suas conexões com as práticas pedagógicas e com as teorias da aprendizagem. Porém, a análise realizada mostrou que continua a ser necessário compreender melhor as relações entre as aprendizagens e as avaliações e muitos outros aspetos relevantes tais como as práticas dos professores e as suas relações com as disciplinas que lecionam.

Palavras-chave: Avaliação Formativa; Avaliação Sumativa; Avaliação Pedagógica; Teoria da Avaliação; Avaliação Educacional.

Abstract

Formative and summative assessments are still key-concepts within the so-called Learning Assessment domain and its conceptualization have evolved since 1967, when Michael Scriven created them. In fact, the definition of these concepts half a century ago, has very little, or even nothing, to do with their current definition. Theory construction has contributed towards the clarification and consolidation of these key-concepts. However, there still is some confusion about the nature, the purposes and defining characteristics of each one of those learning assessment approaches. This text was thought to discuss a conceptual framework of both formative and summative assessments, to clarify their definitions and to analyse relationships between them. A variety of works, which were essentially written in the last 30 years, have been consulted. One could verify that there have been important progresses concerning the theoretical construction of those concepts as well as the relationships between them and their connections with pedagogical practices and learning theories. Nevertheless, the analysis made also revealed that it is still necessary to better understand the relationships between learning and assessments and many other relevant issues such as teachers' practices and their relationships with the subject-matters that they teach.

Key-words: Formative Assessment; Summative Assessment; Pedagogical Assessment; Assessment Theory; Educational Evaluation.

Para um Enquadramento Teórico da Avaliação Formativa e da Avaliação Sumativa das Aprendizagens Escolares

Domingos Fernandes[·]

Introdução

A análise da literatura publicada no domínio da chamada avaliação das aprendizagens mostra que, apesar dos significativos desenvolvimentos que têm permitido fundamentar, clarificar e compreender melhor a sua natureza, os seus propósitos e as suas relações com outros domínios do conhecimento, tais como as teorias da aprendizagem e as teorias curriculares e do ensino, a construção teórica continua a ser sentida como uma necessidade relevante (e.g., Baird *et al*., 2017; Berlak, 1992a; Black e Wiliam, 2006a, 2018; Fernandes, 2006, 2008a; Gipps, 1994; Perrenoud, 1998; Stiggins, 2017; Wiliam, 2017).

Michael Scriven, um dos mais destacados e produtivos teóricos no domínio do conhecimento da avaliação, referiu que a construção e a afirmação da avaliação como disciplina científica enfrentou dificuldades cuja origem foi precisamente identificada no domínio prático ou aplicado da avaliação das aprendizagens onde, segundo este autor, havia problemas vários de natureza conceitual que era necessário enfrentar (Scriven, 1994, 2003; Fernandes, 2008a). Um desses problemas tinha precisamente a ver com a necessidade de clarificar os conceitos de avaliação formativa e de avaliação sumativa. Mais recentemente, ainda relativamente a esta necessidade, Stiggins (2017) referiu que há uma significativa falta de compreensão quer por parte dos responsáveis políticos, quer por parte dos intervenientes mais diretos no processo educativo, como é o caso dos gestores e dos próprios professores, acerca das avaliações internas e externas das aprendizagens. Para este autor é esta a principal razão que pode explicar as comprovadas dificuldades dos sistemas escolares em integrarem a avaliação formativa na vida do dia a dia das escolas e das salas de aula. E esta é uma questão fundamental pois a melhoria da qualidade das aprendizagens e da educação em geral está fortemente relacionada com a qualidade das avaliações. Em particular, a pesquisa realizada nas últimas décadas tem comprovado sucessivamente que a avaliação pedagógica, sobretudo a de natureza formativa, contribui muito para que os alunos aprendam mais e melhor, com mais profundidade e com mais compreensão (Black e Wiliam, 1998; Fernandes, 2008b).

[·] Universidade de Lisboa, Instituto de Educação, Lisboa, Portugal.

É necessário que os propósitos da avaliação pedagógica sejam bem compreendidos por todos os envolvidos. Por um lado, é necessário avaliar para apoiar e para melhorar as aprendizagens dos alunos e, por outro lado, também é necessário avaliar para que se possa fazer uma súmula, um balanço ou um ponto de situação relativamente à qualidade das aprendizagens realizadas pelos alunos num dado momento ou após um dado período de tempo. E isto exige o desenvolvimento de um complexo conjunto de relações entre as avaliações, as aprendizagens e o ensino que contribua para que todos alunos possam aprender (e.g., Fernandes, 2008b; Luckesi, 1995, Vasconcellos, 2005). Note-se que, para efeitos deste texto a designação *avaliação pedagógica* refere-se a todas as avaliações, formativas e sumativas, que se desenvolvem essencialmente no contexto das salas de aula e são da integral responsabilidade dos professores e dos seus alunos. Na verdade, a principal motivação para a elaboração deste trabalho foi a ideia da construção teórica da avaliação pedagógica, processo estreitamente articulado com os processos de aprendizagem e de ensino cujo propósito mais fundamental é, precisamente, apoiar e melhorar as aprendizagens dos alunos e o ensino dos professores.

Nestes termos, o principal objetivo deste capítulo é discutir a construção teórica da avaliação pedagógica, procurando, por um lado, fazer um ponto de situação do atual estado de coisas e, por outro, dar contributos para a clarificação e consolidação conceituais, sobretudo no que se refere à avaliação formativa, à avaliação sumativa e às relações que entre si se podem estabelecer.

O texto foi organizado em cinco partes assim distribuídas: a) *Introdução*; b) *Construção Teórica: Os Anos Iniciais*; c) *Construção Teórica: A Ênfase na Pesquisa Empírica*; d) *Conclusões*; e e) *Referências*. Trata-se de uma estrutura muito simples que obedeceu, de certo modo, a um critério cronológico, cujo principal objetivo foi o de realçar os anos 90 do século passado (*Os Anos Iniciais*) como tendo sido bastante relevantes nas ruturas epistemológicas que se verificaram na construção teórica da avaliação pedagógica. A terceira parte prossegue a discussão da construção teórica que se tem realizado mais recentemente e em que a realização de investigação empírica tem tido um papel bastante assinalável. Finalmente, apresentam-se algumas conclusões e a lista de referências utilizadas.

Construção Teórica: Os Anos Iniciais

Um dos primeiros esforços conhecidos na literatura de teorização destes conceitos teve origem num colóquio sobre avaliação das aprendizagens que teve lugar na Universidade de Genève em 1978 e cujas atas foram posteriormente publicadas num livro que influenciou

significativamente, pelo menos no contexto europeu, a investigação, o desenvolvimento teórico e as práticas pedagógicas (Allal, Cardinet e Perrenoud, 1979). (Note-se que o colóquio se realizou cerca de uma década após a criação dos conceitos de avaliação formativa e de avaliação sumativa.) O livro surgiu numa altura em que as conceções de aprendizagem, avaliação e ensino eram fortemente dominadas pelas perspetivas associacionistas e behavioristas e em que as ideias dominantes sobre o currículo e a avaliação se baseavam nas teorias da eficiência social e da medida. Consequentemente, os conceitos de avaliação formativa e de avaliação sumativa, elaborados por Scriven (1967), há mais de meio século, estavam então fortemente marcados por essas perspetivas. Os textos constantes no livro elaboraram e aprofundaram estes e outros conceitos (e.g., regulação, auto-regulação, diferenciação) a partir de novas e alargadas visões sobre o curículo, a aprendizagem e a avaliação, com contributos relevantes da sociologia e do cognitivismo. Como se discutirá com mais detalhe ao longo deste texto, foi nesta altura que se considerou que o principal propósito da avaliação formativa era apoiar os esforços de aprendizagem dos alunos e, nesse sentido, deveria estar integrada nos processos de ensino e de aprendizagem.

Nos anos 90 do século passado, os anos da avaliação ou *o tempo da avaliação* como então designei essa época (Fernandes, 1992), uma diversidade de autores volta a insistir de forma contundente na necessidade de se desenvolver uma teoria da avaliação que rompesse com os fundamentos ontológicos, epistemológicos e metodológicos de um paradigma que até então vinha sustentando uma avaliação intrinsecamente associada à produção de medidas e de classificações. Esses autores sublinhavam a necessidade de um paradigma de avaliação que, no fundo, estivesse mais preocupado com as questões de natureza pedagógica, nomeadamente com as relações entre o ensino, as aprendizagens e as avaliações e com a necessidade de a avaliação ser, no essencial, um processo destinado a apoiar e a melhorar as aprendizagens (e.g. Berlak *et al*., 1992; Dwyer, 1998; De Ketele, 1993; Gifford e O' Connor (Eds.), 1992; Gipps, 1994). Ainda que a construção de uma teoria da avaliação das aprendizagens ou da avaliação formativa pudesse não ser a prioridade destes autores, a verdade é que, explícita ou implicitamente, todos reconheciam a sua necessidade, elaborando de forma mais ou menos profunda, acerca dos seus fundamentos teóricos e práticos.

Por exemplo, para Berlak *et al*. (1992) uma teoria da avaliação tinha de ser construída com base em pressupostos epistemológicos e ontológicos mais consentâneos com racionalidades críticas e sociocríticas e com o paradigma construtivista, tal como discutido por Guba e Lincoln (1994). Introduzindo aqui um parêntesis, interessa referir nesta altura que, para estes autores, as opções paradigmáticas resultam das respostas que forem dadas a cada

uma das seguintes questões: a) Qual é a forma e a natureza da realidade? O que é que existe que pode ser conhecido? (questões ontológicas); b) Qual é a natureza da relação entre aquele que conhece, ou que quer conhecer, e aquilo que se julga pode ser conhecido? Como é que aquele que quer conhecer se relaciona com a produção de conhecimento? (questões epistemológicas); e c) Como é que aquele que quer conhecer procede para descobrir aquilo que acredita que pode ser conhecido? (questão metodológica). Estas questões são obviamente interdependentes pois a resposta que se der a uma delas vai condicionar as respostas que se podem dar às outras. Por exemplo, se se responder à questão ontológica que há uma realidade objetiva suscetível de ser compreendida e/ou apreendida na sua totalidade (paradigma positivista), então a resposta à questão epistemológica não será, com certeza, a de considerar que o conhecimento resulta da interação entre o investigador e os participantes na investigação (paradigma construtivista). Esta discussão é relevante para o desenvolvimento das perspetivas e das práticas de avaliação pois só a sua compreensão pode contribuir para que se discutam opções de fundo que podem, por exemplo, levar-nos a questionar o paradigma psicométrico, muito associado ao positivismo, no âmbito do qual a avaliação é uma medida objetiva, uma ciência exata, que nos dá uma descrição rigorosa e acima de qualquer suspeita, do que os alunos sabem e são capazes de fazer. Ou, por outro lado, a considerar a relevância de se adotar o paradigma contextual (Berlak, 1992) ou o paradigma construtivista (Guba e Lincoln, 1994), que fundamentam uma avaliação em que, assumidamente, se têm em conta os contextos de ensino, aprendizagem e avaliação, se tem em conta a natureza interativa dos processos de avaliação e em que as atitudes, os comportamentos em geral, as capacidades e os conhecimentos escolares são considerados inseparáveis e, consequentemente, avaliados de forma integrada. Apesar deste contraste entre os paradigmas que é importante para as opções pedagógicas, sobretudo essas, que se podem tomar, a verdade é que, há várias concepções de avaliação que coexistem nos sistemas escolares e, muito provavelmente, continuarão a coexistir. Mas voltemos à construção teórica nos anos 90 do século passado.

Há cerca de três décadas, Berlak (1992a, 1992b) e Archbald e Newmann (1992), entre outros investigadores construtivistas não ortodoxos, questionavam se o paradigma positivista iria ser pura e simplesmente posto de parte para dar lugar ao paradigma construtivista. Na verdade, estes pesquisadores conjeturavam então a possibilidade de os principais fundamentos da teoria da avaliação poderem evoluir através de uma integração ou articulação entre aqueles dois paradigmas. De facto, o que se verifica é que, por exemplo, a conceção de avaliação como medida e a conceção de avaliação como negociação, convivem de forma mais ou menos

articulada, mais ou menos equilibrada, nos atuais sistemas escolares e sociais. A primeira, herdeira das perspetivas psicométricas e positivistas, com particular expressão nas avaliações externas e a segunda inspirada nas perspetivas construtivistas e contextuais, mais comum no contexto das salas de aula. Ou seja, parece ter havido uma evolução no âmbito da qual os dois paradigmas se foram progressivamente articulando e não a rejeição pura e simples de qualquer um deles. Um exemplo interessante desta articulação no contexto da sala de aula, tem a ver com a evolução do conceito de avaliação sumativa, hoje muito valorizada no papel que pode desempenhar para apoiar o ensino e a aprendizagem como veremos mais adiante. Em suma, parece existir uma certa tendência para que a construção teórica vá evoluindo através de formas de articulação entre diferentes paradigmas como parece estar a acontecer nos sistemas escolares do século XXI (Gardner, 2006, Madaus *et al.*, 2000; Mislevy *et al.*, 2003; Torrance, 2003).

Prosseguindo nos anos 90, Gipps (1994) contribuiu igualmente para a teorização da avaliação formativa contrastando os fundamentos epistemológicos e ontológicos da psicometria, com influências significativas do paradigma positivista e das visões da avaliação como medida, com os fundamentos de uma avaliação formativa, baseados nas perspetivas cognitivistas, construtivistas e socioculturais da aprendizagem e da avaliação. Para esta autora, o desenvolvimento teórico da avaliação formativa e das avaliações pedagógicas em geral tem de estar relacionado com a sua qualidade e, por isso, propôs critérios de qualidade alternativos aos conceitos psicométricos de validade e de fiabilidade tais como a *fidelidade curricular*, a *comparabilidade*, a *credibilidade pública*, a *descrição dos contextos* e a *equidade*. Porém, ainda que não seja fácil encontrar alternativas para os conceitos de validade e de fiabilidade utilizados na psicometria, assegurar a qualidade das avaliações formativas deverá passar por outras formulações daqueles conceitos (ver, por exemplo, Guba e Lincoln, 1989). Relativamente a esta importante questão, Stobart (2006) e Black & Wiliam (2006b) atribuem diferentes significados à validade e à fiabilidade das avaliações formativa e sumativa, referindo Stobart que uma avaliação formativa será válida se contribuir efetivamente para a melhoria das aprendizagens e que a validade tem de ser considerada ao nível do contexto das escolas e das salas de aula, pontuando aqui as aprendizagens dos alunos, e ao nível que está para além das escolas, em que deverão ser consideradas as consequências políticas e sociais das avaliações.

Ainda nos finais dos anos 90, dois investigadores do Reino Unido, Paul Black e Dylan Wiliam, publicaram um seminal e incontornável trabalho em que se reforçava a ideia de que o desenvolvimento e a melhoria das práticas de avaliação formativa tinham de ter em conta a

construção teórica, a produção de orientações práticas solidamente fundamentadas e descrições detalhadas e adequadas dos fenómenos de interesse no contexto das salas de aula (Black e Wiliam, 1998). Daqui se infere claramente que a construção teórica é indissociável das práticas que se desenvolvem nas salas de aula pelos professores em interação com os seus alunos. Na verdade, através de publicações posteriores, aqueles autores e seus colaboradores bem como outros investigadores (e.g., Black *et al*., 2002, 2003; Gardner, 2006) evidenciaram a relevância da análise propositada das práticas de avaliação em contexto de sala de aula para a construção teórica.

Em suma, nos anos 90, os anos do *tempo da avaliação*, com a emergência do cognitivismo, do construtivismo e das teorias socioculturais da aprendizagem verificou-se uma rutura epistemológica com a avaliação fundamentada nas teorias associacionistas e behavioristas. Na Europa, essa rutura iniciou-se no início dos anos 80 (Allal, Cardinet e Perrenoud, 1979), em que se defendeu a integração da avaliação nos processos de ensino e aprendizagem e a conceção da avaliação para melhorar e para aprender. Porém, foi ainda um período em que a construção teórica foi mais baseada na retórica do que numa base empírica, a partir da descrição do que acontecia nas salas de aula.

Construção Teórica: A Ênfase na Pesquisa Empírica

Já no presente século, Black & Wiliam (2006a), elaboraram mais profundamente a questão da teorização, referindo que o desenvolvimento de uma teoria da avaliação formativa tinha de estar intrinsecamente associado ao estudo das ações que professores e alunos desenvolvem no contexto das salas de aula. Assim, por exemplo, sublinharam a importância de se estudar como é que os professores se relacionam com a disciplina que ensinam pois consideraram que as suas práticas de avaliação estão relacionadas com a natureza da disciplina e com as concepções epistemológicas que têm acerca dela. Referiram também que era fundamental estudar as interações que se desenvolvem nas salas de aula entre os professores e os alunos, destacando o papel do *feedback* e a relevância das perspetivas socioculturais inspiradas no trabalho de Vygotsky. Para além disso, sublinharam a necessidade de investigar o papel dos professores e das suas ações na regulação das aprendizagens dos alunos e o papel que estes devem igualmente desempenhar nesse processo, destacando a importância do feedback, da autoavaliação e das dimensões metacognitiva e afetiva da aprendizagem.

Nessa altura tive a oportunidade de referir que a construção teórica no domínio da avaliação formativa tinha de passar por cinco ações consideradas fundamentais: a) *clarificar*,

em que se chamava a atenção para a importância de clarificar as designações utilizadas na literatura acerca da avaliação formativa; b) *integrar*, em que se defendia a necessidade de integrar contributos teóricos e conceituais oriundos de diferentes tradições de pesquisa, tais como a francófona, a anglo-saxónica, a iberoamericana e a escandinava; c) *definir*, em que estava implícita a importância de definir detalhadamente os conceitos fundamentais e estruturantes no domínio da avaliação pedagógica; d) *teorizar*, em que se apresentava e discutia uma diversidade de elementos que deveriam ser considerados no desenvolvimento de uma teoria da avaliação formativa; e e) *refletir*, em que foi feito um esforço de avaliação e síntese acerca dos aspetos mais essenciais a ter em conta, nomeadamente no que se referia a possíveis evoluções nos domínios da investigação empírica e da construção teórica (Fernandes, 2006). No fundo, o que parece ter sido mais significativo neste trabalho foi, por um lado, o reconhecimento da necessidade de se desenvolverem esforços para " (...) obter retratos tão nítidos quanto possível das realidades da avaliação nas salas de aula, pois, através deles, poderemos (…) ir elaborando a teoria" (Fernandes, 2006, p. 36). Assim, sugeria-se que a pesquisa a realizar no contexto das salas de aula deveria basear-se numa perspetiva tão integrada quanto possível das realidades e fenómenos relacionados com as aprendizagens, a avaliação e o ensino. Para tal, deveriam ter-se em conta o currículo, os seus domínios e as tarefas ou propostas de trabalho relativas a cada um, a definição de critérios, a seleção de sistemas de recolha e síntese da informação e uma diversidade de aspetos relacionados com a organização e funcionamento pedagógico das dinâmicas da sala de aula (e.g., estrutura das aulas, participação dos alunos, papel do professor). Por outro lado, considerou-se ser necessário estudar uma diversidade de relações envolvendo os professores, os alunos e as tarefas ou propostas de trabalho tendo, naturalmente, como contexto o *ambiente* resultante dos processos de avaliação, ensino e aprendizagem. Nestas condições, foram definidas questões que deveriam orientar pesquisas que pudessem contribuir para a construção teórica tais como:

1. Que critérios de avaliação utilizam os professores e como os articulam com o desenvolvimento das aprendizagens estruturantes que os alunos têm de realizar?
2. Como se poderão caraterizar os papéis dos alunos e dos professores nos processos de ensino, de aprendizagem e de avaliação?
3. Como é que o *feedback* é utilizado por professores e alunos? Qual a sua natureza? De que formas é distribuído pelos alunos?
4. Como é que alunos e professores regulam os processos de ensino e de aprendizagem? Que estratégias são mais utilizadas? Qual a sua natureza? (Fernandes, 2006, pp. 41-42).

As respostas a estas e a outras questões de índole semelhante, foram consideradas fundamentais para a construção de uma teoria da avaliação formativa capaz de fundamentar e orientar as práticas pedagógicas dos professores. Num trabalho posterior, discuti questões teóricas, práticas e metodológicas que deveriam ser consideradas tendo em conta a necessidade de integrar os processos de ensino, aprendizagem e avaliação (Fernandes, 2011). Uma das questões discutidas teve precisamente a ver com a dificuldade em articular estes processos dada a natureza específica das disciplinas científicas que os suportam com as suas questões de investigação próprias, as suas tradições epistemológicas e metodológicas, os seus objetos e propósitos. Esta é, com certeza, uma das razões que pode justificar as dificuldades de articulação teórica e prática entre a avaliação, o ensino e a aprendizagem. James (2006), referiu que um desafio a enfrentar nos próximos anos seria o de trabalhar no sentido da integração das teorias da aprendizagem, do ensino e da avaliação. Ou seja, criar uma teoria das teorias, uma metateoria. Consequentemente, no domínio das metodologias, sugeriu que a unidade de análise a considerar nas pesquisas deveria ser a sala de aula como um todo e não apenas o professor e/ou as suas práticas e/ou as suas perceções. Só desse modo, de acordo com James, poderá fazer sentido falar-se na construção de uma teoria integradora, uma metateoria, como se referiu.

Baird *et al*. (2017) discutiram como é que as teorias da aprendizagem e da avaliação se têm vindo a influenciar mutuamente ao longo dos anos defendendo que se as avaliações devem servir as finalidades da educação, então as teorias da aprendizagem e da avaliação deverão desenvolver-se de forma mais articulada. O que talvez seja singular na perspetiva destes autores é o facto de, na sua discussão acerca das relações entre aquelas teorias, terem considerado as avaliações em larga escala, por um lado, e a chamada avaliação para as aprendizagens por outro. Os domínios do currículo indicam o que os alunos têm de aprender e, por isso, constituem finalidades que ajudam a determinar o que os alunos devem saber e ser capazes de fazer. Consequentemente, as avaliações deverão refletir isso mesmo, isto é, identificar quando é que as aprendizagens ocorrem.

As avaliações internacionais em larga escala (AILE), como o *Programme for Internacional Student Assessemnt* (PISA), têm em conta teorias da aprendizagem e baseiam-se na psicometria. Para Baird *et al.* (2017) a contribuição destes programas internacionais de avaliação em larga escala para a construção das teorias da aprendizagem tem sido modesta. A sua contribuição tem sido mais ao nível do que, e como, se aprende, ou deve aprender, através da sua influência nas políticas públicas de educação. Consequentemente, estes autores consideram que ainda há espaço para aprofundar o estudo das relações entre a avaliação e a

aprendizagem no caso das AILE. Antes de prosseguir, interessa discutir os conceitos de Avaliação Formativa (AF) que, para muitos autores, talvez a maioria, é sinónimo de Avaliação para as Aprendizagens (ApA), e de Avaliação Sumativa (AS), para muitos equivalente a Avaliação das Aprendizagens (AdA).

Como se disse, a ApA e a AF são, para muitos autores, conceitos indiferenciáveis ou equivalentes. No entanto, se pensarmos na AF, tal como era entendida nos anos 70, através de Bloom, Hastings e Madaus (1971), alguns anos após a criação do conceito por Scriven (1967), parece fazer todo o sentido sublinhar que aquelas designações se referem a dois conceitos substancialmente diferentes. Na verdade, na sua primeira conceitualização, a AF era pontual, baseada em testes de papel e lápis, realizada no final de uma dada unidade de ensino, não integrada ou articulada com os processos de ensino e aprendizagem, centrada nos resultados e, num certo sentido, funcionava como uma espécie de preparação dos alunos para o teste de avaliação sumativa (AS) que ocorria após uma qualquer forma de AF, normalmente o chamado *teste formativo*. Tratava-se, como referiu Allal (1986), de uma AF de regulação retroativa, ou seja, não havendo uma avaliação de natureza contínua que acompanhasse o processo de ensino e de aprendizagem, eventuais dificuldades experimentadas pelos alunos ao longo do processo só seriam detetadas no final da unidade, após a realização da AF. Era uma AF muito elementar e redutora, baseada nas conceções da psicologia associacionista e do behaviorismo acerca da aprendizagem e do ensino.

Se, no entanto, pensarmos na AF tal como passou a ser entendida a partir dos anos 80 e, sobretudo, dos anos 90, então talvez não faça muito sentido considerar que a ApA e a AF são conceitos distintos. Na verdade, a partir dos anos 80, uma diversidade de investigadores contribuiu para que o conceito de AF fosse sendo enriquecido com a emergência das perspetivas cognitivistas, construtivistas, socioculturais e social construtivistas da aprendizagem (e.g., Allal, 1986; De Ketele, 1986; Berlak *et al*., 1992; Cardinet (1991); Dwyer, 1998; Gipps, 1994; Harlen e James, 1997; Sadler, 1998). Consequentemente, a AF passou a ser entendida como sendo tendencialmente contínua, articulada com o ensino e a aprendizagem, orientada para a distribuição de *feedback* de elevada qualidade a todos os alunos, acompanhando os processos que utilizavam para aprender, utilizando uma diversidade de processos de recolha de informação e promovendo dinâmicas de regulação e de auto-regulação das aprendizagens, assim como de auto-avaliação, co-avaliação e de avaliação inter-pares. Por isso, os alunos passaram a ser considerados como participantes ativos quer nos processos de aprendizagem, quer nos processos de avaliação. Nestes termos, o principal propósito da AF passou a estar centrado no apoio ao desenvolvimento das aprendizagens por

parte dos alunos, ou seja, a AF passou a ser um processo que, através do *feedback*, deveria permitir que os alunos aprendessem mais e, sobretudo, melhor, com mais compreensão. Assim, o conceito tornou-se muito mais denso e sofisticado pois passou a mobilizar, a integrar e a utilizar uma diversidade de contributos das *novas* teorias da aprendizagem e da avaliação que emergiram a partir do desenvolvimento do cognitivismo, do construtivismo e das visões socioculturais da aprendizagem.

Baird *et al*. (2017), no entanto, ainda que reconheçam que a ApA e a AF têm o propósito comum de melhorar as aprendizagens, consideram que há uma diferença de foco entre ambas. A ApA, na sua opinião, centra-se mais nos processos de auto-regulação e de autonomia dos alunos, sendo tendencialmente contínua e informal. Assim, tal como estes autores referem, pode distinguir-se a ApA das conceções de AF de certas correntes de investigadores norte-americanos, ainda influenciados pelas perspetivas behavioristas da aprendizagem. Continuo a pensar, no entanto, que a distinção não parece justificar-se tendo em conta a caraterização contemporânea de AF que acima se esboçou.

O que parece ser relevante discutir é se a introdução das expressões Avaliação *para as* Aprendizagens e Avaliação *das* Aprendizagens, que de algum modo vieram substituir as expressões Avaliação Formativa e Avaliação Sumativa, têm ou não contribuído para a necessária clarificação teórica e conceitual. Particularmente no que se refere aos propósitos e processos utilizados e também a um problema que se vinha arrastando há anos relacionado com o facto de o termo Avaliação Formativa ser utilizado com pouco critério, de tal forma que qualquer avaliação era considerada formativa ainda que, muitas vezes, não passasse de uma avaliação de *intenção* ou de *vontade* formativa. Noutros casos, designava-se como formativa uma avaliação muito próxima da conceção behaviorista dos anos 70 o que me levou a sugerir uma designação – *Avaliação Formativa Alternativa* – precisamente para sublinhar que a conceção contemporânea de avaliação formativa é alternativa àquela conceção inicial (Fernandes, 2005). Assim, a designação *Avaliação para as Aprendizagens* (ApA) parece ser preferível pois chama a atenção para o seu propósito mais fundamental: apoiar as aprendizagens dos alunos. Avalia-se para que os alunos possam aprender melhor. E, assim sendo, a avaliação é entendida como um processo eminentemente pedagógico e intrinsecamente relacionado com as aprendizagens dos alunos.

Não será fácil definir de forma sucinta o conceito de ApA uma vez que são inúmeras as implicações e relações com as práticas pedagógicas. Por isso, parece-me que a melhor forma de o fazer talvez seja enumerar as suas principais caraterísticas como se fez um pouco mais acima neste texto e como tenho procurado fazer noutros trabalhos em que tenho discutido o

conceito de avaliação formativa (e.g. Fernandes, 1993, 2005, 2008a, 2008b). Assim, o *Assessment Reform Group* (ARG) que, no Reino Unido, criou aquela designação, propôs uma definição simples para a ApA que aqui se traduz livremente:

> A Avaliação para as Aprendizagens é um processo de recolha e interpretação de evidências que professores e alunos utilizam para determinar em que situação se encontram os alunos, onde se pretende que eles cheguem e qual a melhor forma de lá chegarem (ARG, 2002, pp 1-2).

Porém, talvez por reconhecer a dificuldade de caraterizar fielmente a densidade e complexidade da perspetiva de aprendizagem (e de avaliação) que está subjacente a esta definição, o ARG definiu 10 princípios que deverão ser considerados inseparáveis daquela definição e que também se traduzem livremente:

> A Avaliação para as Aprendizagens deve
> 1. integrar uma planificação eficaz do ensino e da aprendizagem.
> 2. estar focada em como os alunos aprendem.
> 3. ser reconhecida como central nas práticas na sala de aula.
> 4. ser entendida como uma capacidade profissional fundamental para os professores.
> 5. ser sensível e construtiva pois qualquer avaliação tem um impacto emocional.
> 6. ter em conta a importância da motivação dos alunos.
> 7. promover o empenho para se alcançarem as finalidades e a compreensão dos critérios através dos quais se avalia a sua consecução.
> 8. orientar os alunos para que aprendam como melhorar.
> 9. desenvolver as capacidades de autoavaliação dos alunos para que estes sejam reflexivos e autónomos.
> 10. reconhecer todo o espetro de aprendizagens de todos os alunos (ARG, 2002, p. 1-2).

A designação Avaliação das Aprendizagens (AdA), por seu lado, refere-se a um balanço, a um ponto de situação que se faz, acerca do que, num dado momento, os alunos sabem e são capazes de fazer. É também uma expressão que me parece substituir bem a designação Avaliação Sumativa (AS) quanto mais não seja por ser mais consistente e fazer mais sentido com a ApA. Mas é igualmente importante referir que a concepção de AS acompanhou a evolução da AF a partir dos anos 70. Na verdade, no início, a AS começou por ter uma função estritamente classificativa mas, a partir dos anos 90, passou a ser considerada como tendo um papel igualmente relevante nos processos pedagógicos relacionados com as aprendizagens dos alunos. Nestas condições, as avaliações de natureza sumativa, as que fazem balanços, pontos de situação, do que foi aprendido, podem ter duas utilizações distintas. Uma utilização formativa, em que os dados obtidos são analisados para distribuir *feedback* pelos alunos tendo em vista a regulação e auto-regulação das suas aprendizagens, não sendo assim mobilizados para efeitos de atribuição de classificações. Uma utilização classificativa, em que os dados obtidos são analisados para atribuir classificações aos alunos.

Assim, tanto a AdA como a ApA devem ter um papel relevante na melhoria das aprendizagens dos alunos, ainda que a sua inserção nas práticas pedagógicas e as suas funções sejam distintas num número de situações. Repare-se que a ApA é, por natureza, tendencialmente contínua e os dados que através dela se obtêm não deverão ser mobilizados para efeitos da atribuição de classificações. É uma avaliação de proximidade, que está muito articulada com o ensino e a aprendizagem, em que a interação social, a comunicação na sala de aula, tem um papel fundamental e é muitas vezes informal e relativamente pouco estruturada. É uma avaliação que acompanha em permanência as dinâmicas de ensino e de aprendizagem que se desenvolvem nas salas de aula e, nessas circunstâncias, tem um papel determinante na distribuição de *feedback* que apoie e oriente os alunos nas suas aprendizagens. A AdA é, por natureza, pontual e os dados que proporciona podem ou não ser mobilizados para classificar os alunos. Num certo sentido, é uma avaliação menos interativa, menos presente no dia a dia das salas de aula, mas deve igualmente ter um papel relevante nas dinâmicas de aprendizagem.

Esta discussão mostra que o problema das relações entre a AF e a AS ou, se quisermos, entre a ApA e a AdA, constitui ainda um importante desafio nos domínios da reflexão e da construção teóricas. Tendo em conta a sua profunda ligação ao que se passa dentro das salas de aula, as práticas pedagógicas e a pesquisa sobre elas têm, naturalmente, uma crucial importância no processo de teorização daquelas relações. Esta questão continua, por isso, ao longo dos últimos anos, a ser objeto de interesse e estudo por uma diversidade de autores (e.g., Black e Wiliam, 2018; Fernandes, 2005, 2006, 2008a; Harlen, 2005, 2006; Harlen e James, 1997; Santos, 2016).

Uma forma simples e elementar para discutir relações entre aquelas duas modalidades de avaliação é a sua referenciação. Na verdade, facilmente se reconhece que a ApA é de referência criterial uma vez que as aprendizagens dos alunos podem ser comparadas com critérios previamente definidos, considerados como uma espécie de ideal a atingir ou, se quisermos, algo que se considera importante para aprender. Assim, não se comparam as aprendizagens de um aluno com as de outros, mas antes com os critérios que foram definidos. A AdA também pode ser, obviamente, de referência criterial mas pode ser, muitas vezes, sobretudo quando é externa, de referência normativa. Isto significa que os seus resultados podem ser utilizados para comparar as aprendizagens dos alunos com uma dada norma (uma média, por exemplo) ou com as aprendizagens de um dado grupo. Estas constatações são suficientes para que uma diversidade de autores considere que as duas modalidades de

avaliação se articulam, pois ambas podem ser referidas a critérios e serem pensadas, planeadas e postas em prática em função deles (Fernandes, 2005; Harlen, 2006).

Para Harlen (2005, 2006), a ApA e a AdA são processos distintos e com implicações diferenciadas mas questiona se tais diferenças, na prática, serão tão significativas para os professores e se os dados obtidos através de avaliações sumativas (AdA) podem ser utilizados para fins formativos, assim como se os dados obtidos através de avaliações formativas (ApA) podem ter utilizações sumativas. Julgo que os propósitos das duas avaliações são distintos e têm implicações pedagógicas igualmente distintas e, por isso, é importante que os professores reconheçam essa diferenciação. Mas, para tal, temos de clarificar do que estamos a falar quando nos referimos à AdA ou à ApA. Assim, estamos perante uma AdA quando os seus resultados nos permitem fazer um balanço, um ponto de situação, uma súmula, acerca do que os alunos sabem e são capazes de fazer. Nestes termos, como já se referiu antes, poderemos utilizar os resultados para efeitos de classificar os alunos ou para efeitos de obter *feedback* para apoiar as aprendizagens e o ensino. Neste último caso, sem quaisquer preocupações quanto às classificações, estamos a dar uma utilização formativa aos dados obtidos através da AdA, podendo proporcionar um *feedback* que oriente os alunos e os ajude no que deverão fazer para aprender. Podemos, por exemplo, administrar um teste, corrigi-lo, classificá-lo e utilizá-lo para dar *feedback* aos alunos e não mobilizar os resultados produzidos para as classificações que se devem atribuir no final de um dado período. Assim, os resultados de um teste podem ser utilizados para apoiar e melhorar as aprendizagens e para proporcionar *feedback* com esses mesmos propósitos. Julgo que Harlen, quando afirmou que tal não é possível, poderia estar a referir-se à AdA com propósitos exclusivamente classificativos.

Relativamente à utilização dos dados da ApA para fins sumativos ou para a AdA, a questão de Harlen põe-se no domínio da validade e da fiabilidade das avaliações formativas que, segundo esta autora parece considerar, não serão as mais desejáveis. No entanto, não põe de lado tal possibilidade desde que se faça a distinção entre as evidências de aprendizagem obtidas pela ApA e as interpretações que se fazem a partir delas. Parece-me igualmente que esta perspetiva pode ser questionada e, mais uma vez, temos de pensar no que entendemos ser a ApA. Se considerarmos que é uma avaliação eminentemente pedagógica, que se desenvolve através do diálogo e da interação social, na proximidade entre alunos e professores, em contextos muito específicos e sem quaisquer propósitos classificativos, então não parece fazer muito sentido utilizar os seus resultados para a AdA em geral e, muito menos, para efeitos da AdA com fins classificativos.

Num outro trabalho (Fernandes, 2008a) considerei que, na reflexão e na investigação relacionadas com o estudo das relações entre a AdA e a ApA, poderíamos ter em conta três perspetivas diferentes: a) considerar que estas duas modalidades de avaliação são dicotómicas, pouco ou nada tendo a ver uma com a outra e, por isso, serem mesmo opostas e com inserções pedagógicas e relações com o ensino e a aprendizagem muito distintas; b) considerar que a AdA e a ApA, tal como sugerido por Harlen (2006), são dimensões distintas de um mesmo constructo em que se considera uma espécie de *continuum* de modalidades de avaliação, desde a *avaliação formativa informal*, talvez do dia a dia, pouco ou nada estruturada, até à *avaliação sumativa formal* com a *avaliação formativa formal*, talvez mais pontual e estruturada, e a *avaliação sumativa informal*, talvez semelhante à anterior, entre aquelas duas. Relativamente a esta perspetiva pode surgir a questão de, na prática, saber quando é que estamos em presença de uma *avaliação formativa formal* ou de uma *avaliação sumativa informal*. Mas trata-se de uma formulação interessante que nos remete para uma outra discussão relacionada com a existência de um conjunto de práticas de avaliação dificilmente consideradas como formativas ou sumativas, mas que poderão ter caraterísticas híbridas ou ainda não caraterizadas. O problema reside em investigá-las, em caraterizá-las e em identificar as práticas e o tipo de dados e evidências que é produzido; e c) considerar que a AdA e a ApA são complementares, ou seja, dois conceitos distintos um do outro mas não dicotómicos que se completam ou complementam nos seus propósitos mas, sobretudo, no que se refere às contribuições que podem dar para apoiar/melhorar as aprendizagens dos alunos.

O que se pode dizer nesta altura é que se a utilização dos dados da AdA para efeitos formativos parece ser aceitável tendo em conta as condições acima discutidas, já a utilização dos dados da ApA para efeitos sumativos parece, à partida, mais problemática tendo em conta a sua natureza e a sua inserção no processo pedagógico. Claro que a ApA pode proporcionar informação de qualidade relativamente ao desenvolvimento das aprendizagens dos alunos mas não podemos transformar a AdA numa espécie de somatório de avaliações para as aprendizagens. No dia a dia das salas de aula, a ApA deve estar presente para apoiar os alunos a melhorar e a vencer as suas dificuldades aceitando eventuais erros e problemas como meios de aprendizagem. Quando chega o momento da AdA, para se fazer um balanço acerca do que os alunos sabem e são capazes de fazer para efeitos de atribuição de uma classificação, parece fazer sentido que os dados da ApA apenas deverão ser mobilizados para efeitos de saber: a) se o aluno aprendeu; b) como é que ultrapassou as dificuldades; c) as razões que poderão ter impedido que assim acontecesse; e d) o que foi efetivamente feito pelo aluno e pelo professor

para dissipar as dificuldades. E nada mais! Em todo o caso, esta questão não está esgotada e necessita de continuar a ser estudada.

Finalmente, julgo que a AdA e a ApA não se devem confundir uma com a outra pois, por definição, têm propósitos distintos, ocorrem em momentos distintos e têm uma inserção pedagógica distinta. Na verdade, consider mesmo que os seus pressupostos epistemológicos são diferentes e, por isso, a informação que se obtém a partir dos dados de cada uma é também diferente. Basta pensarmos no papel e na participação dos alunos e dos professores num e noutro caso, no tipo de conhecimento que é gerado por uma e por outra e nos processos que geram esse mesmo conhecimento. Por isso, a visão de que ambas são dimensões distintas de um mesmo constructo (Harlen, 2006) pode ser questionável do ponto de vista epistemológico. Talvez seja uma ideia melhor compreendida quando a avaliação formativa e a avaliação sumativa foram criadas em 1967 por Scriven e posteriormente, em 1971, aplicadas em contexto escolar por Bloom e colaboradores (Scriven, 1967; Bloom, Hastings e Madaus, 1971). Mas, entretanto, o conceito de avaliação formativa evoluiu muito como tive oportunidade de ir discutindo ao longo deste texto. Por outro lado, parece-me não fazer sentido considerar a AdA e a ApA como dicotómicas uma vez que, por exemplo, ambas contribuem para gerar informação fundamental para distribuir *feedback* de elevada qualidade a todos os alunos e para, através de inferências, conhecer o que, e como, os alunos aprendem. Por isso, deverão ser consideradas complementares.

Black e Wiliam (2018) ao discutirem a natureza inferencial das avaliações, ou seja, ao considerarem que avaliar é um processo através do qual se fazem inferências acerca do que os alunos sabem e são capazes de fazer, referiram que a diferença entre a AdA e a ApA está relacionada com o tipo de inferências que se decide formular a partir de cada uma. Nestas condições, se as inferências forem no sentido de definir em que situação se encontram os alunos em relação a um ou mais domínios do currículo, estamos perante uma utilização sumativa da avaliação. Se, no entanto, tais inferências forem no sentido de saber que orientações devem ser dadas aos alunos para que eles organizem os seus esforços de aprendizagem, então estamos perante uma utilização formativa da avaliação realizada. Isto significa que não são os chamados instrumentos de avaliação que determinam a natureza da avaliação ou vice-versa. Ou seja, um mesmo meio de recolha de informação pode ser utilizado na ApA ou na AdA e, consequentemente, induzir inferências de natureza e conteúdo distintos.

Santos (2016), após ter clarificado a diferenciação entre as duas utilizações ou propósitos que podem ser atribuídos à avaliação e as condições em que as suas relações

podem ser perspetivadas, considerou que se trata de "(...) um imperativo a juntar a tantos outros que cabe ao professor concretizar no seu trabalho junto dos seus alunos." (p. 663) e, para tal, enunciou os seguintes cinco princípios que, na sua opinião, deverão orientar a articulação entre a avaliação formativa e a avaliação sumativa:

> 1. Articular significa estabelecer relações e não simultaneidade de processos.
> 2. Alinhar a avaliação sumativa e formativa entre si com o ensino e com o currículo.
> 3. Restringir a estratégias formais de avaliação formativa para a articulação não elimina a existência de outras mais informais
> 4. Reconhecer que uma prática de avaliação sumativa e formativa exige do professor conhecimento sobre a avaliação, conhecimento do conteúdo e conhecimento do conteúdo para ensinar.
> 5. Atender à cultura de avaliação na sala de aula e em outros contextos em que a avaliação ocorre (Santos, 2016, pp. 659-661).

Para esta investigadora, cujo trabalho está fortemente enraizado na investigação empírica, a avaliação formativa e a avaliação sumativa não são conceitos dicotómicos. Na sua opinião são conceitos e processos complementares, ainda que possam ser considerados como distintos.

Para Baird *et al*. (2017) a ApA ainda que não tenha influenciado de forma significativa e alargada as teorias de aprendizagem, pode permitir o desenvolvimento do conhecimento pedagógico de conteúdo por parte dos professores, contribuindo para que possam melhorar as suas competências para ensinar os chamados conteúdos escolares. Ou seja, uma vez que a ApA é uma perspetiva de avaliação muito orientada para a prática, é natural que a sua influência ao nível das teorias da aprendizagem tenha uma forte ligação às práticas.

Em suma, a Avaliação Formativa, hoje mais vulgarmente designada como *Avaliação para as Aprendizagens* (ApA), que pressupõe uma forte relação entre a avaliação e as aprendizagens, tem no *feedback* o seu conteúdo mais fundamental, se assim se pode dizer. Isto significa que o *feedback*, em cada momento, deve permitir que os alunos saibam o que têm de aprender, em que situação se encontram e que esforços devem fazer para alcançarem o que se espera. A *Avaliação das Aprendizagens* (AdA) é fundamental para complementar a ApA no apoio ao desenvolvimento das aprendizagens dos alunos, em particular quando lhe é dada uma utilização formativa. Os fundamentos da ApA têm vindo a evoluir desde as racionalidades empírico-racionalistas, de natureza positivista, até às racionalidades mais críticas e sociocríticas, de natureza construtivista, mais ou menos relativista. As perspetivas socioculturais e construtivistas da aprendizagem estão muito associadas às perspetivas mais contemporâneas da ApA, conferindo-lhe, como se viu, uma abrangência e uma sofisticação teórica e prática que a diferenciam profundamente da Avaliação Formativa dos anos 70 do século passado. Porém, tal como referem Baird *et al*. (2017), não se pode afirmar que haja

uma relação linear entre as práticas de avaliação e uma dada teoria de aprendizagem mesmo reconhecendo que a avaliação tem sempre algum tipo de impacto na aprendizagem. Porém, a utilização das teorias da aprendizagem na definição das avaliações não está ainda claramente estabelecida pois estas, sejam internas ou externas, são muitas vezes determinadas por outro tipo de *interferências* (e.g., técnicas e tecnológicas, políticas, sociais).

Numa reação ao seminal trabalho de Black e Wiliam (1998), Perrenoud (1998) sugeriu que o desenvolvimento de qualquer teoria no domínio da avaliação formativa tinha de ocorrer no âmbito de um campo teórico mais abrangente, mais alargado, como é, por exemplo, o caso de uma teoria da pedagogia. Vinte anos após esse desafio, Black e Wiliam (2018) discutiram e propuseram um modelo em que as práticas educacionais nas salas de aula e as avaliações que lhes estão associadas estão baseadas em teorias da pedagogia, do ensino, da aprendizagem e ainda pela disciplina específica em que se está a trabalhar, tudo isto, naturalmente, no contexto mais amplo da educação. Na verdade, estes autores afirmam claramente que a teorização da avaliação formativa tem necessariamente de se desenvolver no contexto de uma teoria da pedagogia. Ou seja, no fundo, só poderemos verdadeiramente compreender e apreender a avaliação formativa no contexto de uma teoria da pedagogia. Nestes termos, Black e Wiliam (2018) propuseram um modelo para analisar o papel da avaliação (AdA e ApA) na melhoria da educação, chamando a atenção para as relações e para os aspetos que consideraram mais relevantes. Assim, estes autores consideraram cinco aspetos fundamentais a considerar no processo de construção teórica da avaliação que se desenvolve nas salas de aula: a) pedagogia e ensino; b) teorias da aprendizagem; c) contextos sociais, políticos e culturais; d) planificação e estruturação do ensino e do trabalho pedagógico; e e) práticas de avaliação dos professores (relações entre ApA e AdA).

Aqueles autores desenvolveram uma exaustiva discussão de pendor marcadamente pedagógico, com particular destaque para a natureza do trabalho realizado em contextos de sala de aula, para as práticas de avaliação dos professores e, sobretudo, para as relações entre a AdA e a ApA. A lógica da discussão desenvolveu-se a partir dos cinco aspetos acima enunciados mas com uma grande ênfase e preocupação na melhoria das aprendizagens dos alunos. Neste sentido, destacam-se aqui, telegraficamente, os tópicos que se consideraram mais relevantes neste recente trabalho de Black e Wiliam (2018) orientado para a construção teórica:

- a relevância da linguagem no desenvolvimento do pensamento.
- a seleção de propostas de trabalho que contribuam para mobilizar, integrar e utilizar as aprendizagens que se pretendem desenvolver.

- o papel da pedagogia, das teorias da aprendizagem e da planificação do ensino.
- a particular importância da interação social, do diálogo e da relação pedagógica numa diversidade de dinâmicas de trabalho na sala de aula.
- a influência positiva da avaliação (formativa e sumativa) dos processos e dos resultados das propostas de trabalho nas aprendizagens dos alunos.
- a relevância da *avaliação formativa avaliar mesmo* (*assessment as assessment*), o que não significa classificar.
- o determinante papel do diálogo interativo, componente forte das práticas de avaliação formativa, no desenvolvimento das aprendizagens através do *feedback* que pode proporcionar.
- a relevância da AdA e da sua utilização formativa para proporcionar *feedback* e determinar o que os alunos sabem e são capazes de fazer num dado momento do processo educativo.
- a importância de se perceber que não há uma marcada descontinuidade entre a AdA e a ApA, pois ambas permitem fazer inferências acerca das aprendizagens dos aluno

Ainda que não constitua propriamente uma novidade em relação aos seus trabalhos anteriores (Black e Wiliam, 1998, 2006), este trabalho (Black e Wiliam, 2018), em que discutem a questão da construção teórica no domínio da avaliação formativa, acaba por consolidar e clarificar algumas linhas de força que podem orientar essa construção. Tal como nos trabalhos anteriores, verifica-se que, para estes autores, o trabalho empírico nas salas de aula é incontornável, dando uma relevância muito particular ao trabalho dos professores, às suas perspetivas pedagógicas, ao *feedback* e às relações entre a AdA e a ApA. Neste último aspeto, parece haver um avanço no sentido de clarificar essas relações num enquadramento teórico mais elaborado, com recurso às teorias da aprendizagem de raíz sociocultural, com destaque para a teoria da atividade (Engestrom, 1999), e, em geral, à pedagogia. Mas, para além disso, Black e Wiliam referem como propósito deste seu trabalho contribuir para o desenvolvimento de estudos de índole pedagógica que integrem e enfatizem a importância da avaliação o que, na sua opinião, não tem acontecido.

Finalmente, sublinhe-se que o modelo proposto foi organizado no sentido de contribuir para a convergência das aprendizagens entre os alunos com e sem limitações no domínio da escrita, da língua e dos *códigos* utilizados nas escolas. Assim, é dada particular relevância aos diálogos interativos, ao trabalho de grupo e à aprendizagem em colaboração, ao papel positivo

que a avaliação formativa pode desempenhar nas aprendizagens e à ênfase dada ao envolvimento ativo dos alunos nos processos de aprendizagem e avaliação. Deste modo, as crianças e jovens provenientes de meios sociais, económicos e culturais mais deprimidos podem ter oportunidades reais para ultrapassarem dificuldades decorrentes dessa situação desfavorável.

Conclusões

A construção teórica no domínio da chamada avaliação das aprendizagens é uma condição relevante para o desenvolvimento da pesquisa e das práticas pois pode contribuir para clarificar conceitos e conceções erróneas acerca dos mesmos, ajudar a identificar problemas e contribuir para saber em que ponto nos encontramos e para onde e como estamos a dirigir os nossos esforços de pesquisa. Como se viu ao longo do texto, os esforços de teorização obrigam a sistematizar ideias e teorias, a identificar os seus aspetos mais essenciais, a pesquisar relações entre fenómenos e práticas e a formular e a explorar conjeturas.

Muito há ainda a fazer no que se refere ao processo de construção de uma teoria da avaliação pedagógica. Mas o texto permite concluir que, sensivelmente nas três últimas décadas, foram feitos assinaláveis progressos na clarificação dos conceitos de avaliação formativa e de avaliação sumativa, reconhecendo-se a grande relevância de se estudarem as relações entre eles. Na verdade, aqueles dois conceitos, tal como hoje os entendemos por força da construção teórica, nada têm a ver com as suas conceitualizações dos anos 70 do século passado. Hoje são conceitos teoricamente mais densos, elaborados e complexos, complementares e fortemente associados à melhoria das aprendizagens.

A avaliação formativa está inexoravelmente associada à distribuição de *feedback* de elevada qualidade, é de natureza contínua e tem como fundamental propósito ajudar os alunos a aprender. Por este motivo, passou recentemente a ser designada como *Avaliação para as Aprendizagens*. A avaliação sumativa, ao contrário do que aconteceu durante décadas, não está mais exclusivamente associada à atribuição de classificações. Na verdade, a construção teórica permitiu que ela pudesse ter uma utilização formativa e, nessas circunstâncias, pudesse estar igualmente ao serviço das aprendizagens. Mas continua a ser através dela que se fazem pontos de situação e balanços acerca das aprendizagens realizadas pelos alunos para, a partir daí, se poderem formular evidências passíveis de ser traduzidas por classificações que, neste contexto, terão uma importância cada vez mais residual na vida pedagógica das escolas. Hoje, a avaliação sumativa é designada como *Avaliação das Aprendizagens*.

Estes desenvolvimentos têm muito a ver com os progressos feitos através dos estudos das relações entre aquelas duas modalidades de avaliação no contexto de salas de aula. Essas pesquisas têm permitido clarificar a natureza das relações entre elas, havendo uma forte corrente de pensamento que aponta no sentido de as considerar complementares ainda que possam ser distintas do ponto de vista epistemológico. Repare-se que, no seu início, não só eram consideradas conceitos distintos mas também dicotómicos.

A verdade é que a construção teórica, apesar de poder ser considerada modesta, permitiu fundamentar estas transformações e desenvolvimentos dos pontos de vista epistemológico, ontológico e metodológico. Os modelos e recomendações formulados por alguns dos investigadores cujos trabalhos se discutiram neste texto, mostram que continua a ser necessário compreender melhor as relações entre as aprendizagens dos alunos e as avaliações e outros aspetos relevantes tais como as práticas dos professores e as suas relações com as disciplinas que lecionam e as relações entre as teorias da aprendizagem e as avaliações. Verificou-se que há orientações concretas quanto aos caminhos que é necessário continuar a percorrer para que a avaliação pedagógica seja, cada vez mais, um processo ao serviço das aprendizagens dos alunos e da construção de sistemas escolares mais democráticos, mais éticos, com mais equidade e mais justiça.

Referências

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*, 175-209. Coimbra: Almedina.

Allal, L., Cardinet, J. e Perrenoud, Ph. (1979). *L'évaluation formative dans un enseignement différencié.* (Actes du colloque à l' Université de Genève, Mars 1978.) Berne: Peter Lang.

Archbald, D. e Newmann, F. (1992). Approaches to assessing academic achievement. 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*, 139-180. Albany, NY: State University of New York Press.

Assessment Reform Group (2002). *Assessment for learning: research-based principles to guide classroom practice – 10 principles*. London: Nuffield Foundation.

Baird, J-A., Andrich, D., Hopfenbeck, T. e Stobart, G. (2017) Assessment and learning: fields apart?, *Assessment in Education: Principles, Policy & Practice*, *24*(3), 317-350.

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*, 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*, 181-206. Albany, NY: State University of New York Press.

Berlak, H., Newmann, F., Adams, E., Archbald, D., Burgess, T., Raven, J. e Romberg, T. (Eds.) (1992), *Toward a new science of educational testing and assessment*. Albany, NY: State University of New York Press.

Black, P. e Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551-575.

Black, P. e Wiliam, D. (2006a). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning*, 81-100. London: Sage.

Black, P. e Wiliam, D. (2006b). The reliability of assessments. In J. Gardner (Ed.), *Assessment and Learning*, 119-132. London: Sage.

Black, P. e Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7-74.

Black, P., Harrison, C., Lee, C., Marshall, B. e Wiliam, D, (2003). *Assessment for Learning–putting it into practice*. Buckingham: Open University Press.

Black, P., Harrison, C., Lee, C., Marshall, B. e Wiliam, D. (2002). *Working inside the black box: assessment for learning in the classroom.* London, UK: King's College London.

Bloom, B., Hastings, J. e Madaus, G. (1971). *Handbook of formative and summative evaluation of student learning*. New York: Mac Graw Hill.

Cardinet, J. (1991). L'apport sociocognitif à la régulation interactive. In J. Weiss (Ed.), *L'évaluation: problème de communication*, 199-213. Cousset (Fribourg): Delval.

De Ketele, J-M. (1993). L'évaluation conjuguée en paradigmes. *Revue Française de Pédagogie, 103*, 59-80.

De Ketele, J-M. (1986). A propósito das noções de avaliação formativa, de avaliação sumativa, de individualização e de diferenciação. In L. Allal, J. Cardinet e Ph. Perrenoud (Orgs.), *A avaliação formativa num ensino diferenciado*, 211-218. Coimbra: Almedina.

Dwyer, C. (1998). Assessment and classroom learning: theory and practice. *Assessment in Education: Principles, Policy & Practice, 5*(1), 131-137.

Engestrom, Y. (1999). Activity theory and individual and social transformation. *In* Y. Engestrom, R. Miettinen e R-L. Punamaki (Eds.), *Perspectives on activity theory*, 19-38. New York, NY: Cambridge University Press.

Fernandes, D. (2011). Articulação da aprendizagem, da avaliação e do ensino: questões teóricas, práticas e metodológicas. In J. M. DeKetele e M. P. Alves (Orgs.), *Do currículo à avaliação, da avaliação ao currículo*, 131-142. Porto: Porto Editora.

Fernandes, D. (2008a). Para uma teoria da avaliação no domínio das aprendizagens. *Estudos em Avaliação Educacional*, *19* (41), 347-372.

Fernandes, D. (2008b). *Avaliar para aprender: Fundamentos, práticas e políticas*. São Paulo: Fundação Editora da UNESP.

Fernandes, D. (2006). Para uma teoria da avaliação formativa. *Revista Portuguesa de Educação*, 19(2), 21-50.

Fernandes, D. (2005). *Avaliação das aprendizagens: desafios às teorias, práticas e políticas.* Cacém: Texto Editores.

Fernandes, D. (1993). Complexidade, tensões e mudança na avaliação das aprendizagens. In L. Almeida, J. Fernandes e A. Mourão (Orgs.), *Ensino-aprendizagem da matemática: recuperação de alunos com baixo desempenho*, 43-60. Riba d'Ave: Didáxis.

Fernandes, D. (1992). O tempo da avaliação. *Noesis*, *23*, 18-21.

Gardner, J. (Ed.) (2006). Assessment and learning. London: Sage.

Gifford, B. e O' Connor, M. (Eds.) (1992). *Changing assessments: Alternative views of aptitude, achievement and instruction*. Norwell, MA: Kluwer.

Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. Londres: Falmer.

Guba, E. & Lincoln, Y. (1989). *Fourth Generation Evaluation*. London: Sage.

Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning*, 103-118. London: Sage.

Harlen, W. (2005). Teachers' summative practices and assessment for learning: tensions and synergies. Curriculum Journal, *16* (2), 207-223.

Harlen, W. & M. James (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, *4* (3), 365 – 379.

James, M. (2006). Assessment, teaching and theories of learning. In J. Gardner (Ed.) *Assessment and learning*, 47-60. London: Sage.

Luckesi, C. (1995). *Avaliação da Aprendizagem Escolar*. São Paulo: Cortez.

Mislevy, R.,Wilson, M., Ercikan, K. e Chudowsky (2003). Psychometric principles in student assessment. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, 489-532. Dordrecht: Kluwer.

Madaus, G., Haney, W. e Kreitzer, A. (2000). The role of testing in evaluations. In D. Stufflebeam, G. Madaus e T. Kellaghan, (Eds.) *Evaluation models: Viewpoints on educational and human services evaluation* (2nd Edition), 113-126. Dordrecht: Kluwer.

Madaus, G., Russell, M., e Higgins, J. (2009). The paradoxes of high stakes testing: how they affect students, their parents, teachers, principals, schools, and society. Charlotte, NC: Information Age Publishing.

Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning. Towards a wider conceptual field. *Assessment in Education: Principles Policy and Practice, 5*(1), 85-102.

Sadler, D. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice, 5*, 1, 77-84.

Santos, L. (2016). A articulação entre a avaliação somativa e a formative na prática pedagógica: uma impossibilidade ou um desafio? *Ensaio: Avaliação e Políticas Públicas em Educação*, *24* (92), 637-669.

Scriven, M. (1967). *The methodology of evaluation* (AERA Monograph series on curriculum evaluation, No. 1). Chicago IL: Rand McNally.

Scriven, M. (1994). Evaluation as a discipline. *Studies in Educational Evaluation*, 20, 147-166.

Scriven, M. (2003). Evaluation theory and metatheory. In T. Kellaghan e D. Stufflebeam

(Eds.), *International Handbook of Educational Evaluation*, 15-30. Boston: Kluwer.

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29* (7), 4–14.

Stiggins, R. (2017). *The perfect assessment system*. Alexandria, VA: ASCD.

Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and Learning*, 133-146 London: Sage.

Torrance, H. (2003). Assessment of the national curriculum in England. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, 905-928. Dordrecht: Kluwer.

Vasconcellos, C. (2005). *Avaliação da Aprendizagem: Práticas de Mudança por uma Práxis Transformadora* (7.a Edição). São Paulo: Libertad.

Wiliam, D. (2017). Learning and assessment: a long and winding road? *Assessment in Education: Principles, Policy & Practice*, *24* (3), 309-316.