

Avaliação em Educação: uma discussão de algumas questões críticas e desafios a enfrentar nos próximos anos

Domingos Fernandes*

Resumo

Neste artigo discutiram-se quatro questões críticas da avaliação: a) avaliação formal e informal; b) a avaliação como (trans)disciplina; c) avaliação e qualidade; e d) avaliação e discernimento pragmático. A finalidade da discussão era contribuir para desenvolver práticas de avaliação que pudessem responder melhor aos desafios dos sistemas educacionais. A discussão pareceu mostrar que a avaliação a desenvolver nos próximos anos tem que ser um processo com melhor integração teórica e melhor articulação entre diferentes visões epistemológicas e metodológicas que influenciam as práticas avaliativas. De igual modo, mostrou que, cada vez mais, a avaliação tem que ser um esforço partilhado por investigadores e avaliadores da diversidade de disciplinas que sejam indispensáveis para descrever, analisar e interpretar a realidade a avaliar.

Palavras-chave: Avaliação Educacional. Teoria da Avaliação. Epistemologia da Avaliação. Avaliação e Qualidade.

Educational evaluation: discussion of critical questions and challenges to fare in the coming years

Abstract

Four critical questions on evaluation have been discussed in this paper: formal and informal evaluation; b) evaluation as a transdiscipline; c) evaluation and quality; and d) evaluation and pragmatic discernment. The aim was to contribute to the development of evaluation practices, which can provide better responses to the educational systems challenges. The discussion showed that evaluation, in the coming years will become a process with better theoretical integration and different

* Professor Associado com Agregação em Educação da Universidade de Lisboa, Portugal.
E-mail: dfernandes@ie.ul.pt

perspectives, bringing an impact on evaluation practices. Also, the discussion has shown the importance of bringing new disciplines to be shared by the researchers. Keywords: Educational Evaluation. Evaluation Theory. Evaluation Epistemology. Evaluation and Quality.

Evaluación en Educación: Discusión de Algunas Cuestiones Críticas y Desafíos para los Próximos Años

Resumen

En este artículo se discuten cuatro cuestiones críticas de la evaluación: a) evaluación formal e informal; b) evaluación como (trans)disciplina; c) evaluación y calidad; y d) evaluación y discernimiento pragmático. La finalidad de la discusión es contribuir para el desarrollo de prácticas de evaluación que puedan responder mejor a los desafíos de los sistemas educacionales. La discusión parece mostrar que la evaluación que debe desarrollarse en los próximos años tiene que ser un proceso con mejor integración teórica y mejor articulación entre las diferentes visiones epistemológicas y metodológicas que influyen en las prácticas evaluativas. De igual modo, mostró que, cada vez más, la evaluación tiene que ser un esfuerzo compartido entre investigadores y evaluadores de las distintas disciplinas o asignaturas, todas las que sean necesarias para describir, analizar e interpretar la realidad que será evaluada.

Palabras clave: *Evaluación Educacional. Teoría de la Evaluación. Epistemología de la Evaluación. Evaluación y Calidad.*

Introdução

A avaliação é um domínio fundamental do conhecimento porque nos permite formular juízos acerca de todas as áreas de funcionamento da sociedade e tomar decisões fundamentadas, tendo em vista a sua melhoria. De forma mais ou menos explícita, mais ou menos formal, a avaliação está efetivamente presente em todos os domínios acadêmicos e em todas as áreas da atividade humana, podendo garantir a todos e a cada um dos cidadãos que os bens e serviços de que necessitam são de qualidade e não põem em causa os seus legítimos interesses ou mesmo a sua saúde e segurança. Por isso mesmo, a avaliação, em geral, tem a ver com uma variedade de propósitos tais como:

- a) apoiar a tomada de decisões;
- b) servir a prestação pública de contas;
- c) melhorar práticas e procedimentos;
- d) compreender problemas de natureza social, contribuindo para a identificação de soluções possíveis;
- e) compreender as experiências vividas por quem está envolvido numa dada prática social; e
- f) acreditar e reconhecer programas.

O propósito ou os propósitos que se pretendem alcançar através de um estudo de avaliação determinam, em grande medida, a forma como se planifica e se desenvolve o processo de recolha de informação e como se organiza e divulga o próprio relatório do estudo. Ou seja, os fins a que se destina a avaliação têm necessariamente que ser considerados quando se escolhem os procedimentos de recolha de dados, o tipo de interação que o avaliador vai estabelecer com os participantes ou, em geral, o *design* do estudo que é preciso concretizar. Para que os propósitos de uma avaliação possam ser alcançados utilizam-se, em geral, avaliações sistemáticas ou formais, enquadradas por uma ou mais abordagens teóricas. Mas também se utilizam avaliações informais e combinações de abordagens mais e menos formais. A compreensão da natureza destes tipos de avaliação e das relações que existem entre si é relevante e será abordada neste artigo.

A avaliação educacional vem sendo considerada cada vez mais indispensável para descrever, compreender e agir sobre uma grande variedade de problemas que afetam os sistemas educativos e formativos. Trata-se de um processo social complexo que envolve pessoas que funcionam em determinados contextos, com os seus valores, as suas práticas e políticas próprias e envolve também a natureza do que está a ser avaliado que, por sua vez, também tem as suas finalidades, lógicas e políticas próprias. Nessas condições, não se pode propriamente esperar que os resultados de um número de avaliações sejam consensualmente aceites por todos aqueles a quem, de algum modo, a avaliação possa interessar.

Apesar de os métodos, procedimentos e técnicas utilizados em qualquer avaliação serem exatamente os mesmos que se utilizam em qualquer disciplina das ciências sociais, a verdade é que ainda não há consenso quanto à sua natureza científica. As dimensões valorativas, éticas e políticas da avaliação contribuem para que se ergam obstáculos ao seu cabal reconhecimento como disciplina científica, apesar de poder ser escrutinada e analisada com base em critérios normalmente utilizados em ciência.

O propósito fundamental deste artigo é discutir questões críticas da avaliação em educação inevitavelmente associadas à necessidade emergente de se desenvolverem novas e inovadoras formas de avaliar uma grande diversidade de objetos tais como aprendizagens dos alunos, desempenho de professores, eficácia das escolas, programas educacionais e políticas educativas. Ao empreender esta discussão parto do princípio de que a avaliação em educação, ainda que se tenha desenvolvido apreciavelmente nas últimas décadas dos pontos de vista teórico e prático, necessita de responder melhor às necessidades dos sistemas educacionais. Parece ser evidente que a crescente complexidade dos problemas da educação e da formação exige práticas de avaliação que descrevam, analisem e interpretem melhor, isto é, com maior profundidade, credibilidade, utilidade e rigor, os objetos e fenômenos que nos preocupam, quaisquer que eles sejam.

Tendo em conta estas considerações, este artigo, para além desta introdução está organizado em cinco secções. Nas duas primeiras, discutem-se questões mais diretamente relacionadas com a natureza da avaliação – na primeira, a avaliação formal e a avaliação informal e, na segunda, a avaliação como (trans)disciplina. Na terceira secção, discute-se a questão da qualidade que, como é geralmente reconhecido, está no cerne de qualquer processo de avaliação. A este propósito, Stake (2006) sugere que se uma avaliação não contribuir decisivamente para discernir a qualidade dos objetos, não cumprirá a sua principal missão. Na quarta secção, a partir da idéia do discernimento pragmático, elaboram-se considerações acerca da necessidade de se utilizarem novas racionalidades e de articular e/ou integrar abordagens avaliativas e domínios de conhecimento. Finalmente, na quinta secção faz-se uma síntese e produzem-se algumas reflexões e conclusões.

Avaliação Formal e Avaliação Informal

A investigação recente na área da psicologia e das neurociências evidencia dois resultados interessantes:

- a) os seres humanos tendem a centrar-se mais na elaboração de juízos avaliativos acerca dos objetos que os rodeiam do que na descrição desses mesmos objetos; e
- b) a formulação de juízos avaliativos e a formulação de juízos descritivos parecem estar associadas a diferentes áreas do cérebro (MAIO; OLSON, 2000; MUSCH; KLAUER, 2003).

Estes resultados parecem sugerir que, para os seres humanos, avaliar é uma função mais básica e prioritária do que descrever o que se passa à sua volta, ou seja, parece que, antes do mais, olhamos para o mundo de um ponto de vista avaliativo. No entanto, esta avaliação do dia a dia, esta avaliação informal, aplicada a questões mais críticas como, por exemplo, a avaliação do desempenho de uma escola, é muitas vezes insuficiente e insatisfatória porque é muito influenciada pelas expectativas, experiências, saberes e conceções das pessoas e, por isso, tende a produzir visões muito parciais da realidade (WORTHEN; SANDERS, 1987).

A avaliação, tal como hoje é entendida, é uma exigente prática social e, por isso, é desejável distinguir entre as avaliações do dia a dia que todos vamos fazendo acerca do que nos rodeia, e as avaliações que nos exigem abordagens propositadas, rigorosas, sistemáticas e tão independentes e imparciais quanto possível. No primeiro caso estamos a falar da avaliação informal ou tácita que, neste sentido, é um lugar comum e um processo que parece ser fundamental para que, por exemplo, possamos discernir o que eventualmente nos fará sentir mais confortáveis ou para que possamos ajustar ou regular as nossas ações e decisões numa variedade de

situações pessoais e profissionais. No segundo caso, estamos a falar de avaliações formais que identificam os critérios utilizados, diversificam e mencionam as fontes de dados utilizadas, exibindo assim as evidências que estão na base de um dado juízo avaliativo. Desta forma, a parcialidade associada à avaliação informal pode ser substancialmente reduzida. Além disso, a avaliação formal tem uma abrangência e uma profundidade que a avaliação informal não pode ter porque vai para além da evidência que está ao alcance de um indivíduo e proporciona processos avaliativos que são coletivos (sociais) e não estritamente individuais.

Hoje, ao contrário do que se passava há cerca de duas décadas atrás, aceita-se que tanto a avaliação informal como a avaliação formal geram conhecimentos acerca da realidade que interagem entre si e que se relacionam de formas mais ou menos complexas. Não há consenso na comunidade científica acerca dos diferentes tipos de conhecimento gerados pelas duas formas de avaliação e das relações entre eles. Para vários autores a avaliação formal está mais relacionada com a lógica da ciência enquanto que a avaliação informal se baseia mais na lógica do chamado raciocínio do dia a dia (HAMMERSLEY, 2003; POLKINGHOME, 2000). As diferentes formas de conhecimento prático (MOLANDER, 2002) assumem, nestas circunstâncias, uma especial relevância e devem ser tomadas na devida conta porque contribuem para o desenvolvimento das avaliações formais. Talvez por isso mesmo outros autores, embora reconhecendo que as avaliações formais e informais geram formas alternativas de conhecimento, não deixem de as considerar complementares (SCHMIDT, 1993). Nessas condições, a avaliação formal e a avaliação informal não devem ser vistas como antagónicas, ou mesmo inconciliáveis, mas como processos que se podem e devem complementar porque ambos geram conhecimentos válidos e úteis sobre uma dada realidade.

As realidades a avaliar são, em geral, bastante complexas, integrando uma miríade de elementos (e.g. educacionais, sociais, culturais, psicológicos, estatísticos, económicos, financeiros) que as tornam difíceis de descrever, analisar e interpretar. Por isso, parece pouco prudente considerar apenas um tipo de avaliação, uma única forma de gerar conhecimento ou um único processo de recolher informação avaliativa. As ortodoxias, sejam elas metodológicas, ontológicas ou epistemológicas empobrecem a nossa visão da realidade porque, num certo sentido, limitam as nossas formas de olhar e de compreender essa mesma realidade. Há diferenças significativas entre os dois tipos de avaliação (e.g. grau de profundidade, grau de abrangência, nível das evidências recolhidas) mas parece certo que, no desenvolvimento de uma avaliação formal, é incontornável a sua interação com as atividades inerentes à avaliação informal. Basta pensarmos que, perante um estudo de avaliação de um dado programa, os leitores vão analisar e interpretar os seus resultados tendo em conta as avaliações prévias (informais) que, no dia a dia, foram fazendo do programa em causa.

De modo geral, a avaliação formal e a avaliação informal podem relacionar-se de três formas distintas:

- a) a avaliação formal é considerada um desenvolvimento do tipo de conhecimento que é gerado pela avaliação informal porque o torna mais explícito e mais fundamentado;
- b) a avaliação formal e a avaliação informal geram tipos de conhecimento diferentes mas que se consideram potencialmente complementares; e
- c) a avaliação formal e a avaliação informal desafiam-se e questionam-se mutuamente sem obedecer a qualquer hierarquia e geram formas interativas de conhecimento (STAKE, 2006).

A avaliação que, por norma, se discute nos meios académicos é realizada através de processos deliberados e formais; no entanto, a avaliação formal reflete uma diversidade de aspetos informais e intuitivos e, por isso mesmo, tem pertinência e todo o cabimento o estudo aprofundado das relações entre ambas. Stake (2006) refere com algum humor e oportunidade que a avaliação informal, num certo sentido, poderia reclamar alguma superioridade em relação à avaliação formal. Na verdade, a espécie humana, ao longo da sua existência, foi tomando decisões baseadas em avaliações informais, que garantiram a sua sobrevivência. Assim, segundo aquele autor, uma vez que a avaliação formal, tal como hoje a conhecemos, ainda não tem 100 anos, pode dizer-se que a avaliação informal não só permitiu a sobrevivência da espécie mas também o seu assinalável progresso.

A Avaliação como (Trans)Disciplina

A necessidade de construir uma teoria da avaliação que possa estabelecer as necessárias conexões entre todas as suas áreas práticas, de modo a articulá-las e de forma a desenvolver conceitos e uma linguagem comuns, assim como todo o tipo de mecanismos e de procedimentos próprios de um campo científico (e.g. lidar com *ameaças* à validade, assegurar a identidade), está há muito estabelecida na literatura (DONALDSON; LIPSEY, 2006). No entanto, não têm sido propriamente fáceis os caminhos trilhados pela avaliação para que veja consensualmente consagrado o estatuto de disciplina científica, sobretudo devido ao fato de as afirmações avaliativas, que, por natureza, são valorativas, não serem consideradas cientificamente legítimas por setores do mundo académico. Ao considerar que a ciência não pode admitir quaisquer juízos de valor, uma parte da comunidade científica continua ainda hoje a defender que não pode existir uma ciência da avaliação. No entanto, esta posição, que tem sido sustentada por cientistas e por filósofos da Ciência, que se inspiram na filosofia lógico-positivista, segundo a qual as Ciências Sociais são independentes e livres de quaisquer valores, tem sido refutada por investigadores e avaliadores de diferentes inspirações filosóficas tais como Guba e Lincoln (1989, 1994,

2000), House e Howe (2000, 2003), Scriven (1994, 2000, 2003a), Stufflebeam e Shinkfield (2007) e Stake (2000, 2003, 2004, 2006).

Os defensores da ideia de uma avaliação livre de valores sustentam que não se podem inferir ou deduzir conclusões avaliativas com base em processos científicos legítimos e creíveis. Mas, ao fazê-lo nestes precisos termos, não estarão a autorrefutar a sua própria ideia uma vez que se trata de uma conclusão que tem, claramente, um sentido avaliativo? Na verdade, não será difícil perceber-se que a produção científica, em qualquer domínio, está sempre associada a qualquer forma de avaliação. Seja através da apreciação mais ou menos crítica que se faz da literatura, da investigação existente ou de uma dada abordagem metodológica. E a verdade é que estas apreciações podem ser bem fundamentadas com base em premissas que não são propriamente arbitrárias ou meras questões de gosto ou preferência. Em geral, elas resultam de fatos combinados com definições plausíveis de vários tipos de entidades científicas que acabam por resultar numa espécie de referentes que se constroem de formas mais ou menos complexas.

Parece existir um equívoco lógico na argumentação que está na base da perspectiva da avaliação livre de valores. Trata-se de se considerar erradamente que as conclusões ou afirmações avaliativas são de natureza arbitrária, exprimindo gostos ou preferências pessoais quando, na realidade, muitas delas resultam de fatos e de definições. Mas os argumentos utilizados pelos que consideram que não há valores associados às asserções avaliativas mostram que as conjeturas, inferências e definições concetuais elaboradas no processo de construção do conhecimento necessitam de ser melhor enquadradas e justificadas.

A questão de fundo tem a ver com a afirmação de que é impossível fundamentar e apoiar cientificamente valores sociais. Como se pode inferir da argumentação de Scriven (1994, 2000, 2003a) se, por exemplo, considerarmos a questão de saber se há argumentos sólidos que sustentam a democracia como uma forma de governo, então os cientistas sociais não poderiam, de acordo com aquelas conceções, discutir uma questão desta natureza, que seria relegada para a filosofia política. É evidente que o simples senso comum seria suficiente para que pudéssemos encontrar um conjunto de sólidos argumentos que sustentam a forma democrática de organizar e governar as sociedades. Presentemente, as evidências empíricas, a história política e outros processos de inquérito científico permitem comparar diferentes sistemas de governo e identificar os seus pontos fortes e fracos. Tais processos têm permitido fortalecer a ideia da democracia como o sistema de governo que, até agora, melhor tem respondido a um conjunto de ideais e de valores que se foram consolidando ao longo da história num significativo número de países e civilizações. Nestas condições, será também legítimo que se elaborem conclusões avaliativas acerca de instituições (e.g. hospitais, universidades, escolas do ensino médio) com base num

conjunto de premissas fatuais que, naturalmente, incluem fatos acerca do que as pessoas valorizam ou preferem, que são parâmetros que nos ajudam a definir o problema e não asserções acerca do que é melhor ou do que está certo.

Para um número de investigadores, avaliadores e filósofos (GUBA; LINCOLN, 1989, 1994; HOUSE; HOWE, 1999, 2003; SCRIVEN, 1994, 2000, 2003B; STAKE, 2003, 2004, 2006; STUFFLEBEAM; SHINKFIELD, 2007) a disciplina de avaliação não só é possível e admissível, como se pode comparar a disciplinas tais como a Estatística ou a Lógica, ou seja uma disciplina cuja principal finalidade é a de desenvolver *instrumentos* que possam ser utilizados por outras disciplinas. Para Scriven uma disciplina com estas características é uma *transdisciplina* porque:

- a) está ao serviço de uma grande diversidade de disciplinas e não exclusivamente das chamadas disciplinas clássicas, mais enraizadas academicamente; e
- b) possui num núcleo central de *matérias* tais como as suas teorias organizacionais, as suas metodologias, os seus conceitos e a sua lógica (SCRIVEN, 1994, 2003b).

Uma *transdisciplina* comporta na sua essência, a distinção entre as chamadas disciplinas primárias (e.g. Matemática, Física, Filosofia) e as disciplinas cujo conteúdo é constituído por métodos, abordagens, procedimentos ou algoritmos das disciplinas primárias. Nesse sentido, para aquele autor, a Estatística é considerada uma *transdisciplina*, o mesmo se passando com a Lógica e com a Avaliação. Assim, uma *transdisciplina* possui também as suas *matérias* ou *áreas puras* para além das suas *áreas aplicadas*. Ora, segundo Scriven (1994), o problema da avaliação é que sempre teve áreas aplicadas porque as circunstâncias e os problemas de natureza prática assim o exigiam (e.g. avaliação das aprendizagens dos alunos, avaliação de programas, avaliação de sistemas educativos, avaliação de escolas), mas sempre teve dificuldades várias em determinar a sua área pura, o seu *núcleo central* de definições, de conceitos integradores, de pressupostos básicos ou de noções relativas aos seus próprios limites. Esta será uma das razões que tem vindo retardar a afirmação da avaliação como disciplina.

Avaliação e Qualidade

O fulcro ou o cerne de qualquer processo de avaliação reside no discernimento da qualidade do que se pretende avaliar. Isso significa que a avaliação, na sua essência, tem a ver com a identificação e o reconhecimento da qualidade com base numa diversidade de evidências que permitem comprovar essa mesma qualidade perante outrem. Independentemente das múltiplas e diversificadas questões que se poderão formular numa dada avaliação, há duas que lhe estarão sempre subjacentes ou que, de algum modo, estarão sempre presentes de forma mais ou menos explícita:

- a) como se poderão descrever as percepções das pessoas acerca da qualidade do que se está a avaliar (e.g. escola, currículo, programa para alunos em risco, utilização de novas tecnologias de informação e comunicação)?; e
- b) qual é a qualidade do que se está a avaliar?

Por isso, é necessário caracterizar e discutir as abordagens que têm sido utilizadas para determinar a qualidade de qualquer objeto, nomeadamente aquelas que estão mais associadas à definição de critérios e de *standards*, e as que se baseiam mais nas descrições das percepções dos diferentes intervenientes. No primeiro caso, estamos perante abordagens que, num certo sentido, produzem medidas da qualidade, isto é, a qualidade é determinada através da comparação entre as evidências obtidas no processo de avaliação e os critérios definidos de acordo com uma variedade de processos. Nessas condições, a qualidade pode considerar-se como sendo real e independente do avaliador. No segundo caso, as abordagens utilizadas baseiam-se essencialmente na descrição, análise e discussão das percepções que os intervenientes no processo têm dessa mesma qualidade, isto é, a qualidade é construída socialmente e é reconhecida quando se vê e quando se sente.

Em qualquer dos casos a avaliação da qualidade, baseada no chamado pensamento criterial ou nas experiências vividas e sentidas pelos intervenientes, será sempre uma construção feita por seres humanos e, por isso, estará sempre fortemente relacionada com a experiência pessoal de cada um. Logo, sabe-se que nem todos verão e/ou sentirão o que outros veem e/ou sentem e, por isso mesmo, a qualidade para uns não chega a revelar-se enquanto que, para outros, é possível identificá-la quer através do que parece ser bom, quer através do que parece não o ser.

Como acima já se referiu, o propósito fundamental da maioria das avaliações é formular juízos acerca do valor e do mérito do que é avaliado (e.g. desempenho de uma escola, um programa de prevenção da toxicodependência, aprendizagens em ciência de alunos com 10 anos de idade). Na formulação de tais juízos estarão sempre presentes, de forma explícita ou implícita, apreciações relativas à qualidade do que se avaliou, tais como: "o ensino nesta escola tem em conta a diversidade dos alunos e, em geral, é de bom nível", "o programa destinado a apoiar a integração dos alunos provenientes de países africanos acabou por não responder às suas principais necessidades; segundo o que foi possível apurar junto dos principais intervenientes (alunos, pais, professores e psicólogos da educação) revelou-se aquém do que seria expectável" e "as aprendizagens em biologia dos alunos do 11.º ano com 16 anos de idade, medidas através de um teste estandardizado, podem considerar-se acima da média dos alunos do 11.º ano com 17 anos de idade ou mais". Repare-se que cada uma destas hipotéticas afirmações pressupõe uma certa forma de avaliar a qualidade. Nuns casos é clara a relevância dada ao conhecimento dos significados atribuídos

pelos intervenientes às experiências vividas no âmbito de um dado programa, enquanto que noutros parece incontornável a necessidade de se terem em conta determinados critérios ou *standards*, a partir dos quais será possível fazer apreciações quanto à qualidade do que se avalia. Numas situações, a qualidade é determinada pelo que as pessoas vivem e sentem, com o debate intersubjetivo entre elas a poder desempenhar um papel relevante no seu discernimento. Noutras situações, a qualidade é apreciada de forma mais objetiva, com base em critérios e em *standards* que se definem ou se constroem de formas mais ou menos consensuais.

Os *standards* têm a ver com um certo padrão de excelência, com aquilo que se considera como sendo algo de bom, algo que é desejável alcançar. E quem determina os *standards* são as pessoas - professores, pesquisadores, técnicos dos departamentos da administração, políticos - com base nos seus valores, nos seus sistemas de conceções e em referentes que, supostamente, lhes indicam o que é a qualidade. A definição de *standards* é complexa porque, em muitos casos, não é fácil conseguir que as pessoas se ponham de acordo relativamente à sua relevância, à sua utilidade ou ao seu real significado. Além disso, essa dificuldade também resulta do fato de nem sempre se fazer a necessária distinção entre *standard* e critério (STAKE, 2006). Segundo este autor, um critério indica-nos um descritor, um atributo ou uma característica de um dado objeto de avaliação (e.g. conhecimentos académicos; altura de uma pessoa), enquanto que um *standard* se refere à quantidade dessa característica ou desse atributo que é necessária para uma determinada valoração (e.g. 10 valores para poder transitar de ano numa dada disciplina; 175 cm de altura mínima para se poder ingressar numa dada equipa de basquetebol).

Ao dizermos que o desempenho de um aluno na resolução de problemas de Matemática é excelente, fazemos esta afirmação baseados num espetro mais ou menos alargado de critérios, tais como: a organização e clareza do trabalho escrito, a seleção e utilização de estratégias de resolução apropriadas; e os conhecimentos de Matemática. Mas, como nos diz Stake (2006), isso não significa que todos sejam utilizados ou que a cada um seja atribuído o mesmo peso, ou seja, "o pensamento baseado em critérios é importante, mas a interpretação também é igualmente importante" (STAKE, 2006, p. 46).

Podemos então dizer que a qualidade de um dado objeto pode ser determinada através de, pelo menos, duas abordagens distintas que decorrem de pressupostos epistemológicos igualmente distintos. Assim, numa abordagem, a qualidade pode ser determinada com base num conjunto de critérios e de *standards* através do qual, em geral, se encontra uma medida. Estamos perante uma perspetiva da qualidade como medida. Neste caso, uma das tarefas mais complexas a ser realizada pelos

avaliadores, é definir os critérios e *standards* a utilizar ou identificar o critério que melhor poderá determinar a qualidade de um dado objeto. A avaliação da qualidade é, nestas circunstâncias, um processo caracterizado pelo que Stake (2006) e Stake e Schwandt (2006) designam por *pensamento criterial*. Trata-se de uma visão muito difundida nos sistemas educativos e nas sociedades em geral, que vem sendo teorizada por autores e investigadores tais como Scriven (1991) e Schwartz e Mayne (2005), e que, no fundo, está subordinada a uma lógica de avaliação que se baseia nas seguintes ações:

- a) definir critérios que caracterizem o mérito;
- b) definir *standards*;
- c) medir o desempenho e compará-lo com os *standards*; e
- d) sintetizar e integrar os resultados de forma a que se produza um juízo acerca do mérito, valor ou significado do objeto.

Os programas internacionais de avaliação externa das aprendizagens dos alunos seguem esta perspectiva, particularmente os que têm sido desenvolvidos pela Organização para a Cooperação e o Desenvolvimento Económico (OCDE) como é o caso do *Programme for International Student Assessment (PISA)*.

Apesar de essa concessão de qualidade através da sua medida ser considerada a ideal, particularmente quando se está perante a avaliação de objetos cuja dimensão e características não permitem, ou tornam muito difícil, o estabelecimento de interações avaliador-intervenientes (e.g. avaliação das aprendizagens em ciências naturais dos alunos do 7.º ano de escolaridade de um dado país; avaliação do grau de satisfação em relação à profissão dos professores do ensino básico), subsistem problemas que não se podem ignorar. Em geral, os objetos de avaliação são complexos, não sendo fácil caracterizá-los rigorosamente com base em alguns indicadores. Além disso, definir *standards* e estabelecer determinadas medidas que, supostamente, traduzem o sucesso de um programa pode ser problemático e, em si mesmo, implica sempre uma decisão arbitrária.

Noutra abordagem, a qualidade é essencialmente determinada pelas experiências que uma diversidade de intervenientes viveram com os diferentes aspetos do objeto cuja qualidade se pretende avaliar. A perspectiva da qualidade como experiência baseia-se na proximidade do avaliador com o objeto de avaliação no seu contexto. Dessa forma procura-se compreender *in loco* e de forma mais ou menos profunda as vivências, a linguagem ou os significados subjetivos e intersubjetivos atribuídos aos fenômenos por parte daqueles que estão a viver *por dentro* um determinado programa. A interpretação das experiências pessoais dos intervenientes será um dos elementos essenciais na determinação da qualidade do que quer que esteja sob avaliação e, por isso, é dada relevância ao seu conhecimento prático.

Para Molander (1992), o conhecimento prático apresenta três características principais:

- a) baseia-se na participação e no diálogo com outras pessoas, assim como nas experiências vividas com todo o tipo de materiais e de circunstâncias físicas de uma situação concreta;
- b) não há propriamente uma separação, uma distinção, entre o conhecimento e a sua aplicação, entre o saber e o fazer; isto é, não constituem diferentes etapas de um processo; e
- c) não representa nem retrata a realidade traduzindo-se na resposta imediata às questões que se colocam ou na pronta execução das tarefas a realizar numa diversidade de contextos e de atividades humanas.

O conhecimento prático não deve ser sobrevalorizado porque é um tipo de conhecimento que é intrínseco à existência das pessoas e faz parte do mundo físico e social específico em que elas vivem, não tendo, por isso, a ver primeiramente com processos intelectuais e racionais que exijam um esforço propositado e deliberado de elaboração. No entanto, é uma fonte importante, ou mesmo crucial, de discernimento da qualidade e, nesse sentido, deve ser seriamente considerado, pois as experiências pessoais e o conhecimento que delas resultam são elementos essenciais para as apreciações que se fazem acerca da qualidade de um programa, de um currículo ou da organização e funcionamento pedagógico de uma escola.

As evidências destinadas a analisar a *qualidade como experiência* e as evidências que permitem estudar a *qualidade como medida* não são da mesma espécie. Nesse último caso as evidências são indicadores observáveis e mensuráveis que se podem comparar com critérios e *standards* que se definiram e/ou construíram previamente, enquanto que no primeiro caso a percepção da qualidade é construída a partir das narrativas obtidas dos diferentes intervenientes (STAKE, 2006).

De acordo com Stake e Schwandt (2006), também a perspectiva da *qualidade como experiência* tem, pelo menos, duas limitações importantes:

- a) há uma excessiva dependência da capacidade do observador para ouvir, ver e pensar com rigor e clareza assim como da sua reputação e credibilidade; não é fácil reconhecer as competências e o conhecimento prático de um avaliador ou avaliadora no que se refere à sua capacidade para observar, à profundidade e à abrangência com que o faz e à forma como controla os enviesamentos; e
- b) em situações de confrontação e de desacordo quanto à qualidade de um dado objeto de avaliação, serão poucas as pessoas que estarão disponíveis para aceitar as percepções pessoais de qualidade daqueles e daquelas cujas ideias se opõem às suas.

Stake (2006) e Stake e Schwandt (2006) consideram que a experiência humana está na origem da qualidade e que, sem a primeira, a segunda não fará real sentido. Por isso, poderá dizer-se que há qualidade porque as pessoas a vivem e experimentam e lhe dão um sentido e um significado num dado contexto. Atentemos nas palavras de Stake (2006, p. 47) que aqui se traduzem livremente:

O que reconhecemos de imediato como um melão ou um ensino de boa qualidade é algo que, no meu entender, tem a sua raiz em experiências vividas com eles, agora ou no passado. Apesar de, atualmente, já termos desenvolvido *standards*, convenções ou tradições, formais ou informais, para atribuir uma classificação aos melões ou ao ensino, a verdade é que as raízes do significado da qualidade estão na experiência que esta última proporcionou ao longo do tempo. Um relógio é de boa qualidade, em parte, porque não se atrasa nem se adianta e as suas peças são de muito bom fabrico, mas também porque as pessoas se sentem muito satisfeitas com ele e o consideram superior a outros relógios que conheceram. Isto faz com que nós, avaliadores e avaliadoras, na nossa atividade avaliativa, sejamos recetivos à idéia de reconhecer que a qualidade depende em grande medida de quem a tenha experimentado.

A discussão anterior parece evidenciar que a qualidade é um conceito complexo e multifacetado cuja natureza torna muito difícil, ou mesmo impossível, a sua integral representação. Além disso, é necessário ter em conta que, no domínio da educação e da formação (e.g. programas de apoio ao ensino da Matemática, programas de formação de adultos, programas de apoio à utilização de novas tecnologias de informação), a determinação da qualidade de algo é sempre datada e situada num dado contexto. Na realidade, muito dificilmente se poderá, num dado momento, determinar a qualidade de algo para sempre. Os juízos avaliativos são contingentes e interpretativos por natureza e não podemos impedir que assim seja. Num certo sentido, afirmar que algo tem qualidade não deixa de ser uma conjectura, uma suposição ou um juízo que se emite com mais ou menos convicção, com mais ou menos fundamentação. Por isso, a questão da avaliação da qualidade tem muito a ver com o que se poderá considerar ser o melhor e mais apropriado juízo que se faz do que se está a avaliar, num determinado contexto e circunstâncias, na posse de certo tipo de evidências e com determinado tipo de intervenientes. Em qualquer circunstância, a afirmação de que o que se avaliou tem qualidade estará sempre relacionada com o sistema de conceções e com as teorias sociais dos avaliadores.

Em qualquer processo de avaliação será sempre possível recolher dados relativos às experiências vividas pelos participantes e dados provenientes de processos de medida. É natural que numa situação em que ao que se está a avaliar ao

alcance físico do avaliador (e.g. avaliar a organização e o funcionamento de um departamento curricular de reduzida dimensão, avaliar o ambiente de avaliação, de ensino e aprendizagem numa sala de aula), seja dada mais ênfase à recolha de dados relativos às experiências vividas pelos diferentes intervenientes sendo muito reduzida a recolha de dados quantitativos. Por outro lado, em situações em que o que se está a avaliar esteja fora do alcance físico imediato do avaliador (e.g. avaliação das necessidades de formação dos professores de Matemática da zona centro do País; avaliação das aprendizagens dos alunos do 12.º ano no domínio da Filosofia) a tendência será para que a recolha de dados se baseie mais em processos em que se comparam as medidas obtidas indiretamente, através de qualquer tipo de instrumento adequado, com *standards* e critérios previamente determinados.

Em todo o caso, como nos dizem Stake e Schwandt (2006) se a avaliação de um dado ente que está ao alcance físico do avaliador se baseia fundamentalmente em dados quantitativos será designada como demasiado mecanicista. Se, por outro lado, a avaliação de um ente que está para além do alcance físico e imediato do avaliador é essencialmente construída com base nas experiências pessoais de um número indeterminado, mas sempre necessariamente reduzido de participantes, dir-se-á que estamos perante um trabalho excessivamente subjetivo.

Do que até agora se discutiu poder-se-á inferir que há duas conceções de qualidade de um dado ente que se pretende avaliar: uma que se baseia na definição prévia de critérios e *standards* formais; e outra que se determina a partir do encontro e da interação que se estabelece com o ente a avaliar. A partir daqui, o que parece constituir um importante desafio para os avaliadores é conjugar, equilibrada e inteligentemente estas duas visões, tendo em conta a natureza do que se pretende avaliar e as suas circunstâncias. Por exemplo, numa avaliação em larga escala destinada a avaliar a qualidade das aprendizagens dos alunos de um dado ano de escolaridade numa dada disciplina, é evidente que a definição de critérios e *standards* será primordial, enquanto que a interação social com os alunos tenderá a ser reduzida. A propósito deste exemplo e de outros de semelhante natureza, importa transcrever nesta altura algumas ideias de Stake e Schwandt (2006, p. 411) que parecem relevantes e que aqui se traduzem livremente:

O discernimento da qualidade é um comportamento humano. Pode basear-se em toda a panóplia de instrumentos usados na exploração espacial, mas não há reconhecimento da qualidade ou da má sorte até que uma mente o reconheça. Sem experiência não se pode reconhecer a qualidade ou a falta de qualidade de algo, mesmo se incluirmos a medida de condições e de perceções como oposição à referida base de experiência. A experiência não nos garante uma interpretação correta mas a sua ausência é quase uma garantia para uma interpretação incorreta.

Para estes autores parece clara a relevância das experiências vividas para efeitos da avaliação da qualidade de um dado objeto, mas fica também a ideia de que é necessário ponderar devidamente as circunstâncias a partir das quais será legítimo fazer inferências relativas a essa mesma qualidade. Efetivamente, as experiências ou as medidas obtidas podem não nos garantir, isoladamente, e por si sós, o melhor discernimento, devendo considerar-se a possibilidade de as conjugar para que eventualmente seja possível fazer inferências mais sustentadas e mais credíveis.

Avaliação e Discernimento Pragmático

A ideia do discernimento pragmático que discuti em trabalhos anteriores (FERNANDES, 2010, 2011), inspira-se em racionalidades decorrentes das perspectivas dos pragmatistas norte-americanos, de que John Dewey é a figura de referência, dos teóricos escandinavos contemporâneos da teoria da atividade (ENGSTROM, 1999; ENGSTROM; MIETTINEM, 1999) e ainda nos trabalhos filosóficos de alguns avaliadores (HOUSE, 2000; HOWE, 2003; HOUSE; HOWE, 1999). Estas racionalidades têm emergido como alternativas ao lógico-positivismo e ao construtivismo, paradigmas que, de algum modo, surgem na literatura como preponderantes e alternativos. A título de exemplo, refere-se que os teóricos da teoria da atividade não subscrevem as dicotomias com que normalmente somos confrontados na literatura das ciências sociais e da educação, rejeitando categoricamente uma visão dicotômica da realidade (e.g. quantitativo vs. qualitativo; objetivo vs. subjetivo; local vs. global). Além disso, em vez de conceptualizar a vida social organizada como uma pirâmide de estruturas rígidas todas dependentes de um centro de poder, veem-na antes como uma rede de sistemas de atividade que se sobrepõem e se inter-relacionam, propondo uma racionalidade alternativa à racionalidade do controlo e da generalização dos positivistas e à racionalidade relativista dos construtivistas. Tal racionalidade baseia-se na aceitação da relação dialética que existe entre cada indivíduo e a estrutura social em que está inserido, no reconhecimento da necessidade de se formularem juízos acerca do valor e do mérito dos entes e fenómenos sociais que nos rodeiam, assim como na admissão da existência de uma realidade objetiva que pode ser rigorosamente estudada e compreendida. No entanto, os pragmatistas consideram que podem existir diferentes perspectivas sobre uma dada realidade e, nesse sentido, defendem a ideia de que não há uma verdade absoluta acerca dessa mesma realidade. O que acontece é que, num dado momento, uma das perspectivas pode sobrepor-se a todas as outras.

No processo de recolha de informação, o método a utilizar deve estar dependente dos propósitos e das questões de avaliação, sendo importante articular métodos qualitativos e quantitativos. Além disso, os pragmatistas também advogam a articulação entre os métodos dedutivo e indutivo na construção do conhecimento e reconhecem a relevância das relações de causa e efeito; no entanto, consideram que é impossível afirmar com certeza absoluta que um dado efeito é o resultado

direto de uma dada causa. Finalmente, tal como os construtivistas, os pragmatistas defendem que a avaliação e a investigação não são livres de valores e que estes fazem parte dos processos de construção do conhecimento (CHRISTIE; FLEISHER, 2009).

House (2000) refere que as discussões paradigmáticas nos chamam a atenção para a natureza da avaliação e, muito particularmente, para o fato de não a podermos considerar uma ciência exata, no sentido clássico do termo. Na verdade, nas práticas de avaliação, em vez da demonstração e da lógica dedutiva e indutiva, tendo em vista a procura da verdade e dos resultados certos, utilizam-se formas de argumentação e racionalidades que visam à credibilidade, à plausibilidade e a utilidade dos resultados obtidos. Teremos, pois, que retirar as devidas consequências do fato de a avaliação ser um domínio do conhecimento que, em geral, não produz resultados certos ou exatos porque não nos pode dar a certeza da prova. Apesar disso, há abordagens de avaliação em que se utilizam "provas cartesianas" e métodos indutivos. Porém, facilmente se compreende que os resultados das avaliações muito dificilmente serão definitivos. Essas considerações remetem-nos para as seguintes palavras de Ernest House que aqui traduzo livremente:

[...] a avaliação persuade mais do que convence, argumenta mais do que demonstra, é mais credível do que certa e a aceitação que suscita é mais variável do que o que seria desejável. Isto não significa, porém, que se reduza à mera retórica ou que seja completamente arbitrária. O facto de não estar limitada ao raciocínio dedutivo e indutivo não significa que seja irracional (HOUSE, 2000, p.72).

A partir dos anos 80, perante a diversidade de abordagens de avaliação e depois de ter passado o período mais conturbado do debate paradigmático, começou a discutir-se a importância de articular perspectivas, procedimentos e metodologias nos estudos de avaliação (e.g. FERNANDES, 1994, 2010; FERNANDES; BORRALHO; AMARO, 1994; GREENE; CARACELLI; GRAHAM, 1989; HOWE, 2003; TALMAGE, 1982). Na verdade, começou a sublinhar-se a ideia de que, para obter descrições mais profundas e precisas dos objetos de avaliação (e.g. programa, projeto, aprendizagens dos alunos, medida de política), era importante articular, de forma inteligente e seletiva, o que se considerasse mais adequado das diferentes abordagens avaliativas. Para um número de teóricos do domínio da avaliação essa posição era impraticável uma vez que não se podiam articular perspectivas com pressupostos ontológicos, epistemológicos e metodológicos tão díspares e até antagónicos (GUBA; LINCOLN, 1981, 1989, 1994; LINCOLN; GUBA, 1985; STAKE, 1978). Mas, apesar destas posições críticas, a verdade é que os estudos de avaliação começaram, cada vez mais, a compatibilizar e a articular abordagens, métodos e procedimentos. Stake (2006), autor que vem do relativismo ontológico e epistemológico, chamou a atenção para a vantagem de se articularem abordagens de avaliação baseadas no chamado

pensamento criterial, mais objetivas, quantitativas e centradas essencialmente nos resultados, com perspectivas baseadas nas práticas e nas experiências pessoais que, em geral, são mais subjetivas, qualitativas e orientadas para os processos. Apesar de serem duas abordagens com pressupostos teóricos, metodológicos e epistemológicos distintos, a sua articulação inteligente, considerando os propósitos e as questões da avaliação, permite obter melhores avaliações e melhores juízos acerca do mérito e do valor do que quer que seja que for objeto de avaliação.

A discussão que tenho vindo a fazer sugere que o discernimento pragmático é indissociável de um conhecimento profundo das abordagens existentes no domínio da avaliação. Na verdade, elas constituem o patrimônio teórico e conceptual existente e traduzem o que tem sido possível aprender através da reflexão teórica, dos estudos empíricos e, em geral, das práticas. Não cabe no âmbito deste artigo discutir as abordagens de avaliação que têm vindo a ser analisadas, sistematizadas e organizadas segundo várias lógicas por uma diversidade de autores (ALKIN; CHRISTIE, 2004; FERNANDES, 1992, 1994; SCHWANDT; BURGON, 2006; SHADISH; COOK; LEVITON, 1991; STUFFLEBEAM, 2000; STUFFLEBEAM; SHINKFIELD, 2007; VIANNA, 2000). O que talvez seja importante referir nesta altura é que os esforços de racionalização, de sistematização e de discernimento são fundamentais para desenvolver a teoria e as práticas de avaliação.

A opção por uma perspectiva eclética, articulando e integrando, sempre que considerado necessário, uma ou mais abordagens de avaliação de programas, poderá ser a que melhor se ajusta à complexidade de muitos contextos em que as avaliações se desenvolvem. Tal opção deve depender mais dos propósitos e questões da avaliação e das suas potencialidades para gerar a informação que é necessária, e muito menos de ideias pré-concebidas acerca do que são abordagens legítimas de avaliação.

Em suma, o discernimento pragmático está associado, entre outras, às seguintes ideias fundamentais:

1. na avaliação de um dado objeto é importante ter em conta que não há uma verdade absoluta acerca de uma dada realidade;
2. as dicotomias (e.g. quantitativo vs. qualitativo, processo vs. produto, local vs. global) não contribuem para compreender profunda e cabalmente os fenómenos e as realidades que são objeto de avaliação, sendo necessário considerar possibilidades de articulação;
3. a avaliação, realizada a partir da integração de múltiplas perspectivas e da cooperação interdisciplinar, é socialmente mais útil no sentido em que pode contribuir para resolver mais eficazmente os problemas, melhorando a vida das pessoas, das instituições e da sociedade; e
4. a avaliação tem que estar baseada num forte enquadramento teórico que

resulte da integração e da utilização de abordagens que se revelem mais adequadas, tendo em conta os propósitos e as questões da avaliação e o contexto concreto em que se desenvolve.

Síntese e Reflexão Final

Este artigo foi escrito partindo do pressuposto que, sem uma discussão e uma compreensão aprofundada acerca das chamadas questões críticas da avaliação, dificilmente poderemos responder mais eficazmente às necessidades, cada vez mais prementes, de informação acerca do que acontece, a muitos níveis, nos sistemas educativos. Dessa forma, foi decidido abordar quatro questões críticas:

- a) avaliação Formal e Avaliação Informal;
- b) avaliação como (Trans)Disciplina;
- c) avaliação e Qualidade; e
- d) discernimento Pragmático e Desenvolvimento da Avaliação.

Nas duas primeiras discutiram-se aspetos mais associados à natureza da avaliação, a terceira abordou a questão do discernimento da qualidade dos objetos, questão central da avaliação, e, finalmente, a quarta, discutiu a idéia do discernimento pragmático que, no essencial, tem a ver com a necessidade de se abrirem novos caminhos ao desenvolvimento da avaliação, baseados noutras racionalidades, na integração de abordagens de avaliação e na relevância e utilidade social dos resultados da avaliação.

Vejamos, muito sucintamente, para cada uma das questões, os aspetos que pareceu importante sublinhar. No que se refere à avaliação formal e à avaliação informal, é importante reconhecer-se que ambas produzem conhecimento acerca dos objetos e se relacionam com formas mais ou menos complexas, podendo ser consideradas complementares. Ou seja, há uma espécie de evolução epistemológica que tem levado os investigadores a considerarem que o conhecimento produzido pelas avaliações tácitas, do dia a dia, não deve ser desprezado. Consequentemente, parece ser necessário ter em conta esta evolução.

A credibilidade da avaliação como disciplina científica parece estar dependente da consolidação de um núcleo central de definições, de conceitos integradores e de pressupostos básicos. Esta perspetiva contraria a de um número de investigadores para quem a avaliação é uma prática que se limita a utilizar os métodos e as ferramentas heurísticas das ciências sociais. Nessas condições, pode fazer sentido retomar a discussão empreendida por Michael Scriven há aproximadamente duas décadas, acerca da avaliação como (trans)disciplina (SCRIVEN, 1994) e consolidar os argumentos aí aduzidos relativamente às interpelações daqueles que sustentavam, e ainda sustentam, que a investigação e a avaliação são livres de valores. Por outro

lado, parece igualmente importante clarificar que a avaliação, como disciplina, tem como finalidade produzir conhecimento sobre determinados objetos e realidades, que possa ser utilizado por outras disciplinas. E é precisamente nesse sentido que Scriven a chamou uma (trans)disciplina.

A questão crítica por excelência em qualquer processo de avaliação tem a ver com o discernimento da qualidade. Apesar da dificuldade em definir o que se entende por qualidade do que quer que seja, parece relevante sublinhar que, dado um determinado objeto, será, em geral, possível avaliar se tem ou não qualidade. Obviamente que o poderemos fazer com maior ou menor rigor, reunindo maiores ou menores consensos, mas, em princípio, poderemos fazê-lo de forma credível e socialmente útil. A discussão evidenciou duas visões de qualidade que se revelam fundamentais para que possamos ser cada vez mais rigorosos na avaliação: a qualidade como medida, baseada em critérios e em *standards*, e que, num certo sentido, é independente do avaliador, e a qualidade como experiência, baseada na interpretação que o avaliador faz das experiências pessoais e das práticas daqueles que, de algum modo, vivenciaram um determinado objeto (e.g. programa, projeto, intervenção pedagógica). Assim, outro importante desafio é o desenvolvimento de estudos de avaliação que articulam estas duas visões de qualidade.

Finalmente, no que se refere ao discernimento pragmático, parece que se devem destacar três idéias fundamentais. A primeira, de natureza ontológica, diz-nos que há diferentes perspectivas sobre a realidade que se pretende conhecer e compreender, não havendo verdades absolutas sobre essa mesma realidade. A segunda está relacionada com a possibilidade de se explorarem racionalidades alternativas ao lógico-positivismo e ao relativismo ontológico e epistemológico dos construtivistas. Esta possibilidade passa por considerar que as dicotomias e as ortodoxias ontológicas, metodológicas e epistemológicas acabam por produzir visões redutoras da realidade, limitando a profundidade e a abrangência das análises. A terceira idéia tem a ver com o esforço que é necessário fazer para se articularem e/ou integrarem as múltiplas abordagens de avaliação que proliferam na literatura, fazendo depender a sua utilização dos propósitos e questões da avaliação em vez de qualquer visão acerca do que é a "boa" abordagem de avaliação. Nessa terceira idéia, é importante sublinhar que a teoria e os fundamentos da avaliação estão traduzidos na miríade de abordagens que têm vindo a ser desenvolvidas e propostas pelos investigadores e avaliadores ao longo dos anos. Logo, é fundamental conhecer profundamente tais abordagens.

Tendo em conta a discussão aqui desenvolvida, parece oportuno, em jeito de conclusão dar destaque às seguintes afirmações:

1. a avaliação não é uma ciência exata nem uma mera técnica, e é necessário retirar desses fatos as devidas ilações;

2. a avaliação, em geral, não se reduz a uma medida, sendo uma prática social sofisticada que exige participação e interação social. Note-se que as medidas são indispensáveis para se obterem boas descrições das realidades educacionais, mas não são boas para descrever tudo porque não é possível, ou é extremamente difícil, medir tudo;
3. a avaliação não produz, em geral, resultados exatos nem definitivos; porém, deverá ser credível, rigorosa e útil para todos os intervenientes no processo;
4. a avaliação permite-nos discernir a qualidade de qualquer objeto, sendo muitas vezes desejável que esse discernimento seja feito com base na utilização complementar de avaliações baseadas em critérios e de avaliações baseadas na experiência e nas práticas das pessoas; e
5. a avaliação deve ser utilizada para melhorar a vida das pessoas, das organizações e das sociedades, sem que dela se deixem de retirar as devidas consequências que, normalmente, implicam a regulação e a autorregulação do ente avaliado.

As questões discutidas neste artigo foram designadas como críticas porque, efetivamente, o entendimento que se desenvolver acerca delas é determinante para o tipo de avaliação que se vai pôr em prática. A discussão pareceu mostrar que a avaliação a desenvolver nos próximos anos tem que ser um processo com melhor integração teórica e melhor articulação entre diferentes visões epistemológicas e metodológicas que influenciam as práticas avaliativas. De igual modo, mostrou que, cada vez mais, a avaliação tem que ser um esforço partilhado por investigadores e avaliadores da diversidade de disciplinas que se revelem indispensáveis para descrever, analisar e interpretar rigorosamente uma dada realidade. E a verdade é que essa não tem sido a prática mais comum no mundo ainda algo difuso da avaliação em educação.

Referências

ALKIN, M. ; CHRISTIE, C. An Evaluation theory tree. In: ALKIN, M. C. (Org.), *Evaluation roots: Tracing theorists' views and influences*. London: Sage, 2004. p. 381-392.

CHRISTIE, C.; FLEISHER, D. Social inquiry paradigms as a frame for debate on credible evidence. In: DONALDSON, S. I., CHRISTIE, C. E MARK, M. (Ed.). *What counts as credible evidence in applied research and evaluation practice?* London: Sage, 2009, p. 19-30.

DONALDSON, S. ; LIPSEY, M. Roles for theory in contemporary evaluation practice: Developing practical knowledge. In: SHAW, I.; GREENE, J. ; MARK, M. (Ed.). *The Sage handbook of evaluation*. London: Sage, 2006. p. 56-75.

ENGESTROM, Y. Activity theory and individual and social transformation. In: ENGESTROM, Y., MIETTINEN, R.; PUNAMAKI, R-L. (Ed.). *Perspectives on activity theory*. New York, NY: Cambridge University Press, 1999. p. 19-38.

ENGESTROM, Y.; MIETTINEN, R. (1999). Introduction. In: ENGESTROM, Y., MIETTINEN, R.; PUNAMAKI, R-L. (Eds.). *Perspectives on activity theory*. New York, NY: Cambridge University Press, 1999.p. 1-18.

FERNANDES, D. Avaliação de programas e projetos educacionais: das questões teóricas às questões das práticas. In: FERNANDES, D. (Org.), *Avaliação em educação: olhares sobre uma prática social incontornável*. Pinhais: Editora Melo, 2011.p. 185-208.

_____. Acerca da articulação de perspectivas e da construção teórica em avaliação educacional. In: ESTEBAN, M. T. e AFONSO, A. J. (Org.). *Olhares e interfaces: Reflexões críticas sobre a avaliação*. São Paulo: Cortez, 2010.p. 15-44.

_____. Contornos de uma experiência de avaliação desenvolvida no Instituto de Inovação Educacional (1990-1993). *Boletim da Sociedade Portuguesa de Educação Física*, Lisboa,n.10/11, p. 7-32, 1994.

_____. *Prática e perspectivas de avaliação: Dois anos de experiência no Instituto de Inovação Educacional*. Lisboa: [s.n.],1992 Documento policopiado, 1992.

FERNANDES, D., BORRALHO, A. ; AMARO, G. Unicidade ou multiplicidade paradigmática na investigação em resolução de problemas? In: FERNANDES, D., BORRALHO, A.; AMARO, G. (Org.). *Resolução de problemas: processos cognitivos, concepções de professores e desenvolvimento curricular*. Lisboa: Instituto de Inovação Educacional, 1994.p. 303-308.

GREENE, J., CARACELLI, V. ; GRAHAM, W. Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, Thousand Oaks, CA,v.11,n. 3, p. 255-274, 1989.

GUBA, E.; LINCOLN, Y. Epistemological and methodological bases of naturalistic inquiry. In: STUFFLEBEAM, D., MADAUS, G. ; KELLAGHAN, T. (Ed.), *Evaluation models: viewpoints on educational and human services evaluation* (2nd. ed). Dordrecht: Kluwer, 2000.

_____.; _____. *Fourth generation evaluation*. Competing paradigms in qualitative research. In: DENZIN, N.; LINCOLN, Y. (Ed.). *Handbook of qualitative research*. London: Sage, 1994. p. 105-117.

GUBA, E. ; LINCOLN, Y. *Fourth generation evaluation*. Newbury Park, CA: Sage, 1989.

_____.; _____. *Effective evaluation*. San Francisco, CA: Jossey Bass, 1981.

HAMMERSLEY, M. Social research today: some dilemmas and distinctions. *Qualitative Social Work*, London, v.2, n. 1, p. 25-44, 2003.

HOUSE, E. *Evaluación, ética y poder*. 3.ed. Madrid: Morata, 2000.

HOUSE, E. ; HOWE, K. Deliberative democratic evaluation. In: KELLAGHAN, T., ; STUFFLEBEAM, D. (Ed.). *International handbook of educational evaluation*. Dordrecht: Kluwer, 2003.p. 79-102.

_____.; _____. Deliberative democratic evaluation in practice. In: STUFFLEBEAM, D.; MADAUS, G. ; KELLAGHAN, T. (Ed.). *Evaluation models: viewpoints on educational and human services evaluation*. 2nd. ed. Dordrecht: Kluwer, 2000. p. 409-422.

_____.; _____. *Values in education and social research*. London: Sage, 1999.

HOWE, K. *Closing methodological divides: toward democratic educational research*. Dordrecht: Kluwer, 2003.

LINCOLN, Y. ; GUBA, E. *Naturalistic inquiry*. London: Sage, 1985.

MAIO, G. ; OLSON, J. *Why we evaluate: Functions of attitudes*. Mahwah, NJ: Lawrence Erlbaum, 2000.

MOLANDER, B. Tacit knowledge and silenced knowledge: fundamental problems and controversies. In: GORANZON, B. ; FLORIN, M. (Ed.). *Skill and education: Reflection and experience*. London: Springer-Verlag, 1992.p. 9-31.

MUSCH, J. e KLAUER, K. *The psychology of evaluation: affective processes in cognition and emotion*. Hillsdale, NJ: Lawrence Erlbaum, 2003.

POLKINGHORNE, D. Psychological inquiry and the pragmatic and the hermeneutical traditions. *Theory and Psychology*, London, v. 10,n.4, p.453-479, 2000.

SCHMIDT, M. Grout: Alternative kinds of knowledge and why they are ignored. *Public Administration Review*, v.53, n.6, p.525-530, 1993.

SCHWANDT, T. ; BURGON, H. Evaluation and the study of living experience. In: SHAW I., GREENE, J. ; MARK, M. (Ed.). *The Sage handbook of evaluation*. London: Sage, 2006.p. 98-117.

SCHWARTZ, R.; MAYNE, J. Assuring the quality of evaluative information: theory and practice. *Evaluation and Program Planning*, Oxford, v.28, p.1-14, 2005.

SCRIVEN, M. Evaluation theory and metatheory. In: KELLAGHAN, T.; STUFFLEBEAM, D. (Ed.). *International handbook of educational evaluation*. Dordrecht: Kluwer, 2003a. p. 15-30.

_____. Evaluation in the new millennium: the transdisciplinary vision. In: DONALDSON, S. e SCRIVEN, M. (Ed.). *Evaluating social programs and problems: visions for the new millennium*. Mahwah, NJ: Erlbaum, 2003b, p.19-42.

_____. Evaluation ideologies. In: STUFFLEBEAM, D., MADAUS, G. E; KELLAGHAN, T. (Ed.). *Evaluation models: viewpoints on educational and human services evaluation* (2nd. Edition). Dordrecht: Kluwer, 2000, p. 249-278.

_____. Evaluation as a discipline. *Studies in Educational Evaluation*, v.20, p. 147-166, 1994.

SHADISH, W., COOK, T. ; LEVITON, L. *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage, 1991.

STAKE, R. Advocay in evaluation: A necessary evil? In: CHELIMSKY, E.; SHADISH, W. (Eds.). *Evaluation for the 21st century: A handbook*. London: Sage, 2004, p. 470-476.

_____. *Evaluación comprensiva y evaluación basada em estándares*. Barcelona: Graó, 2006.

_____. Program evaluation, particularly responsive evaluation. In: STUFFLEBEAM, D., MADAUS, G. ; KELLAGHAN, T. (Ed.). *Evaluation models: viewpoints on educational and human services evaluation*. 2nd.ed. Dordrecht: Kluwer, p.343-362, 2000.

_____. Responsive evaluation. In: KELLAGHAN, T.; STUFFLEBEAM, D. (Eds.). *International handbook of educational evaluation*. Dordrecht: Kluwer, 2003, p. 63-68.

_____. The case study method in social inquiry. *Educational Researcher*, Thousand Oaks, CA v.7, n.2, p.5-8, 1978.

STAKE, R. ; SCHWANDT, T. On discerning quality in evaluation. In: SHAW, I.; GREENE, J. E; MARK, M. (Ed.). *The Sage handbook of evaluation*. London: Sage, 2006. p. 404-418.

STUFFLEBEAM, D. Foundational models for 21st century program evaluation. In: STUFFLEBEAM, D.; MADDAUS, G.; KELLAGHAN, T. (Ed.). *Evaluation models: viewpoints on educational and human services evaluation*. 2nd. ed. Dordrecht: Kluwer, 2000. p. 33-83.

STUFFLEBEAM, D. ; SHINKFIELD, A. *Evaluation theory, models & applications*. San Francisco, CA: Jossey-Bass, 2007.

TALMAGE, H. Evaluation of programs. In: MITZEL, H. (Org.). *Encyclopedia of educational research*. 5th ed. New York, NY: The Free Press, 1982. p. 131-154.

VIANNA, H. M. *Avaliação educacional: teoria, planejamento, modelos*. São Paulo: Ibrasa, 2000.

WORTHEN, B. ; SANDERS, J. *Educational evaluation*. NY: Longman, 1987.

Recebido em: 19/12/2012

Aceito para publicação em: 18/04/2013