

NOVOS ESTUDOS EM AVALIAÇÃO EDUCACIONAL

HERALDO MARELIM VIANNA¹

1. AVALIAÇÃO E DECISÃO: Daniel L. Stufflebeam

1.1 Introdução

O modelo da avaliação educacional estruturado por Stufflebeam, Guba, Hammond, Provas, entre outros, (1971) resulta da proveitosa experiência que tiveram na década de 60, quando enfrentaram a tarefa de avaliar o sistema público de ensino elementar e secundário das escolas do distrito de Columbus, Ohio, nos Estados Unidos. A idéia nuclear do modelo centra-se no dimensionamento da avaliação com o objetivo de permitir tomada de decisões. Houve um posicionamento favorável a essa colocação, que passou a influenciar especialistas de diferentes países, inclusive na Inglaterra (Norris, 1993), como se pode constatar mesmo em críticos das metodologias de avaliação norte-americanas, alguns bem ácidos, a exemplo de Parlett e Hamilton (1972). As repercussões das idéias de Stufflebeam e seus companheiros foram bastante tímidas no Brasil, limitando-se a alguns poucos círculos acadêmicos [1].

¹ Pesquisador da Fundação Carlos Chagas, São Paulo – SP.

1.2 Concepção do modelo e definição

Ao conceberem o modelo, que ficou conhecido pelo anagrama CIPP — contexto, **input** (insumo), processo e produto —, Stufflebeam et al. (1971) deram-lhe um caráter analítico e racional, abrangendo diversos momentos: **Planejamento das decisões; Estruturação das decisões; Implementação das decisões e Reciclagem das decisões;** havendo para cada um desses momentos, respectivamente, uma forma específica de avaliação: (1) — avaliação do contexto; (2) — avaliação dos insumos (**input**); (3) — avaliação do processo e, finalmente, (4) — avaliação do produto. Verificamos, pois, que para Stufflebeam et al. a cada tipo de decisão corresponde um tipo de avaliação. É necessário ressaltar que a abordagem apresenta a avaliação segundo o papel que desempenha em um sistema de mudança social planejada. É a avaliação para melhorar e não para provar, conforme Stufflebeam et al. dizem no início do seu trabalho (1971).

O ponto central do modelo de Stufflebeam et al. (1971) é a própria definição que apresentam de avaliação, como um processo para descrever, obter e proporcionar informação útil para julgar decisões alternativas. Analisando-se esta definição, observa-se que destaca três elementos que são da maior importância para a compreensão do modelo que foi proposto:

1. A avaliação é um processo sistemático, contínuo.
2. O processo de avaliação pressupõe três momentos da maior importância:
 - 2.1 esboçar as questões a serem respondidas,
 - 2.2 obter informações que sejam relevantes para responder às questões propostas;
 - 2.3 proporcionar aos responsáveis pela tomada de decisões todas as informações necessárias.
3. A avaliação serve para a tomada de decisões.

Rose e Nyre (1977) mostram que o modelo CIPP desenvolvido por Stufflebeam et al. procura responder basicamente a quatro importantes indagações:

1. Quais os objetivos que devem ser concretizados?
2. Quais os procedimentos a adotar para realizar os objetivos?
3. Os procedimentos estabelecidos funcionam com a devida propriedade?
4. Os objetivos estão sendo realmente alcançados?

1.3 Tipos de decisão e tipos de avaliação

O modelo CIPP gira em torno da tomada de decisão e procura, desse modo, estabelecer quatro tipos de **situações de decisão** em educação, distinguindo-os dos quatro correspondentes **tipos de decisão**, além dos quatro **tipos de avaliação** (contexto, insumo, processo e produto). As situações de decisão são uma consequência direta da própria definição de avaliação, isto é, a abrangência de uma avaliação e o seu rigor são em grande parte determinados pela importância da decisão a ser tomada. Por sua vez, a importância da decisão está ligada ao significado da mudança que se pretende introduzir. Alguns tipos de decisão exigem avaliações de maior complexidade e realizadas dentro do maior rigor; desse modo, seus custos são elevados; outras, porém, que têm um menor impacto, não demandam a realização de avaliações com grande detalhamento [2].

A consideração da disponibilidade de informação e a capacidade da pessoa responsável pela decisão de usá-la é outro aspecto a considerar em situações de tomada de decisão. Quando há poucas informações disponíveis ou quando o responsável pela decisão não usa as informações da forma como as mesmas se encontram no momento, a avaliação deve, necessariamente, ter uma maior abrangência. Os dois fatores — significância das mudanças pretendidas e a disponibilidade de informações, assim como a capacidade de usá-las pelo responsável por tomar a decisão — quando cruzados, geram novas situações, entre outras as chamadas **pequenas versus grandes mudanças**, conforme Stufflebeam et al. (1971). A distinção entre pequena e grande mudança está na **intensidade da controvérsia** provocada pela mudança; assim, quanto mais controvérsia provocar, maior ou

mais importância terá a mudança. As maiores mudanças geralmente demandam uma maior reestruturação do sistema. As pequenas mudanças, contrariamente, não têm um impacto maior sobre o que a sociedade considera importante, assim, suas consequências são menores e praticamente não provocam maiores consequências. Esclarecendo a questão, a mudança de um livro didático no início do ano letivo é um exemplo bastante típico de pequenas mudanças, pois não costumam provocar controvérsias; alterações curriculares, sobretudo se se tratar de inovações, demandam maiores informações para fundamentar uma decisão, segundo a exemplificação de Stufflebeam et al. (1971).

O modelo de Stufflebeam et al. (1971) apresenta situações de decisão com denominações especiais: **homeostática**, **incremental**, **neomobilística** e **metamórfica**, referindo-se cada uma delas a uma mudança pretendida². Decisões homeostáticas têm como objetivo a manutenção do **status quo**. A maioria das decisões em educação é homeostática. As decisões incrementais referem-se a atividades de desenvolvimento, sobretudo as que constituem parte de um *continuum* de inovações. Ao contrário do que julgam Stufflebeam e associados, muitas inovações em educação são atividades incrementais, visam algum tipo de melhoria sem que isso represente um risco de contestação, sem criar espírito de revolta. As **decisões neomobilísticas** denotam atividades grandes e inovadoras, que são desenvolvidas com a finalidade expressa de solucionar grandes problemas. Uma programação educacional a longo prazo demanda decisões neomobilísticas. Finalmente, as **decisões metamórficas** visam a produzir mudanças completas e radicais, em um sistema educacional, sendo as idéias de Ivan Illich — **deschooling society** — um exemplo característico, na visão de Rose e Nyre (1977). Obviamente, as decisões metamórficas são inteiramente assintóticas, utópicas, e a probabilidade de implementá-las praticamente nula.

² Stufflebeam et al. (1971) usam vocábulos de outras áreas do conhecimento, como a biologia, a economia, a informática etc, o que demandou a elaboração de um glossário (p. 353/355), que foi utilizado no presente capítulo.

Há em cada uma dessas situações de decisão um número considerável de decisões educacionais que, segundo Stufflebeam et al. (1971), podem ser classificadas como:

1. **decisões de planejamento** para determinar os objetivos;
2. **decisões estruturais** para planejar os procedimentos e alcançar os objetivos;
3. **decisões de implementação** para examinar e reformular os procedimentos, se for o caso;
4. **decisões de reciclagem** para julgar e tomar posição em relação aos resultados dos objetivos.

A cada tipo de decisão corresponde, naturalmente, um dos quatro tipos de avaliação — CIPP — , conforme a discussão que segue.

1.4 Avaliação do contexto

A avaliação do contexto é o tipo geralmente mais usado em avaliação, sendo seu objetivo estabelecer necessidades, especificar população/amostra de indivíduos a considerar, e estabelecer os objetivos que devem concretizar as necessidades. É um tipo de avaliação que serve para o planejamento das decisões. Os procedimentos para a avaliação do contexto compreendem, especialmente:

1. definição e descrição do ambiente em que as mudanças ocorrerão;
2. identificação das necessidades e dos recursos disponíveis;
3. identificação de problemas e deficiências em relação às necessidades;
4. predição de futuras deficiências, levando em consideração o desejável, o possível e os resultados prováveis.

A avaliação do contexto proporciona elementos que servirão para justificar um determinado programa; desse modo, ainda segundo Stufflebeam et al. (1971), a avaliação do contexto procura dar respostas às seguintes indagações:

1. quais as necessidades não concretizadas que existem no contexto e que não são atendidas por uma determinada instituição?
2. quais os objetivos que devem ser alcançados para atender a essas necessidades?
3. quais os objetivos que devem receber apoio específico da comunidade?
4. qual o conjunto de objetivos com maiores possibilidades de concretização?

A análise das discrepâncias entre os objetivos da instituição (escola) e o desempenho dos alunos vai possibilitar a identificação das necessidades que não foram concretizadas, sendo preciso analisar as condições responsáveis pelas diferenças refletidas nas discrepâncias, com a finalidade de caracterizar os objetivos que devem ser alcançados. A consulta à comunidade, por sua vez, poderá dizer qual será a sua reação aos objetivos e qual o apoio que dará à avaliação. A factibilidade de muitos objetivos implica levantar custos e recursos disponíveis. A avaliação do contexto, assim, segundo os ensinamentos de Stufflebeam et al., além de contribuir para decisões relacionadas ao meio, possibilita, ainda, analisar o inter-relacionamento das várias partes do programa.

1.5 Avaliação de insumos

A **avaliação de insumos** procura estabelecer como usar os recursos para alcançar os objetivos definidos pelo programa. O seu produto final é o planejamento de procedimentos alternativos ou estratégias, considerando o potencial de custos-benefícios. No caso da avaliação dos insumos, que objetiva a **estruturação de decisões**, Stufflebeam et al. (1971) apresentam cinco indagações a serem respondidas:

1. a estratégia do projeto possibilita uma resposta lógica a um conjunto de objetivos específicos?
2. uma determinada estratégia obedece a todos os requisitos legais?

3. quais as estratégias já existentes e que são potencialmente importantes para realizar os objetivos pré-determinados?
4. quais os procedimentos específicos e o tempo (cronograma) necessários para implementar uma certa estratégia?
5. quais as características operacionais e os efeitos das diferentes estratégias que estão sendo consideradas?

A avaliação dos insumos resulta na especificação de materiais, procedimentos, cronograma de realização, condições materiais, pessoal, orçamento; enfim, oferece todos os elementos necessários à concretização dos objetivos.

1.6 Avaliação de processo

A **avaliação do processo**, destinada à implementação de decisões, realimenta, periódica e continuamente, os responsáveis pelo programa em todas as fases do desenvolvimento dos projetos desde o seu início. O seu objetivo é detectar deficiências de planejamento ou implementação, e monitorar vários aspectos do projeto, a fim de identificar e corrigir possíveis problemas. A avaliação do processo, mostra Stufflebeam et al. (1971), do mesmo modo que a avaliação formativa, promove o levantamento freqüente de informações e apresenta relatórios aos responsáveis pelo projeto com a finalidade de garantir o prosseguimento dos trabalhos. Stufflebeam et al. (1971) sugerem algumas perguntas para a avaliação do processo:

1. o projeto está seguindo o cronograma?
2. o staff precisa ser retreinado ou reorientado antes do término da presente fase do projeto?
3. as instalações e os materiais estão sendo usados adequada e apropriadamente?
4. quais os principais obstáculos a serem vencidos na atual fase do projeto?

A avaliação do processo, além de permitir o aprimoramento do programa de avaliação, possibilita, também, o registro de informações úteis para análises posteriores do programa em desenvolvimento ou de futuros programas.

1.7 Avaliação do produto

A **avaliação do produto** mede e interpreta os resultados obtidos em certos momentos pré-definidos do programa e ao seu término. É uma avaliação destinada a servir à reciclagem das decisões. Este tipo de avaliação procura, de acordo com Stufflebeam et al.,

1. identificar consequências e discrepâncias entre os objetivos pretendidos e os que foram realmente alcançados;
2. identificar os resultados não pretendidos, mas desejáveis, e outros que ocorram;
3. reciclar o programa para poder concretizar os objetivos que não foram alcançados;
4. dar informações ao responsável pela tomada de decisão quanto ao futuro do programa, sua continuação, término, modificações na sua estrutura etc.

A avaliação do produto, em síntese, determina discrepâncias entre o pretendido e o real, e analisa os fatores determinantes dessa diferença.

1.8 Modelo CIPP — procedimentos

O modelo de Stufflebeam et al. (1971) procura orientar o avaliador, na medida do possível. Apesar de não esclarecer maiores detalhes sobre o planejamento de cada tipo de avaliação (CIPP), oferece, entretanto, um **checklist** de procedimentos que são aplicáveis aos quatro tipos de avaliação:

1. enfocar a avaliação, identificando e definindo as situações de decisão com os objetivos da avaliação, as situações a serem conduzidas e a política a usar na sua operacionalização;
2. planejar a coleta de dados;
3. planejar a organização dos dados;
4. planejar a análise dos dados;
5. especificar a audiência a que serão destinados os relatórios, planejar sua formatação e cronograma;

6. administrar a avaliação desenvolvendo um plano geral da execução do planejamento³.

1.9 Abordagens alternativas

Stufflebeam (1974), ao discutir abordagens alternativas em avaliação educacional, estabeleceu uma diferença entre avaliação para tomada de decisão e avaliação relacionada à responsabilidade educacional (**accountability**). A primeira é proativa, semelhante na sua conceituação e na sua prática, à avaliação formativa. A avaliação relacionada à **accountability** é retroativa e tem um papel somativo. Se bem considerarmos o modelo CIPP, verificamos que os quatro tipos de avaliação — contexto, insumo, processo e produto — são formativos quando oferecem informações que possibilitam o aprimoramento do programa, sendo somativos no caso das informações estarem relacionadas ao futuro do programa.

Stufflebeam preocupou-se em deixar bastante claro os propósitos da avaliação, ressaltando que somente através de um processo que identifique suas necessidades, selecione as melhores estratégias, monitore as mudanças desejadas e meça o impacto dessas mudanças é possível evitar as distorções que podem invalidar todo um programa com vistas à promoção de mudanças na escola. A avaliação é a chave mestra desse processo, como bem destacou.

A questão da relação pesquisa e avaliação não poderia deixar de ser abordada por Stufflebeam, que, como outros, apresenta uma classificação dos vários tipos de avaliação. Vendense esses vários tipos de esquemas que mostram semelhanças e diferenças, como foi discutido anteriormente, nota-se que existe um traço comum entre pesquisa e avaliação, que é a obtenção de conhecimento para a compreensão dos fenômenos educacionais e a solução de seus problemas.

³ Os diversos passos para cada uma das fases do planejamento de uma avaliação foram detalhados por Stufflebeam e são apresentados de forma resumida em um texto extraído de *Evaluation as enlightenment for decision making*, publicado em Worthen e Sanders (1973) — *Educational Evaluation: Theory and Practice*. Belmont, CA: Wadsworth.

A proliferação de abordagens para a realização de estudos avaliativos também foi considerada por Stufflebeam e outros atuantes na área, como, por exemplo, Cronbach, Stake, Popham, Guba e House. Foi um esforço visando a evitar a instalação de um caos, mas ficou bastante claro, na elaboração desses esquemas taxonômicos, que a avaliação é uma atividade com múltiplos aspectos, a ser usada em diversos momentos do processo educacional e que o mesmo modelo pode ser classificado de diversas maneiras, dependendo, naturalmente, da ênfase que lhe é dada.

A análise da obra de Stufflebeam mostra a sua constante preocupação com as necessidades de uma definição apropriada de uma metodologia para cada um dos casos da avaliação educacional. Por outro lado, considera, em diferentes momentos, a importância do planejamento da coleta de dados, que nem sempre é feita de forma adequada, prejudicando, forçosamente, a tomada de decisões e o aprimoramento da educação. Lembremos o pensamento que apresenta na página inicial de seu livro básico (1971): **The Purpose of Evaluation is Not to Prove, But to Improve.** Realmente, a avaliação não visa a provar, mas sim a melhorar a educação.

Stufflebeam em diferentes momentos dos seus vários escritos apresenta uma séria preocupação com os problemas éticos envolvidos em uma avaliação e com a necessidade de um programa de pesquisas sobre as avaliações realizadas. Apesar de preocupar-se com ocorrências do contexto americano, acreditamos que essas preocupações sejam igualmente válidas para outros ambientes educacionais, inclusive o brasileiro. Seguindo esta linha de pensamento, é fundamental para os que se dedicam à avaliação uma certa familiaridade com os **Padrões para Avaliações de Programas, Projetos e Materiais Educacionais**, que não é obra exclusiva de Stufflebeam, pois foi elaborada por um amplo comitê de educadores, por ele presidido, na qualidade de representante do **National Council on Measurement in Education (NCME)** [3].

1.10 Padrões de Avaliação

Os **Padrões para Avaliação de Programas, Projetos e Materiais Educacionais**⁴ surgiram como decorrência de uma sugestão que fora feita pelos **Standards for Educational and Psychological Tests** (1974), envolvendo um grande número de pessoas e vários anos de árduo trabalho cooperativo, em que importantes organizações científicas norte-americanas se fizeram representar. O trabalho resultante somente foi divulgado após amplo processo de revisão crítica.

O desenvolvimento desse projeto, como seria previsível, demandou a superação de diversos pontos controversos relacionados com o seu conteúdo, a sua formatação e a experimentação a que deveria ser submetido, antes de sua divulgação na sociedade educacional. O grupo de cientistas, inicialmente formado por especialistas ligados a grandes instituições e centros de pesquisa, foi, depois, acrescido de outros elementos, incluindo diretores e superintendentes de escola, estatísticos, psicométristas, avaliadores, psicólogos, especialistas em currículo e professores, além de teóricos da educação e membros de diferentes conselhos escolares. Essa heterogeneidade veio a constituir-se em um elemento extremamente positivo, dando inclusive força política ao grupo junto a diferentes segmentos da sociedade.

A elaboração de **Padrões para Avaliação** era uma imposição geral, tendo em vista abusos cometidos e que foram denunciados por G. Madaus, conforme acentuaram Ridings e Stufflebeam (1981). Além disso, prestigiosas instituições, como a **American Educational Research Association** (AERA), a **American Psychological Association** (APA) e o **National Council on Measurement in Education** (NCME) consideravam imprescindível e inadiável a definição de marcos orientadores para a prática de avaliações, ainda que plenamente conscientes da natureza bastante controversa do assunto, sujeito a ser analisado sob

⁴ *Standards for Evaluations of Educational Programs, Projects and Materials*. New York. McGraw-Hill Book Co. 1981.

diferentes perspectivas, com a apresentação de pontos de vista muitas vezes irredutivelmente conflitivos [4].

As pessoas envolvidas com o problema, no contexto da sociedade norte-americana, apresentavam uma certa concordância quanto aos benefícios que adviriam da definição de padrões, como resumem Ridings e Stufflebeam (1981):

- linguagem comum que facilitaria a comunicação e a colaboração entre avaliadores;
- existência de um conjunto de princípios gerais que possibilitaria a manipulação de grande variedade de problemas ligados à avaliação;
- possibilidade de ter um quadro referencial de conceitos que permitiria estudar as questões de avaliação;
- estabelecimento de definições que orientariam pesquisas e o desenvolvimento de avaliações;
- bases para definir e caracterizar as questões relacionadas ao problema da responsabilidade educacional;
- possibilidade de apresentar uma descrição do estado da arte em avaliação;
- geração de um espírito de credibilidade do público em relação às avaliações a serem realizadas e disseminadas.

A memória do projeto revela que o seu processo de desenvolvimento foi consideravelmente longo, pois o grupo responsável buscou envolver diferentes elementos, conforme assinalamos anteriormente, procurou testar em campo a factibilidade dos vários elementos desenvolvidos, e, através de discussões e debates com a comunidade científica, esforçou-se para ouvir diferentes correntes de pensamento, inclusive as que tinham posições divergentes, a fim de esclarecer e solucionar os problemas surgidos, mostrando, inclusive, os limites do trabalho que estavam empreendendo, bastando, para esse fim, atentar para o próprio título do projeto: — **avaliação de programas, projetos e materiais educacionais**. A questão do uso da própria palavra **padrões** foi esclarecida pelo grupo para evitar que alguém pudesse interpretá-la como um conjunto de regras puramente mecânicas ou níveis de desempenho. A visão do grupo era a de que os padrões fossem princípios amplamente compartilhados

entre os especialistas na área e que serviriam para determinar a qualidade de uma avaliação, como, inclusive, destacaram Ridings e Stufflebeam (1981). Os padrões definidos ficariam, assim, na visão do grupo responsável por sua elaboração, restritos ao campo da avaliação, ainda que, eventualmente, pudessem ser aplicados a outras áreas, como a da saúde, por exemplo, o que depois veio efetivamente a ocorrer.

A elaboração dos padrões levou em consideração os atributos de uma avaliação: A) utilidade, B) praticabilidade, C) propriedade e D) precisão. Os padrões de **utilidade** pretendem assegurar que uma avaliação satisfaça as necessidades de informações práticas de determinados públicos. Os padrões de **praticabilidade** buscam garantir que a avaliação seja realista, realizada com a devida prudência e economia. Os padrões de **propriedade** visam a fazer com que a avaliação seja conduzida de acordo com as normas legais e éticas, além de levar em conta o devido bem-estar dos envolvidos na avaliação e dos afetados por seus resultados. Os padrões de **precisão** procuram fazer com que uma avaliação revele e transmita informações tecnicamente adequadas sobre as características do objeto estudado e que determinem seu mérito ou valor. O conjunto dos padrões, reunidos em quatro grupamentos, de acordo com os atributos de uma avaliação adequadamente conduzida deixa evidenciada a sua importância na meta-evaliação. (Vide **Standards for Evaluation of Educational Programs, Projects and Materials**, 1981).

Os **Padrões de Avaliação**, desenvolvidos no período de 1976 a 1980, refletem, como acentuam Ridings e Stufflebeam (1981), a preocupação da avaliação educacional, especialmente a partir do final dos anos 60, década em que houve grande intensificação das suas atividades, sobretudo no contexto norte-americano, quando começou a surgir, então, da parte de alguns segmentos, o interesse de verificar se as avaliações de programa realizadas eram responsivas às necessidades da comunidade. Viu-se, assim, no conjunto de padrões relativos à utilidade, que o avaliador não pode desconhecer os destinatários do seu trabalho, deve dar as informações que lhes são necessárias e apresentar os

resultados com elementos relevantes para atender às expectativas e às necessidades dos que vão utilizar os resultados da avaliação. O conjunto relativo à praticabilidade mostra que a avaliação deve considerar a questão do custo/benefício, pois uma avaliação pode ter um custo considerável e oferecer, ao final, resultados pouco significativos, que não justificam o empreendimento. A avaliação deve, ainda, em função desses padrões, ser planejada para ser realizada no contexto natural, em tempo dimensionado com precisão e usando um grupo técnico estritamente necessário, ou seja, a avaliação tem que ser realista, prudente, politicamente viável e realizada com parcimônia financeira.

Os padrões relacionados à propriedade mostram que a avaliação apresenta efeitos que vão afetar diferentes pessoas — administradores, professores, alunos e pais — de diferentes maneiras; desse modo, deve ser preocupação do avaliador não violar os direitos dessas pessoas, antes precisa ter em mente estratégias que protejam esses direitos; por conseguinte, a avaliação deve ser conduzida levando em conta os preceitos legais e éticos que asseguram os direitos dos que estão envolvidos no processo e dos que são afetados por seus efeitos, destacam com bastante propriedade Ridings e Stufflebeam (1981).

Será que as informações geradas pela avaliação revestem-se realmente de importância? Foram obtidas de acordo com os procedimentos técnicos? Os dados oferecidos permitem conclusões relevantes, que respondem às necessidades que levaram à avaliação? Os padrões relacionados à precisão constituem o conjunto mais numeroso, pois, basicamente, traduzem a preocupação maior do avaliador, relacionada com a validade de todo o processo de avaliação.

Os padrões foram apresentados de maneira simples e sem usar jargão técnico que pudesse dificultar o seu entendimento e a sua aplicação. Ainda que os responsáveis por sua elaboração acreditem que sejam aplicáveis a todas as avaliações, devemos, entretanto, observar que, dependendo dos aspectos a serem objeto de consideração na meta-avaliação, alguns aspectos são mais relevantes do que outros, ou seja, os Padrões possuem uma importância relativa na análise de uma avaliação. Precisamos

considerar, também, que esses Padrões somente serão úteis na medida em que a aplicação implicará melhoria da qualidade do trabalho de avaliação.

Alguns outros pontos devem ser considerados, destacando-se inicialmente, o fato de que, apesar dos esforços, os Padrões não visaram a perfeição. A sua aplicação faz-se necessária para identificar os pontos que precisam reparos e as revisões que devem ser feitas, como reconhecem seus próprios elaboradores, sem o que a sua qualidade não será aprimorada. Além disso, é necessário, também, que esses padrões sejam objeto de discussões entre os avaliadores e eliminado o comportamento habitual, o de ser a avaliação um trabalho solitário, da parte do avaliador, e isolado, sem uma discussão mais aprofundada com a comunidade educacional.

É preciso considerar, também, que esses Padrões pressupõem da parte dos avaliadores, aspirantes ao seu uso⁵ e aplicação, um treinamento, com a participação de quem já vivenciou o problema e enfrentou, assim, as dificuldades que costumam oferecer aos que estão realizando uma avaliação de avaliações.

A partir desses Padrões, algumas pesquisas ainda se fazem necessárias sobre questões relevantes ligadas à sua própria natureza, destacando-se, entre outras, pesquisas que determinem a sua validade em relação ao aprimoramento das avaliações; comprovem a importância efetiva de seu uso na decisão de avaliar e em aspectos do seu planejamento, entre outros; e mostrem que, apesar de seus custos, existem benefícios que justifiquem o seu emprego, como, por exemplo, o aumento da credibilidade e da fidedignidade das avaliações. Estas e outras pesquisas, como explicitam Ridings e Stufflebeam (1981), podem contribuir para a revisão e atualização dos Padrões, que, indiscutivelmente, são de grande utilidade para os avaliadores, os quais, lamentavelmente, no nosso contexto, e também em outros, continuam a ignorá-los.

⁵ Ver exemplo de sua aplicação em Worthen e Sanders (1987).

1.11 NOTAS E COMENTÁRIOS

- [1] O modelo CIPP, proposto por Stufflebeam e sua equipe no final dos anos 60, representa uma alternativa para as idéias sobre avaliação que existiam àquela época, que, no geral, tinham como características os seguintes aspectos: **orientação por objetivos; uso de instrumentos de medida e emprego de um design experimental**. Os conceitos apresentados por Stufflebeam (1969) e, especialmente, as metodologias oferecidas por Scriven (1967) e Stake (1967), dão novas dimensões à avaliação educacional.

A metodologia apresentada por Stufflebeam et al. (1971) parte do princípio de que “**The most importante purpose of program evaluation is not prove but to improve**”. Aprimorar e não provar é o ponto central do modelo CIPP. A fim de compreender o processo de desenvolvimento do modelo CIPP, precisamos considerar alguns aspectos do contexto em que ele surgiu, como já o fizemos anteriormente ao nos referirmos à evolução da avaliação. A lei que definiu a educação compulsória nos Estados Unidos (1965) — **Elementary and Secondary Education Act (ESEA)** — estabeleceu que todos os projetos financiados por órgãos federais deveriam ser obrigatoriamente avaliados, o que gerou uma crise, tendo em vista que os educadores não tinham experiência em planejamento e realização de projetos de avaliação (Stufflebeam, In: Madaus et al., 1993), o que vem ocorrendo, atualmente, no Brasil, na década de 90, em que se observa um **boom** de avaliações nem sempre bem planejadas e implementadas.

Além da falta de treinamento e experiência, ficou patente que não existiam modelos que atendessem às necessidades dos novos projetos. Centros educacionais propõem diferentes modelos, inclusive o **Ohio State University Evaluation Center**, onde vai ser desenvolvida a versão inicial do modelo CIPP. O Centro foi contratado para avaliar as escolas públicas de Columbus, Ohio. A avaliação inicialmente proposta foi a tradicional, seguindo as idéias apresentadas por Tyler no seu magnífico artigo — **General Statement on Evaluation** (Tyler, 1942) —, que vai influenciar a avaliação norte-americana por mais de 30 anos e que ainda hoje se reveste de grande atualidade, sendo lamentável que alguns segmentos da atual geração de avaliadores, no contexto brasileiro, não leiam e não possam, assim, meditar sobre a problemática da avaliação, a partir dos pressupostos de Tyler.

Stufflebeam e sua equipe logo descobriram que o modelo não seria adequado para avaliar os diversos projetos do sistema público de ensino de Columbus (Stufflebeam, In: Madaus et al., 1993), entre outras razões, por que os objetivos definidos, escritos por consultores educacionais e administradores, que tinham uma experiência bastante limitada da rede de ensino, não refletiam a realidade vivenciada pelos estudantes. Além da

falta de adequação dos instrumentos (testes), o enfoque tyleriano somente permitiria ter os relatórios conclusivos no final de cada projeto, o que não correspondia à expectativa da comunidade educacional.

Algumas providências foram imediatamente tomadas, iniciando-se pela definição de avaliação, considerada como a congruência entre desempenhos e objetivos (Tyler, 1942), e que passa a ter um novo enfoque, com a redefinição em termos de um processo destinado a fornecer informações úteis para a tomada de decisões (Stufflebeam, 1969), definição esta que foi rapidamente assimilada pelo mundo da avaliação norte-americana e por outros contextos educacionais, sendo que no caso brasileiro a aceitação é meramente formal, teórica. Tendo associado avaliação à tomada de decisões, Stufflebeam e sua equipe preocuparam-se em identificar os principais **tipos de decisão** que deveriam tomar e, a partir desse ponto, estabelecer as **estratégias** mais adequadas, o que constitui o ponto focal do seu sistema.

Inicialmente, foram enfocadas as decisões de **implementação** do projeto (envolvimento dos professores, alocação de recursos, adaptação de material instrucional, envolvimento da comunidade, comunicação entre os participantes da avaliação etc.) e a elaboração de relatórios sobre os resultados do **processo**. Outra grande decisão seria a relacionada com a continuidade ou término do projeto, aumento ou diminuição do apoio financeiro, institucionalização do projeto etc, sendo estas decisões caracterizadas como de **reciclagem** e a sugestão de que fossem apoiadas por informações sobre o que o projeto tinha **produzido**. Assim, nesse momento, o projeto tinha procurado conceituar a avaliação como uma **avaliação de processo** para orientar a **implementação** e a **avaliação de produto** para servir às decisões de **reciclagem**. O novo, e que fugia à proposta de Tyler, estava na **avaliação do processo**.

A avaliação do modelo por diferentes grupos sofreu críticas por não ter considerado a avaliação de objetivos. Stufflebeam e sua equipe mais uma vez intervieram, propondo a avaliação e o relatório das necessidades dos estudantes e dos problemas do sistema, como forma de auxiliar os educadores na escolha de objetivos relevantes (Stufflebeam, In: Madaus et al., 1993), e que para as decisões sobre planejamento fosse realizada a **avaliação do contexto**. O esquema avaliativo ainda não estava completo, porquanto não tinham sido especificadas decisões sobre os **meios** (horários, procedimentos, planos administrativos etc) para alcançar um determinado conjunto de objetivos. Estas decisões foram denominadas de **decisões estruturais** e relacionadas à **avaliação do input**.

O esquema geral do modelo CIPP, nesse momento de sua evolução, conforme a apresentação de Stufflebeam, (1993) apresentava as seguintes características: avaliação do contexto para informar às decisões de **planejamento**; avaliação do **input**, para servir às decisões **estruturais**;

avaliação do **processo** para orientar as decisões de **implementação**, e avaliação do **produto** a ser utilizada nas decisões de **reciclagem**.

O modelo de Stufflebeam não se concentra apenas na realização de um estudo de avaliação. Simultaneamente, procura prestar serviços aos responsáveis pela tomada de decisão em uma instituição educacional, como mostra o resumo a seguir, que procura destacar a relevância dos quatro tipos de avaliação para a tomada de decisão e a "accountability", segundo Stufflebeam (In: Madaus et al., 1993).

	Orientação formativa (tomada de decisão)	Orientação somativa (accountability)
Avaliação do CONTEXTO	Orientação na escolha dos objetivos e fixação de prioridades	Registro de objetivos e base para a sua escolha ao longo de necessidades, oportunidades e problemas
Avaliação do INPUT	Orientação para a escolha de estratégias do programa. Especificação para procedimentos do planejamento.	Registro da estratégia escolhida e seu planejamento; razões para sua escolha entre várias alternativas.
Avaliação do PROCESSO	Orientação para implementação	Registro do processo real
Avaliação do PRODUTO	Orientação para o término, continuação, modificação ou instalação	Registro dos resultados e decisões de reciclagem

Ao mesmo tempo que Stufflebeam desenvolvia o modelo CIPP, Stake trabalhava na abordagem **Countenance**, que depois seria incorporada à Avaliação Responsiva, como veremos na discussão sobre Stake. Há uma grande identidade metodológica entre estes modelos, com mínimas diferenças. As diferenças são maiores se compararmos a abordagem Formativa/Somativa de Scriven com o modelo CIPP. Scriven dá maior destaque à avaliação somativa, enquanto Stufflebeam caracteriza a avaliação para a tomada de decisão como formativa e a avaliação para a responsabilidade educacional como somativa. As divergências parecem maiores entre Stufflebeam e Scriven, se lermos o magnífico trabalho deste último — **Evaluation Ideologies**, sobre o qual fizemos algumas observações no capítulo relacionado a Scriven; entretanto, diga-se, o próprio Stufflebeam (1993) é o primeiro a reconhecer que as divergências entre ambos são mais aparentes que reais e decorreram de diferentes perspectivas e experiências.

[2] Avaliação para tomada de decisão, conforme o modelo de Stufflebeam (1993), somente ocorre se: a) serve para uma decisão; b) implica escolha entre alternativas; e c) é usada para atribuir recursos para um próximo intervalo de tempo entre uma e outra decisão, segundo chama a atenção Borich, In: Walberg e Haertel (1990). Não é um método ou uma técnica para tomar decisões, é mais uma perspectiva para organizar e enfocar conceitos da avaliação em relação a uma futura decisão. A abordagem procura maximizar a utilização dos resultados da avaliação. O avaliador, na abordagem, passa a ser uma extensão do responsável pela tomada de decisão. Assim, nessa abordagem, além do seu papel técnico, o avaliador assume uma posição de interface com o **decisor** no delineamento das alternativas de decisão e oferece informações que possibilitam selecionar alternativas de ação. O avaliador trabalha em colaboração com o responsável pela tomada de decisão não só para desenvolver uma percepção das decisões a tomar e da definição as alternativas, mas, também, ao definir os critérios avaliativos a serem usados na avaliação. A responsabilidade do avaliador, na abordagem, é grande, cabendo-lhe elaborar um sumário dos dados para decisão de forma prática e apresentar proposições que não sejam ambíguas nos seus termos. O posicionamento, entretanto, não é pacífico. Há quem defende a tese de que a tomada de decisão é responsabilidade única e exclusiva do agente contratante da avaliação.

A questão do paradigma da abordagem **decision-making** é complexa, tendo em vista que há uma diversidade de contextos e de diferentes abordagens, em que os responsáveis pelas decisões muitas vezes têm diferentes valores, prioridades e posicionamentos políticos diferentes. Contextos diversos, estilos administrativos diferentes, objetivos organizacionais diferentes acabam, entre outros elementos, gerando uma diversidade de referenciais de procedimento na condução de uma avaliação **decision-making**, onde se destaca, por exemplo, o modelo de Provus (1972/73), que se desenvolve em cinco estágios para avaliação de processo: planejamento (descrição do input, processo e resultado), instalação (operações de campo), processo (medidas), produto (resultados terminais) e comparações do programa, última fase em que o avaliador compara o programa experimental com um outro programa. Em cada fase do processo, segundo Provus, há três atividades: a) concordância sobre os padrões (objetivos) do programa; b) determinação de discrepâncias existentes entre aspectos do programa e c) uso das informações sobre as discrepâncias para determinar as deficiências do programa. As discrepâncias são levadas ao conhecimento do responsável pela tomada de decisão para saber se continua com o programa, na sua fase seguinte, ou revê os padrões ou as operações. A questão do modelo **Adversary Evaluations** já foi anteriormente analisada e a ela remetemos o leitor.

O avaliador nessa abordagem não pode ser ingênuo, devendo acautelar-se contra as distorções do uso das informações pelo responsável pela tomada de decisão, que pode favorecer certos aspectos bem sucedidos de um programa, e ignorar possíveis falhas existentes nos mesmos, gerando, assim, uma falsa imagem favorável ao programa junto ao público (alunos, professores, pais etc) e às autoridades governamentais, adverte com bastante propriedade Borich no seu ensaio.

- [3] A questão da definição de padrões em educação é das mais complexas, como bem mostra Casassus⁽¹⁾, e estamos apresentando esta nota para chamar a atenção para o assunto, que não está diretamente ligado aos padrões discutidos a seguir em [4]. Os padrões são construtos teóricos, são referenciais elaborados e aceitos por pessoas com experiência e autoridade na área. Sem a definição de padrões é impossível agir com segurança e, ao mesmo tempo, ter confiança nas ações que empreendemos. A partir desse ponto, podemos dizer que os padrões são construtos teóricos que permitem atuar com confiança (e segurança) nas ações educacionais. É impossível realizar uma avaliação do rendimento escolar sem a perspectiva dos padrões educacionais que orientam qualquer análise e discussão.

Os padrões são internalizados por quem os adota, mas necessitam ser externalizados para que sejam do conhecimento público, a fim de que possam ser definidas responsabilidades e possa ser exigida uma prestação de contas (*accountability*). Os padrões adotados têm uma relação estreita com a qualidade das nossas ações, indicam o nível de qualidade daquilo que empreendemos em educação.

Ainda que os padrões sejam de grande relevância em educação, somente nos dias que fluem começam a ser discutidos e definidos (o que não é o caso do Brasil). Assim, é preciso que de modo bastante claro fiquem definidas as **competências conceituais e práticas** esperadas dos alunos, o que é de suma importância para o trabalho educacional, inclusive no campo da avaliação e deve ser de conhecimento de toda a sociedade, incluindo alunos, pais e professores, o que já foi feito, por exemplo, nos Estados Unidos, que definiram os padrões educacionais para o ensino básico obrigatório até o ano 2000. É preciso não confundir padrões educacionais com objetivos instrucionais, curriculares. Os padrões são em número limitado e referem-se ao essencial; enquanto os objetivos são específicos e podem ser utilizados para alcançar os padrões. Entre esses padrões não devemos esquecer que existem **padrões básicos**, que devem ser alcançados por todos; por outro lado, existem, ou devem ser considerados, os **padrões de excelência**, que são ideais, alcançáveis por alguns poucos, mas que precisam ser previstos. Os padrões não podem ficar restritos ao mínimo (as célebres competências mínimas), precisamos considerar igualmente o máximo, ainda que reconhecendo

que essa é uma temática bastante complexa e polêmica. Devemos pensar, também, que os padrões devem ser expressos de forma a que possam ser **observados**, **medidos** (se for o caso) e **avaliados**, para que sejam definidas responsabilidades e prestação de contas.

A presente nota objetivou chamar a atenção para o problema, especialmente em nosso contexto educacional, que apresenta uma grande diversidade sócio-cultural. Sem uma definição desses padrões, as análises comparativas de macroavaliação, como as que ora vêm sendo feitas neste final da década de noventa, não se revestem de sentido, apesar de todo aparato tecnológico utilizado e a construção de escalas matematicamente corretas, mas que não fazem sentido para a comunidade escolar e para a sociedade.

(1) CASASSUS, J. *Distinciones conceptuales relativas a estandares en educación*. Paper (1996).

- [4] Os 30 padrões de avaliação resultaram do esforço hercúleo de representantes de 12 instituições, totalizando aproximadamente 200 pessoas, sob a coordenação de Stufflebeam. Esses padrões têm uma dupla finalidade: — possibilitar o desenvolvimento da avaliação de programas, projetos e materiais, e permitir o julgamento da eficiência dessa avaliação, ou seja, orientar o processo de meta-avaliação. Há um consenso entre as grandes instituições norte-americanas (AERA - American Educational Research Association, APA - American Psychological Association e NCME - National Committee on Measurement in Education, entre outras) que esses padrões são indispensáveis na avaliação de avaliações (**meta-avaliação**), por constituírem, como destacam Stufflebeam e Madaus (In: Madaus et al., 1993), uma linguagem comum que facilita a comunicação e a colaboração em avaliação; um conjunto de princípios gerais que permite lidar com grande diversidade de problemas de avaliação; um referencial teórico para a realização de estudos de avaliação; um conjunto de definições operacionais para orientar a pesquisa e o desenvolvimento do projeto de avaliação; uma descrição pública do estado da arte em avaliação educacional; uma base para os avaliadores se reportarem à responsabilidade educacional; uma ajuda no sentido de desenvolver a credibilidade pública em relação à área da avaliação educacional. Ressaltemos, como inclusive o fazem Stufflebeam e Madaus (In: Madaus et al., 1993), que os **padrões** referem-se **exclusivamente** à avaliação de programas, projetos e materiais, sendo excluídos outros tipos de avaliação.

Os padrões de avaliação estão agrupados em quatro grupos, de acordo com os atributos de uma avaliação: — sua **utilidade**; sua **factibilidade**, sua **propriedade** (adequação; correção) e sua **precisão**. Os padrões de **utilidade** refletem o consenso na literatura relacionado à

necessidade de a avaliação de programas ser responsiva às indagações das clientelas. Os padrões, em geral, exigem que os avaliadores conheçam as várias audiências interessadas nos resultados da avaliação, estabeleçam as informações de que precisam, e apresentem clara e concisamente informações relevantes no tempo próprio. Os padrões relacionados a **factibilidade** em termos materiais e de custos-benefícios. A avaliação exige um plano definido, operativo, prático e parcimonioso em relação aos recursos financeiros. Ou seja, o padrão factibilidade significa ser: — realista, prudente, politicamente viável e parcimonioso, no dizer de Stufflebeam e Madaus (In: Madaus et al., 1993). O terceiro grupo de padrões — sua **propriedade** — reflete o fato das avaliações afetarem as pessoas de diferentes modos; assim, são padrões que garantem os direitos das pessoas e os protegem, considerando que muitas avaliações têm implicações éticas. A avaliação deve, assim, ser realizada de acordo com as normas legais, levar em consideração o bem-estar de todos os envolvidos, e dos que podem ser afetados por seus resultados. Os padrões relacionados com a **precisão** demandam que as informações levantadas sejam tecnicamente adequadas e que as conclusões estabelecidas estejam logicamente relacionadas aos dados. A análise conjunta dessas características servirá para demonstrar a **validade global** da avaliação. (Ver o resumo dos padrões de avaliação coordenados por Stufflebeam et al. e uma tabela relativa à consideração de algumas características de uma suposta avaliação em Madaus et al., 1993).

2. AVALIAÇÃO RESPONSIVA: Robert E. Stake

*Les mathématiques n'épuisent pas le monde:
elles ne sont qu'un élément de la nature et
des modes de connaissance de l'homme.
(pensait Nicolas de Cuse).*

Karl Jasper⁶

2.1 Introdução

Stake é incontestavelmente uma das grandes figuras da educação norte-americana, ao lado de Tyler, Cronbach, Scriven e Stufflebeam, entre outros, no campo da avaliação, área a que deu enorme contribuição, espalhada, entretanto, em ensaios e artigos publicados em diferentes revistas de nível internacional. Uma

⁶ Jasper, Karl (1994) – *Initiation à la méthode philosophique*. Éditions Payot & Rivages. Paris, p. 16.

síntese de sua carreira como educador e avaliador — doutorado com tese sobre análise fatorial, professor de matemática na Academia Naval de Annapolis (Maryland, USA), professor, pesquisador e avaliador, chegou ao ápice de sua carreira como **full professor**, lecionando e orientando alunos nacionais e estrangeiros, inclusive brasileiros, na University of Illinois (Urbana), especificamente no **Center for Instructional Research and Curriculum Evaluation** (CIRCE). A obra que produziu tem um caráter multiforme e aborda diferentes problemas, sendo que, no presente trabalho, inicialmente, nos fixaremos na apresentação de seu modelo **countenance**, na avaliação responsiva, já apresentada na parte relacionada a considerações teóricas e posicionamentos, e, por último, na metodologia do estudo de caso, sua grande contribuição para a avaliação qualitativa. Eventualmente, no final, na parte de **notas e comentários**, enfocaremos outros aspectos, mas brevemente e na medida do possível [1].

2.2 Comunalidade e singularidade

A visão de mundo de muitas pessoas se direciona para a falsa suposição de que tudo se categoriza como comunalidade ou singularidade, para usar as expressões de Stake (1983a) retiradas de uma de suas áreas de especialização, a análise fatorial. Ou seja, as pessoas concentram suas atenções sobre aquilo que é comum a todos os sujeitos ou no que é singular a cada um dos elementos estudados. Stake (1983a) lembra que a argumentação relacionada à singularidade e à subjetividade é considerada como se referindo à pesquisa e à avaliação qualitativa, e que a discussão ligada à objetividade e à comunalidade relacionar-se-ia à pesquisa e à avaliação quantitativa. As expressões quantitativa e qualitativa, ainda segundo Stake (1983a), dão margem a interpretações que podem ser consideradas duvidosas.

Stake (1983a) dimensiona toda essa problemática com bastante propriedade ao dizer: "O argumento a favor e contra singularidade e subjetividade é muitas vezes visto como uma diferença entre as abordagens quantitativas e qualitativas na pesquisa. A distinção é importante ainda que esses termos dêem margem a falsas interpretações. Todos os pesquisadores

quantitativos, é claro, fazem discriminação qualitativas, assim como todos os pesquisadores qualitativos descrevem importantes quantidades em seus relatórios sobre educação.”

“Quando nos referimos a quantitativo colocamos o problema em termos de uma questão de ênfase. A pesquisa quantitativa caracteriza-se por dados extraídos de um grande número de casos sobre um pequeno número de variáveis. A pesquisa qualitativa é caracterizada por dados obtidos a partir de um pequeno número de casos sobre um grande número de variáveis.” (Stake, 1983a)

2.3 Avaliação quantitativa e qualitativa

Stake encaminha a sua discussão sobre o assunto de uma forma a demonstrar que a diferença entre pesquisa/avaliação quantitativa e qualitativa é de natureza epistemológica, envolvendo as generalizações que os dois tipos de avaliação e/ou pesquisa podem proporcionar.

Ao discutir com maior amplitude o problema da avaliação e/ou pesquisa quantitativa e qualitativa, Stake (1983a) faz considerações de grande pertinência com o objetivo de caracterizar as duas abordagens. Assim, mostra que, a partir da percepção de um fenômeno educacional, a avaliação e/ou pesquisa quantitativa: — identifica algumas poucas variáveis ou propriedades; — seleciona as variáveis intuitivamente (redução e delimitação que são necessárias à avaliação e/ou pesquisa); — concentra-se nas variáveis ou propriedades que mostram grande variação; — realiza medidas e estabelece correlações, procurando identificar padrões de covariação; — explica genericamente os fenômenos, a partir da idéia que tem a respeito da população; — seleciona uma amostra para estudo, compara medidas, faz análises estatísticas, procurando explicar a variação e a covariância; — prepara tabelas e quadros para explicar essa variação e covariância; — generaliza, formalmente, sobre a relação entre as variáveis estudadas, elaborando relatório. O produto é uma explicação que destaca as características das variáveis para a população. A generalização, desse modo, assemelha-se a leis.

Igualmente, sempre a partir da percepção de um fenômeno educacional, a avaliação e/ou pesquisa qualitativa: — identifica eventos ou casos que sejam de interesse (pessoas, grupos, produtos etc); — considera propriedades ou variáveis, que, entretanto, não são o ponto nuclear do estudo, na verdade, momentos, lugares, pessoas são o centro do enfoque; — procura a singularidade individual, base para a compreensão do que é típico e, depois, do geral; — identifica seqüências importantes de eventos, testemunhas para eventos passados, e observa como os eventos são determinados; — registra e interpreta as observações, procura padrões de covariação, selecionando-os intuitivamente; — verifica com rigor as observações, validando seu significado via triangulação, ou seja, as observações são repetidas por múltiplos observadores, usando múltiplos métodos e múltiplas interpretações teóricas; — valida as observações em diferentes fases e momentos à medida que o processo se desdobra; — elabora relatório final que é um retrato descritivo dos fenômenos observados, apresentando generalizações naturalistas (informais, explicações pessoais do mundo externo) e propicia aos usuários da avaliação e/ou pesquisa condições para que cheguem às suas próprias generalizações.

As generalizações a que as avaliações e/ou pesquisas quantitativas e qualitativas chegam não são as mesmas, ressalta Stake (1983a), considerando que possuem diferentes conceituações dos dados e diferentes bases para validação e interpretação; por isso, acreditamos que ambos os tipos de avaliação/pesquisa se completem para a estruturação de novos conhecimentos, a partir de um mesmo fenômeno visto por intermédio de óticas diferentes.

2.4 Avaliação naturalista

Ao longo do presente texto referir-nos-emos com grande freqüência às expressões “pesquisa naturalista” e “avaliação naturalista”. É bom que procuremos deixar claro o que essas expressões significam, na linha de Stake (1983b). Atribui-lhe Stake um significado bastante amplo, querendo dizer que, na avaliação/pesquisa naturalista, os sujeitos são observados na sua

atividade habitual, em seu habitat usual e que as observações são apresentadas em uma linguagem não técnica, empregando palavras e conceitos com os quais os usuários (clientes) da avaliação estão familiarizados. Assim, o avaliador naturalista, na visão de Stake (1983b), deve observar, minimizando qualquer intervenção, e deve relatar, em linguagem coloquial, o que constatou sobre os sujeitos observados nos seus procedimentos usuais, nas suas atividades do dia-a-dia e no seu ambiente natural.

Há quem advogue um maior número de avaliações naturalistas (L. Cronbach, E. Guba, por exemplo), mas não apenas em caráter exploratório, e sim em estudos mais amplos, abrangendo diferentes aspectos da educação; outros (F. Kerlinger e J. Popham, por exemplo) adotam uma posição parcialmente contrária, achando a avaliação/pesquisa naturalista útil, mas numa fase de exploração preliminar, pois partem do pressuposto de que as conclusões baseadas nesse tipo de avaliação seriam subjetivas e não possuiriam fidedignidade, como relata Stake (1983b). Os opositores da pesquisa naturalista, por outro lado, argumentam que esse tipo de avaliação poderia comprometer o prestígio que a teoria das medidas e as tecnologias de análise estatística desfrutam no momento. Os adeptos da avaliação naturalista vêem as coisas sob um outro enfoque, sugerem "*um sentido diferente da missão e um conjunto diferente de percepções. A esperança centra-se na confiança de que descrições naturalistas serão úteis, especialmente se a validade basear-se na triangulação e numa revisão dialética*", como propõe Stake (1983b).

A grande problemática a enfrentar para o sucesso da avaliação naturalista, segundo a colocação de Stake (1983b), com a qual concordamos plenamente, não está na subjetividade, mas nos seus custos, que são altíssimos. A questão, entretanto, não pode ser vista, exclusivamente, na perspectiva dos custos. Necessário se faz que, além dos custos, sejam considerados, também, os benefícios, sobretudo se houver "*um aumento valioso do conhecimento existente*" a respeito dos complexos problemas da educação, de acordo com o destaque de Stake (1983b).

Ao realizar uma avaliação ou pesquisa é preciso pensar na sua destinação, no seu objetivo principal. Stake (1983b) diz muito bem que *"a maioria das pesquisas educacionais deve estar a serviço da educação e, claramente, proporcionar uma melhor compreensão de seus problemas práticos. É admirável que algumas pesquisas sejam abstratas, sobretudo visando a formação de pesquisadores..."*. Referindo-se à situação nos Estados Unidos, Stake (1983b), candidamente, declara: *"Não seria realista acreditar que a maioria das atuais pesquisas... possa contribuir de maneira significativa para a solução de problemas práticos da educação, nos limites da minha existência"*. Acreditamos que o mesmo ocorra no Brasil, país que ainda vive na pré-história da avaliação e, também, da pesquisa educacional. Precisamos, no atual contexto, produzir avaliações e pesquisas que sejam educacionalmente significativas e úteis àqueles que precisam enfrentar os problemas do dia-a-dia na educação.

Após a presente introdução, em que alguns dos posicionamentos básicos de Stake foram apresentados, a partir das conferências proferidas na Pontifícia Universidade Católica do Rio de Janeiro (1983a, 1983b), passaremos ao modelo de avaliação que apresentou⁷.

2.5 Modelo de Avaliação

O ensaio de Stake (1967) — **The Countenance of Educational Evaluation⁸**—, ao lado do escrito por Scriven e

⁷ Ver Depresbiteris, L. Avaliação de Programas e Avaliação de Aprendizagem. *Educação e Seleção*, 19. Fundação Carlos Chagas, São Paulo, SP. 1989.

⁸ A palavra **countenance** apresenta diferentes significados: semblante, rosto; proteção; apoio; favorecer (Ferreira, Pe. J. A. **Dicionário Inglês-Português**. Edição do Autor. Porto. s/d), fisionomia, semblante, feições; expressão (facial), face, rosto, aprovação, apoio, compostura, calma, controle, aprovar, dar apoio, sustentar (Houaiss, A. e Avery, C. A. (Ed.) - **Dicionário Barsa das Línguas Inglesa e Portuguesa**. Appleton - Century - Crofts, New York, 1964), expressão, semblante, fisionomia, calma, seriedade, compostura, aprovação, proteção, estímulo, apoio. Santos, A. S. dos - **Guia Prático de Tradução Inglesa**. Editora Cultrix, São Paulo, 1^a ed. 4^a tiragem, 1995. Alguns textos em língua vernácula traduzem o título do artigo de Stake como **"O Semblante da Avaliação Educacional"**, o que não nos parece perfeitamente adequado, tendo em vista o seu conteúdo, donde conservarmos o título original.

publicado no mesmo ano, representa um grande passo no dimensionamento da questão da avaliação educacional, que, a partir de Tyler, passa a ter novos enfoques e procura, até a presente data, mas sem sucesso, criar uma possível teoria geral. Stake terá um papel sobremodo relevante na teorização da avaliação; inicialmente, ressalta que o homem em geral, independentemente das suas qualificações, acredita na educação, mas valoriza a avaliação de diferentes maneiras. Os educadores, por seu lado, apresentam diferenças nos seus posicionamentos relativos à essência e ao valor dos programas educacionais, sendo poucos os que conseguem posições iguais em relação aos mesmos, tendo em vista a multiplicidade dos modelos de avaliação. As idéias de Stake (1967) vão dar origem à avaliação responsiva e possibilitam o surgimento da avaliação naturalista.

Stake (1967) destaca que a avaliação apresenta um lado formal e outro informal. A **avaliação informal** depende de observações casuais, objetivos implícitos, normas intuitivas e julgamento subjetivo. Todas estas características traduzem o dia-a-dia educacional, não sofrendo, por isso, grandes contestações. A avaliação informal varia bastante de qualidade, algumas vezes bastante percuciente; mas outras, completamente superficial. A **avaliação formal** depende de outros elementos, inclusive de comparações controladas e do uso de testes padronizados. Apesar de referências constantes, não existem testes padronizados no contexto brasileiro, dado o pouco desenvolvimento das técnicas psicométricas empregadas no contexto nacional. Voltando a Stake, ele acentua que, realmente, o que se faz é a avaliação informal, com base na opinião de professores sobre o programa, a sua lógica e/ou a reputação profissional dos responsáveis pela sua elaboração. Tudo isso provoca, naturalmente, uma grande insatisfação em relação à avaliação formal, existindo poucos estudos desse tipo de avaliação que sejam realmente significativos na percepção de Robert E. Stake, que também põe em dúvida a contribuição dessa avaliação para a educação.

Stake (1967) chama a atenção para o fato de que no exame de avaliações formais não existe um esforço para relacionar as

condições antecedentes e as interações (*transactions*)⁹ em sala de aula, nem uma preocupação em estabelecer uma relação entre estes procedimentos e os resultados alcançados, geralmente representados pelo que Stake chama de escores convencionais de testes. Pouco se faz também para associar o que os educadores pretendem fazer e aquilo que eles realmente realizam. Stake (1967) critica o excesso de preocupação psicométrica de alguns avaliadores em relação à fidedignidade dos escores e à validade preditiva dos instrumentos, quando, no caso da avaliação de um currículo ou programa, as diferenças individuais dos estudantes devem ser substituídas por outras preocupações, como a contingência entre condições anteriores, atividades de sala e resultados.

A partir dessas colocações, Stake (1967) destaca que o seu objetivo não é dizer o que medir ou como medir, mas apresentar os elementos necessários para o desenvolvimento de um plano de avaliação, especialmente em relação a programas educacionais, e não de materiais, pois a avaliação daqueles pressupõe que se faça a destes últimos. Procura, assim, introduzir uma conceituação da avaliação orientada para a complexa e dinâmica natureza da educação, cujos propósitos variam de uma situação para outra. O que serve para uma escola talvez não sirva para outra, diz Stake (1967), que chega a um ponto crucial quando afirma que o especialista vê a si mesmo como alguém que descreve aptidões e desempenhos, e que o professor e o administrador esperam que o avaliador atribua um grau/escore ao mérito de alguém ou de alguma coisa, que julgue as coisas em relação a padrões externos ou a critérios que pouco têm a ver com a escola e seus objetivos. Ninguém vê que descrição e julgamento são, na verdade, os dois atos básicos da avaliação (Stake, 1967). É possível que um avaliador como indivíduo se abstenha de julgar, ou até mesmo deixe de coletar o julgamento dos outros. É igualmente provável que o avaliador só queira destacar o que o programa apresenta de valor. A sua avaliação será sempre incompleta, porque é preciso

⁹ Alguns estudiosos da avaliação traduzem *transactions* como transações, parecemos, entretanto, que o vocábulo interações representa melhor o sentido da palavra usada por Stake.

que o programa seja inteiramente descrito e inteiramente julgado para que se tenha uma avaliação educacional.

Os dados levantados pelo avaliador são de diferentes fontes e coletados de diferentes maneiras. Segundo Stake (1967), devem ser agrupados em três conjuntos de informações:

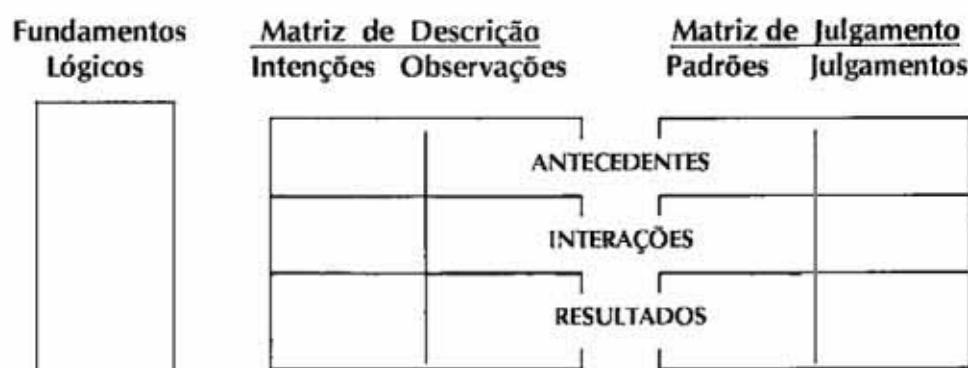
1. **Antecedentes** — condições existentes antes do ensino e da aprendizagem que podem se relacionar com os resultados, ou seja, aquilo que os especialistas em instrução programada chamam de comportamento de entrada.
2. **Interações** — conjunto das inúmeras interações de estudante com professor, estudante com estudante, pais com orientadores etc., ou seja, a sucessão de compromissos que constituem o processo educacional.
3. **Resultados** — conhecimentos, habilidades e atitudes dos estudantes, como consequência da experiência educacional. São os aspectos que, tradicionalmente, interessam à avaliação formal. Stake (1967) chama a atenção para o fato de que o impacto do processo instrucional sobre os professores, administradores e outros, como conjunto de informações, também deveria ser incluído entre os resultados. A questão dos resultados é complexa porque deve incluir não apenas aqueles que são evidentes mas também os que somente mais tarde podem ser medidos, como, por exemplo, transferência e efeitos da reaprendizagem (Stake, 1967).

Antecedentes, interações e resultados são elementos que participam da **Descrição** e do **julgamento**. Stake (1967), no seu modelo, mostra que os julgamentos podem ser classificados como padrões gerais de qualidade ou julgamentos específicos de um dado programa. Os dados descriptivos são classificados em **intenções** e **observações**. Ao avaliador cabe preparar um registro do que os educadores pretendiam (intenções), e das percepções dos observadores. O próprio Stake apresenta um exemplo imaginário de 12 conjuntos de dados, admitindo, contudo, que os educadores e avaliadores não esperam registros com todos esses detalhes, mesmo em futuro distante.

As intenções ou objetivos, segundo Stake (1967), são todos os efeitos desejados, os que podem ser antecipados, e mesmo aqueles que despertam algum receio. Estes tipos de objetivos (**intenções**) são o que os outros esperam, sobretudo os estudantes. Ao final, a relação das intenções é uma lista de prioridades sobre o que deve ocorrer ao longo do desenvolvimento do programa (objetivos), sendo um desafio para o observador conseguir a relação de todos os objetivos pretendidos.

A figura (1) ajuda a compreender as colocações de Stake (1967).

Figura 1 - Afirmações e dados a serem coletados por um avaliador de um Programa Educacional, segundo R. Stake (1967)



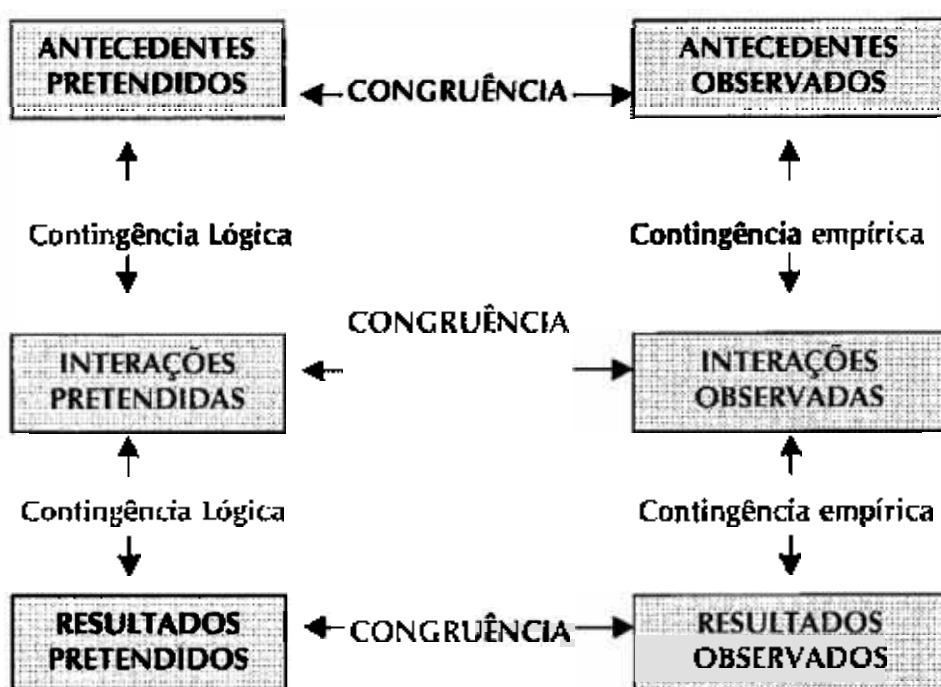
A maioria dos dados descritivos são classificados como observações. As observações podem ser coletadas diretamente pelo avaliador ou através de instrumentos (inventários, dados biográficos, entrevistas, **checklists**, opinionários, e testes psicométricos), segundo as colocações de Stake. Ainda que o avaliador dê especial atenção a resultados de medidas, não pode ignorar os resultados de suas observações, nem as condições anteriores (antecedentes) e nem os procedimentos instrucionais. Ao longo de seu ensaio, Stake chama a atenção para o fato de que o avaliador, na escolha de variáveis a estudar, deve tomar várias decisões subjetivas e, obviamente, limitar o número de elementos a estudar, descartando tudo aquilo que não contribui diretamente para o seu trabalho. A atenção do avaliador deve estar voltada principalmente para as variáveis ligadas aos

objetivos estabelecidos pelo educador, incluindo, naturalmente, outras que por ventura tenha observado. O problema dos efeitos colaterais e o dos ganhos incidentais, nota Stake (1967), devem ser aspectos a considerar.

Uma avaliação não está completa se não considerar os fundamentos do programa, ainda que, às vezes, não sejam perfeitamente explícitos no programa. Estes fundamentos lógicos são a base filosófica do programa, e refletem seus propósitos, sendo, portanto, de capital importância para o avaliador, ao considerar as intenções do programa, além de ser importante para a futura escolha daqueles que vão participar dos julgamentos.

Para fins de processar os dados descritivos da avaliação, discute Stake (1967), é preciso determinar as contingências (o que pode ou não suceder) entre antecedentes, interações e resultados, estabelecendo, ainda, a congruência (relação) entre intenções e observações. A representação gráfica a seguir dá uma idéia do processamento dos dados da matriz descritiva :

DADOS DESCRIPTIVOS



Os dados de um programa são **congruentes** se o pretendido realmente acontece. A consequência fatal entre antecedentes, interações e resultados raramente acontece. É preciso comparar o pretendido com o observado a fim de positivar discrepâncias, e descrever a congruência que por acaso existe, observando-se, entretanto, que a congruência não é indicativa de fidedignidade ou validade dos resultados. Indica, na verdade, se o que se pretendeu teve oportunidade de ocorrer efetivamente. A descrição dos pares de variáveis numa coluna da matriz de dados informa mais do que as variáveis isoladamente. Avaliação, ressalta Stake (1967), é uma procura das relações (congruências) que permitem a melhoria da educação, cabendo, desse modo, ao avaliador identificar resultados que são contingentes em relação a determinados antecedentes e procedimentos instrucionais. A contingência entre aspectos pretendidos é de natureza lógica e essa lógica deve apoiar-se em experiências anteriores, ou em pesquisas anteriores, ou, então, em observações semelhantes que foram anteriormente realizadas. A avaliação da contingência das observações depende de provas empíricas. A avaliação de um único programa não proporciona dados para que se possam emitir enunciados sobre a contingência, ficando-se na dependência de experiências anteriores. As contingências e congruências que os avaliadores conseguirem identificar devem ficar sujeitas aos julgamentos de peritos, lembrando-se que a não congruência pode ser importante fator na avaliação de um programa e esclarecer possíveis problemas de relevância (Stake, 1967).

A necessidade de estudos comparativos é algo que se impõe, tendo em vista o fato de que nenhuma escola pode avaliar o impacto do seu programa sem o conhecimento do que outras escolas estão realizando para concretizar seus objetivos (Stake, 1967). A questão é bastante complexa quando se considera, por exemplo, que os escores/notas dos alunos se baseiam em critérios personalíssimos dos professores, que têm padrões bastante diferenciados entre si. Stake (1967) chama atenção para as descrições pouco informativas que os professores fazem do grau de excelência dos seus alunos, ainda que esses mesmos

professores sejam competentes nas suas áreas substantivas e na identificação das dificuldades que os alunos apresentam.

A questão dos padrões de comparação é bastante complexa pela multiplicidade dos próprios padrões e a sua variação de estudante para estudante, de professor para professor e de grupo de referência para grupo de referência (Stake, 1967). Cabe, assim, ao avaliador identificar os melhores padrões para o seu caso especial, o que é uma grande responsabilidade. A definição de padrões é indispensável no caso da avaliação formal de um programa, e a grande dificuldade na realização desse trabalho está na definição de uma metodologia adequada para processar os diferentes julgamentos apresentados.

Os julgamentos podem basear-se em padrões absolutos, refletindo julgamentos pessoais, ou padrões relativos, considerando características alternativas dos programas. Ou seja, avaliar o programa em relação a opiniões sobre como deve ser um "bom" programa, ou em comparação a outros programas já existentes. Antes de qualquer julgamento, de acordo com Stake (1967), é preciso verificar se cada um dos padrões definidos foi efetivamente alcançado. Julgar vai consistir em atribuir um peso, um valor a cada padrão. Assim, um julgamento racional em avaliação educacional é uma tomada de decisão relativa aos padrões de referência para decidir se toma ou deixa de tomar uma determinada decisão administrativa, segundo Stake (1967). Vê-se, pois, que a decisão sobre quais variáveis estudar ou decidir sobre quais padrões empregar são comprometimentos subjetivos inevitáveis na avaliação educacional, assim como subjetiva é também a decisão sobre o tipo de avaliação a realizar, objeto da preocupação de Cronbach (1965) e também do próprio Scriven (1967), na sua **Methodology of Evaluation**. Stake (1967) discrimina a questão do formal (**avaliação objetiva**) do informal (**avaliação subjetiva**), sugerindo procedimentos formais se

julgamentos racionais devem ser apresentados¹⁰.

2.6 Avaliação responsiva

Anteriormente, ao definirmos certos posicionamentos, demos um crédito à avaliação responsiva. Voltamos, novamente, ao mesmo tema, usando como referencial o importante texto de Stake e Pearsol (1981) para uma discussão mais longa sobre esse tipo de avaliação que, parece-nos, deveria merecer uma prática mais entusiástica em nosso contexto educacional, como pode ser depreendido da importância dos conceitos.

Aqueles que não têm um maior trato com a sua metodologia reagem, inicialmente, como se uma verdadeira avaliação não tivesse ocorrido, apesar do árduo trabalho realizado pelos especialistas. Os relatórios muitas vezes, no caso de uma avaliação responsiva, podem causar uma grande surpresa à clientela, que o julgará severamente, considerando-o completamente **insuficiente**, por carência de dados, por falta de "informações" resultantes de uma metodologia baseada em complexos programas de computação, que, supostamente, forneceriam os elementos que dariam o caráter de suficiência aos relatórios.

¹⁰ Procuramos fazer uma apresentação do modelo de avaliação de Stake (*The Countenance of Educational Evaluation*) que é importante e extremamente útil para o desenvolvimento de uma avaliação. Não se trata de um **cookbook** (receita), mas um quadro referencial que destaca os dois elementos fundamentais da avaliação: **descrição e julgamento**. A discussão sobre **contingência e congruência** chama a atenção para a questão da análise dos dados, especialmente se são volumosos. A questão do julgamento merece toda a atenção de Stake (1967) por ser um momento crítico, sobretudo tendo em vista que uma escolha inadequada de padrões pode invalidar os julgamentos e, consequentemente, a própria avaliação, impossibilitando estudos comparativos, que são cruciais em avaliação educacional. Fizemos uma descrição de um modelo, mas a riqueza da obra de Stake precisa ser examinada, tendo em vista a sua contribuição para a avaliação **responsiva/naturalista**, o que será feito posteriormente em dois momentos do presente trabalho. O objetivo do presente capítulo foi, reiteramos, apresentar algumas idéias básicas de Stake, que devem ser objeto de reflexão aprofundada, independentemente da orientação metodológica do avaliador.

Analisemos, no caso de uma avaliação responsiva, os diversos procedimentos que caracterizam uma avaliação fundamentada em princípios científicos.

Para iniciarmos uma avaliação responsiva é preciso que haja uma perfeita interação entre a instituição interessada no trabalho, através de sua diretoria, equipes técnicas e corpo docente, e o avaliador. Isso se justifica, porque uma avaliação não pretende bloquear atividades e nem o desenvolvimento de seus programas, pois, se necessário, no caso de uma avaliação responsiva, há uma simultaneidade entre o desdobrar do programa, com vistas a futuras alterações, e a própria avaliação.

O início de uma avaliação responsiva centra-se no aprofundamento do conhecimento de todos os elementos da cultura institucional e dos posicionamentos em relação ao que deva ser, é ou não deva ser objeto da avaliação, que, no caso, pensamos em um programa. Isso é importante, tendo em vista que a atuação eficiente de um programa não pode ser medida de forma ortodoxa, por meio de instrumentos psicométricos, e com o emprego de estatísticas, por mais sofisticadas que possam ser. A reunião acima proposta — instituição e avaliador — deve ser consequente e esclarecer de modo irredutível os limites da avaliação, inclusive do ponto de vista administrativo. É o momento mais apropriado para o esclarecimento de dúvidas, incertezas e de respostas às várias questões que porventura se apresentam às partes.

Uma cronologia de atividades e do envolvimento do avaliador deve ser estabelecida, assim como a duração do projeto. Tudo no sentido de conciliar o trabalho do avaliador com as atividades e os interesses institucionais. O avaliador deve apresentar suas dúvidas iniciais, a serem esclarecidas pela instituição, que deve ficar ciente do surgimento de dúvidas e problemas os mais diversos ao longo dos trabalhos. Essa fase inicial, assim como o desenrolar da avaliação, centra-se em um grande e transparente diálogo entre o avaliador e a parte contratante, onde se discutirá, certamente, o impacto que a avaliação passa a ter no corpo discente, assunto que, por sua importância, não pode ser ignorado nesse momento inicial, tendo

em vista suas repercussões. O impacto, no caso, traduz-se por modificações no **saber** (conhecimento), no **saber — fazer** (habilidades) e no **saber — ser**, em novas atitudes, por conseguinte¹¹.

A instituição contratante, nesse momento inicial, também terá suas dúvidas, suas interrogações e, possivelmente, suas inquietações. Ou seja, muito provavelmente a primeira indagação será — como serão coletados os dados?, seguindo-se-lhe outra: de que maneira alunos, professores e toda a comunidade se envolverão no programa de avaliação? Haverá manifestação para o conhecimento das informações dos participantes, sendo prudente da parte do avaliador, como bem lembram Stake e Pearsol (1981), que se estabeleçam prazos adequados para o bom andamento do trabalho, sem atropelos, sem sobrecarga e sem criar uma atmosfera estressante, como sói acontecer em alguns casos. O perfil do alunado é quase sempre uma manifestação solicitada através do rendimento escolar. Uma avaliação responsiva não se vê obrigada, necessariamente, a aplicar testes ou a submeter a amostra estudantil a provas, instrumentos muitas vezes não adequados quando se deseja identificar novas idéias, novas percepções e novos questionamentos resultantes da implantação de um novo programa.

O assunto apresentado é de uma grande complexidade, tendo em vista que diferentes alunos em um mesmo programa apresentam diferentes idéias, percepções e questionamentos, não se podendo, muitas vezes, fazer um juízo de valor estabelecendo que esta idéia é melhor do que aquela outra. O mesmo ocorre com percepções e questionamentos. O perfil do aluno, por mais importante que seja, não é suficiente em uma avaliação responsiva, é fundamental que seja feito um contraste (comparação) entre o perfil do aluno e a observação decorrente da interação aluno/professor, de modo a verificar até que ponto o professor leva o aluno à reflexão e se os objetivos propostos pelo programa estão sendo efetivamente atingidos e traduzidos por

¹¹ Ver o artigo de Depresbiteris, L. e Taurino, M. do S. O difícil percurso de um educador no mundo dos critérios de avaliação. In: *Estudos em Avaliação Educacional*, nº 14, pp. 45-63. Fundação Carlos Chagas, São Paulo, SP. 1996.

experiências relevantes (Stake et al., 1981). Em função desses resultados é que se pode afirmar, categoricamente, se o programa estaria tendo algum impacto.

Uma fase fundamental na avaliação responsiva é a das entrevistas envolvendo todos os participantes do programa. Além do mais, toda uma programação de observações, por pessoas devidamente qualificadas, deve ser planejada para ser realizada em sala de aula. Aqui se nos depara um sério problema: — o da confidencialidade das informações proporcionadas por professores e alunos, que nunca devem ser individualizados em exposições orais e/ou relatórios escritos. A discussão do relatório final com a parte contratante é uma prática salutar, para fins de correção de possíveis erros factuais, destacando-se, entretanto, que essa prática de correção não se deva transformar em censura, o que seria inteiramente descabível. Se houver discordâncias, as mesmas deverão constar em anexo ao relatório final.

Uma avaliação, qualquer que seja a sua natureza, precisa ser planejada, inclusive a avaliação responsiva, incluindo, nesse caso, uma análise do programa, um exame do material didático, discussão crítica dos planos de aula, conversas com professores responsáveis pelo desenvolvimento dos programas e um levantamento exaustivo de todo o material pertinente ao programa a ser avaliado. Somente, então, depois da conversa e discussão inicial e do exame exaustivo de todo o material levantado e analisado, o avaliador define suas atividades e estabelece os elementos que servirão de guia para as futuras observações, peças fundamentais na avaliação responsiva.

As pensarmos em uma avaliação responsiva, devemos propor uma série de interrogações (Stake et al., 1981) e encontrar respostas que nos permitam desenvolver um plano de atividades coerentes de forma profícua. Quais algumas dessas indagações?

1. Qual a justificativa fundamentada para a inclusão do programa no currículo?
2. Qual a influência que o programa exercerá sobre os estudantes, ou o que eles obterão com a inclusão desse novo programa?

3. Qual a história de vida desse programa, ou seja, quais suas diferentes fases até a sua forma atual?
4. Quais as diferenças desse programa em relação aos outros que existiam na instituição?
5. Quais as recomendações estabelecidas para avaliar os alunos ao longo do desenvolvimento do programa?
6. Quais os pré-requisitos necessários aos alunos para acompanharem o programa?

A partir dessas colocações de Stake et al. (1981) fica evidente que o avaliador, ao iniciar o seu trabalho avaliativo/investigativo, não possui respostas para esse conjunto de indagações e de outras perguntas que por acaso possam surgir no desenvolvimento de suas atividades. Fica claro, entretanto, que essas questões anteriormente apresentadas se impõem do ponto de vista conceitual e representam um roteiro para o desenvolvimento de todo um plano de trabalho. Ainda como condição liminar, é necessário que se examinem os vários relatórios que porventura existam sobre o desenvolvimento do programa e que o avaliador tenha conversas bastante candidas com os atores responsáveis pela implementação do mesmo.

É extremamente importante, diríamos mesmo ser fundamental, que o avaliador tenha o maior número possível de dados diferenciados sobre a experiência que os alunos estão tendo com o programa posto em ação, ou seja, dados da sua vivência em face de uma nova realidade. Seria de toda conveniência, ainda, que os pais dos alunos fossem também ouvidos, por intermédio de seus representantes no Conselho da Escola ou, se for o caso, na Associação de Pais e Mestres. É claro que, na fase de implementação de um programa, sempre surjam críticas e vozes discordantes ou existam pessoas que não se mostrem entusiasmadas a seu respeito. Também essas devem ser ouvidas e não podem ser de forma alguma esquecidas, porque, possivelmente, poderão dar contribuição valiosa para os avaliadores. É bom reiterarmos que, na avaliação de um programa ou curso, devem ser considerados os padrões fixados para a constatação do progresso que os estudantes estariam tendo em relação ao programa que está sendo desenvolvido e avaliado.

Stake e Pearsol (1981) encarecem a importância de que sejam adotados procedimentos a fim de determinar o impacto do programa nos alunos e que o avaliador, com esse objetivo em mente, concentre os seus esforços e toda a sua atenção em relação aos mesmos. Isoladamente, é de toda conveniência que os alunos sejam submetidos a uma avaliação bem aprofundada, com vistas a positivar as reais aquisições de novos conceitos e a formação de atitudes em decorrência da influência do programa. A avaliação dessas dimensões, estamos certos, é uma atividade bastante complexa e, às vezes, até mesmo impossível. Desse modo, é interessante que o avaliador procure outras pessoas que conhece e possam atuar como observadoras junto às crianças para que verifiquem se as modificações constatadas podem ser realmente atribuídas aos programas. O avaliador precisa estar plenamente consciente de que nem sempre uma simples aula ou até mesmo todo um curso (programa) produz alterações nos participantes que sejam facilmente perceptíveis.

As informações sobre os alunos existentes nos arquivos podem ser (e são) úteis para uma avaliação responsiva (Stake et al., 1981); no entanto, o problema, pelo menos em nosso contexto, está na natureza dessas informações, que muitas vezes não são adequadas para definir o perfil do estudante. São, habitualmente, escores globais (brutos) que não oferecem a caracterização do aluno e representam um dado quase sempre perdido, por não resultarem de uma avaliação analítica do estudante.

Ao final, mostram Stake et al. (1981) que uma avaliação responsiva deve ter elementos suficientes para poder oferecer informações consistentes relacionadas com:

- (1) as preocupações ligadas à avaliação do programa e possíveis discordâncias, porque sempre haverá qualquer forma de oposição ao trabalho, ostensivamente ou de forma velada;
- (2) as necessidades manifestas pelo conjunto curricular em que o programa está inserido;
- (3) a formatação que esse programa deva efetivamente ter;

(4) a possível falta de alternativas para o programa (ou as possíveis alternativas).

É utópico acreditar que, ao término de uma avaliação, todos os pontos inicialmente definidos tenham sido alcançados e respostas conclusivas sejam dadas. Stake (1981), por exemplo, lembra algumas interrogações passíveis de não terem uma resposta cabal e inteiramente satisfatória — quais os principais aspectos instrucionais que diferenciam os novos programas de outros na mesma escola? qual o nível de maturidade (capacitação) para que o aluno tenha sucesso no programa?

Antes da elaboração do relatório final, convém que o avaliador procure analisar as duas questões propostas e identifique, se for o caso, as razões que o impediram de respondê-las, fazendo, inclusive, uma autocrítica e uma análise da influência de seus possíveis vieses. Queremos lembrar que, para chegar à fase de elaboração de um relatório final, o avaliador, segundo a lição bastante clara de Stake et al. (1981), deve ter passado por diferentes etapas:

- 1) negociar com a parte contratante e/ou agências financeiras os parâmetros necessários para a realização do estudo;
- 2) estabelecer os diferentes tópicos ou questões que possam ser objeto de preocupação para os que propõem e/ou financiam a avaliação;
- 3) participar do contexto em que vai realizar a avaliação e observá-lo;
- 4) identificar assuntos e aspectos que sejam objeto de controvérsias, a partir de informações levantadas por intermédio de entrevistas e questionários;
- 5) apresentar, para discussão inicial, os primeiros resultados em um relatório de caráter preliminar;
- 6) investigar, em profundidade, e o mais amplamente possível, as controvérsias, preocupações e dúvidas que possam existir sobre o programa;
- 7) identificar elementos que possam entrar em conflito com os dados encontrados, invalidando-os.

O relato de uma avaliação responsiva, como mostra Stake (1981), é feito a partir de respostas levantadas pelo avaliador e que refletem as experiências das pessoas envolvidas com o objeto da avaliação. A descrição apresentada leva em conta as manifestações dos professores, administradores e pessoal técnico sobre a escola, seu currículo, seus alunos e, naturalmente, no caso, o programa avaliado. Ao fazer a descrição operacional do programa, o avaliador utiliza-se das suas observações sobre a interação professor/aluno, transcrevendo alguns dos seus registros mais significativos, inclusive citando manifestações, apreciações dos alunos. Ocorrências verificadas são igualmente registradas para que os próprios leitores do relatório vivenciem a experiência de sala de aula.

A apresentação do programa e seus elementos devem ser analisados com a indicação de pontos negativos e positivos sobre sua organização. O impacto do programa sobre o estudante é importante que seja ressaltado, como faz Stake (1981). Esse impacto não é apresentado em termos de notas/escores, mas descrevendo as experiências vividas pelos próprios alunos e que demonstram a sua compreensão dos principais pontos e conceitos do programa avaliado. Exemplos de trabalhos dos alunos, trechos de relatório de entrevistas, de observações, de manifestações dos professores, e finalmente, de escores e outros aspectos quantificados são elementos que servem para demonstrar em que medida o programa está tendo efetivo impacto. O relatório final também apresenta as dificuldades encontradas na implantação do programa e os elementos adversos que precisaram ser vencidos.

A presença do avaliador na escola visa a observar as atividades e o contexto, registrar e questionar o que está vendo, validar interpretações e relatar as controvérsias que predominam, fazendo tudo isso através das palavras e das imagens do pessoal envolvido na implantação do programa (Stake, 1981). Ao final, é possível que o avaliador não se sinta inteiramente satisfeito, que muitas oportunidades não tenham sido devidamente exploradas, às vezes por razões orçamentárias, que comparações talvez pudesse ser realizadas considerando-se diferentes níveis de escolaridade, mas possivelmente este passo fugisse ao objetivo a

que a avaliação se propunha. A avaliação realizada, ainda que apresente aspectos ligados a estudo de caso, não é efetivamente um estudo de caso, tendo em vista que muitos aspectos não foram considerados. A avaliação procurou responder a alguns problemas específicos que foram colocados pelas partes interessadas, assim, o que efetivamente foi feito caracteriza-se como uma avaliação responsiva, lembram Stake et al. (1981) ao final de seu ensaio sobre a metodologia que desenvolveram e pode ser extremamente útil no atual contexto brasileiro. [2] [3] [4].

2.7 Estudo de caso

The purpose of case study is not to represent the world, but to represent the case, diz Stake (1994) em seu magnífico trabalho sobre estudo de caso, definindo com simplicidade, precisão, objetividade e clareza o que a metodologia em questão se propõe a realizar. Outro aspecto que procura esclarecer é que alguns estudos de caso são qualitativos e que outros não se revestem dessa característica, desfazendo, desse modo, a falsa idéia prevalecente em alguns segmentos da avaliação educacional de que todos os estudos de caso devem ser necessariamente qualitativos. O estudo de caso, como uma forma de pesquisa, lembra Stake (1994), é definido por seu interesse em casos individuais e não pelo método de pesquisa utilizado.

O crucial em estudo de caso é de natureza epistemológica, como situa Stake (1994), porque surge a questão: — o que efetivamente se pode apreender a partir de um único caso, simples ou complexo, sobre o qual investigamos em um prazo de tempo curto ou longo? A expressão estudo de caso pode gerar celeuma, como coloca o próprio Stake (1994), tendo em vista que não existe um consenso sobre o que seja caso, além de a palavra estudo se revestir de uma certa ambigüidade. O estudo de caso acaba sendo um processo e um produto da nossa aprendizagem, lembra o mestre de Illinois.

Além do mais, existem diferentes tipos de estudo de caso. O estudo de caso **intrínseco** tem por objetivo obter melhor compreensão de um caso particular. A sua finalidade, lembra

Stake (1994), não é compreender um construto, um fenômeno genérico, construir uma teoria, ainda que, eventualmente, isso possa ocorrer. O estudo resulta de um interesse intrínseco em um aluno, em uma determinada instituição ou em um certo programa curricular. O estudo de caso **instrumental** visa a possibilitar uma melhor compreensão de um aspecto controverso ou a aprimorar uma teoria. O caso, segundo Stake (1994), em si mesmo tem um interesse secundário, serve de suporte para facilitar a compreensão do que é objeto de interesse. Finalmente, o estudo de caso **coletivo** é, na verdade, um estudo instrumental que se amplia para vários casos, pois o interesse não se limita à compreensão de um único caso, estuda um certo número de casos para pesquisar fenômenos, populações ou condições gerais.

As três categorias de estudo de caso definidas por Stake (1994) são extremamente úteis do ponto de vista didático, mas nem sempre se ajustam à realidade dos interesses do avaliador. O próprio Stake (1994) mostra que trabalhos de etnógrafos, demógrafos e muitas outras categorias de especialização poderiam ampliar uma possível taxonomia de diferentes tipos de caso, o que, entretanto, foge ao seu interesse imediato.

Stake (1994) mostra que o estudo de caso procura identificar o que é comum e o que é singular em um caso, ainda que muitos estudos de caso acabem apresentando dados sobre a natureza do mesmo, suas origens históricas, seu contexto físico e outros contextos (econômico, político, legal e até mesmo estético), tendo em vista que o caso é uma entidade de natureza complexa, apesar de sua singularidade, e justamente tendo em vista essa complexidade o estudo de caso deve ser de natureza holística.

Os estudos de caso com freqüência tratam de matérias controversas, circunstâncias problemáticas de áreas diversas, que a pesquisa qualitativa não analisa em função de resultados experimentais com o emprego de testes de hipóteses. Ainda que as generalizações não sejam o objetivo da avaliação qualitativa, o avaliador, no estudo de caso, segundo Stake (1994), procura estudar a complexidade dos elementos de modo a que possa fazer generalizações a partir da especificidade de cada caso.

É evidente que o estudo de caso proporciona diferentes elementos que podem levar alguém a aprender a partir de um único caso e, nesse sentido, há quem chegue a falar na possibilidade de uma epistemologia do particular. Acreditamos, como Stake (1994), que um relatório de estudos de caso possa oferecer ao leitor uma experiência vicária, que leva o leitor a aprender o que lhe foi relatado, como se fosse uma experiência vivida por ele próprio. O avaliador, entretanto, nem sempre consegue passar para o seu leitor o significado de muitos acontecimentos, porque são de natureza personalíssima e intransferível. A disseminação dos resultados de estudos de caso é um problema crítico, assinala Stake (1994), considerando que a questão da transferência de conhecimento é um processo extremamente complexo.

Uma outra questão crítica no estudo de caso refere-se à validade das comunicações feitas pelos avaliadores e à possível distorção dos dados e das observações realizados. Essa possibilidade realmente existe, ninguém as nega, mas pode ser superada mediante a triangulação, ou seja, usando-se múltiplas percepções e o envolvimento de vários observadores, a fim de que diferentes significados possam ser esclarecidos. É evidente que a repetição de uma observação não é tarefa fácil, mas é uma forma de verificar o significado das várias formas apresentadas por um fenômeno ao ser estudado.

A maioria dos estudos de caso, observa Stake (1994), procura dar ênfase às descrições e a interpretações personalistas, qualquer que seja a sua orientação — naturalista, etnográfica ou outra qualquer. O fundamental em um estudo de caso é a observação reflexiva e interpretativa para que seja possível identificar o significado de uma variável e suas influências (Stake, 1994). O avaliador, no estudo de caso, especialmente, é atuante, envolve-se e se relaciona de forma a participar de todas as operações relativas ao caso, o que lhe vai possibilitar uma reflexão mais intensa sobre as ocorrências observadas. Nem tudo o avaliador vai poder obter diretamente, daí a necessidade de muitas vezes precisar dedicar parte do seu tempo a entrevistas e a estudo de documentos, comportamentos típicos de avaliação

qualitativa, na qual se inclui o estudo de caso. Tudo isso, no estudo de caso, objetiva possibilitar uma melhor compreensão do fenômeno estudado. Ressalte-se que um estudo de caso não procura generalizar (validade externa), o seu ponto central não é fazer generalizações, como acentua Stake (1994), mas **compreender** o universo avaliado.

Um problema sério no estudo de caso está na seleção do próprio **caso** a ser avaliado. O avaliador, em certas situações, já tem um caso pré-definido, não o escolhe. Outras vezes, ao avaliador é apresentada uma amostra de problemas para selecionar um único e estudá-lo. É bom lembrar que, na seleção de um caso, faz-se necessário identificar o potencial de informações que esse caso pode fornecer ou, em outras palavras, impõe-se que o caso escolhido nos proporcione uma oportunidade para **aprender**, que é o objetivo primeiro de um estudo de caso.

Ao nos empenharmos em um estudo de caso, devemos resolver diferentes e complexos problemas, entre os quais o da seleção da amostra. Ainda que em certos casos a amostra já esteja definida, é necessário que o avaliador se preocupe com relação a pessoas, lugares e aos acontecimentos que precisam ser prioritariamente estudados. A seleção da amostra pelo avaliador vai depender da sua experiência e da sua capacidade de intuição (Stake, 1994). O problema da amostragem se complica quando o avaliador, na amostra selecionada, precisa selecionar grupos específicos para observação de um determinado aspecto. Tudo isso pressupõe que, antes de selecionar a amostra, o avaliador discuta o problema com diferentes pessoas e obtenha dados bem definidos, para, então, extrair a amostra, que deve ser, obviamente, representativa para que informações relevantes possam ser obtidas.

O estudo de caso em avaliação educacional é uma metodologia bastante promissora, mas que ainda não foi devidamente explorada, pelo menos com o devido rigor, salvo, naturalmente, as exceções de sempre. A decisão de realizar um estudo de caso implica considerar diferentes aspectos, como mostrou Stake (1994), destacando-se, principalmente,

1. a conceituação do objeto do estudo;
2. a definição da prioridade das questões avaliativas;
3. o estabelecimento de padrões relativos aos dados;
4. a necessidade da triangulação das observações;
5. a escolha de interpretações alternativas a serem investigadas;
6. o desenvolvimento de afirmações sobre o caso, ou, talvez, generalizações.

Se atentarmos para os passos anteriores, aperceber-nos-emos que guardam bastante semelhança com as fases de uma pesquisa qualitativa; entretanto, como ressalta Stake (1994), outros aspectos precisam ser considerados, como, por exemplo, a natureza descritiva dos relatórios e o estabelecimento de comparações com outros estudos de caso, deixando ao leitor o encargo de fazer as suas próprias generalizações, o que nos parece medida prudente. Stake (1994), com o bom senso que sempre demonstra em seus trabalhos teóricos, mostra que o estudo de caso integra a metodologia científica, e que pode ser usado para sugerir problemas para investigação, contribuindo, assim, para maior refinamento do conhecimento.^[5] [6].

2.8 NOTAS E COMENTÁRIOS

[1] A amostragem das obras de Stake em língua portuguesa não é representativa, apesar da feliz iniciativa de Goldberg e Souza, que, em obra editada sob sua responsabilidade, traduziram quatro textos bastante significativos (1982a; 1982b; 1982c; 1982d), em que nos apoiamos para apresentar alguns subsídios de Stake em relação a assuntos importantes, que não podem ser ignorados. As traduções de Vianna (1983a; 1983b), constantes da bibliografia, serão utilizadas no corpo do capítulo ou em notas e comentários, eventualmente.

Stake (1982a), partindo de Scriven (1967) e de uma taxonomia elaborada por Worthen e Sanders (1987), faz uma série de colocações a respeito de modelos de avaliação, que vale a pena considerar, sobretudo porque nos auxilia na compreensão de muitos aspectos de seu pensamento teórico. Inicialmente, é enfocada a questão da avaliação formativa x avaliação somativa, cuja distinção inicial foi feita por Scriven, ao caracterizar os estudos realizados no decorrer da elaboração de um programa (avaliação formativa) e aqueles levados a efeito ao término de um programa (avaliação somativa). Stake (1982a) acha que a distinção

entre um e outro tipo de avaliação não é bastante nítida, levando em conta, nas suas palavras, que: “É evidente que o programa que está sendo criado aumenta dia a dia em número de componentes. É difícil distinguir entre avaliação somativa de um componente concluído e a avaliação formativa de uma parte do programa.” A questão central na avaliação estaria no porquê e não no quando, na visão de Stake, o que nos leva a refletir que ambas as questões nos conduzem à tomada de decisões, mas a decisões diferentes.

O artigo de Stake, traduzido por Goldberg e Souza (1982a), é extremamente rico de idéias que costumam preocupar avaliadores e pesquisadores no seus problemas cotidianos: avaliação oficial x avaliação oficiosa; análise do caso particular x generalização; avaliação de processo x produto; descrição x julgamento; avaliação pré-determinada x avaliação reacional; avaliação global x avaliação analítica, aspectos que, em diferentes momentos, foram considerados ainda que ligeiramente, e constituíram objeto de nossa preocupação no decorrer do presente trabalho.

Ao fazermos estas reflexões, sem dúvida inspiradas no texto de Stake (1982a), não podemos fugir ao momento existencial, quando indagações surgem diante do contraste entre avaliação interna x avaliação externa. A avaliação interna é realizada pelo pessoal integrante da instituição interessada na avaliação, por avaliadores que integram os quadros institucionais, e que à instituição estão ligados funcionalmente. É uma avaliação pré-determinada quase sempre, é uma avaliação oficial. O oposto, uma avaliação realizada por investigadores estranhos à instituição ou ao projeto, seria uma avaliação externa. Em princípio, sim, porque, muitas vezes, a parte contratante exerce pressões sobre a contratada (avaliador), impondo seus objetivos, estabelecendo suas diretrizes, controlando procedimentos e outros aspectos que desfiguram inteiramente a avaliação, que, desse modo, perde o seu caráter externo. Stake (1982a) diz que “uma avaliação interna provavelmente será mais formativa do que somativa.” Uma palavra — provavelmente — salvou a afirmação, que deve ser compreendida no contexto norte-americano, não sendo possivelmente verdadeiro em nosso meio, no qual, predominantemente, as avaliações internas e pseudo-externas são somativas.

Stake (1982d), considerando a divergência de posição doutrinária entre os avaliadores, propôs a eminentes personalidades na área da avaliação educacional⁽¹⁾ um conjunto de perguntas sobre aspectos fundamentais. O nosso objetivo não se centra na análise dessas respostas, que foram feitas de forma lúcida e inteligente, como seria de prever, mas apresentar o nosso posicionamento pessoal em relação a cada uma delas: 1) liberdade do avaliador na escolha da metodologia para realizar o estudo avaliativo, 2) realização de pesquisa básica no decorrer de uma

avaliação, 3) limitações ao direito de divulgação dos resultados da avaliação em revistas especializadas, 4) problemas éticos em avaliação e a questão da confidencialidade, 5) a questão das conclusões apresentadas pelo avaliador no seu relatório, 6) os possíveis desvirtuamentos das conclusões do avaliador, e, finalmente, 7) procedimentos e cautelas a adotar ao aceitar um projeto de avaliação.

1) A questão da liberdade em avaliação precisa ser examinada com bastante cautela, para que problemas não acabem por invalidar o projeto. É necessário lembrar inicialmente que há limitações à liberdade, que são decorrentes de normas estabelecidas pela própria sociedade. Na avaliação, o avaliador é livre para escolher a metodologia adequada, os tratamentos estatísticos que se fizerem necessários. É livre, também, para deixá-los de utilizar, se sua abordagem não exigir o seu emprego, e é livre para recusar um projeto se influências estranhas desejarem impor seu suposto *know how*.

Outro aspecto a considerar no âmbito da liberdade do avaliador está no fato de que o seu poder decisório não é individual e personalíssimo, mas resulta, freqüentemente, da confluência de diferentes manifestações, especialmente dos vários públicos. Um projeto da avaliação implica negociações entre partes diversas, que afetam os graus de liberdade de que dispõe o avaliador. A liberdade do avaliador não deve constituir um elemento de pressão para impor decisões pessoais. Um projeto de avaliação resulta de um consenso entre várias partes, com interesses diversos, às vezes inicialmente conflitantes. Se o avaliador não concordar com os termos aceitos pela maioria, resta-lhe o direito que lhe assiste de ter a liberdade de recusar o projeto ou dar ao mesmo a continuidade desejada, e neste caso ainda dispõe da liberdade de demitir-se, por uma questão de coerência.

2) A realização de uma pesquisa básica simultaneamente com um projeto de avaliação é um assunto que merece reflexão. Em princípio, essa pesquisa não deve ser realizada, porque, muito possivelmente, não é de interesse da instituição patrocinadora e quase nunca é da parte da clientela que está interessada em respostas específicas. Por outro lado, a realização de uma pesquisa básica no decorrer da própria avaliação determinará, sem sombra de dúvida, o afastamento do avaliador dos seus interesses imediatos, perdendo-se, possivelmente, no mundo das conjecturas.

É preciso ficar claro que a realização de pesquisas básicas é fundamental para o desenvolvimento de qualquer campo científico; contudo, no campo da avaliação, faz-se necessário constatar se essa pesquisa trará alguma contribuição metodológica e possibilitará encaminhar e solucionar problemas educacionais de relevância, ou se representará apenas uma especulação teórico-abstrata. No primeiro caso, sem dúvida, a pesquisa deverá ser realizada; no segundo, poderá ser

adiada para um outro momento mais adequado. O certo, porém, é que uma instituição que se dedica à avaliação deve necessariamente possuir um staff com suficiente **expertise** para realizar pesquisas sobre assuntos que muitas vezes esclarecem pontos obscuros que a avaliação não consegue deslindar, ao mesmo tempo que contribui para o aumento do acervo de conhecimentos.

3) Parece-nos que o avaliador deva ter o direito de divulgar os resultados de sua avaliação em artigos e livros. É necessário, entretanto, que tenha cautela e não divulgue elementos que devam ficar restritos à instituição — é preciso muitas vezes usar de bom senso —, o que não quer dizer que a instituição interessada na pesquisa de avaliação tenha o direito irrestrito de dizer o que deve ou não deve ser publicado. A questão deve ser amplamente discutida, constituindo tópico específico do contrato de locação de serviços, a fim de evitar questões que possam ter implicações jurídicas. A questão não deve ser posta em termos de censura. Certas informações, por questões éticas, são de propriedade exclusiva da instituição; outras, ao contrário, precisam ser do domínio público, sem sombra de dúvida, e merecem ser divulgadas.

4) Há ocasiões em que o avaliador age de acordo com a sua consciência, quando se trata de problemas éticos que não admitem invocar questões de confidencialidade. É um momento de decisão personalíssima, porque, ainda que o trabalho do avaliador seja em aberto, é preciso que o avaliador decida sobre a conveniência da divulgação dos resultados, especialmente quando estes ainda não refletem conclusões finais. A divulgação seria, portanto, prematura. As avaliações não devem ser sigilosas, verdadeiros segredos de estado. Os resultados parcelados ou finais de uma avaliação devem ser divulgados. Uma coisa, no entanto, é certa, não se deve, sob o pretexto de escrúpulos de consciência, sonegar informações que são importantes para a compreensão do fenômeno avaliado.

5) As conclusões do avaliador muitas vezes são ignoradas, o que é lamentável, mas essa é a realidade. Não se pode impor nem mesmo à instituição que concorde com as conclusões do avaliador. O assunto é preocupante, mas, quando se considera a linha de Stufflebeam, não se pode deixar de admitir que a avaliação deve ter descrições e julgamentos para a tomada de decisões. Um relatório meramente conclusivo corre o risco de ser recusado ou, o que é pior, inteiramente ignorado. Parece-nos, portanto, que o avaliador deva saber apresentar seus resultados de forma que não sejam impositivos, mas ofereça à instituição contratante, à clientela interessada, todo um conjunto de elementos indispensáveis para um julgamento sobre o problema que foi avaliado.

6) O desvirtuamento das informações de uma avaliação é algo que não se consegue impedir e contribui para que seus dados sejam ignorados ou desacreditados. Haja vista a ocorrência com a divulgação do relatório

IAEP-92 — International Assessment of Educational Progress, que foi feita indevidamente apresentada por revistas não especializadas, a partir de um simples **press-release**, gerando informações sem respaldo nos dados da avaliação, especialmente no relativo ao Brasil. A mídia, com grande freqüência, ao informar acaba por deformar. Uma vez deformadas as informações, é praticamente impossível restaurar a verdade dos fatos. O importante, portanto, é que o relatório final não ofereça condições de censura e nem possibilite alterações no seu conteúdo, mesmo sabendo que, por mais cuidadoso que esse relatório seja, haverá interpretações as mais diversas, o que é inevitável.

7) Um procedimento recomendável, se houver uma tradição de avaliação, é conhecer a memória escrita e não-escrita de projetos anteriores. Saber um pouco da história é uma forma de tomar cautelas para que os mesmos não se repitam. Uma conversa com outros avaliadores parece-me extremamente útil, assim como um diálogo permanente com a instituição interessada na avaliação também é fundamental. Esta conversa é de suma importância porque define todas as condições, estabelecem-se todos os acertos necessários ao trabalho. Sem um entendimento entre todas as partes — instituição, avaliador, pais, professores e alunos — uma avaliação bem sucedida é inviável.

(1) M. Alkin, D. Campbell, D. Cohen, D. Hamilton, W. Harlen, E. House, M. Levine, T. Owens, M. Parlett, M. Provus, R. Rippey, R. Stake, J. Stanley, D. Stufflebeam, A. Thomas e R. Tyler (In: Stake, 1982d).

[2] Há uma grande rarefação de estudos qualitativos, mesmo em países com ampla tradição de pesquisa e avaliação. Isto resulta em parte da crença, aliás falsa, de que as pesquisas e avaliações educacionais devam ser isentas de subjetividade. Stake (1982b) procurou demonstrar que há lugar para a subjetividade tanto na pesquisa como na avaliação, estabelecendo com esse objetivo o seu posicionamento em relação aos **dados** da pesquisa e/ou avaliação e aos procedimentos metodológicos. Mostra, dessa forma, que os dados da pesquisa/avaliação devem provir de experiências pessoais, e os procedimentos metodológicos basearem-se na observação participante. O problema está, na visão de Stake (1982b), em saber escolher bons métodos de observação e interpretação subjetiva, razão pela qual nem sempre os relatórios de pesquisa naturalista se apresentam como de boa qualidade. Além do mais, o controle de possíveis vieses é uma tarefa árdua, ainda que não impossível.

Outro problema está em nos aproximarmos da realidade do fenômeno estudado, o que é possível apesar da subjetividade da observação, enquanto a objetividade dos procedimentos pode nos afastar da real dimensão do que está sendo estudado e pesquisado. A questão crucial não está, portanto, no subjetivo x objetivo, no qualitativo x quantitativo, ambos têm seu lugar na pesquisa/avaliação educacional, observa com justa razão Stake (1982b). O que importa não é tanto a precisão

(fidedignidade), mas o significado (validade), em que pese a procura quase obsessiva de muitas avaliações em demonstrarem que seus resultados são precisos, fidedignos e confiáveis. A pesquisa ou avaliação na perspectiva qualitativa procura muito mais que a precisão dos dados. Na verdade, busca captar a realidade de um fenômeno, suas múltiplas dimensões e seu valor. O julgamento de valor é sempre subjetivo. Stake (1982b) lembra muito bem que “*O julgamento subjetivo constitui parte central e essencial do ato avaliativo. Não apenas o programa (ou outra entidade qualquer) deve ser julgada em termos de suas excelências e deficiências. Também seus resultados, seus custos, seus processos e outras propriedades devem ser submetidas a uma revisão valorativa*”.

Julgamos que o problema da subjetividade x objetividade é uma questão de crença (fé docimológica, digamos). Se acreditamos na significância estatística (ou não significância) da diferença de duas médias ou na significância de F (ou não significância) na análise da variância das médias de diferentes variáveis, medidas com um instrumento que julgamos (acreditamos) proporcionar resultados precisos, ainda que não tenhamos certeza, por que, então, não acreditar, também, em uma análise subjetiva, em uma introspecção ou em um **insight** que nos permite compreender o sentido, a razão de ser, o valor de um determinado fenômeno avaliado? A dedução não invalida a indução; o empírico objetivo não exclui o subjetivo. As vezes, o alcance da verdade independe de uma estrutura rígida, supostamente científica.

- [3] Scriven (1967) faz uma distinção entre avaliação formativa e avaliação somativa, conforme apresentação anterior. A distinção é realmente significativa, contudo não é suficiente, propondo Stake (1983) que se faça, também, uma diferença entre estudos de avaliação preordenada e estudos de avaliação responsiva. É necessário destacar, inicialmente, que existem diferentes formas de avaliar programas, mas que não existe a **melhor**. Algumas avaliações recomendadas em uma determinada situação não respondem às perguntas de diferentes clientelas possíveis em outras situações. Existem diferentes modelos de avaliação, entretanto, Stake (1983) deixa bem claro que, qualquer que sejam os modelos adotados para a avaliação, sempre deixam de destacar algum aspecto que é realmente importante para a compreensão em profundidade daquilo que se está avaliando; desse modo, Stake declara-se inclinado a preferir planejamentos de avaliação que desempenhem um serviço determinado e ofereçam respostas úteis a grupos específicos de pessoas interessadas.

A avaliação responsiva é proposta por Stake (1983) como sendo a que melhor se adapta à realização de serviços específicos e a que permite considerar aspectos controversos que precisam ser analisados. A proposta responsiva talvez sacrifique um pouco a fidedignidade das medidas, mas, em compensação, aumentaria a utilidade das informações. Voltamos,

assim, ao início do presente comentário. A avaliação preordenada preocupa-se com a apresentação de objetivos, usa testes objetivos, destaca padrões estabelecidos pelos programas e apresenta relatórios bem semelhantes aos de uma pesquisa. A avaliação responsiva supera formalismos de comunicação, prefere um relacionamento mais natural com a clientela interessada. Stake (1983) reconhece que a avaliação responsiva não é uma abordagem nova, mas, ao longo dos tempos, por força de sua subjetividade, vem encontrando resistências à sua utilização; contudo, o problema pode ser superado (ver **comentário 2**), usando-se métodos adequados de observação por diferentes pessoas e sua triangulação.

Uma avaliação é responsiva na medida em que seu desenvolvimento procura:

1. concentrar-se com especial ênfase nas atividades do programa e não exclusivamente nas suas intenções;
2. coletar o máximo possível de informações para atender aos reclamos da audiência a que se destina;
3. apresenta diferentes perspectivas de valor que procuram traduzir o sucesso ou fracasso do programa.

A fim de realizar uma avaliação responsiva, um elenco de providências deve ser estabelecido, segundo a orientação de Stake (1983):

1. elaborar um plano detalhado de observações;
2. selecionar pessoas qualificadas para realizar as observações;
3. preparar, com auxílio dos observadores, para rápidas apresentações, diferentes relatos, gráficos etc;
4. procurar identificar o que é mais importante para os vários tipos de audiência e coletar expressões de indivíduos que apresentem pontos de vista discordantes;
5. verificar a qualidade dos registros apresentados pelos observadores e verificar junto aos participantes do programa como reagem às descrições; e
6. identificar pessoas com experiência na área e coletar sua reação à importância dos elementos encontrados; ao mesmo tempo, procurar estabelecer a reação da audiência a que os dados se destinam, procurando estabelecer a sua relevância.

A avaliação responsiva procura interagir informalmente e registrar as ações e reações dos que se acham envolvidos no processo de avaliação, procurando assegurar o máximos de fidedignidade na comunicação com os vários tipos de audiência. Ao final, um **relatório** pode ser apresentado ou não, em função do que foi estabelecido inicialmente entre as partes.

Acreditamos, como Stake (1983), que o avaliador se deva questionar sobre as razões que determinaram a avaliação e, assim, concentrar sua atenção no programa, observando sucessos e insucessos, para, então,

propor questões e critérios, sem se impressionar com objetivos pré-definidos e a elaboração *a priori* de instrumentos para coleta de informações, ou seja, preocupar-se com aspectos instrumentais da educação, sem dúvida importantes, mas que não devem suplantar os valores intrínsecos da observação.

Ouvir e saber ouvir é um dos aspectos mais importantes da avaliação responsiva. É através de conversas que o avaliador começa a identificar os problemas, diz-nos Stake (1983), ou problemas em potencial. Assim, a avaliação responsiva parte de problemas que, na realidade, refletem a complexidade do fenômeno educacional. Os problemas oferecem uma perspectiva para, numa avaliação responsiva, estabelecer discussões, orientar observações sistemáticas, realizar entrevistas e, se for o caso, definir instrumentos de mensuração.

Stake (1967) representou uma matrix no artigo *The Countenance of Educational Evaluation* que apresenta treze categorias de informações que não mostram grande diferença das que ofereceram, por exemplo, Stufflebeam (1971) e Provus (1971). Fica evidente nos três casos citados que o importante é coletar informações a partir de fontes independentes e confiáveis.

Uma avaliação responsiva não é feita por fases, porque muitos de seus aspectos são recorrentes, como mostra o planejamento de Stake (1985). As 12 etapas da sua planificação podem seguir no sentido horário, anti-horário e até mesmo transorário: — conversar com clientes e organizadores dos programas; identificar escopo do programa; rever atividades do programa; descobrir propósitos e preocupações; conceituar problemas; identificar necessidade de dados; selecionar observadores, juízes e instrumentos, se necessários; observar antecedentes, interações e resultados; preparar descrições, estudos de caso; validar, confirmar, tentar desconfirmar; informações para as várias audiências; relatórios formais, se for o caso. Estas etapas são às vezes apresentadas sob a forma circular, para facilitar a percepção do todo; no entanto, como mostra Stake, muitos desses acontecimentos são simultâneos, devendo o avaliador, se necessário, voltar a eventos anteriores sempre que a situação exigir.

- [4] A realização de avaliações de forma intensiva, como ocorreu na década de 60 no contexto americano, nem sempre trouxe os resultados esperados. Uma atitude crítica em relação à produção supostamente científica fez-se necessária. Cuba, por exemplo, em vários momentos de seus trabalhos, adotou uma posição bastante crítica e pôs em dúvida o valor dessas avaliações, realizadas após 1965, que se preocupavam apenas com o produto e nem sempre forneciam dados indispensáveis às decisões para a reformulação ou interrupção de programas educacionais.

Cuba e Lincoln, em *Effective Evaluation* (1981), analisam com cuidade a maioria dos problemas ligados à avaliação educacional.

Quanto à grande diversidade de modelos, nem sempre relevantes, depois de compará-los, mostram que entre eles há algo importante, possivelmente a única coisa que têm em comum, a idéia de que a avaliação deve preocupar-se em determinar o valor daquilo que está sendo avaliado, como no caso dos programas educacionais.

As pesquisas quantitativas nem sempre satisfizeram a um número expressivo de educadores, levando-os à procura de diferentes alternativas; desse modo, trabalhos qualitativos e naturalistas, que nos anos 60 nem sempre foram aceitos com tranqüilidade no mundo acadêmico, passaram a usufruir de um maior prestígio a partir da década de 70, graças à reação surgida em diferentes momentos por elementos expressivos no mundo da educação. Stake (1967), Parlett e Hamilton (1976), Eisner (1978), e Guba e Lincoln (1981), entre outros avaliadores, preocupados com questões metodológicas, passaram a dar maior destaque à análise indutiva, à descrição e ao estudo das percepções das pessoas (julgamentos) sobre fenômenos educacionais. O posicionamento de Guba e Lincoln em relação ao quantitativo e ao qualitativo não é extremado (1981), acreditam que uma certa complementariedade possa existir entre os dois paradigmas, ainda que suas posições sejam bastante claras em relação à avaliação qualitativa. Aliás, uma distinção inflexível entre os dois métodos era um dos dogmas do positivismo lógico, conforme H. R. Howe, citado por Worthen e Sanders (1987).

Ao nos situarmos em relação à abordagem naturalista da avaliação educacional, não podemos deixar de reconhecer que Stake (1967), com o seu ensaio — *The Countenance of Educational Evaluation* — apresentou a primeira reação teórica significativa, valorizando o julgamento dos avaliadores no processo de avaliação, o que veio a influir grandemente nas novas abordagens que posteriormente surgiram, quase sempre contrárias ao objetivismo e ao mecanismo dos modelos que então dominavam a avaliação educacional. Guba adota uma posição contrária à abordagem racionalista e, juntamente com Yvonna Lincoln (1981), ao fazer uma revisão crítica dos principais modelos em avaliação educacional, acaba por rejeitá-los, excetuando a proposta de avaliação responsiva de Stake. Guba e Lincoln (1981) incorporaram a avaliação responsiva de Stake à avaliação naturalista, que julgam superior às demais abordagens existentes. Assim o fazem porque acreditam que a avaliação responsiva possa englobar todos os outros modelos, sendo uma abordagem mais flexível para responder aos interesses de diferentes grupos.

A questão da abordagem naturalista foi apresentada por diversos teóricos, entretanto, coube a Guba, no final da década de 70, mostrar de forma mais completa as suas vantagens na área educacional. Guba fez, inicialmente, uma análise comparativa entre pesquisa naturalista, com base na fenomenologia e na etnografia, e a pesquisa realizada com

fundamento no paradigma positivista experimental. A partir desse confronto, iniciou, realmente, a metodologia naturalista. Guba e Lincoln (1981), após estabelecerem as ligações entre a avaliação responsiva delineada por Stake e a naturalista, apresentaram a descrição dos procedimentos para a realização dessa abordagem, trabalho que desenvolveram com extrema competência em obra posterior — *Naturalistic Inquiry* (1985), fundamental para o dimensionamento e a realização da avaliação segundo esta nova abordagem.

Quando consideraram a avaliação, Guba e Lincoln (1981, 1985) destacaram que sua principal função consistiria em dar respostas às informações solicitadas por vários tipos de clientela, devendo levar em consideração as diversas perspectivas de valor dos seus membros. Ao adotar uma abordagem naturalista o avaliador analisa uma determinada atividade *in loco*, sem manipulações ou controle. O avaliador nessa posição desempenharia o papel de quem aprende a partir daqueles que são estudados, os informantes, no dizer de Guba e Lincoln (1985). O que o avaliador aprende? Aprende conceitos que descrevem o mundo, aprende a usar as definições desses conceitos, aprende a traduzir o mundo dos informantes para que possa, assim, compreendê-lo.

A avaliação naturalista, na perspectiva de Guba e Lincoln (1981, 1985), preocupa-se mais com a credibilidade dos seus achados do que propriamente com a sua validade interna, usando para estabelecer essa credibilidade os métodos da checagem cruzada e da triangulação com o objetivo de confirmar os dados encontrados. A adequação dos dados, mais do que a sua validade externa, é outra preocupação da abordagem naturalista, quando se trata de realizar uma avaliação em outros contextos ou em outros sujeitos (Guba e Lincoln, 1981).

O avaliador naturalista, no caso de uma preocupação maior com a consistência dos dados da avaliação, concentra-se mais na sua auditoria do que na fidedignidade, sendo aquela característica obtida por intermédio da atuação de um segundo grupo de avaliadores que fariam a revisão de toda a documentação e da fundamentação dos raciocínios subjacentes à avaliação. O avaliador, dessa forma, poderia verificar a concordância dos seus dados com os que foram determinados pelo grupo de auditoria. Ainda na pesquisa naturalista, a preocupação do avaliador não estaria na neutralidade da avaliação, mas na possibilidade de confirmação dos resultados. Sua meta não é, pois, a objetividade, como na avaliação quantitativa. Os dados da avaliação naturalista devem ser, segundo a lógica dessa abordagem, factuais e confirmáveis.

O pesquisador/avaliador naturalista, ao coletar seus dados, procura determinados tipos de informação (Guba e Lincoln, 1981), que podem ser resumidos da seguinte forma:

1. informações descritivas sobre o objeto da avaliação e seu contexto;

2. informação responsiva às indagações, documentando-as, procurando relações de causa e efeito, e identificando possíveis ações a empreender;
3. informação responsiva aos problemas, esclarecendo-os e identificando possíveis linhas de ação para resolvê-lo;
4. informações sobre valores (esclarecendo-os, procurando encontrar suas fontes e até que ponto são os mesmos convincentes);
5. informação sobre padrões a serem usados na avaliação (identificando critérios, expectativas e necessidades).

Usando entrevistas, observações, referências não-verbais, documentos, registros, e medidas não-obstaculizantes (*unobtrusives*), anotações de campo e o registro das fontes, o avaliador, empregando a descrição como uma técnica de registro (Guba e Lincoln, 1981), faz o levantamento das informações para, numa fase posterior, apresentar a descrição final da globalidade dos dados coletados.

A introdução da metodologia naturalista provocou reações críticas que a julgaram inconsistente nos seus fundamentos e de difícil uso no exercício prático da avaliação educacional. As contracriticas, por sua vez, procuraram demonstrar a superioridade da abordagem responsiva/naturalista e sua maior flexibilidade relativamente às demais abordagens, que, inclusive, admitiria a possibilidade de ser usada com outras metodologias, para a avaliação de aspectos específicos.

A avaliação naturalista, sem sombra de dúvida, deu maior relevo ao elemento humano, e passou a direcionar a visão do avaliador para determinados aspectos que não são considerados na avaliação objetiva. Os usuários da avaliação responsiva/naturalista ou de outras metodologias participativas vêem a educação como uma atividade humana extremamente complexa e tentam refletir sobre essa complexidade a fim de extrair dessa reflexão o máximo de elementos explicativos sobre o objeto (fenômeno) estudado. A possibilidade de novos *insights* talvez seja um dos grandes méritos da avaliação responsiva/naturalista. O advento dessa metodologia trouxe, como decorrência, o desenvolvimento de novas técnicas de coleta de dados em relação a elementos menos tangíveis de aspectos diversos dos seres humanos (Guba e Lincoln, 1981, 1985; Stake, 1967; Worthen e Sanders, 1973, 1987).

A questão da subjetividade como argumento contrário a esse tipo de abordagem ficou definitivamente esclarecido por Guba e Lincoln (1981), sendo uma restrição — a subjetividade — igualmente aplicável a abordagens que têm uma orientação objetivista. As críticas à supervalorização da observação e às análises baseadas em perspectivas individuais, assim como ao fato de a pesquisa naturalista apresentar pouco ou mesmo nenhum interesse pelo emprego de instrumentos de

medida ou no uso de dados de grupo, constituem os pontos centrais do ataque por parte dos que seguem abordagens não-subjetivistas.

Achamos que os argumentos contrários à avaliação responsiva / naturalista possam ser adequadamente respondidos, como efetivamente já o foram por diversos especialistas na área (Guba e Lincoln, 1981). A subjetividade das formulações, argumento maior contra a avaliação naturalista, não nos impressiona e nem preocupa. O que constitui real preocupação para nós é a possibilidade de improvisação de falsas abordagens naturalistas em nosso contexto educacional. Queremos explicitar que muitos desses improvisados avaliadores usam a Fenomenologia sem nunca terem tido um trato mais íntimo com o pensamento dos que a desenvolveram; ou, então, usam ou pretendem usar o método etnográfico sem o conhecimento de trabalhos fundamentais. A improvisação parece-nos, ao lado da tendência a digressões filosóficas pouco coerentes, o grande problema. Outros aspectos da pesquisa responsiva/naturalista, como, por exemplo, seus altos custos e o longo tempo de execução que costuma demandar, são questões menores e perfeitamente controláveis.

- [5] O estudo de caso é um exemplo bastante comum de avaliação/pesquisa naturalista. A metodologia empregada para a sua concretização é muitas vezes chamada de qualitativa, ou etnográfica ou fenomenológica. Stake (1983b) destaca que essas palavras significam diferentes coisas, entretanto, globalmente, estão procurando destacar a singularidade e a contextualidade de cada situação educacional. Apesar de amplamente utilizado em diferentes áreas científicas, o estudo de caso não vem recebendo o merecido destaque em educação, como lembra Lüdke (1983).

A metodologia de estudo de caso na área educacional recebeu um grande avanço a partir do final da década de 70, graças, especialmente, à contribuição teórica e aos trabalhos realizados por Robert E. Stake, bastando lembrar o seu artigo seminal — *The Case Study Method in Social Inquiry* —, citado por Lüdke (1983), e a obra conjunta — *Case Studies in Science Education* —, realizada para a National Science Foundation (NSF), em 1978, trabalho que sem dúvida constitui um marco na avaliação educacional.

Ao comentar o trabalho de Stake (1983b), Lüdke (1983), com a sua experiência em pesquisas qualitativas, consegue sintetizar com rara felicidade o que vem a ser o estudo de caso, ao escrever que essa metodologia “*procura retratar, naturalmente, a realidade do fenômeno educacional em sua inteireza, sem depender muito de artifícios analíticos utilizados por outros métodos mais tradicionais da investigação científica. No esforço de retratar a realidade, o pesquisador dela procura se aproximar ao máximo, tanto pela sua maneira de agir durante o*

estudo, como pelo relato final. Este deverá ser um discurso menos acadêmico, mais natural, coloquial, para que as pessoas envolvidas possam não só se encontrar e se reconhecer no estudo como também encontrar contribuições para um melhor equacionamento dos problemas em questão".

Stake (1983b), ao considerar o estudo de caso, chama a atenção para o fato de que o mesmo pode ser conceituado em termos de limites do caso, problemas do estudo e padrões dos dados. O contraste entre a teoria e a realidade a ser avaliada ou pesquisada cria uma situação desafiante que o avaliador precisa superar. Stake (1983b) sugeriu aqueles três elementos básicos para a concretização efetiva de um estudo de caso:

- 1º estabelecimento de limites rigorosos para o caso;
- 2º definição dos problemas ou questões básicas para o estudo;
- 3º e a busca de padrões emergentes dos dados.

O avaliador (ou pesquisador), sempre dentro da linha estabelecida por Stake (1983b), "precisa especificar o caso a ser estudado, o que geralmente é uma tarefa difícil", "o que é e o que não é incluído no caso?" Ou seja, o limite que circunscreve o próprio caso. Stake (1983b) chama a atenção para o seguinte: "Os limites, mais talvez do que as tendências centrais, ... merecem maior atenção, pois muito do significado do caso é encontrado em suas extremidades, tendo em vista a influência especialmente exercida pelos elementos circundantes, por seu contexto. Os limites são freqüentemente variáveis e indistintos.... A reflexão sobre os limites auxilia-nos a progredir na compreensão do caso."

A questão do estabelecimento de limites provoca inúmeras indagações, como, por exemplo, as que foram propostas por Lüdke (1983): — "Como... estabelecer limites para o caso que assegurem ao mesmo tempo sua exequibilidade e sua validade aos olhos da comunidade acadêmica? Que critérios devem ser usados para pensar e justificar esses limites? Devem prevalecer as questões de custo, duração, disponibilidade de pessoal, ou as de interesse científico mais específico? Quem deve ser ouvido na tomada de decisões sobre os limites do caso: — os envolvidos no estudo ou os que o financiam? Em que ordem? Como estabelecer uma equação satisfatória entre o desejado e o possível, sem comprometer o interesse científico?" A questão dos limites a serem definidos é realmente complexa, mostrando, por outro lado, a necessidade de **expertise** e vivência da realidade para levar a efeito um estudo de caso.

Uma pergunta se nos ocorre de imediato: por que Stake procurou chamar a atenção, dando bastante destaque, sobre a questão de limites, problemas e padrões em estudos de caso. O próprio Stake (1983b) esclareceu a sua preocupação ao escrever que é preciso trabalhar "no sentido de estabelecer meios mais efetivos de provar a complexidade e a

contextualidade dos fenômenos educacionais. Ainda que, atualmente, muita coisa fique para a intuição — não obstante seja a intuição um ingrediente essencial à pesquisa naturalista —, continuamos a pesquisar rotinas que farão esses métodos mais seguros".

Uma avaliação ou pesquisa apresenta problemas que precisam ser solucionados. Stake (1983b) diz: "A palavra problema lembra-nos de que os assuntos ou questões para indagações possuem componentes de valor e são potencialmente contenciosos". "Freqüentemente, identifico problemas que 'influenciam' a instrumentalidade de uma determinada decisão...". "O número de problemas aumenta com o desenvolvimento do estudo e na medida em que os assuntos apresentam diversos níveis de importância."

A procura de padrões nos dados merece da parte de Stake (1983b) algumas considerações sobre as quais o avaliador, ou o pesquisador, deve refletir ao realizar um estudo de caso: "a palavra 'padrões' deve ser entendida como configurações recorrentes. Em estudos mais quantitativos, são chamados de 'covariações' ou 'tendências cíclicas'. Padrões são regularidades que nos possibilitam um discernimento sobre a natureza do problema".

Os dois últimos aspectos-chaves do estudo de caso: definição dos problemas e a busca de problemas emergentes dos dados constituem aspectos que somente podem ser dimensionados adequadamente se o avaliador (ou o pesquisador) se valer da experiência já adquirida no campo da sociologia e da etnografia, áreas de conhecimento que desenvolveram e aprofundaram metodologias qualitativas para a análise e a compreensão de fenômenos sociais, entre os quais podem ser incluídos os educacionais. Lüdke (1983), com justa razão, argumenta que "Urge superar o falso conflito entre análises quantitativas e qualitativas, e desenvolver métodos que realmente procurem responder às necessidades sentidas pelos que trabalham em educação e querem melhorá-la. O estudo de caso será certamente um desses métodos."

Não se pode esquecer, como reitera Lüdke (1983), em diferentes momentos de sua discussão sobre o artigo de Stake (1983b), que o estudo de caso "privilegia o enfoque do todo, em detrimento de variáveis preestabelecidas; assume a importância da subjetividade para o conhecimento científico, estimula mais o desenvolvimento de padrões sugeridos pelos próprios dados, do que a testagem de hipóteses pré-fixadas". E mais adiante, no mesmo trabalho, Lüdke (1983) volta ao assunto, colocando o problema de maneira enfática, mas com bastante clareza: "Ao privilegiar a visão do todo e estimular o uso da interpretação pessoal, o estudo de caso permite ao pesquisador ir além do que a simples apresentação de dados numéricos permitiria. O pesquisador é instado a integrar esses dados numa perspectiva muito mais ampla, composta por sua própria experiência em educação e o contexto onde o

caso se situa. Sua prática educacional e seu conhecimento de teoria educativa, aliados à sua capacidade de interpretação e à habilidade de comunicação serão integrados na elaboração de um quadro coerente a respeito de uma determinada instância da realidade educacional. Esta não será apenas representada por um conjunto de tabelas, com dados significativos ou não, mas um relato que, embora de cunho pessoal, se preocupará em transmitir uma imagem total do fenômeno estudado. O leitor receberá o relato e a interpretação desse fenômeno e naturalmente acrescentará suas próprias interpretações."

- [6] Há uma concordância, sem sombra de dúvida, de que a metodologia do estudo de caso é perfeitamente adequada para trabalhos na área das ciências do homem, inclusive na educação. Voltamos, entretanto, novamente, a enfocar a principal restrição feita à metodologia, que, segundo a manifestação de alguns, não permitiria generalizações. Ainda que o assunto já tenha sido anteriormente abordado, a ele voltamos, tendo em vista a sua importância, reiterando a nossa posição.

Stake, In: Madaus et al. (1993), propõe-se a mostrar que o estudo de caso, como metodologia da pesquisa, está epistemologicamente em perfeita harmonia com a experiência do leitor e é a partir dessa experiência que realiza suas generalizações, quando for o caso. As pesquisas/avaliações são realizadas visando a que as pessoas aumentem a sua compreensão a respeito de algo (um ser, um fenômeno, um objeto). Essa compreensão resulta de uma experiência que tanto pode ser direta ou vicária (indireta). Assim, segundo Stake (1993), uma maneira eficaz de proporcionar uma melhor compreensão é por intermédio de palavras e exemplos que traduzam uma experiência natural adquirida pelo envolvimento pessoal na situação. A partir de W. Dilthey, citado por Stake (1993), nem sempre o objetivo e o "científico" são o melhor meio de levar o homem à compreensão de si mesmo. O somatório de suas experiências permite uma compreensão e um saber que levam o homem a compreender a si mesmo e aos outros; desse modo, o estudo do homem se baseia nas relações entre experiências vividas, expressão e compreensão.

Stake (In: Madaus et al. 1993) faz uma distinção entre **conhecimento proposicional**, — composto de afirmações que podem ser compartilhadas e que para a maioria das pessoas decorre da observação de objetos e acontecimentos —, e **conhecimento tácito**, que também envolve objetos e acontecimentos, mas resulta da experiência com os mesmos, e do pensamento reiterado sobre os objetos e acontecimentos. Através da sua racionalidade, o homem observa a si mesmo, mas o conhecimento de si mesmo decorre da sua consciência. O conhecimento tácito inclui associações que nem sempre podem ser expressas, mas que dão origem a novos significados e idéias. A maior parte do nosso conhecimento, e com

o qual criamos novas formas de compreensão, argumenta Stake a partir da filosofia do espírito de Dilthey, é o conhecimento tácito.

A questão da generalização, por intermédio de estudo de caso, é posta por Stake (1993) nos seguintes termos: ela se desenvolve como produto da experiência, a partir do conhecimento tácito, e geralmente nunca são previsões. A generalização conduz à ação e dela é inseparável, podendo vir a transformar-se em conhecimento proposicional, desde que submetida a testes que caracterizam a generalização formal. A educação é uma área em que pouquíssimas "leis" foram efetivamente validadas; assim sendo, Stake, In: Madaus et al. (1993), considera ser pedante dar constante atenção às "leis" na área da educação e, naturalmente, a supostas "generalizações".

3. AVALIAÇÃO E O AVALIADOR EDUCACIONAL

The time is ripe for broadening the base from which inquiry in education can go forward. It is time for a more catholic sense of possibility; we need, in my opinion, to widen our epistemology.

Eisner, 1993

3.1 Introdução

Ao planejamos o presente trabalho, fizemos, inicialmente, algumas colocações, sob a forma de perguntas, que nos orientaram no desenvolvimento dos diversos capítulos. Agora, à guisa de conclusão, passaremos a analisar cada uma das indagações propostas, examinando o caráter sistematizado do conhecimento em avaliação, os diversos problemas relacionados à construção e delimitação do espaço da avaliação e, finalmente, consideraremos os aspectos mais significativos na formação de um avaliador educacional, com vistas à sua atuação na prática avaliativa.

3.2 Avaliação como conhecimento sistematizado

Ao fazer um levantamento do progresso da avaliação educacional nos Estados Unidos, Smith (In: Madaus et al., 1993), partiu de duas interrogações que, em princípio, poderemos adotar, considerando o problema no Brasil, por corresponderem

de certa forma às proposições orientadoras anteriormente apresentadas: *Qual o desenvolvimento da avaliação educacional até a presente data?* e *Quais as necessidades de crescimento no futuro?* A história da avaliação no Brasil, com exceção da precária avaliação do rendimento escolar que há séculos vem sendo realizada, é bastante recente, como mostram os trabalhos de Gatti e Vianna (vide nota [1]), preocupados com a memória e a análise do que tem acontecido nessa área. A partir da bibliografia, constatarmos que a experiência brasileira está limitada no tempo, porque somente na década de 80 e, especialmente, no início dos anos 90 é que os estudos de avaliação começaram a ser realizados, mas de uma forma restrita, envolvendo, especialmente, a avaliação de sistemas de ensino sob a ótica do seu produto.

O conhecimento que temos da avaliação educacional é limitado e enviesado, pouco ou quase nada foi feito relacionado com a avaliação de programas, materiais e produtos, assim como a avaliação de amplos sistemas de educação foi ignorada, salvo, neste último caso, experiências episódicas como as de Minas Gerais e São Paulo. Por outro lado, os trabalhos incidem, na sua quase totalidade, sobre o sistema público e, eventualmente, sobre a avaliação na área privada, exceto algumas raras experiências, como a do **Sistema Nacional de Avaliação do Ensino Básico** (SAEB), em 1995, que além de informações sobre a rede pública, levantou, também, dados do ensino privado; entretanto, o impacto dessa avaliação em nível nacional foi bastante restrito. Além disso, praticamente inexistem avaliações relacionadas ao ensino supletivo e às diferentes formas de educação especial. O mesmo se pode dizer sobre a educação em nível de 3º grau, que apenas nos últimos anos tem procurado desenvolver metodologias de avaliação institucional, especialmente na esfera oficial. Destaque-se, mais uma vez, que a avaliação está presa ao conhecimento cognitivo, apesar de envolver outras variáveis (titulação de corpo docente, laboratórios, biblioteca, produção científica etc). Vemos, portanto, com a amostra restrita que possuímos, no presente momento, que nos falta uma experiência mais ampla no campo da avaliação educacional; contudo, a

persistirem as atuais tendências, é possível que venhamos a explorar novos caminhos, se superarmos nossas limitações em termos de recursos humanos.

Analisando a questão segundo uma perspectiva histórica, vemos que a avaliação educacional ficou limitada na sua prática, envolvendo quase sempre o rendimento escolar. Não houve, salvo alguns poucos casos, grandes especulações teóricas, um refletir sobre a avaliação¹² visando a criação de novos modelos e a elaboração de uma teoria da avaliação. Ou seja, não tivemos um Tyler, um Cronbach, um Scriven, para mencionarmos apenas alguns dos grandes teóricos da avaliação no contexto norte-americano. Além do mais, não foram apresentadas alternativas para os modelos elaborados por Stake, Stufflebeam, Guba, Parlett e Hamilton, entre outros. No caso brasileiro, houve a influência de vários desses modelos, nem sempre aplicados adequadamente, porque nenhum deles diz como fazer efetivamente uma avaliação, que assim fica sujeita a inspirações do momento.

Quando consideramos que, em outros contextos, formas diversas de avaliação estão se desenvolvendo — **responsiva** (Stake), **naturalista** (Guba), **iluminativa** (Parlett e Hamilton), entre outras — apercebemo-nos de que estamos presos ao tradicional e rotineiro. A figura do avaliador educacional não se cristalizou, não existe como identidade profissional — haja vista, por exemplo, o número limitado de instituições que se dedicam à avaliação educacional e o número extremamente parcimonioso de periódicos voltados exclusivamente para a avaliação, e com tiragens regulares. A avaliação e o avaliador, pelo menos no presente momento brasileiro (1999), estão à procura de uma identidade [2].

A leitura das apresentações das idéias básicas de Tyler, Cronbach, Scriven, Stufflebeam e Stake permite-nos extrair

¹² Precisamos, por ser de justiça, fazer referência a dois modelos pioneiros desenvolvidos e utilizados na prática para a avaliação de programas educacionais: GOLDBERG, M. A. A. Era uma vez uma avaliação: lições de vida. *Educação e Avaliação*, nº 2. Ed. Cortez. São Paulo. 1981; SAUL, A. M. *Avaliação Emancipatória - desafio à teoria e à prática da avaliação e formulação de currículo*. São Paulo, Cortez: Autores Associados. 1988.

algumas lições. Sente-se, inicialmente, que a avaliação é um campo aberto à reflexão teórica e, ao mesmo tempo, que esse pensamento teórico é, necessariamente, a infra-estrutura que possibilita a prática da avaliação. Sem teoria não é possível uma prática consequente. As teorias refletem diferentes posicionamentos epistemológicos, o que valoriza a avaliação como prática da atividade intelectual. A avaliação, assim como a ação educacional, exige constante reflexão sobre o ente avaliado e as implicações do ato de avaliar.

A reflexão sobre o ser avaliado mostra, primeiramente, a complexidade do campo da avaliação, pela inclusão de diferentes sujeitos. Pensemos apenas no ente/aluno, o que nos leva a duas dimensões da avaliação: cognitiva e não-cognitiva. A avaliação cognitiva desdobra-se em sub-dimensões, digamos assim, apresentando uma visão quantitativa e outra qualitativa, visões estas que se completam para que possamos fazer um juízo sobre o sujeito. A dimensão cognitiva exige diferente instrumentação — por critério ou por norma, dependendo da natureza da avaliação: formativa ou somativa, com diferentes linhas de interpretação. A dimensão não-cognitiva, nem sempre considerada, abrange um amplo espetro de características, que demandam pessoal especializado, técnicas e instrumentos variados, exigindo, freqüentemente, a construção de um instrumental **ad-hoc**, face à carência do mercado interno. Toda essa reflexão leva-nos a constatar que a avaliação exige diferentes tipos de capacitação, revelando-se, assim, um campo de conhecimentos integrados, que necessita, forçosamente, da contribuição, como se vê no exemplo apontado, de elementos provenientes de diferentes áreas científicas, ligadas ou não às ciências do homem, sendo, por via de consequência, uma área interdisciplinar que demanda uma ação solidária de vários especialistas. A situação é igualmente complexa se considerarmos os sistemas educacionais, as instituições (universidades e seus departamentos, escolas), programas curriculares, professores, administração, entre outros entes que podem ser objeto de uma avaliação na área educacional [3].

Creamos, assim, que, ao longo dos vários capítulos, especialmente os publicados em **Estudos em Avaliação Educacional** nº 16 (1997) e 17 e 18 (1998), ficou claro que a avaliação educacional é um campo de conhecimentos sistematizados que lança mão de outras áreas (Psicologia, Sociologia, Antropologia, Demografia, entre outras, não esquecendo a Filosofia (Epistemologia) e a própria Filosofia da Ciência). É um campo em constante transformação que sofre o impacto de diferentes tecnologias e, dessa forma, vem refinando cada vez mais os seus procedimentos de análise e interpretação, no campo quantitativo e qualitativo. *Podemos, desse modo, afirmar, positivamente, em relação à pergunta inicial, quando indagamos se a avaliação educacional é um espaço possível de conhecimento sistematizado [4].*

3.3 Problemas relativos à construção e delimitação do espaço da avaliação

A avaliação educacional — e vamos circunscrever a nossa reflexão à avaliação de programas, projetos e materiais — tem seus problemas, como aliás, todas as áreas do conhecimento humano. Eventualmente, alguns desses problemas podem ser extrapolados para outras áreas da avaliação, mas seguiremos o pensamento de Stufflebeam e de seus colaboradores na definição de padrões de qualidade (parte final do primeiro estudo).

Um problema possível de ocorrer refere-se à identificação do público envolvido ou afetado pela avaliação. Nem sempre é possível, ainda que extremamente desejável, caracterizar a clientela, sobretudo a que será afetada pelo resultado da avaliação, especialmente quando consideramos o nosso contexto, problema que, entretanto, está sendo superado graças à informatização, à melhoria das técnicas de coleta de informações e à organização de banco de dados confiáveis. A credibilidade e a aceitação de uma avaliação estão associadas à credibilidade do avaliador, que deve possuir competência técnica e experiência suficiente para empreender a coordenação de uma equipe de avaliação. Isso pode ocorrer — e estamos sempre pensando no caso brasileiro — em decorrência da falta de tradição de

pesquisas e da ausência de profissionais na área, dando margem a improvisações, sobretudo quando se nos deparamos com um **boom** de avaliações, como no momento presente, ou quando a avaliação passa a ser um modismo, na vã esperança de que a avaliação possa resolver todos os problemas educacionais, quando a sua real finalidade é a de identificá-los para que possam ser dimensionados, analisados e solucionados, o que não constitui tarefa do avaliador. A falta de pessoal com treinamento adequado pode ser em parte solucionada com a contratação de assessores com **expertise** no campo, como vem ocorrendo em alguns projetos nacionais. A questão da seleção de informações constitui outro problema que às vezes decorre da inexperiência do avaliador ou da sua ânsia de coletar o máximo possível de dados, que muitas vezes serão depois desprezados por não se ajustarem aos objetivos da avaliação. É necessário, pois, tendo em vista o objetivo da avaliação, selecionar os dados que são efetivamente cruciais para a avaliação. É evidente que uma avaliação procura obter o máximo possível de elementos informativos, levando-se em conta, entretanto, a sua pertinência para o estudo avaliativo. Um problema comum, mas que também pode ser superado, está relacionado à interpretação dos dados, como pode ser identificado em algumas avaliações. Sente-se que falta ao processo uma estrutura lógica que permita descrever, analisar e interpretar com validade as informações. Às vezes, elementos relevantes são abandonados e outros menos valiosos são considerados. A deficiência na interpretação, assim como os outros problemas anteriormente apresentados, comprometem a qualidade de uma avaliação.

As questões relacionadas à clareza e disseminação dos relatórios são fatores limitativos do impacto de uma avaliação. É preciso considerar que os relatórios se destinam a diferentes níveis de audiência e que, portanto, devem considerar esse aspecto. As tecnalidades devem ficar restritas ao pessoal especializado, capaz de entender os procedimentos necessários à realização da pesquisa/avaliação em nível sofisticado; no outro extremo, para os que têm acesso à mídia, uma simples descrição do que foi constatado, em nível de sumário executivo, mas com

elementos que inibam distorções, é suficiente. Tudo isso com a máxima clareza possível, evitando-se linguagem hermética e obscura, que dificulta o entendimento dos próprios especialistas. O problema pode ser resolvido com uma boa revisão, evitando-se o **copy-desk**, que aprimora a qualidade da linguagem, mas às vezes distorce as colocações científicas do texto.

A nossa preocupação no momento é com problemas que possam prejudicar a qualidade da avaliação, por isso nos vamos referir a três outros aspectos — disseminação, tempo oportuno e impacto — que podem constituir futuros problemas, gerando impedimentos para a consolidação do espaço da avaliação. A avaliação não deve ter um caráter endóginio, explicamo-nos: é necessário que, realizada uma avaliação, suas constatações sejam imediatamente divulgadas, disseminadas entre diferentes segmentos da sociedade. A avaliação não é apenas para autoconsumo; há um compromisso tácido do avaliador (e das partes contratantes) com a sociedade, a qual deve ser participante de todo o processo, — do planejamento à elaboração dos relatórios. Uma das características fundamentais da avaliação é a sua transparência, que vai ter consequências, sobretudo no seu impacto, como veremos. A divulgação de uma avaliação pode ser um elemento limitativo, se não for feita em tempo oportuno, que permita a utilização dos dados para a solução de problemas vigentes. O fator tempo pode contribuir para a perda da validade da avaliação; seus dados, ainda que, teoricamente, tenham valor científico, assumem um ponto na perspectiva histórica. Uma pesquisa avaliativa deve produzir impacto, mas não o impacto momentâneo da divulgação dos órgãos da mídia. Esse é um impacto fugaz, que satisfaz vaidades pessoais, mas não é consequente. Apesar de difícil de medir e não existir metodologias adequadas, o impacto de uma avaliação se traduz por mudanças no **pensar**, mudanças no **agir**, mudanças nas **atitudes**, enfim, mudanças no **ser**. O impacto está relacionado a mudanças nos seres humanos, nas instituições, nos produtos, nos materiais educacionais. Isso, a nosso ver, é impacto.

Outros problemas poderiam ser considerados e analisados, limitamo-nos, entretanto, a citar alguns deles — viabilidade

política, relação custo-benefício, conflito de interesses, questões éticas, relações humanas, responsabilidade fiscal, validade e fidedignidade instrumental, adequação das análises (qualitativas e quantitativas), **objetividade informativa**, entre outras — porque julgamos que esses problemas possam ser superados, se devidamente gerenciados. A discussão inicial e a presente enumeração permite-nos afirmar que não há avaliação sem problemas, mas isso não determina que a *construção do espaço da avaliação fique restrito, sofra limitações*. O espaço da avaliação é amplo e dá margem à criatividade e à fantasia, como veremos ao discutir a construção do avaliador.

3.4 A função do avaliador

Antes de discutirmos os aspectos relevantes na construção de profissionais de avaliação, procuraremos situar a função do avaliador, que se torna mais clara quando certos aspectos são considerados, como a definição de avaliação, no pensar de Nevo, In: Walberg e Haertel (1990), na sua discussão do papel da avaliação e o envolvimento do avaliador. Se considerarmos a copiosa literatura sobre o assunto poderemos agrupar os modelos de avaliação em três grandes conjuntos: avaliação por objetivos (Tyler); avaliação descritivas e **nonjudgemental** (Stufflebeam; Cronbach) e avaliação que implique julgamento (Scriven; **Joint Committee**, 1981). As definições de avaliação refletem várias e diferentes perspectivas, inclusive em relação ao papel que o avaliador vai desempenhar. Assim, numa avaliação por objetivos, o avaliador é um controlador, um auditor; numa avaliação descritiva, o avaliador é um cientista, um educador que procura informar às várias audiências sobre aspectos relevantes a fim de aumentar a sua compreensão sobre diferentes problemas, e no caso de uma avaliação com vistas a um julgamento, o avaliador assemelha-se a um juiz, um crítico literário ou a um crítico de arte. Stake, juntamente com Dale Kerr, da University of Northern British Columbia, escreveu primoroso artigo sobre o

pesquisador (1994), a partir de comentários sobre sete quadros de Magritte¹³.

Algumas de suas apreciações são transcritas no original e, posteriormente, comentadas em relação ao avaliador, como uma introdução à discussão que faremos sobre a construção do avaliador.

"Research can be designed so that as much as its power to express conclusions is its power to stimulate thinking".

"Like the artist, the researcher is a 'provider of images', images to hang in the mind, to confront images already formed in the complexity of our innate, tacit experience, to draw attention to our expectations and assumptions, to shock us out of our day-to-day complacency".

"Similarly, the researcher is the provocateur of understanding, portraying the common in problematic ways. Our words and images are not the experience or the object they represent — they are as humanoids from another world".

"In research, all representations are misrepresentations; data can but imperfectly disclose their objects. The appearance of precision in our indicators of teaching and schooling is our curse".

"The researcher, like the artist, is a producer of the impossible. His or her perception is colored by tacit knowledge, both personally and culturally constructed. The effort to represent, to explain, is forever incomplete".

"... a reader of a research report comes to understand that the newly reported case is not necessarily typical of cases in general but in certain ways gives insight into the small population of cases he or she already knows".

"Given all the interpretational privilege, good interpretation is cautious, minding the blackness of wishful thinking just beyond the researcher's window".

¹³ René Magritte (1898-1967), pintor belga, influenciado, inicialmente, pelo Cubismo e pelo Futurismo, a partir de 1925 abraça o Surrealismo. Sua obra caracteriza-se pela "ironia, agudeza e por um espírito de inteligente debate". *História Geral da Arte - Dicionário Biográfico de Artistas* — vol. II. Madrid, Ediciones del Prado. 1997.

O trabalho de Stake e Kerr (1994) é extremamente provocativo e apresenta algumas idéias que não podemos desconsiderar, por referirem-se a questões nucleares da avaliação, assim como da pesquisa educacional, considerando que, sem dúvida, no momento atual, ocorrem mudanças em nossas percepções do conhecimento. Há uma revolução epistemológica **in progress** que afeta ambas as perspectivas quantitativa e qualitativa, ainda que muitos não se dêem conta do que está efetivamente ocorrendo. As nossas percepções, quer queiramos ou não, são compreendidas em função de nossas experiências passadas, pessoais, do nosso conhecimento tácito, para empregar a expressão de Stake e Kerr (1994).

A partir dessas reflexões, somos levados a dizer que ao planejarmos uma avaliação pretendemos expressar nossas conclusões ao final do trabalho, avaliamos, assim, com vistas ao estabelecimento de manifestações que possam contribuir para a definição e posterior solução de um problema; contudo, o planejamento de uma avaliação tem um outro objetivo bastante explícito, que é o de promover o pensamento em profundidade sobre o fenômeno avaliado. É evidente que as conclusões a partir dos dados são importantes, mas o realmente relevante é que a avaliação seja um estímulo para o pensamento na procura de conhecimentos que levem à solução dos problemas educacionais.

3.5 Aspectos significativos em um avaliador

Que é o avaliador? É um cientista, é um pesquisador, mas bem considerando o que deveria ser, acreditamos que poderíamos dizer, usando um verdadeiro achado de Stake e Kerr (1994), que o avaliador é um “**provedor de imagens**”, alguém que proporciona elementos essenciais para o espírito daqueles que devem confrontar diferentes imagens resultantes da experiência tácita, imagens que vão despertar a atenção para novas expectativas e suposições de possíveis novas verdades, e são essas imagens que muitas vezes nos chocam, nos trazem para um novo mundo bem diverso do nosso dia-a-dia. Além de cientista e pesquisador, o avaliador, como o artista, deve produzir uma obra que nos leve à construção de novas realidades.

Há alguns anos (1982) preocupamo-nos com o problema do perfil do avaliador educacional¹⁴ e procuramos, na literatura (Milman, 1975, entre outros) e na experiência com outras pessoas e instituições envolvidas com os problemas da avaliação, determinar quais as características necessárias a um avaliador atuante. Tentamos agrupá-las em 6 (seis) conjuntos:

1.0 O avaliador deve ser capaz de:

- 1.1 **especificar** informações necessárias para o desenvolvimento de programas de avaliação;
- 1.2 **localizar, ler e integrar** informações existentes na literatura técnica;
- 1.3 **analisar** avaliações anteriores sobre o assunto de seu interesse e discutir suas implicações;
- 1.4 **definir** com precisão os objetivos gerais e específicos de uma avaliação.

2.0 O avaliador deve ser capaz de:

- 2.1 **examinar**, criticamente, estratégias de avaliação e **selecionar** a adequada para o seu problema;
- 2.2 **formular** questões ou hipóteses a serem respondidas ou verificadas na avaliação;
- 2.3 **desenvolver** planejamento adequado para a coleta de dados que possibilitem realizar 2.2.

3.0 O avaliador deve ser capaz de:

- 3.1 **selecionar** amostra capaz de atender os objetivos de um projeto;
- 3.2 **identificar**, a partir dos objetivos, as variáveis a considerar na avaliação.

4.0 O avaliador deve ser capaz de:

- 4.1 **selecionar** e/ou **construir** os instrumentos necessários à avaliação;
- 4.2 **estabelecer** a validade dos instrumentos e **determinar** o grau de fidedignidade dos resultados;
- 4.3 **dominar** técnicas qualitativas de coleta de dados;

¹⁴ VIANA, H. M. Avaliação educacional - problemas gerais e a formação do avaliador. *Educação e Seleção*, nº 5. Fundação Carlos Chagas, São Paulo, 1982.

- 4.4 **gerenciar** o programa de avaliação e **corrigir** seus desvios;
 - 4.5 **selecionar** as técnicas estatísticas necessárias, se for o caso;
 - 4.6 **definir** os procedimentos de análise em termos de processamento de dados.
- 5.0 O avaliador deverá ser capaz de:
- 5.1 **interpretar** e **estabelecer** conclusões fundamentadas;
 - 5.2 **elaborar** relatórios e **discutir** suas conclusões.
- 6.0 O avaliador deverá ser capaz de:
- 6.1 **estabelecer** relações interpessoais adequadas com a equipe de avaliação;
 - 6.2 **administrar** recursos humanos e materiais ligados à avaliação.

É óbvio que esse é um perfil desejável, mas assintótico, ideal, sendo impossível encontrar todas essas características em uma única pessoa. Lembremo-nos de que há o divino (perfeição) e o humano (limitação), mas todas as qualidades seriam desejáveis, especialmente a fixada em 6.1 — relações interpessoais adequadas, tendo em vista que a avaliação é um trabalho de grupo e o seu êxito depende fundamentalmente de um esforço cooperativo, sob uma liderança segura, forte, porém humana. Acrescentando a tudo isso, diríamos que o avaliador deve ter um bom treinamento na avaliação quantitativa, inicialmente, e, depois, em avaliação qualitativa, conforme a linha seguida por Robert Stake na Universidade de Illinois (Urbana), com um **minor** envolvendo Filosofia (lógica, epistemologia e fenomenologia) e Filosofia da Ciência, a fim de possuir as características idealizadas anteriormente. Pensamos, assim, ter atendido à interrogação que nos propusemos inicialmente: — se essa é uma área com especificidade (— a avaliação —), que aspectos são relevantes na construção dos profissionais que atuam ou atuarão nesse espaço? [5] [6].

3.6 O futuro da avaliação

Falamos do passado e do presente, mas e o futuro da avaliação? O que acontecerá? O que deve acontecer? O que pode acontecer? Sem pretendermos fazer um exercício de futurologia — ciência (?) sem grande credibilidade —, podemos, a partir das tendências atuais, pressentir que, em nível de escola e mesmo de sistema de ensino, a avaliação se concentrará enfaticamente no uso de instrumentos de medida, especialmente para a tomada de decisões administrativas. É possível, entretanto, que, com o treinamento de avaliadores, formas alternativas de avaliação venham a surgir (avaliação individualizada, qualitativa), a médio prazo, especialmente se instituições especializadas se dedicarem à pesquisa e à avaliação.

A segunda interrogação — o que deve acontecer? — acreditamos que, se a idéia de avaliação prosperar e se essa área passar a ter um caráter interdisciplinar, trabalhando em conjunto com outras ciências do homem, a sua produção passará a exigir um trabalho conexo, indispensável ao seu desenvolvimento, ou seja, a avaliação da avaliação, que pode ser expressa simplesmente pela palavra meta-avaliação. Ao lado disso, a prática da avaliação gerará, certamente, a necessidade de estudos conceituais e empíricos sobre a avaliação — estudos descritivos da prática da avaliação; estudos sobre o impacto das avaliações, definindo-se, inicialmente, o entendimento de impacto como procuramos mostrar anteriormente; e estudos empíricos sobre a metodologia da avaliação. Enfim, a avaliação será uma área de pesquisas.

Quando pensamos sobre o que pode acontecer, julgamos, sob risco de sermos temerários, que a escola saberá julgar potencialidades não exploradas da avaliação; que a pesquisa sobre avaliação não será episódica, mas terá uma linha contínua de seguimento, e que não apenas a avaliação mas, também, a própria pesquisa da avaliação serão temas prioritários para os governos e as agências financiadoras de projetos. Tudo isso é extremamente ambicioso, visionário e fantasioso sem dúvida, mas olhar o futuro exige um certo grau de fantasia, e nesse momento, gostaríamos de transcrever parte de um trecho do artigo de

Kenneth E. Boulding¹⁵, citado por Smith (1993) — “a fantasia é parte das metodologias adequadas de todos os campos do conhecimento humano”.

3.7 NOTAS E COMENTÁRIOS

- [1] A avaliação no Brasil ainda tem um longo caminho a percorre, como se infere dos trabalhos que se preocupam em resgatar a sua memória: GATTI, B. A. - Avaliação da Jornada Única em São Paulo. **Estudos em Avaliação Educacional**, nº 5, Fundação Carlos Chagas, São Paulo, 1992; GATTI, B. A.; VIANNA, H. M. e DAVIS, C. - Problemas e Impasses da Avaliação de Projetos e Sistemas Educacionais. **Estudos em Avaliação Educacional**, nº 4. Fundação Carlos Chagas, São Paulo, 1991; GATTI, B. A. - O rendimento escolar em distintos setores da sociedade. **Estudos em Avaliação Educacional**, nº 7. Fundação Carlos Chagas, São Paulo; GATTI, B. A. - Avaliação Educacional no Brasil: experiências, problemas, recomendações. **Estudos em Avaliação Educacional**, nº 10, Fundação Carlos Chagas, São Paulo, 1994; VIANNA, H. M. - Avaliação do Ciclo Básico de Alfabetização em Minas Gerais. **Estudos em Avaliação Educacional**, nº 5, Fundação Carlos Chagas, São Paulo, 1992. VIANNA, H. M. - Avaliando a Avaliação: da prática à pesquisa. **Estudos em Avaliação Educacional**, nº 5, Fundação Carlos Chagas, São Paulo, 1992; VIANNA, H. M. - Avaliação do desempenho em Matemática e Ciências - uma experiência em São Paulo e Fortaleza. **Estudos em Avaliação Educacional**, nº 5, Fundação Carlos Chagas, São Paulo, 1992; VIANNA, H. M. - Evasão, repetência e rendimento escolar - a realidade do sistema educacional brasileiro. **Estudos em Avaliação Educacional**, nº 4, Fundação Carlos Chagas, São Paulo, 1991; VIANNA, H. M. e FRANCO, G. T. - Avaliação do Rendimento dos Alunos do 1º Grau da rede privada: pontos críticos e convergências. **Estudos em Avaliação Educacional**, nº 7, Fundação Carlos Chagas, São Paulo, 1993; VIANNA, H. M. - Avaliação Educacional - uma perspectiva histórica. **Estudos em Avaliação Educacional**, nº 12, Fundação Carlos Chagas, São Paulo, 1995 (parte final relativa ao Brasil); VIANNA, H. M.; ANTUNES, A. L.; SOUZA, M. A. - Desenvolvimento de um programa de avaliação do Sistema Estadual de Ensino - o exemplo de Minas Gerais. **Estudos em Avaliação Educacional**, nº 8, Fundação Carlos Chagas, São Paulo, 1993.
- [2] O destaque dado a modelos que refletem uma posição não-quantitativa leva-nos a considerar algumas idéias de Stake, que apresentou na reunião do **Interamerican Congress of Educational**

¹⁵ BOULDING, K. E. Science: Our common heritage. **Science**, 207 (4433), 1980, 831-36.

Administrators, em Brasília (1984), relacionando pesquisa qualitativa e compreensão. Afirmou Stake que há uma preocupação de muitos em fazer distinção entre pesquisa (avaliação) qualitativa e quantitativa, o que, na sua opinião, é pouco claro. Lembremo-nos que em qualquer estudo qualitativo — etnográfico, naturalista, fenomenológico, hermenêutico ou holístico — há lugar, como acentua Stake (1984), para a enumeração e o reconhecimento de diferenças de quantidade. Por sua vez, quando se realiza um estudo de natureza estatística, por exemplo, um **survey**, ou um experimento que demanda controle, portanto, quando empreendemos uma pesquisa ou avaliação quantitativa, vemo-nos às voltas com descrições e interpretações.

A questão não pode ser colocada simplesmente em termos do quantitativo **versus** qualitativo. É um problema bem mais complexo, por ser de natureza epistemológica. Assim, com justa razão, Stake sugere que se proponha o problema de outra forma e que se dimensione a pesquisa (avaliação) em novos termos **apresentar explicações x promover o conhecimento**. É evidente que as explicações visam a levar ao conhecimento, devendo-se, entretanto, conforme o raciocínio de Stake, baseado, nas idéias de G. H. von Wright¹¹, fazer uma distinção entre **explicação formal e compreensão experiencial**.

O campo da avaliação é grandemente influenciado pela idéia de que a pesquisa (avaliação) deva levar, necessariamente, a **generalizações**, na suposição de que somente assim se estaria criando uma estrutura científica. A idéia do que é **situacional** é eliminada, o contexto é eliminado em busca de grandes relações explicativas (Stake, 1984). É a procura da **explicação** por intermédio de processos quantitativos que leve à **generalização** a partir de um grande número de casos dissimilares, coloca Stake. Temos, assim, a explicação formal.

Não se pode deixar de considerar que o **situacional** é igualmente importante. Uma situação particular, específica, na lógica de Stake, é muitas vezes mais importante do ponto de vista científico (e educacional) do que situações em geral que possam levar à construção de um princípio ou até mesmo de uma teoria. Uma escola, uma criança ou um currículo podem ser objetos de estudos complexos, envolvendo diferentes **backgrounds**. Ambas as abordagens são possíveis no caso, sendo que a abordagem qualitativa vai se interessar pela singularidade do caso, a variedade de percepções. Isso não significa que essa avaliação não se preocupa com a questão da generalização, apenas não é uma preocupação dominante no observador.

A visão do avaliador qualitativo, no dizer de Stake (1984) é **mais holística e menos analítica** em relação a um determinado fenômeno observado. Situações múltiplas são examinadas, mas sem comparações. A preocupação maior do avaliador é a de identificar entre os diversos aspectos observados qual o de maior importância para ser aprendido em

relação ao fenômeno estudado. O importante em avaliação (pesquisa), conforme o pensamento de Stake (1984), não é discutir **generalizações formais ou situações particulares**. O que importa é a qualidade da avaliação (pesquisa), não existindo hoje em dia, pelo menos em países de larga experiência em avaliação, aquele estado de espírito que se predisponha a aceitar um tipo de avaliação e a negar o valor do tipo oposto.

Isso não significa isentar a avaliação qualitativa de alguns problemas, como reconhece o próprio Stake (1984), que aponta deficiências inquestionáveis em muitos casos: 1) excesso de subjetividade das observações; 2) imprecisão de linguagem; 3) dificuldade em descrever o planejamento da avaliação; 4) tamanho volumoso dos relatórios; 5) generalização incorreta dos dados; 6) custos elevados; 7) dilemas éticos na avaliação de sujeitos humanos. Stake (1984) destaca que os problemas existem, é evidente, mas podem ser controlados e até mesmo superados. (1) George Henrik von Wright (1916), filósofo finlandês de origem sueca, interessado em estudos sobre indução e probabilidade, sofreu grande influência de Ludwig Wittgenstein, a quem sucedeu como professor de Filosofia na Universidade de Cambridge (1948). Editou várias obras de Wittgenstein, além de escrever material biográfico e de análise crítica sobre a obra do filósofo austro-africano. Os seus *Philosophical Papers*, em 3 vols., indicam o seu interesse: *Practical Reason* (vol. I, 1983), *Philosophical Logic* (vol. II, 1983) e *Truth, Knowledge and Modality* (vol. III, 1984). Vide *The Oxford Companion to Philosophy*, Edited by T. Honderich, Oxford, Oxford University Press, 1995, p. 903.

- [3] A proposição relacionada a **um espaço possível do conhecimento humano sistematizado** em avaliação educacional pode ser considerada a partir da categorização das obras que subsidiaram o presente trabalho e arroladas na Bibliografia. Procuramos categorizar essa bibliografia segundo a temática dominante:
1. **Avaliação em geral:** House, 1989a, 1989b, 1989c; Nevo, 1989, 1990; Popham, 1975; Rose e Nyre, 1977; Scriven, 1972a, 1972b, 1974; Tyler, 1942, 1978, 1982, 1993; Wahlstrom e Traub, 1972.
 2. **Avaliação — teoria:** Alkin, 1969; Alkin e Daillak, 1990; Cronbach, 1963; Lewy, 1990; Madaus, 1993; Scriven, 1974; Wolf, 1990.
 3. **Avaliação e objetivos:** Atkin, 1974, 1978; De Landshere, 1990; Eisner, 1967, 1969; Eraut, 1990; Popham, 1978, 1990; Stake, 1970; Tyler, 1974.
 4. **Avaliação e currículo:** Alkin, 1990; Hastingser, 1972; Kemmins, 1989; Payne, 1974, 1975; Sanders, 1990; Schubert, 1990; Sosniak, 1990; Taylor e Cowley, 1972; Walberg e Haertel, 1990; Walken, 1989; Wiseman e Pidgeon, 1970.
 5. **Avaliação e Modelos:** Airasian, 1993; Alkin, 1972; Alkin e Ellet, 1990; Bertam e Childers, 1974; Borich, 1990; Glaser, 1970; Glass, 1972; Hammond, 1973; Katz e Morgan, 1974; Luft et al., 1974; Metfessel e

Michael, 1967; Provus, 1969, 1972, 1973; Scriven, 1973; Stecher, 1990; Steinmetz, 1976; Taylor e Maguire, 1966; Walberg e Haertel, 1990; Wolf, 1990; Worn, 1982; Wright e Hess, 1974.

6. Avaliação de Programas e Produtos: Borich, 1974; Coleman, 1971; Cronbach, 1982; Reynolds e Walberg, 1990; Rossi, 1971; Scriven, 1971, 1974; Riding e Stufflebeam, 1975; Suchman, 1971; Walker, 1974; Welch e Walberg, 1974.

7. Avaliação e Pesquisa: Caro, 1971, 1982; Ross e Cronbach, 1976; Schubert, 1990; Scriven, 1989; Strickland, 1990; Welty, 1972; Willems, 1972; Worthen, 1990.

8. Avaliação e Tomada de Decisão: Atkin, 1978; Scheers, 1990; Stufflebeam, 1969, 1971, 1974; Stufflebeam et al., 1971; Stufflebeam e Webster, 1980.

9. Avaliação Qualitativa: Dorr-Bremme, 1990; Eisner, 1978, 1978a, 1978b, 1978c, 1979, 1993; Eraut, 1990; Cuba, 1969, 1975; Cuba e Lincoln, 1981, 1989; Hamilton, 1978; Lincoln e Cuba, 1985; Parlett, 1990; Stake, 1973, 1975, 1978, 1982a, 1982b, 1982c, 1982d, 1983a, 1983b, 1984, 1990, 1994; Stake e Pearsol, 1981; Stake e Trumbul, 1984; Willis, 1990.

10. Avaliação e Problemas: Atkin, 1963, 1967; Choppin, 1990; Stake, 1973, 1974; Straw e Cook, 1990.

Fica evidenciado que a avaliação possui um **corpus** teórico, que, entretanto, não constitui uma teoria geral única, tendo em vista que as posições refletem diferentes abordagens epistemológicas, cada uma com seus fundamentos lógicos, seus valores e objetivos. Acontece com a avaliação o que ocorre com outras áreas, como, por exemplo, a geometria euclidiana e não-euclidiana, não se podendo dizer que a geometria não seja um conhecimento sistematizado, ou com a física (Newton, Einstein, dos quanta), que são visões diferentes, mas formam um sistema de conhecimentos. A avaliação é uma área nova, articulada a diversas outras áreas de conhecimento, e reflete diferentes concepções de mundo, mas nem por isso deixa de ser um conjunto organizado de conhecimentos, com várias abordagens teóricas.

- [4] O conhecimento na área da avaliação é bastante amplo, como se pode ver se consultarmos a encyclopédia de Anderson (1981) e a de Walberg e Haertel (1990), ambas citadas na bibliografia, e o monumental trabalho coordenado por Torsten e Postlethwaite (1985), mencionado no texto sobre Scriven, no rodapé da nota [1]. A nova edição da obra coordenada por Keeves, J. P. **Educational Research, Methodology and Measurement — an International Handbook**. Second Edition. Pergamon. New York/London, 1997, deixa clara a complexidade da avaliação, que passou a demandar especializações em campos mais restritos:

aprendizagem, currículo, professor, instituição etc, face à impossibilidade do domínio do conhecimento acumulado em cada uma das áreas.

- [5] A informática faz parte da nossa cultura e o computador passou a ser instrumento de trabalho indispensável em todos os campos, inclusive na avaliação. Essa situação influencia, naturalmente, a formação do avaliador. Em alguns aspectos, deve ser eminentemente conceitual e de compreensão da estrutura lógica de procedimentos e de como analisar os elementos fornecidos pelo processamento dos dados. A existência de **softwares** facilita a parte operacional, mas não dispensa a formação substantiva do avaliador. Exemplificando, o TESTFACT, desenvolvido por D. T. Wilson, R. Wood e R. D. Gibbons, possibilita uma análise de itens segundo a perspectiva clássica, e, ao mesmo tempo, permite a análise fatorial dos itens, usando matrizes de correlação tetracórica, além de outros procedimentos próprios da análise fatorial. O programa liberta o avaliador da exausta tarefa da parte computacional, mas não elimina a necessidade de conhecer, no caso específico, a teoria clássica da análise de itens e a formulação teórica da análise fatorial. A idéia fundamental é que o computador e seus *softs* sejam indispensáveis, mas como auxiliares do avaliador e não elidem a necessidade de uma formação teórica fundamentada.
- [6] Anteriormente, ao discutirmos o momento presente da avaliação no Brasil, fizemos menção à experiência do **SAEB** em 1995, que representou um grande avanço em termos metodológicos, com a adoção da **Teoria da Resposta ao Item** (TRI) para procedimentos ligados à calibragem dos itens, ao **equating** dos instrumentos e para a análise dos resultados. A chamada TRI é realmente importante porque, entre outros aspectos, resolve problemas relacionados à construção de testes e à realização de avaliações em larga escala⁽²⁾. A Teoria da Resposta ao Item não invalida os métodos clássicos de análise, mas por ser um instrumento com maiores possibilidades e oferecer resultados mais precisos o seu uso está se generalizando, sobretudo nos Estados Unidos, por influência do **Educational Testing Service** (ETS). A sua matemática é extremamente complexa e seus fundamentos estão nos capítulos escritos por Birnbaum (cap. 17 e segs.) na obra fundamental de Lord e Novick⁽³⁾, a que poucos têm acesso pelo fato de exigir do leitor sofisticação matemática e estatística. Vários **softwares** foram criados para atender às exigências da nova metodologia — BILOG-W (R. J. Mislevy e R. D. Bock); BILOG-MG (M. F. Zimowski; E. Muraki; R. J. Milesvy; R. D. Bock); MULTILOG (D. Thiessen); PARSCALE (E. Muraki; R. D. Bock), entre outros. Os programas **por si** não resolvem os problemas se o avaliador não tiver uma formação básica que sirva de suporte às suas necessidades teóricas. A idéia ou a tese que defenderemos é a de que, sem uma base teórica que fundamente a prática da avaliação, o uso de metodologias complexas

carecem de sentido, sobretudo quando usadas na disseminação dos resultados a audiências que não têm condições (e nem necessitam) de perceber as implicações de modelos matemáticos usados na solução de problemas complexos no tratamento dos dados. A forma como os resultados são divulgados influenciam consideravelmente o grau de impacto de uma avaliação.

- (2) Mislevy, R. J. Foundations of a New Test Theory. In, Frederiksen, N.; Mislevy, R. J. e Bejar, I. I. *Test Theory for a New Generation of Tests*. Lawrence Erlbaum Associates Publishers. Hillsdale, New Jersey, 1993.
- (3) Lord, F. M. e Novick, M. R. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley, 1968.

BIBLIOGRAFIA

- AIRASIAN, P. W. Societal Experimentation. In: MADAUS, G. F. et al. (Ed.) *Evaluation Models-viewpoints on educational and human services evaluation*. Boston, Kluwer-Nijhoff Publishing, 1993.
- ALKIN, M. C. Evaluation Theory Development. *Evaluation Comment*, 2, 2-7, 1969.
- ALKIN, M. C. Towards an Evaluation Model: a Systems Approach. In: TAYLOR, P. A. e COWLEY, D. M. *Readings in Curriculum Evaluation*. W. M. C. Brown Co. Publishing. Dubuque, Iowa, 1972.
- ALKIN, M. C. Curriculum Evaluation Models. In: WALBERG, H. J. et al. *The International Encyclopedia of Educational Evaluation*. Oxford, Pergamon Press. 1990.
- ALKIN, M. C. e DAILLAK, R. H. Impact of Evaluation Studies. In: WALBERG, H. J. et al. *The International Encyclopedia of Educational Evaluation*. Oxford, Pergamon Press. 1990.
- ALKIN, M. C. e ELLETT Jr., F. S. Development of Evaluation Models. In: WALBERG, H. L. et al. *The International Encyclopedia of Educational Evaluation*. Oxford. Pergamon Press. 1990.
- ATKIN, J. M. Some evaluation problems in a course content improvement project. *Journal of Research in Science Teaching*, 1, p. 129-132, 1963.
- ATKIN, J. M. Research Styles in Science Education. *Journal of Research in Science Teaching*, 5, p. 338, 1967.

- ATKIN, J. M. Behavioral Objectives in Curriculum Design: a cautionary note. In: PAYNE, D. A. **Curriculum Evaluation: Commentaries on Purpose, Process, Product.** Lexington, Mass. D. C. Heath and Company, 1974.
- BERTRAM, C. L. e CHILDERS, R. D. A multistage model for evaluating educational products. In: BORING, G. D. **Evaluating Educational Programs and Products.** Englewood Cliffs, New Jersey. Educational Technology Publications. 1974.
- BORICH, G. D. **Evaluating Educational Programs and Products.** Englewood Cliffs, New Jersey. Educational Technology Publications. 1974.
- BORICH, G. D. Decision-oriented evaluation. In: WALBERG, H. L. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.
- CARO, F. G. **Readings in Evaluation Research.** New York, Russell Sage Foundation. 1971.
- CARO, F. G. Pesquisa avaliativa: — uma visão panorâmica. In: GOLDBERG, M. A. A. e SOUZA, C. P. (Org.) **Avaliação de Programas Educacionais: vicissitudes, controvérsias, desafios.** EPU. São Paulo. 1982.
- CHOPPIN, B. H. Evaluation, Assessment and Measurement. In: WALBERG, H. L et al. **The International Encyclopedia of Educational Evaluation.** Oxford, Pergamon Press. 1990.
- COLEMAN, J. S. Evaluating Educational Programs. In: CARO, F. G. **Readings in Evaluation Research.** New York, Russell Sage Foundation. 1971.
- CRONBACH, L. J. Course Improvement through Evaluation. **Teachers College Record**, 64, p. 672-83. 1963.
- CRONBACH, L. J. **Designing Evaluation of Educational and Social Programs.** San Francisco, Jossey-Bass. 1982.
- DE LANDSHEERE, D. V. Taxonomies of Educational Objectives. In: WALBERG, H. J. et al. **The International Encyclopedia of Educational Evaluation.** Oxford, Pergamon Press. 1990.
- DORR-BREMME, D. W. Naturalistic Evaluation. In: WALBERG, H. L. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.

- EISNER, E. W. Educational Objectives: help or hindrance? **School Review**, 75, nº 3. 1967.
- EISNER, E. W. Instructional expressive educational objectives: their formulation and use in curriculum. In: W. J. POPHAM (Ed.) **Instructional Objectives**. AERA Monograph Series on Curriculum Evaluation, nº 3. Chicago: Rand McNally. 1969.
- EISNER, E.W. Establishing a direction. In: HAMILTON, D. et al. **Beyond the Numbers Game**. MacMillan Education Ltd. London. 1978a.
- EISNER, E. W. Implications for Evaluation. In: HAMILTON, D. et al. **Beyond the Numbers Game**. MacMillan Education Ltd. London. 1978b.
- EISNER, E. W. Thick description. In: HAMILTON, D. L. et al. **Beyond the Numbers Game**. MacMillan Education Ltd. London. 1978c.
- EISNER, E. W. The use of qualitative forms of evaluation for improving educational practice. **Educational Evaluation and Policy Analysis**. 1(6), 11-19. 1979.
- EISNER, E. W. Educational Connoisseurship and Criticism: their Form and Functions. In: Educational Evaluation. **Journal of Aesthetic Education**, nº 3-4, July-October, 1976, p. 135-50. In: MADAUS, G. F. (Ed.) **Evaluation Models**. Boston. Kluwer-Nijhoff Publishing, 1993.
- ERAUT, M. R. Educational Objectives. In: WALBERG, H. J. et al. **The International Encyclopedia of Educational Evaluation**. Oxford. Pergamon Press. 1990.
- ERAUT, M. R. Intrinsic Evaluation. In: WALBERG, H. J. et al. **The International Encyclopedia of Educational Evaluation**. Oxford. Pergamon Press. 1990.
- GLASER, R. Evaluation of instruction and changing educational models. In: M. C. WITTROCK e D. E. WILEY (Ed.) **The Evaluation of Instruction: Issues and Problems**. New York. Holt, Rinehart and Winston, 1970.
- GLASS, G. V. e WORTHEN, B. R. Evaluation and research: similarities and differences. **Curriculum Theory Network**, Fall. 1971.

- GLASS, G. V. Two Generations of Evaluation Models. In: TAYLOR, P. A. e COWLEY, D. M. **Readings in Curriculum Evaluation.** W. M. C. Brown Co. Publishing, Dubuque, Iowa. 1972.
- GOLDBERG, M. A. A. e SOUZA, C. P. (Org.) **Avaliação de Programas Educacionais: vicissitudes, controvérsias, desafios.** EPU. São Paulo. 1982.
- GUBA, E. G. The failure of educational evaluation. **Educational Technology**, 9, 1969, p. 29-38. 1969.
- GUBA, E. G. Problems in utilizing the results of evaluation. **Journal of Research and Development in Education.**, 8, 42-54. 1975.
- GUBA, E. G. e LINCOLN, Y. S. **Effective Evaluation.** San Francisco: Jossey-Bass. 1981.
- GUBA, E. G. e LINCOLN, Y. S. **Fourth Generation Evaluation.** Newbury Park, California. Sage Publications. 1989.
- GUBA, E. G. e LINCOLN, Y. S. Epistemological and Methodological Bases of Naturalist Inquiry. **Educational Communications and Technology Journal**, nº 4, 30, Winter, 1982. In: MADAUS, G. F. et al. **Evaluation Models.** Boston. Kluwer-Nijhoff Publishing. 1993.
- HAMILTON, D. et al. **Beyond the Numbers Game: a reader in educational evaluation.** London: MacMillan Education. 1978.
- HAMMOND, R. L. Evaluation at the local level. In: WORTHEN, B. R. e SANDERS, J. R. **Educational Evaluation: Theory and Practice.** Worthington. Ohio. Charles A. Jones Publishing Co. 1973.
- HASTINGS, J. T. Evaluating Change. In: TAYLOR, P. A. e COWLEY, D. M. **Readings in Curriculum Evaluation.** W. M. C. Brown Co. Publishing, Dubuque, Iowa. 1972.
- HOUSE, E. R. Assumptions underlying evaluation models. **Educational Researcher**, March, 1978, p. 4-12. In: MADAUS, F. G. et al. (Ed.) **Evaluation Models.** Boston. Kluwer-Nijhoff Publishing. 1993.
- HOUSE, E. R. (Ed.) **New Directions in Educational Evaluation.** London/Philadelphia. The Falmer Press. 1989.

- HOUSE, E. R. Evaluation and Legitimacy. In: HOUSE, E. R. **New Directions in Educational Evaluation**. London/Philadelphia. The Falmer Press. 1989.
- HOUSE, E. R. How we think about evaluation. In: HOUSE, E. R. **New Directions in Educational Evaluation**. London/Philadelphia. The Falmer Press. 1989.
- KATZ, D. S. e MORGAN, R. L. A holistic strategy for the formative evaluation of educational programs. In: BORICH, G. D. **Evaluating Educational Programs and Products**. Englewood Cliffs, N. J. Educational Technology Publications. 1974.
- KEMMIS, S. Seven Principles for Programme Evaluation in Curriculum Development and Innovation. In: HOUSE, E. R. **New Directions in Educational Evaluation**. London/Philadelphia. The Falmer Press. 1989.
- LEWY, A. Formative and Summative Evaluation. In: WALBERG, H. L. et al. **The International Encyclopedia of Educational Evaluation**. Oxford. Pergamon Press. 1990.
- LINCOLN, Y. S. e GUBA, E. G. **Naturalistic Inquiry**. Newbury Park, California. Sage Publications. 1985.
- LÜDKE, M. Discussão do trabalho de R. E. Stake: Estudos de caso em pesquisa e avaliação educacional. **Educação e Seleção**, jan/jun, nº 7. São Paulo. Fundação Carlos Chagas. 1983.
- LUFT, M. et al. A quality assurance model for process evaluation. In: BORICH, G. D. **Evaluating Educational Programs and Products**. Englewood Cliffs. New Jersey. Educational Technology Publications. 1974.
- MADAUS, G. F. et al. **Evaluation Models - Viewpoints on Educational and Human Services Evaluation**. Boston. Kluwer-Nijhoff Publishing, 1993.
- MADAUS, G. F. The Clarification Hearing: A Personal View of the Process. **Educational Researcher**. January 1982, p. 4-11. In: MADAUS, G.F. (Ed.) **Evaluation Models**. Boston. Kluwer-Nijhoff Publishing, 1993.

- METFESSEL, N. S. e MICHAEL, W. B. A paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs. **Educational and Psychological Measurement**, 27, 1967, p. 931-43.
- MILMAN, J. Selecting educational researchers and evaluators. **ERIC Clearinghouse on tests, measurement and evaluation. ETS**. Princeton. New Jersey, 1975.
- NEVO, D. Conceptualization of Educational Evaluation: An Analytical Review of the Literature. In: HOUSE, E. R. **New Directions in Educational Evaluation**. London/Philadelphia. The Falmer Press. 1989.
- NEVO, D. Role of Evaluation. In: WALBERG, H. J. et al. **The International Encyclopedia of Educational Evaluation**. Oxford. Pergamon Press. 1990.
- NORRIS, N. **Understanding Educational Evaluation**. Kogan Page. London. 1993. Published in Association with CARE — School of Education, University of East Anglia.
- PARLETT, M. R. Illuminative Evaluation. In: WALBERG, H. L. et al. **The International Encyclopedia of Educational Evaluation**. Oxford. Pergamon Press. 1990.
- PARLETT, M. e HAMILTON, D. Evaluation as illumination: a new approach to the study of innovative programs. In: GLASS, G.V. (Ed.) **Evaluation studies review annual**. Vol. 1. Beverly Hills, California. Sage. 1976.
- PAYNE, D. A. (Ed.) **Curriculum Evaluation: Commentaries on Purpose, Process, Product**. Lexington, Mass. D.C. Heath and Company. 1974.
- PAYNE, D. A. Toward a characterization of curriculum evaluation. In: Payne, D. A. (Ed.) **Curriculum Evaluation**. Prentice-Hall. Englewood Cliffs. New Jersey, 1975.
- POPHAM, W. J. **Educational Evaluation**. Englewood Cliffs, New Jersey. Prentice-Hall Inc. 1975.
- POPHAM, W. J. Must all objectives be behavioural? In: HAMILTON et al. **Beyond the Numbers Game**. MacMillan Ltd. London. 1978.

- POPHAM, W. J. A Twenty-year Perspective on Educational Objectives. In: WALBERG, H. J. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.
- PROVUS, M. M. Evaluation of ongoing programs in the public school system. **National Society for the Study of Evaluation - NSSE 68th Yearbook**, Part II, 1969, p. 242-283.
- PROVUS, M. M. The Discrepancy Evaluation Model. In: TAYLOR, P. A. e COWLEY, D. M. **Readings in Curriculum Evaluation.** W. M. C. Brown Co. Publishing. Dubuque, Iowa. 1972.
- PROVUS, M. M. **Discrepancy Evaluation.** Berkeley, California. McCuthan. 1973.
- REYNOLDS, A. J. e WALBERG, H. J. Program Theory in Evaluation. In: WALBERG, H.L. et al. **The International Encyclopedia of Educational Evaluation.** Oxford Pergamon Press. 1990.
- RIDINGS, J. M. e STUFFLEBEAM, D. L. The Project to Develop Standards for Educational Evaluation: its past and future. **Studies in Educational Evaluation**, vol. 7. Pergamon Press Ltd. Great Britain. 1981.
- ROSE, C. e NYRE, G. F. **The Practice of Evaluation.** ERIC — Clearinghouse on Tests, Measurement and Evaluation. Educational Testing Service. Princeton. 1977.
- ROSS, L. e CRONBACH, L. J. (Ed.) Handbook of evaluation research: Essay review by a task force of the Stanford Evaluation Consortium. **Educational Researcher**, 5, 10, p. 9-10. 1976.
- ROSSI, P. H. Evaluating Educational Programs. In: CARO, F. G. **Readings in Evaluation Research.** New York. Russell Sage Foundation. 1971.
- SANDERS, J. R. Curriculum Evaluation Models. In: WALBERG, H. J. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.
- SCHEERENS, J. Beyond decision-oriented evaluation. In: WALBERG, H. L. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.

- SCHUBERT, W.H. Curriculum Validation. In: WALBERG,H.J. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.
- SCRIVEN, M. **The Methodology of Evaluation.** Perspectives of Curriculum Evaluation. AERA. Monograph 1, Rand McNally and Co., Chicago. 1967.
- SCRIVEN, M. Evaluating Educational Programs. In: CARO, F. G. **Readings in Evaluation Research.** New York. Russell Sage Foundation. 1971.
- SCRIVEN,M. Pros and Cons about Goal-free Evaluation. **Evaluation Comment**, 3 (4), p.1-7. 1972a.
- SCRIVEN, M. An Introduction to Meta-Evaluation. In: TAYLOR, P. A. e COWLEY, D. M. **Readings in Curriculum Evaluation.** W. M. C. Brown Co. Publishing. Dubuque, Iowa. 1972b.
- SCRIVEN, M. Goal-Free Evaluation. In: HOUSE, E. R. (Ed.) **School Evaluation — The Politics and Process.** Berkeley, California. McCutchan Publishing Corporation. 1973.
- SCRIVEN, M. The Concept of Evaluation. In: APPLE, M. W. et al. (Ed.) **Education Evaluation: — Analysis and Responsibility.** Berkely, California. McCutchan Publishing Corporation. 1974.
- SCRIVEN, M. Standards for the evaluation of educational programs and products. In: G. D. BORICH (Ed.) **Evaluating educational programs and products.** Englewood Cliffs, N. J. Educational Technology Publications. 1974.
- SCRIVEN, M. Evaluation perspectives and procedures. In: W. J. POPHAM (Ed.) **Evaluation in Education.** Berkeley, California. McCutchan.1974.
- SCRIVEN, M. Evaluation as a Paradigm for Educational Research. In: E. R. HOUSE. **New Directions in Educational Evaluation.** London/Philadelphia. The Falmer Press. 1989.
- SMITH, N. L. The Progress of Educational Evaluation: — Rounding the First Bends in the River. Proceedings of 1980 Minnesota Evaluation Conference on Educational Evaluation: Recent Progress, Future Needs. In: MADAUS, G. F. et al. (Ed.) **Evaluation Models.** Boston. Kluwer-Nijhoff Publishing. 1993.

- SOSNIAK, L. A. Curriculum Feasibility Studies. In: WALBERG, H. J. et al. **The International Encyclopedia of Educational Evaluation**. Oxford. Pergamon Press. 1990.
- STAKE, R. E. The countenance of educational evaluation. **Teachers College Record**, 68, nº 7, p.523-540. 1967.
- STAKE, R. E. Objectives, priorities, and other judgment data. **Review of Educational Research**. 40 (2), p. 181-212. 1970
- STAKE, R. E. Measuring what learners learn (with special look at performance contracting). In: E. HOUSE (Ed.) **School Evaluation**. Berkeley, California, McCutchan. 1973.
- STAKE, R. E. Program Evaluation, Particularly Responsive Evaluation. (paper) Institute of Education at Göteborg University. 1973.
- STAKE, R. E. Language, Rationality, and Assessment. In: PAYNE, D. (Ed.) **Curriculum Evaluation: Commentaries on Purpose, Process, Product**. Lexington, Mass. D. C. Heath and Company, 1974.
- STAKE, R. E. (Ed.) To Evaluate an Arts Program. In: **Evaluating the Arts in Education. A Responsive Approach**. Columbus, Ohio. Merrill, 1975.
- STAKE, R. E. The Case Study Method in Social Inquiry. **Educational Researcher**, February, 7, 1978, p.5-8. In: MADAUS, G. F.(Ed.) **Evaluation Models**. Boston, Kluwer-Nijhoff Publishing. 1993.
- STAKE, R. E. Novos Métodos para Avaliação de Programas Educacionais. In: GOLDBERG, M. A. A. e SOUZA, C. P. (Org.) **Avaliação de Programas Educacionais:- vicissitudes, controvérsias, desafios**. EPU. São Paulo.1982a.
- STAKE, R. E. Uma Subjetividade Necessária em Pesquisa Educacional. In: GOLDBERG, M. A. A. e SOUZA, C. P. (Org.) **Avaliação de Programas Educacionais: vicissitudes, controvérsias, desafios**. EPU. São Paulo.1982b.
- STAKE, R. E. Conversando sobre Avaliação. In: GOLDBERG, M. A. A. e SOUZA, C. P. (Org.) **Avaliação de Programas Educacionais: vicissitudes, controvérsias, desafios**. EPU. São Paulo.1982c.

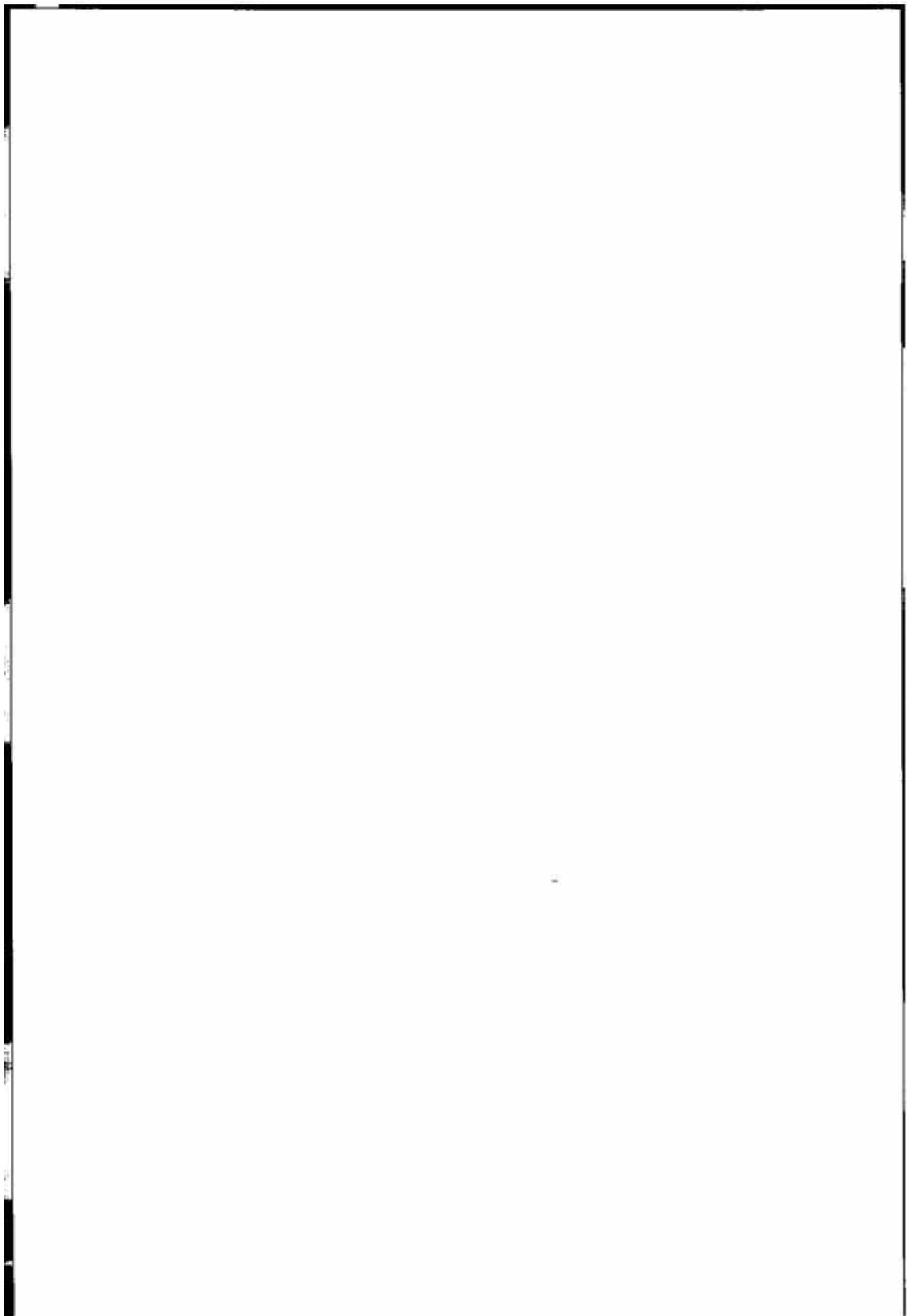
- STAKE, R. E. Como os Avaliadores Responderam a Sete Questões Essenciais. In: GOLDBERG, M. A. A. e SOUZA, C. P. (Org.) **Avaliação de Programas Educacionais: vicissitudes, controvérsias, desafios.** EPU. São Paulo. 1982d.
- STAKE, R. E. Pesquisa qualitativa/naturalista — problemas epistemológicos. **Educação e Seleção**, jan/jun., nº 7. 1983a. Trad. de H. M. Vianna. São Paulo. Fundação Carlos Chagas.
- STAKE, R. E. Estudos de caso em pesquisa e avaliação educacional. **Educação e Seleção**, jan/jun., nº 7, 1983b. Trad. de H. M. Vianna. São Paulo. Fundação Carlos Chagas.
- STAKE, R. E. **Qualitative Inquiry to Promote Understanding.** (paper) **Interamerican Congress of Educational Administrator.** Brasília, 1984.
- STAKE, R. E. An Evolutionary View of Program Improvement. In: HOUSE, E. R. **New Directions in Educational Evaluation.** London/Philadelphia. The Falmer Press. 1989.
- STAKE, R. E. Responsive Evaluation. In: WALBERG, H.J. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.
- STAKE, R. E. Retrospective on "The Countenance of Educational Evaluation". (paper). Version of 27/09/90.
- STAKE, R. E. Program Evaluation, Particularly Responsive Evaluation. (Paper) Conference on New Trends in Evaluation, Göteborg, Suécia, 1973. In: MADAUS, G. F. et al. (Ed.) **Evaluation Models.** Boston. Kluwer-Nijhoff Publishing. 1993.
- STAKE, R. E. e KERR, D. **René Magritte, Constructivism and the Researcher as Interpreter.** American Educational Research Association. AERA. 1994
- STAKE, R. E. Case Studies. In: DENZIN, N. e LINCOLN, Y. **Handbook of Qualitative Research.** Sage Publications. Thousand Oaks, California, 1994.
- STAKE, R. E. e PEARSOL, J. A . Evaluating Responsively. In: BRANDT, R. S. (Ed.) **Applied Strategies for Curriculum Evaluation.** ASCD. 1981.

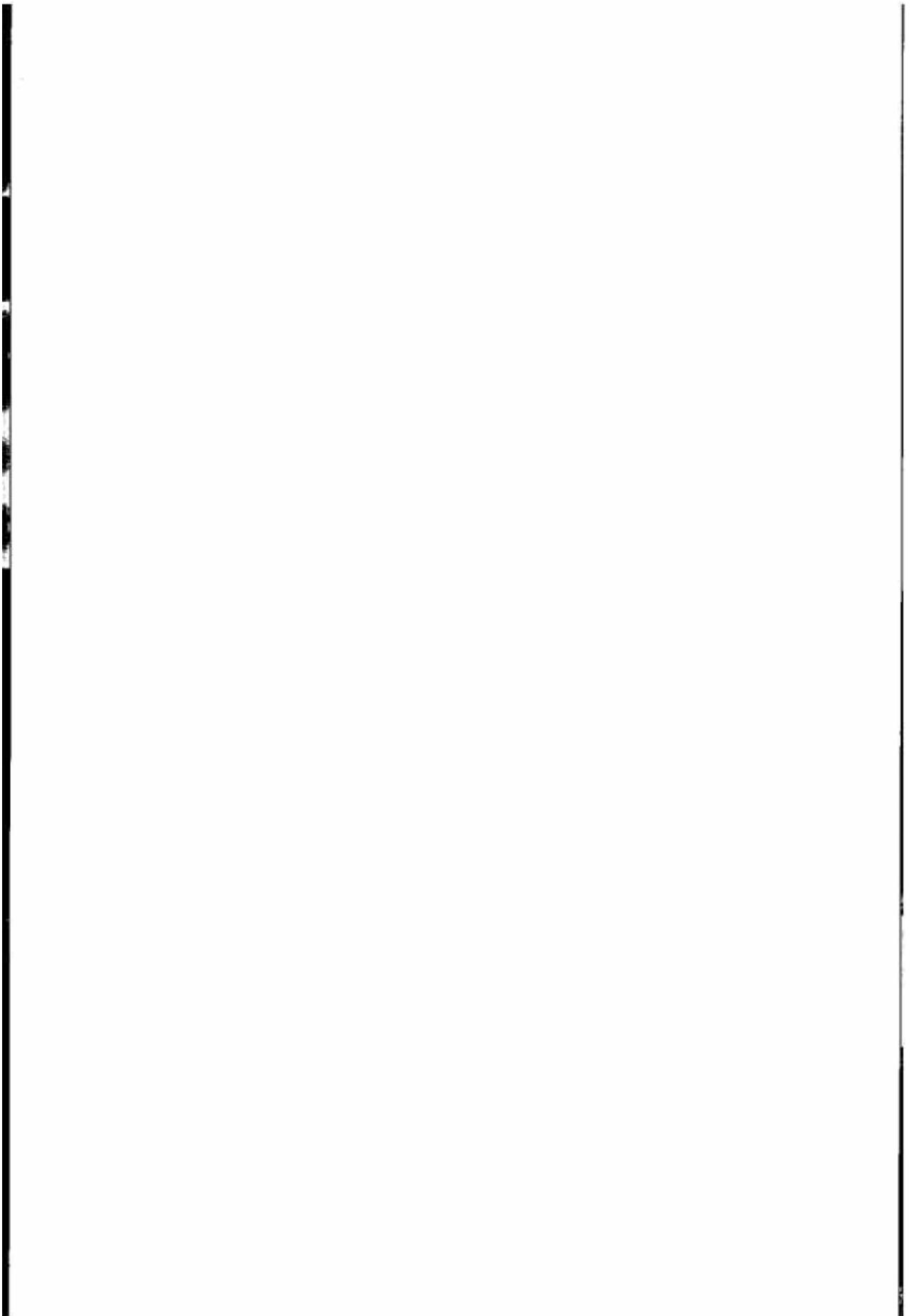
- STAKE, R. E. e TRUMBULL, D. B. Naturalistic Generalization. **Review Journal of Philosophy and Social Science**. vol. VII, n^os 1 e 3, India, 1982.
- STANDARDS for EVALUATIONS of EDUCATIONAL PROGRAMS, PROJECTS, and MATERIALS.** The Joint Committee on Standards for Educational Evaluation. New York. McGraw-Hill Book Co., 1981
- STECHER, B. Goal-free evaluation. In: WALBERG, H. L. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990
- STEINMETZ, A. The Discrepancy Evaluation Model. Measurement in Education. n^o 1, 7, 1976, p. 1-7; n^o 2, 7, 1976, p. 1-6. In: MADAUS, G.F. (Ed.) **Evaluation Models.** Boston, Kluwer-Nijhoff Publishing. 1993.
- STRAW, R. B. e COOK, T. D. Meta-evaluation. In: WALBERG, H.L. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.
- STRICKLAND, K. Eight-year Study. In: WALBERG, H. J. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.
- STUFFLEBEAM, D. L. Evaluation as enlightenment for decision making. In: BEATTY, W. H. (Ed.) **Improving Educational Assessment and an Inventory of Measures of Affective Behavior.** Washington, D.C. NEA. 1969.
- STUFFLEBEAM, D. L. The relevance of the CIPP Evaluation Model for Educational Accountability. **Journal of Research and Development in Education**, 5 (1), p. 19-25. 1971.
- STUFFLEBEAM, D. L. Alternative Approaches to Educational Evaluation: a self-study guide for educators. In: POPHAM, W. J. (Ed.) **Evaluation in Education.** Berkeley, California, McCutchan Publ. Corporation. 1974.
- STUFFLEBEAM, D. L. The CIPP Model for Program Evaluation. In: MADAUS, F.G. et al. (Ed.) **Evaluating Models — viewpoints on Educational and Human Services Evaluation.** Boston. Kluwer-Nijhoff Publishing. 1993.

- STUFFLEBEAM, D. L. e MADAUS, G. F. The Standards for Evaluation of Educational Programs, Projects, and Materials: a description and summary. In: MADAUS, G. F. et al. **Evaluation Models**. Boston. Kluwer-Nijhoff Publishing. 1993.
- STUFFLEBEAM, D. L. et al. An Introduction to the PDK book: **Educational Evaluation and Decision-Making**. Itasca, Ill. F.E. Peacock Publishers, Inc. 1971.
- STUFFLBEAM, D. L. e WEBSTER, J. An Analysis of Alternative Approaches to Evaluation. **Educational Evaluation and Policy Analysis**, 2 (3), p. 5-19. 1980.
- SUCHMAN, E. A. Evaluating Educational Programs. In: CARO, F. G. **Readings in Evaluation Research**. New York. Russell Sage Foundation. 1971.
- TAYLOR, P. A. e COWLEY, D. M. **Readings in Curriculum Evaluation**. W.M.C. Brown Co. Publishing. Dubuque,Iowa. 1972.
- TAYLOR, P. A. e COWLEY, D. M. New Dimensions of Evaluation. In: TAYLOR, P. A. e COWLEY, D. M. **Readings in Curriculum Evaluation**. W. M. C. Brown Co. Publishing. Dubuque, Iowa, 1972.
- TAYLOR, P. A. e MAGUIRE, T. O. A Theoretical Evaluation Model. **Manitoba Journal of Educational Research**. June, 1966.
- TYLER, R. W. General Statement on Evaluation. **Journal of Educational Research**, 35, p. 492-501. 1942.
- TYLER, R. W. Considerations in Selecting Objectives. In: PAYNE, D. (Ed.)**Curriculum Evaluation: Commentaries on Purpose, Process, Product**. Lexington, Mass. D.C. Heath and Company.1974.
- TYLER, R. W. Avaliando Experiências de Aprendizagem. In: GOLDBERG, M. A. A. e SOUZA, C. P. (Org.) **Avaliação de Programas Educacionais: vicissitudes, controvérsias, desafios**. EPU. São Paulo, 1982.
- TYLER, R. W. Evaluating Learning Experiences. In: HAMILTON, D. **Beyond the Numbers Game**. Mac Millan Ltd. London. 1978.

- TYLER, R. W. A Rationale for Program Evaluation. In: MADAUS, G. F. et al. (Ed.) **Evaluation Models. Viewpoints on Educational and Human Services Evaluation.** Boston. Kluwer-Nijhoff Publishing. 1993.
- WAHLSTROM, M. M. e TRAUB, R. E. Evaluation du rendement scolaire. Documentation et information pédagogique. **Bulletin International d'Éducation.** n° 184. UNESCO, Paris. 1972.
- WALBERG, H. J. e HAERTEL, G. D. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.
- WALBERG, H. J. e HAERTEL, G. D. Evaluation Approaches and Strategies-Introduction. In: WALBERG, H. L. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.
- WALBERG, H. J. e HAERTEL, G. D. Curriculum Evaluation. In: WALBERG, H. J. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.
- WALKEN, R. Three Good Reasons for Not Doing Case Studies in Curriculum Research. In: HOUSE, E. R. **New Directions in Educational Evaluation.** London/Philadelphia. The Falmer Press. 1989.
- WALKER, J. P. A Quality Assurance Model for Educational Research and Development. In: BORICH, G. D. **Evaluating Educational Programs and Products.** Englewood Cliffs, New Jersey. Educational Technology Publications. 1974.
- WELCH, W. W. e WALBERG, H. J. A Course Evaluation. In: WALBERG, H. J., (Ed.) **Evaluating Educational Performance.** Berkeley, California. McCutchan Publishing Corporation. 1974.
- WELTY, G. Evaluative Research and Research Design. In: TAYLOR, P. A. e COWLEY, D. M. **Readings in Curriculum Evaluation.** W. M. C. Brown Co. Publishing. Dubuque, Iowa. 1972.

- WILLEMS, E. P. Implications of Viewing Educational Evaluation as Research in the Behavioral Sciences. In: TAYLOR, P. A. e COWLEY, D. M. **Readings in Curriculum Evaluation.** W.M.C. Brown Co. Publishing. Dubuque, Iowa. 1972.
- WISEMAN, S. e PIDGEON, P. **Curriculum Evaluation.** National Foundation for Educational Research (NFER) in England and Wales. London. 1970.
- WOLF, R. M. A Framework for Evaluation. In: WALBERG, H. L. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.
- WOLF, R. M. The Nature of Educational Models. In: WALBERG, H. L. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.
- WORN, B. R. Visão Geral do Mosaico Formado pela Avaliação e Controle Educational. In: GOLDBERG, M. A. A. e SOUZA, C. P. (Org.) **Avaliação de Programas Educacionais: vicissitudes, controvérsias, desafios.** EPU. São Paulo. 1982.
- WORTHEN, W. R. Competencies for educational research and evaluation. **Educational Research**, 4, 1975.
- WORTHEN, W. R. Program Evaluation. In: WALBERG, H. L. et al. **The International Encyclopedia of Educational Evaluation.** Oxford. Pergamon Press. 1990.
- WORTHEN, B. R.; SANDERS, J.R. **Educational Evaluation: theory and practice.** Belmont, California. Wadsworth. 1973.
- WORTHEN, B. R.; SANDERS, J. R. **Educational Evaluation: — alternative approaches and practical guidelines.** New York-London. Longman, 1987.
- WRIGHT, W. J. e HESS, R. J. A criteria acquisition model for educational product. In: BORICH, G. D. **Evaluating Educational Programs and Products.** Englewood Cliffs. New Jersey. Educational Technology Publications. 1974.





Apresentação

Promoção Automática e Adequação do Currículo ao Desenvolvimento do Aluno
Dante Moreira Leite

Avaliando Impactos Sociais de uma Política Educacional Democrática
Márcio da Costa

A Experiência de Avaliação Educacional em Minas Gerais
Maria Alba de Souza

Novos Estudos em Avaliação Educacional
Heraldo Marelim Vianna