

# ESTUDOS EM AVALIAÇÃO EDUCACIONAL

HERALDO MARELIM VIANNA<sup>1</sup>

## 1. AVALIAÇÃO E OBJETIVOS: Ralph W. Tyler

### 1.1 Introdução

A avaliação evolui diferentemente nos vários ambientes educacionais e suas histórias são, conseqüentemente, bem diversas umas das outras, mesmo quando os valores e as crenças que deram origem às várias metodologias apresentam pontos comuns (Norris, 1993). O exame da questão na perspectiva norte-americana possibilita perceber que a estrutura social determinada pelo desenvolvimento econômico, no princípio do século XX, exigiu a reformulação de todos os níveis educacionais, uma atuação mais eficiente dos educadores no gerenciamento da administração institucional, e novas estruturas curriculares, assim como novas estratégias de ensino, segundo mostrou Norris (1993), da Universidade de East Anglia (Inglaterra), e Vianna (1995)<sup>2</sup>.

### 1.2 Avaliação e o processo sócio-econômico

Há uma preocupação, nas primeiras décadas do século XX, tanto nos Estados Unidos como na Inglaterra, em associar de maneira bastante estreita o processo sócio-econômico a valores e conhecimentos transmitidos por intermédio da educação. Essa visão

---

<sup>1</sup> Pesquisador Senior da Fundação Carlos Chagas – FCC.

<sup>2</sup> Avaliação Educacional: — uma perspectiva histórica. *Estudos em Avaliação Educacional*, nº 12, São Paulo, Fundação Carlos Chagas, 1995.

do problema exigiu, naturalmente, um sistema de controle de todas as atividades mais diretamente ligadas ao processo educacional, o que possibilitou, conseqüentemente, o desenvolvimento da pesquisa, da avaliação educacional e, naturalmente, o da tecnologia dos instrumentos de medida e das técnicas de análise quantitativa, graças ao trabalho pioneiro e profícuo de E. Thorndike e K. Pearson, entre outros, nos Estados Unidos e na Inglaterra, respectivamente. Nessa fase, avaliação e medidas educacionais apresentam-se confundidas até o surgimento, nos anos 30, do importante trabalho empreendido por Ralph W. Tyler, como reconhece Norris (1993), apesar de suas discordâncias conceituais com o grande sistematizador da avaliação educacional.

### **1.3 Influência de Binet**

Antes de considerar o trabalho de Tyler, analisar os pressupostos em que baseou o seu modelo de avaliação e o seu significado, assim como as reações geradas ao modelo de avaliação por objetivos, é necessário destacar, como já o fizeram Madaus, Airasian e Kellaghan (1980), a influência de Alfred Binet e seus estudos (1905) para a medida da inteligência nos Estados Unidos e, inclusive, na Inglaterra, especialmente em relação às medidas educacionais, com repercussão bastante profunda na avaliação dos alunos e seu rendimento escolar. Os resultados dos chamados testes de inteligência eram utilizados na interpretação do desempenho dos estudantes nos testes de escolaridade, sobretudo quando se tratava de um baixo desempenho. A idéia de que qualquer pessoa poderia aprender começou a ser posta à prova e, em conseqüência, houve uma mudança nos testes de rendimento, que eram referenciados a critério e o desempenho do aluno era comparado a um valor absoluto, passando a testes do tipo referenciados a normas, em que o desempenho do estudante é relacionado ao comportamento do seu grupo (Madaus, Airasian e Kellaghan, 1980). Foi, assim, o início da idolatrização da curva normal, que domina a área da avaliação até meados dos anos 60, quando os teóricos da avaliação, entre os quais Cronbach e Bloom, começaram a contestar a idéia de que o fracasso do aluno era

responsabilidade exclusivamente sua, sem participação dos sistema, da escola e do próprio professor.

#### 1.4 The Eight-Year Study

O modelo das diferenças individuais, surgido logo após o conflito de 1914-18, vai dominar o mundo da educação e influenciar as formas de avaliação educacional, inclusive tornando-se uma idéia fixa, quase obsessiva, diríamos, no espírito dos avaliadores de diferentes áreas, que passaram por longo tempo a acreditar quase que piedosamente no mito da curva normal de probabilidade, simples representação gráfica de uma função matemática, o que teve profundas implicações na interpretação dos resultados de avaliações que envolviam o componente rendimento escolar. A primeira grande exceção aos testes padronizados referenciados a normas foi o projeto **Eight-Year Study** (1942), idealizado, planejado e desenvolvido por Ralph W. Tyler, na avaliação da eficiência diferencial de vários tipos de escola, como claramente demonstrado por Madaus, Airasian e Kellaghan (1980).

Os anos 20 e 30 assistem a um desenvolvimento considerável dos testes padronizados de escolaridade, de acordo com Worthen e Sanders (1987). Ao mesmo tempo, a controvérsia entre currículo tradicional e currículo progressista, este na proposta de John Dewey, vai ter grandes repercussões. Tyler, ao lado de Dewey, no **Movimento para a Educação Progressista**, vai atuar ativamente no primeiro grande estudo de avaliação longitudinal – **The Eight-Year Study (1932-40)** – para dirimir dúvidas sobre a eficiência diferencial de vários tipos de escola, no período que antecede à Segunda Guerra Mundial (1939-45). O estudo planejado por Ralph Tyler procurou responder ao questionamento da eficácia da escola tradicional em relação à escola secundária progressista, tendo em vista que muitos **colleges** e universidades recusavam estudantes oriundos das escolas que seguiam a proposta de Dewey, sob a alegação de que as mesmas não ofereceriam créditos em áreas curriculares supostamente importantes (Madaus et al., 1993) [1].

O estudo experimental de Tyler, com base na análise dos resultados de programas educacionais elaborados de acordo com as

necessidades dos alunos, independentemente dos pré-requisitos definidos pelas universidades em relação aos exames de ingresso nesse nível de ensino (Norris, 1993), vai mostrar uma nova concepção de avaliação educacional, que consiste em comparar os objetivos pretendidos aos que foram realmente alcançados. O esquema de Tyler supera, assim, o que vinha sendo até então adotado (Madaus et al., 1993), que simplesmente comparava o desempenho entre grupos experimentais e de contraste. A sua proposta, por outro lado, conforme discutiremos, partiu da medida de objetivos comportamentais, preocupando-se com os resultados decorrentes da aprendizagem. Tyler, nos 25 anos seguintes (1957), terá grande atuação não apenas em atividades relacionadas a testes, medidas e avaliação, mas influenciará toda a educação norte-americana, com amplos reflexos em outras regiões, inclusive no próprio Brasil, ainda que hoje existam algumas reações ao seu pensamento, em função de “modismos” educacionais [2].

### 1.5 Modelo de Tyler

O modelo de Tyler (1942) é bastante simples e parte do princípio de que educar consistiria em gerar e/ou mudar padrões de comportamento, devendo, em consequência, o currículo ser construído com base na especificação de habilidades desejáveis expressas em objetivos a serem alcançados<sup>3</sup>. A avaliação, na concepção de Tyler, verificaria a concretização dos objetivos propostos, a congruência entre resultados e objetivos. Seria, pois, uma forma de **validar** os pressupostos em que se baseariam os programas curriculares (**construtos**). As suas idéias têm grande repercussão e influenciam outros modelos, como, por exemplo, o de Hammond (s/d) e, mais tarde, o de Metfessel e Michael (1967) e o de Provus (1973). [3] [4].

---

<sup>3</sup> O pensamento de Tyler está expresso em seus numerosos trabalhos; acreditamos, contudo, que a leitura de seu artigo seminal sobre avaliação (1942), suas considerações sobre objetivos educacionais (In: Payne (ed.), 1974), e a parte relacionada a experiências de aprendizagem (Goldberg e Souza (Eds.), 1982; Hamilton (Ed.), 1978), consignados na bibliografia, sob a rubrica de TYLER, ofereçam um quadro bem amplo de suas concepções sobre educação, currículo e avaliação.

Tyler, em seu ensaio clássico — **General Statement on Evaluation** — (1942) deixa claro que cabe à avaliação verificar, periodicamente, até que ponto a escola demonstra eficiência como instituição responsável pela promoção da educação. A avaliação, desse modo, proporcionaria subsídios para uma análise crítica da instituição, possibilitando a reformulação de sua programação curricular. A avaliação, segundo essa perspectiva, ofereceria elementos para uma crítica fundamentada da instituição, baseada em dados empíricos, e permitiria, ao mesmo tempo, uma discussão sobre a eficiência da sua atuação. A avaliação, ainda de acordo com o pensamento de Tyler, possibilitaria o aprimoramento dos programas, com a eliminação do inoperante e o desenvolvimento daqueles aspectos que se tivessem revelado positivos.

A avaliação teria, ainda, como objetivo covalidar as hipóteses formuladas na estruturação dos novos currículos, que refletiriam o sistema de crenças e valores sociais e culturais da sociedade em que está inserida a escola. O currículo, no pensamento de Tyler, deve corresponder a essas aspirações da sociedade, que se concretizariam pela atuação da comunidade escolar no desenvolvimento dos vários programas. A idéia de congruência é uma das tônicas do pensamento de Tyler e, no aspecto exposto, a avaliação estaria constatando em que medida as hipóteses levantadas para a elaboração dos currículos estariam sendo congruentes com a realidade social.

A orientação do estudante, mais do que a simples transmissão de conteúdos nem sempre relevantes, seria outro dos objetivos propostos no plano de avaliação de Tyler. A orientação do aluno somente seria realizável depois que uma avaliação criteriosa fizesse o levantamento de todas as informações sobre o seu desempenho escolar, a fim de caracterizar possíveis problemas no seu desenvolvimento como alguém que está sendo conduzido por intermédio de um processo educativo previamente determinado e com objetivos delineados. A avaliação, desse modo, não deveria ficar restrita a apenas alguns aspectos, como geralmente ocorre, que se limita à verificação do rendimento escolar, mas abranger outras dimensões de forma a oferecer uma orientação segura ao aluno, ao longo de sua escolaridade [5].

A avaliação, qualquer que seja o modelo apresentado visa, sem sombra de dúvida, a uma tomada de decisão, que envolve professores, administradores, pais e os próprios alunos, que, assim, precisam de elementos de informação relativamente à eficiência dos sistemas, especialmente no que se relaciona com a aprendizagem. A avaliação, na proposta de Tyler (1942), dado o seu carácter sistemático, possibilita eliminar com segurança todos aqueles elementos que possam gerar desconfiças da parte da comunidade, porque os dados levantados mostram o sucesso operacional do currículo ou, então, apontam aqueles elementos que devem ser imediatamente corrigidos com o fito de restaurar a credibilidade que se deve depositar em uma escola bem orientada. Tyler (1942) deixa claro que a sociedade global – apesar da redundância da expressão – deve atuar em consonância com a sociedade educacional, a fim de garantir a continuidade do saber humano em suas diferentes formas, responsabilidade prioritária da escola. A avaliação, portanto, serve denexo entre a escola e a sociedade, que passa a ser informada dos problemas e pode intervir, superando-os [6].

### **1.6 Interação professor/aluno**

Há um aspecto da proposta de Tyler (1942) que não pode ser ignorado, que é o das relações professor/aluno. O professor deve deixar claramente exposto aquilo que pretende dos seus alunos através de diversas práticas curriculares. Os alunos, por sua vez, ficariam em condições de apresentar respostas satisfatórias às demandas da escola, dos currículos e dos professores. Depreende-se, por conseguinte, que o modelo de Tyler (1942) contém um apelo ao solidário, em que a educação não surge como um trabalho isolado, resultante da atividade de alguns poucos, mas um esforço cooperativo, com o envolvimento de diferentes segmentos sociais. Tyler (1942), ao definir seu modelo de avaliação, deixou bastante claro o seu posicionamento em relação a diversos pontos capitais de sua filosofia da educação, que não podem ser olvidados, tendo em vista que constituem o elemento de sustentação de estrutura metodológica do seu modelo, dando-lhe a coerência interna sem dúvida necessária para a sustentação de

qualquer proposta operacional, como a que pretendeu realizar em avaliação educacional.

Quais, então, esses pontos expostos por Tyler (1942), apresentados de forma abreviada, e que já foram objeto de cogitação em outro trabalho anteriormente divulgado (1982)<sup>4</sup>? Sem dúvida, os pontos-chaves concentram-se nos seguintes aspectos:

- 1) A educação é um processo que visa a criar padrões de conduta, ou a modificar padrões anteriores, nos indivíduos.
- 2) Os padrões de conduta desenvolvidos na escola são, na realidade, os objetivos educacionais.
- 3) O êxito de um programa educacional, verificado através da avaliação, depende da concretização desses objetivos.
- 4) A avaliação deve incidir sobre o aluno como um todo, nos seus conhecimentos, habilidades, modos de pensar, atitudes e interesses, sem se concentrar em apenas elementos isolados, como, na realidade, acontece nos dias fluentes.
- 5) A avaliação pressupõe diversidade de instrumental para avaliar múltiplos comportamentos, não devendo ficar restrita, apenas, a exames escritos, como geralmente ocorre.
- 6) A avaliação não se concentra apenas no estudante, como acentua Tyler (1942), não é um ato isolado, mas um trabalho solidário que deve envolver, além de alunos, claro, os professores, administradores e, sem sombra de dúvida, os próprios pais, que devem ter voz ativa no processo.

### 1.7 Eficiência e pontos críticos

O modelo de avaliação proposto por Tyler (1942) é aparentemente simples e mudou, inteiramente, o enfoque da avaliação, que se concentrava, inicialmente, nas habilidades dos indivíduos, passando, então, a preocupar-se com o julgamento do currículo. A avaliação, nessa concepção, supera a idéia, aliás equivocada, de que seria uma simples tecnologia com vistas a ordenar os indivíduos, segundo um determinado critério, discriminando-os uns dos outros. Na

---

<sup>4</sup> Vianna, H. M. Avaliação Educacional - algumas idéias precursoras. **Educação e Seleção**, São Paulo. Fundação Carlos Chagas, nº 6, 1982.

verdade, com Ralph W. Tyler, a avaliação passou a ser um meio de verificar em que medida os objetivos curriculares, assim como os processos instrucionais, se concretizaram na prática, de acordo com o destaque bastante procedente de Norris (1993).

Acreditamos seja necessário ressaltar a posição de Tyler em relação aos testes e aos objetivos, muitas vezes confundidos por pessoas nem sempre integradas no seu pensamento. Tyler não admitia a identificação dos conceitos de **avaliação** e **medida**, usados quase como expressões sinônimas. Isso, em oposição ao pensamento de Tyler, seria admitir que a avaliação visaria a mensurar apenas diferenças individuais, o que não estava de acordo com o seu pensamento. Sua posição era frontalmente de recusa diante desse posicionamento, sendo a medida, especialmente a do rendimento escolar, apenas um momento do complexo processo de avaliação. Isso, entretanto, não significava que negasse a importância dos exames, dos testes e das provas. A sua preocupação com objetivos, por outro lado, ligava-se à sua própria concepção de objetivo, que seria uma transformação desejável nos padrões de comportamento dos alunos. Tyler, na análise do desempenho dos estudantes, no projeto **Eight-Year Study**, dá destaque à palavra avaliação como substituta de outras usadas, como as que foram mencionadas, e que ainda continuam a ser empregadas. A avaliação, para Tyler, está identificada com um processo de investigação de valores, devendo verificar, periodicamente, a eficiência das escolas e identificar os pontos críticos dos vários programas curriculares, a fim de aperfeiçoá-los e, naturalmente, validar os princípios que alicerçam a atividade de uma instituição escolar.

### 1.8 Reações e críticas iniciais

Os anos 50, nos Estados Unidos, além da figura de Tyler, sofreram a influência dos trabalhos de E. F. Lindquist. As estatísticas desenvolvidas graças aos seus esforços estabeleceram princípios para o uso do **experimental design** no campo da educação e foram usadas intensamente em relação à quantificação de variáveis educacionais. Tyler, por sua vez, nessa mesma época, obteve o concurso de outros educadores, como foi o caso de Bloom, na hierarquização dos

objetivos instrucionais. As idéias de Tyler consolidaram-se nesse momento, os testes padronizados passaram a ter maior disseminação, profissionais receberam solicitações para que expressassem suas opiniões e trabalhos experimentais foram realizados para a avaliação de novos currículos (Madaus, Airasian e Kellaghan, 1980; Madaus, Stufflebeam e Scriven, 1993).

É lógico, como expõem Madaus et al. (1993), que toda inovação gere resistências, e críticas, às vezes bastante exacerbadas, são feitas ao planejamento experimental de Lindquist, à teoria da congruência de Tyler e à atuação de Bloom na categorização das capacidades a serem medidas (objetivos), iniciando-se uma reação ao uso da Psicometria em avaliação, com fundamento em diferentes posições epistemológicas relativas à pesquisa e, naturalmente, à avaliação, conforme esboçaremos em suas linhas gerais.

A importância de Tyler aprofundou-se em toda a vida educacional norte-americana, ao projetar, nos anos 60, o **National Assessment of Educational Progress - NAEP**, avaliação periódica da situação do ensino nos Estados Unidos e que ainda subsiste, fornecendo importantes elementos para as correções necessárias e a identificação de pontos de excelência ou pontos críticos a serem enfrentados<sup>5</sup>. Um desdobramento do NAEP, na década de 90, no plano internacional, foi o **IAEP - International Assessment of Educational Progress**, que teve, inclusive, a participação parcial do Brasil, com amostras de escolas de São Paulo e Fortaleza.

O **Eight-Year Study**, delineado por R. W. Tyler, domina a cultura pedagógica norte-americana até os dias de hoje, apesar das reações surgidas e que determinaram o aparecimento de novos modelos. Uma geração de avaliadores, entretanto, segue o pensamento de Tyler, ainda que em alguns casos com alterações, podendo-se indicar, entre outros, os nomes bastante expressivos de B. S. Bloom; H. Taba; D. R. Krathwohl; B. B. Masia; N. M. Metfessel; W. B. Michael; J. T. Hastings e G. F. Madaus. A influência de Tyler no Brasil não ficou restrita à área de currículo, com a tradução de seu livro básico sobre o assunto, mas ocorreu de forma indireta graças à

---

<sup>5</sup> O Sistema de Avaliação do Ensino Básico (SAEB), no Brasil, a partir de sua última aplicação em 1995, procurou seguir, em alguns de seus aspectos, o modelo do NAEP.

disseminação da obra de Bloom e seus associados sobre a avaliação do rendimento escolar e as Taxonomias dos objetivos educacionais, divulgadas a partir do início dos anos 70 no País.

O modelo de Tyler, tendo em vista seu delineamento, baseado em objetivos comportamentais, e sua metodologia de análise quantitativa, ficou, naturalmente, sujeito à crítica no debate das posições epistemológicas dos vários teóricos da avaliação, sendo-lhe imputada uma orientação geral do individualismo metodológico, porquanto a validação dos currículos dar-se-ia por meio do contraste entre predições logicamente estabelecidas (objetivos) e os resultados coletados (scores), na linha dos delineamentos experimentais, conforme as proposições de D. T. Campbell e J. C. Stanley, em **Experimental and Quasi-Experimental Designs for Research** (Rand McNally & Company, Chicago, 1966). Ora, tudo isso entrava em confronto com a reação estabelecida inicialmente pelos teóricos sociais, na Europa, a partir de cientistas alemães, com repercussões posteriores nos Estados Unidos, discussões que são analisadas em profundidade, alguns anos depois, por Habermas<sup>6</sup> e, mais recentemente, por Guba<sup>7</sup> [7].

A discussão remonta ao possível conflito entre os objetivos das ciências sociais e os das ciências físicas, que, dessa forma, por via de consequência, deveriam seguir, forçosamente, metodologias diversas. A abordagem nas ciências físicas teria um caráter nomotético, visaria ao estabelecimento de leis e generalizações estatísticas, enquanto as ciências do social estariam interessadas na análise intensiva dos indivíduos ou de casos particulares, adotando, portanto, metodologias ideográficas, segundo a exposição de Norris (1993). Questionava-se, assim, a possibilidade de estudar da mesma forma fenômenos naturais e sociais a partir de uma unidade metodológica, determinando, em decorrência, um reducionismo (identidade de conteúdo e de método).

A idéia de que o que é válido em uma área é igualmente válido em outra representava, assim, o pensamento ortodoxo do empirismo reducionista e durante muito tempo foi a temática central das

---

<sup>6</sup> HABERMAS, J. **Conhecimento e Interesse**. Rio de Janeiro. Zahar Editora. 1982.

<sup>7</sup> GUBA, E. G. (Ed.) **The Paradigm Dialog**. Newbury Park, California. Sage Publications. 1990.

discussões sobre qual seria a orientação epistemológica a ser seguida em avaliação educacional. A idéia de que a linguagem em avaliação e em pesquisa deveria ser operacional, sem ambigüidades; as diversas variáveis a serem observadas bem definidas e seus conceitos claramente delimitados, sendo o objetivo estabelecer conclusões generalizáveis, eram idéias contestadas pelos que se propunham a criar um novo posicionamento científico de sentido humanístico, ou seja, sociológico, nas ciências do homem, inclusive na área da avaliação educacional, reagindo, dessa forma, às explicações nomotéticas, que visavam a explicar o fenômeno para predizer e, depois, generalizar.

### **1.9 Posições de Cronbach, Scriven, Stake e Parlett/ Hamilton**

O trabalho de Cronbach — **Course Improvement through Evaluation** (1963) — apresenta uma posição crítica face ao modelo objetivo de Tyler, quando mostra que a avaliação não pode ficar presa a simples aspectos rotineiros e ritualísticos da mensuração. A avaliação, na sua concepção, e com inteira procedência, tinha por finalidade não apenas fazer um julgamento final, o que seria limitá-la nos seus objetivos, mas oferecer meios que possibilitassem o aprimoramento dos currículos, o que acreditamos está de acordo com o espírito do trabalho de Tyler ao longo de seu desenvolvimento. Cronbach (1963), com a sua perspicácia habitual, soube imediatamente compreender o caráter multidimensional dos resultados da aprendizagem e, a esse respeito, demonstra claramente que essa situação não pode ser constatada em um único escore compósito, agregando diferentes elementos. Assim, uma avaliação conseqüente exige uma coleta diversificada de informações para que seja viável uma descrição dos currículos ou programas educacionais e possam ser identificados os elementos que devem sofrer a intervenção dos especialistas, no processo de aprimoramento das atividades educacionais. A crítica bastante procedente de Cronbach incorporou-se ao trabalho de Tyler, como pode ser depreendido do seu trabalho igualmente fundamental publicado na antologia de Madaus et al. (1993).

As posições de Scriven, em sua obra clássica — **Methodology of Evaluation** (1967) —, dão continuidade e ampliam pontos levantados por Cronbach (1963) e introduzem um novo elemento de grande significado: a obrigação do avaliador determinar o valor (mérito) do que está sendo objeto da avaliação — programas, projetos, materiais ou quaisquer outras inovações educacionais. Scriven parte, conforme será posteriormente analisado, da necessidade de diferenciar papéis (**roles**) de objetivos (**goals**), sem o que não se pode determinar a eficácia das práticas educacionais. A posição de Scriven é, assim, de restrição às avaliações baseadas exclusivamente em **goals**, por não possibilitarem procedimentos adequados para o julgamento do valor ou mérito dos objetivos.

Stake, nessa época (1967), em obra capital — **The countenance of educational evaluation** —, apresenta críticas às práticas convencionais da avaliação. A sua preocupação imediata concentra-se na discussão dos diferentes tipos de evidências que os avaliadores deveriam levantar no decorrer do seu trabalho. A avaliação tal como estava sendo feita, segundo o modelo de Tyler, e na opinião de Stake, gerava a suspeição dos educadores e administradores, insatisfeitos com a linha seguida e tendo em vista as implicações políticas das avaliações realizadas no contexto norte-americano, daí a sua proposta de um novo modelo, que possibilitasse um levantamento bastante exaustivo de aspectos complexos e particulares de um determinado programa. A avaliação, desse modo, ainda segundo Stake (1967), não deveria ter um caráter meramente episódico, mas permitir uma compreensão de todo o processo relacionado a um programa, por intermédio de um amplo levantamento de informações para uma tomada de decisões em bases realistas.

A reação mais veemente, se é que podemos usar o termo, contra a avaliação por objetivos, segundo o modelo estabelecido por Ralph W. Tyler, partiu de **Parlett e Hamilton** (1972), na Inglaterra, repercutindo, depois, em outras áreas, inclusive nos Estados Unidos, com uma repercussão bastante tardia no Brasil, assim mesmo em círculos limitados. A reação procurou refletir a insatisfação de muitos educadores aos modelos tradicionais de avaliação, por um lado, e, por outro, rejeitava, em muitos aspectos, a influência da tradição

psicológica que dominava a pesquisa educacional. Os críticos da situação então dominante pretendiam que a pesquisa educacional e, naturalmente, sem sombra de dúvida, a própria avaliação educacional tivessem uma abordagem sociológica, e que ambas — pesquisa e avaliação — se envolvessem na consideração de valores humanos e adotassem paradigmas que fossem para seres humanos.

Parlett e Hamilton, em **Evaluation as illumination: a new approach to the study of innovatory programs** (1972), reagiram à psicométrica e ao experimentalismo, propondo uma avaliação na tradição humanista, que fosse responsiva às necessidades de diferentes grupos, iluminativa do complexo organizacional, do ensino e do processo de aprendizagem, relevante em relação às diferentes decisões a serem tomadas, e que fosse divulgada por intermédio de uma linguagem acessível aos diferentes grupos interessados no assunto. A reação de Parlett e Hamilton opunha-se, assim, ao quantitativo do modelo por objetivos e procurava apresentar uma visão mais pragmática do planejamento e da concretização da avaliação, pretendendo, desse modo, por via de uma avaliação responsiva, atenuar os problemas que decorriam daquilo que acreditavam como resultante de um dogmatismo metodológico, como destaca Norris (1993). Pretendiam a valorização do verbal e não do quantitativo, que decorria de uma tradição psicométrica que se impunha desde os primeiros anos do século XX, na Inglaterra e nos Estados Unidos.

As idéias de Parlett e Hamilton foram duramente criticadas por muitos, e ainda o são, apesar de possuírem um número relativamente expressivo de seguidores, tendo em vista que, na realidade, elas não oferecem condições para um controle acurado do grau de subjetividade dos pesquisadores/avaliadores e, por acréscimo, complicando ainda mais o problema e não evitam as digressões personalíssimas dos avaliadores e/ou pesquisadores. A chamada avaliação iluminativa que estava sendo proposta, na visão de seus críticos, constituía, na realidade, um simples método e não tinha um embasamento suficientemente fundamentado, sendo, na realidade, simples reação à tradição psicométrica em avaliação.

### 1.10 Possíveis limitações

O modelo de avaliação por objetivos, mesmo na apreciação de seus opositores, é racional, e suas críticas se concentraram não exatamente na tecnologia da avaliação, mas no posicionamento de Tyler e seus seguidores em relação à natureza do conhecimento e nas formas de sua aquisição. A importância da avaliação e sua utilidade não são discutidas, o confronto estaria na limitação do modelo para a reformulação das experiências curriculares e, especialmente, quanto aos objetivos, que seriam colocados em termos bastante pedestres, ou seja, de uma forma por demais trivial. Alguns teóricos opositores, por outro lado, como é o caso de Eisner (1993), — ver observações em Notas e Comentários [5] —, preocupados com problemas ligados à arte e ao seu processo de criação, envolvendo a complexa questão da criatividade, colocaram em dúvida a possibilidade de aplicar a avaliação por objetivos a todo e qualquer tipo de conhecimento, via a especificação dos resultados esperados, tendo em vista a imprevisibilidade da criação artística.

### 1.11 NOTAS E COMENTÁRIOS

- [1] O pensamento de Tyler, suas experiências e os caminhos que o levaram à avaliação estão apresentados em substancial artigo (Madaus et al., 1993), no qual explica o surgimento de uma de suas obras mais interessantes — **Constructing Achievement Tests** (1934) <sup>(1)</sup> — fruto de um trabalho desenvolvido a partir de 1929, na Ohio State University, em pesquisas sobre ensino e avaliação nos cursos de graduação, na área de Biologia. A má qualidade dos testes utilizados, que exigiam sobretudo o reconhecimento de fatos e conhecimentos simples, além da incapacidade de elaborar objetivos comportamentais (conteúdo e habilidade), entre outros aspectos críticos, contribuíram para que redigisse o trabalho indicado, que, apesar de editado na década de 30, ainda oferece uma orientação bastante sensata na construção de instrumentos de medida.

A partir de 1934, Tyler passou a vivenciar um novo problema, ao ser nomeado diretor do **Eight-Year Study** e sua atuação foi na área curricular das escolas de 2º Grau (**high-school**), que, em termos de conteúdo e de atividades, ainda refletiam o pensamento dos anos 20. Havia uma grande pressão para mudança, o mundo se tinha modificado após a depressão de 29 e a escola e

seu currículo continuavam os mesmos. Graças à ação pioneira da **Progressive Education Association**, um pequeno número de escolas e sistemas (30) foi autorizado a desenvolver programas educacionais que atendessem aos interesses dos alunos e ao seu desejo de continuar seus estudos. Um programa que levasse em consideração os interesses, as necessidades, as atividades e as características de aprendizagem dos jovens. Iniciado em setembro de 1933, o projeto piloto teria a duração de oito anos (**Eight-Year Study**). Especialistas em currículo e em avaliação começaram a trabalhar juntos. A fim de fazer face à complexidade dos problemas, e estimulado por Hulda Taba, Tyler escreveu um **syllabus** intitulado **Basic Principles of Curriculum and Instruction**<sup>(2)</sup>, que seria discutido em vários seminários, e mais tarde publicado sob a forma de livro, influenciando o pensamento pedagógico de educadores nos Estados Unidos e em outros países, inclusive no Brasil. Currículo deixa de ser um conjunto de tópicos para ser experiências de aprendizagem associadas a objetivos, e organizadas de forma a maximizarem seu efeito cumulativo. Ao formular suas idéias sobre currículo, Tyler, como ele próprio reconhece (In: Madaus, et al., 1993), sofreu a influência das concepções educacionais de John Dewey (1859-1952) e de Alfred North Whitehead (1861-1947), sobretudo no relativo aos objetivos da educação e às teorias do ensino e da aprendizagem. Os trabalhos de Tyler tiveram grande repercussão. A partir de 1938, a Universidade de Chicago, por exemplo, seguiu sua orientação na elaboração de exames compreensivos, e na década de 40 passaram a constituir o fundamento teórico de numerosos testes. As suas idéias sobre currículo serviram para a elaboração e/ou revisão de numerosos currículos profissionais (medicina, enfermagem, assistência social, engenharia e agricultura). Entretanto, modestamente, confessa não poder precisar a influência de suas idéias nos currículos das escolas primárias e secundárias (Tyler, In: Madaus et al., 1993).

(1) Ralph W. Tyler **Constructing Achievement Tests**. Bureau of Educational Research, Columbus, Ohio: State University Bureau of Educational Research, 1934.

(2) Ralph W. Tyler **Basic Principles of Curriculum and Instruction**. Chicago: University of Chicago Press, 1949.

- [2] Ao discutir programas destinados à melhoria dos sistemas sociais, Airasian (1993) mostra que nem sempre os mesmos produzem as melhorias pretendidas, sendo difícil estabelecer se um programa vai ter, efetivamente, algum impacto, inclusive tendo em vista a precariedade das metodologias para determinar o impacto. Apesar do otimismo que envolveu certos programas no contexto norte-americano, que abordaram aspectos sociais e econômicos, incluindo entre aqueles o educacional, a repercussão do seu impacto nem sempre correspondeu às expectativas com que os programas foram criados. Apesar dos esforços empreendidos, a solução para a medida do impacto talvez esteja no desenvolvimento de uma metodologia que use métodos quase-

experimentais, possibilitando conciliar a realidade do mundo às exigências da pesquisa avaliativa, como assinala Airasian (In: Madaus et al., 1993).

- [3] O texto cita alguns modelos neotylarianos, que, entretanto, não tiveram maior repercussão, pelo menos no Brasil. Ficaremos limitados ao que tem maiores possibilidades de divulgação entre nós, o de Provus (1973), que apresentou o modelo de **Discrepancy Evaluation**, mais tarde desenvolvido por vários teóricos e praticantes da avaliação, inclusive Steinmetz (In: Madaus et al., 1993), o qual ressalta o fato de que a avaliação abrange diferentes atividades e propósitos, e que a sua prática gera formas mais ou menos intensas de preocupação. Isso porque a avaliação está associada a valor, mérito e julgamentos, o que determina reações emocionais, pois tudo isso tem significados sociais e profissionais, podendo representar perda de **status** e alterações em projetos de vida e de carreiras, entre outras conseqüências. A abordagem da **Discrepancy Evaluation** propõe-se a reagir de forma construtiva às expectativas criadas pelo ato de avaliar, apresentando, com esse objetivo, algumas idéias e procedimentos que procuraremos resumir.

A idéia básica dessa abordagem centra-se no fato de que, ao avaliarmos, fazemos **comparações**, que somente são possíveis se **padrões** forem fixados, entendendo-se por padrão a descrição de como alguma coisa — objeto, pessoa, programas etc. — **deva ser**. Fixados os padrões, é necessário obter medidas do desempenho (**performance**); portanto, nessa abordagem, avaliar é comparar desempenhos a padrões estabelecidos, e essas comparações nos vão fornecer possíveis **discrepâncias** entre ambos — desempenhos e padrões. Avaliação seria, assim, o julgamento do valor de algo, estabelecido a partir da discrepância entre os **padrões** e a **performance**.

O avaliador na **Discrepancy Evaluation** não fixa os padrões, nem julga a comparação entre o padrão e o desempenho, ainda que seja o responsável pela coleta dos dados da performance. O avaliador atua como auxiliar do cliente interessado na avaliação, que é solicitado pelo avaliador para apresentar perguntas em relação ao que é avaliado. A partir dessas perguntas, são estabelecidos os padrões, identificadas as fontes de informação, selecionados ou construídos os instrumentos, se necessários, e coletados, então, os dados pelo avaliador.

Outros aspectos a serem definidos em colaboração — clientes e avaliador — referem-se às informações coletadas sobre os desempenhos e a fonte que vai proporcionar esses elementos. Precisamos lembrar que, ao fixar os padrões, o cliente deve considerar o conhecimento que tem sobre o objeto a ser avaliado, assim como suas próprias experiências, valores e objetivos. A fixação de padrões é uma atividade orientada por objetivos. Um procedimento habitual no estabelecimento de padrões é a realização de uma análise de componentes, no caso de programas educacionais, que são divididos e às vezes até mesmo subdivididos em vários de seus aspectos, com vistas à determinação de diversos níveis de detalhe.

A partir desse ponto, cada componente e subcomponente sofre uma análise para determinar o que é **input**, processo e produto. Os objetivos são os produtos (**outputs**); o que será feito, o processo; e tudo que é necessário para o processo (recursos financeiros, pessoal, material, pré-requisitos etc.) constitui o **input**. A **discrepancy evaluation** incluída entre as abordagens de avaliação orientadas por objetivos, seria, digamos, um modelo quase-tyleriano, e teria um caráter sistêmico, visando a dar destaque à auto-avaliação de programas educacionais. O modelo tem suas potencialidades a explorar, fazendo-se necessário o conhecimento de sua teoria e prática (Provus, 1973), ou um contato com obras de síntese e aplicação (Provus, In: Taylor e Cowley, 1972; Provus, 1969).

- [4] O fato de existir um número elevado de modelos de avaliação, como pode ser positivado na obra de Madaus et al. (1993), não significa que sejam os mesmos, necessariamente, aplicáveis ao contexto educacional. Podem despertar um certo entusiasmo inicial, mas, depois, a realidade revela dificuldades múltiplas que traduzem a sua inaplicabilidade ao contexto das ciências humanas. A experiência vivida por Popham e Carlson (In: Madaus (Ed.), 1993) para participarem da avaliação do **Hawaii 3-on-2 Program**, dirigido por uma empresa de grande reputação, **Northwest Regional Education Laboratory (NWREL)**, segundo a metodologia do **Adversary Model**, que segue, nas suas linhas gerais, procedimentos semelhantes ao que ocorrem nos tribunais norte-americanos (equipe de defesa, equipe de acusação e corpo de jurados - árbitros), mostrou-se a mesma ineficiente, sujeita a freqüentes distorções e custos mais elevados do que a de outros modelos, para decepção de Popham e Carlson, conforme se referem no artigo de relato da mencionada experiência, que conclui com a afirmação: **The Adversary Evaluation Model brings with it not only a few real (and many imaginary) dividends; it brings also far too many deficits** (p. 212). A escolha de um modelo deve ser criteriosa, sujeita a uma análise isenta e crítica, sem impulsos emocionais.
- [5] As formas usadas para avaliar, como deixou claro Eisner (In: Madaus (Ed.), 1993), influenciam o sistema escolar, sua filosofia e a sua prática. A perspectiva histórica, pelo menos no contexto norte-americano, a partir do início do século XX, mostra que a educação passou a acreditar que um dia a avaliação teria suas "leis", que regeriam os fenômenos educacionais e fariam pela educação aquilo que as idéias de Einstein e outros fizeram pela física. Eisner (1993), ao analisar o contexto norte-americano, mostra-se crítico ácido da avaliação educacional nos moldes praticados, destacando que o problema atual estaria ligado às suas origens, com os trabalhos de E. L. Thorndike, que geraram a crença na possibilidade de uma ciência da educação com suas leis da aprendizagem, que substituiriam a intuição e a **habilidade artística (artistry)** por **conhecimentos e métodos**.

O grande objetivo da pesquisa educacional estaria em identificar os melhores métodos de ensinar as várias áreas curriculares. Segundo a visão de Eisner (1993) tudo isso coincide com o movimento favorável à eficiência da educação e à introdução de procedimentos gerenciais científicos nas escolas, visando à eficiência (sucesso) e o efetivo na educação (produção de bons resultados nas escolas). Eisner antecipa, assim, a questão da qualidade em educação, objeto de debate nos dias fluentes, e, ao mesmo tempo, preocupa-se com a educação como atividade nomotética (estabelecimento de leis) e atividade ideográfica (estudo descritivo da individualidade), rejeitando o nomismo positivista e a idéia de que as ciências naturais exatas seriam o ideal para a compreensão da realidade.

Eisner, ligado às artes e ao problema da criatividade, ainda que reconheça que as ciências e tecnologias trouxeram alguma contribuição para a educação, no seu provocativo ensaio (In: Madaus et al., 1993), aponta que as mesmas tiveram influências negativas, atribuindo às ciências e às tecnologias quatro conseqüências deletérias, na sua visão:

1ª) como a pesquisa se baseia em pressupostos científicos e visa a identificar leis ou generalizações semelhantes a leis (**law-like**), sua tendência seria a de tratar qualidades de situações particulares como **instrumentalidades**, como conseqüência haveria uma supersimplificação do particular através de um processo de redução que traduziria a complexidade de uma caracterização por um único conjunto de escores, transformando qualidade em quantidade: um número, que é incapaz de expressar o simbolismo da criação artística, na visão de Eisner;

2ª) a orientação tecnológica tende a focar o desempenho numa **situação futura** e, no processo, a enfraquecer o **significado do presente**. Daí, certamente, a preocupação com a **formulação de objetivos comportamentais**, a serem alcançados ao término (no futuro); o presente é sacrificado, valorizando-se o futuro e contribuindo, assim, para que alunos e professores sejam atormentados por recompensas extrínsecas, prêmios, altos escores e, no caso dos professores, compensação financeira em função do alto número de sucessos (objetivos concretizados). Eisner diz que se o futuro se torna muito importante (**all-important**), deve ser alcançado a qualquer custo, e o presente é sacrificado no altar do amanhã;

3ª) a abordagem científica e tecnológica da educação, segundo Eisner (In: Madaus (Ed.), 1993), leva à tentativa de objetivar o conhecimento; ou seja, a qualidade deve ser empiricamente manifesta, e possível de quantificação, com vistas à garantia da fidedignidade e à segurança da precisão. A educação estaria, portanto, influenciada pelo behaviorismo radical. Assim, no dizer de Eisner, citando um de seus alunos, para saber o que alguma pessoa sente, saber o que significa comportamento, deve ir além do comportamento;

4ª) quando se procura estabelecer leis que governem o controle do comportamento humano, diz Eisner, ainda no mesmo ensaio (In: Madaus

(Ed.), 1993), não causa surpresa o fato de que também se procure a efetivação de um conjunto de objetivos de forma quantitativa, que, no pensar de Eisner, leva, inevitavelmente, ao uso de testes (padronizados). O teste padronizado é o **padrão**, acentua Eisner, é o mesmo para todos, os objetivos são padronizados, assim como são padronizados os julgamentos. Os testes, escreve Eisner, **de facto**, tornam-se objetivos. Eisner afirma que, diante dessa situação, os instrumentos não permitem avaliar a personalização do ensino e da aprendizagem.

Eisner é crítico das atuais formas de avaliação, orientada por objetivos, que influenciariam, negativamente, o caráter do ensino, o conteúdo dos currículos, e os objetivos que as escolas procurariam atingir; por outro lado, acha que os procedimentos atuais de avaliação, ainda que científicos na sua epistemologia e tecnológicos na sua aplicação, além de terem conseqüências limitadas, não se aplicam aos objetivos que as ARTES podem concretizar. A preocupação de Eisner é, especificamente, com os problemas relacionados à arte e à criação artística, e devemos ressaltar, daí sua posição anticomportamental na linha tyleriana. A posição de Eisner é coerente com as suas idéias, pois diz claramente que ensinar exige **artistry** (ensinar para ele é uma arte) e que a educação é um processo que difere de um contexto para outro. Eisner, na sua crítica à avaliação, baseia-se em dois conceitos: competência educacional (**educational connoisseurship**) e criticismo educacional. O ensino para Eisner tem um caráter ideográfico e que a vida na sala de aula não pode ser explicada e nem controlada por leis comportamentais. O professor/avaliador deve ser um **connoisseur** capaz de distinguir qualidades, discriminar diferenças, comparar e contrastar, perceber o significado do que foi vivenciado. A partir dessa percepção, avaliar. Quanto ao criticismo, Eisner parte da tese de John Dewey de **arte como uma experiência**, em que o criticismo tem como fim a reeducação da percepção, e a função do crítico é auxiliar a ver a obra de arte — na literatura, na música, nas artes visuais, no drama e no filme. **Educational criticism** para Eisner seria uma nova forma de observar o fenômeno que constitui a vida educacional na sala de aula, apresentando seus aspectos descritivo e interpretativo, a fim de compreender o significado, o **valor** e as alternativas das formas de ação em um contexto social.

Ao final Eisner (In: Madaus (Ed.), 1993) diz algo que merece reflexão, e reproduzimos suas próprias palavras: "*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*". (p. 346)

[6] Há um grande número de modelos de avaliação e a melhor forma de compará-los é examinando seus pressupostos teóricos, para conhecimento de sua lógica e suas possibilidades. Há várias taxonomias de modelos, como, por exemplo, as de Worthen e Sanders (1973), Popham (1975) e Stake (1967); entretanto, para fins da presente nota, limitar-nos-emos à de House (1978; 1993), e ficaremos adstritos apenas aos modelos de Tyler, Scriven, Stufflebeam e Stake, objeto de análise e discussão no presente trabalho.

Os modelos variam na medida em que os pressupostos em que se baseiam são divergentes e muitos modelos são variações modificadas de outros modelos. Veremos quatro modelos, entre os definidos por House (In: Madaus et al. (Ed.), 1993).

1. **Objetivos comportamentais** — os objetivos são expressos em termos de desempenhos específicos do estudante e são medidos por testes referenciados a normas ou a critério. A abordagem foi proposta inicialmente por Tyler. O objetivo do modelo é a produtividade; **accountability**.
2. **Tomada de decisão** — a avaliação é estruturada levando em consideração a decisão a ser tomada. A metodologia tende a usar questionários e entrevistas (*survey*). A grande figura na área é Stufflebeam, com o modelo CIPP. Eficiência e controle da qualidade são seus objetivos.
3. **Independente de Objetivo (Goal-Free)** — A idéia básica de Scriven foi a de minimizar os efeitos de vieses em avaliação. O avaliador não se orienta a partir dos objetivos do elaborador do programa; cabe-lhe, ao contrário, através da avaliação, determinar esses objetivos. A utilidade social é o objetivo do modelo.
4. **Interação (Transation)** — A abordagem parte dos próprios processos educacionais: — a sala de aula, a escola, o programa, na percepção de Stake. Utiliza-se de métodos informais de avaliação e uso do estudos de caso como principal figura nessa área. Podemos dizer que o modelo visa a compreensão e a identificação da diversidade.

House (In: Madaus, et al. (Ed.), 1993) chama a atenção para o fato de que os atuais modelos de avaliação derivam de filosofia do liberalismo e que o liberalismo é muitas vezes radicalmente empiricista, e que os modelos de avaliação têm um **flavor** empiricista. House analisa a epistemologia liberal objetivista e mostra que os modelos têm uma orientação gerencial (**management**) e suas maiores preocupações concentram-se na **accountability**, na eficiência e no controle de qualidade. A análise de House parte da obra dos empiricistas ingleses (Locke (1632-1704), Berkeley (1685-1753) e Hume (1711-1776), detendo-se, depois, nas idéias de J. Stuart Mill (1806-73) e na sua obra *A System of Logic* (1893). Apoiando-se em diferentes estudos, House mostra que Francis Galton (1822-1911), contemporâneo de Stuart Mill, emprestou sua genialidade à divulgação da

obra do grande lógico, e criou as bases da Psicometria com seus estudos sobre as diferenças individuais, estudos que, por sua vez, tiveram grande influência sobre Cattell, que vai introduzir os testes mentais nos Estados Unidos, e contribuem para o despontar da figura de E. L. Thorndike, principiando, assim, uma longa história, em meados do século XVII e que chega até nós, no século XX, atravessando grandes controvérsias e reações, as quais deram margem ao surgimento de uma epistemologia subjetivista e uma metodologia naturalista, preocupada em comunicar-se com públicos que não são técnicos (professores e o público em geral), usando uma linguagem comum, baseando-se numa lógica informal, e apoiando-se em entrevistas e observações para a coleta de informações. A história dos modelos ajuda a compreender a própria evolução da avaliação. A distância entre uma questão típica de um modelo baseado em análise de sistema, cujo público é predominantemente preocupado com aspectos econômicos — **Quais os programas mais eficientes?** e a indagação proposta por um avaliador envolvido com o conhecimento de individualidades — **Qual a percepção de diferentes pessoas sobre o programa?** mostra que há um longo caminho a percorrer, e somente a reflexão crítica ajudará a compreender. O ensaio de House é, sem dúvida, excelente subsídio.

- [7] Um modelo de avaliação, apesar de bem estruturado, precisa ser modificado, para atendimento de especificidades de certas situações. A experiência de Madaus (In: Madaus (Ed.), 1993), na avaliação do programa de **Testes de Competência Mínima**, nesse sentido, é bem elucidativa. O modelo original foi o **Adversary Model**, que segue procedimentos da prática jurídica norte-americana. A principal modificação sugerida foi a eliminação do júri ou painel, que teria por incumbência tomar decisões ou fazer recomendações sobre o objeto avaliado, ficando eliminado o veredito a favor ou contra os grupos envolvidos na avaliação, não haveria “perdedores” nem “ganhadores”. Os grupos passariam a esclarecer os aspectos ligados à eficiência ou ineficiência do programa. Outros aspectos ligados à realização de “tapes”, constituição de grupos de debates, depoimentos e seu tempo de duração, aspectos gerenciais do processo de avaliação, forma de apresentação do produto final, entre outros, deveriam sofrer modificações para que o modelo adotado fosse efetivamente eficiente na tomada de decisão sobre o programa em questão. Um modelo, por mais cuidadoso que seja a sua estruturação, nem sempre é aplicável a qualquer situação avaliativa

## 2. AVALIAÇÃO E SUA PRÁTICA: L. J. Cronbach

### 2.1 Introdução

Ao escrever o seu célebre artigo — **Course improvement through evaluation** (1963) —, Cronbach não pretendeu apresentar um modelo para a realização de avaliações, conforme destacamos anteriormente (Vianna, 1986). As idéias que ventilou tiveram ampla ressonância e *contribuíram em grande parte para o surgimento* de dois trabalhos fundamentais na área da avaliação educacional, publicados ambos em 1967: o de Stake — **The countenance of educational evaluation** —, e o de Scriven, — **The methodology of evaluation** —, que serão objetos de consideração nos próximos capítulos. O artigo de Cronbach concentra-se na discussão de quatro aspectos de extrema importância: 1) — a associação entre avaliação e o processo de tomada de decisão; 2) — os diferentes papéis da avaliação educacional; 3) — o desempenho do estudante como critério de avaliação de cursos; e 4) — algumas técnicas de medida à disposição do avaliador educacional.

### 2.2 Idéias fundamentais

As idéias de Cronbach, no seu artigo de 1963, são resumidas nos seus pontos essenciais (Vianna, 1986) a fim de oferecer uma visão geral dos problemas que suscitaram e que, decorridos 35 anos, ainda são relevantes e constituem um roteiro de *questões de grande importância*: relação entre os especialistas em currículo e o avaliador; avaliação e seus diferentes entendimentos; decisões possíveis em *decorrência da avaliação*; avaliação no aprimoramento e na revisão de cursos; avaliação e comparação de cursos; procedimentos de medida em avaliação educacional; avaliação de cursos e a opinião de especialistas; o uso da observação em avaliação educacional; e emprego do **item sampling**; a importância da análise dos itens; problemática dos instrumentos de medida na avaliação de currículos; o emprego de testes na medida de processos complexos; a especificidade dos itens e a questão da transferência de aprendizagem; e, finalmente, o uso de medidas de atitude e os estudos de **follow-up** na avaliação de cursos. A importância das idéias de Cronbach levou-nos a divulgá-las

em 1986; agora, decorridos mais de dez anos, tendo em vista a atualidade das mesmas, voltamos a discuti-las, porque constituem um **corpus** bastante coerente e sobre o qual é preciso refletir, especialmente por aqueles que praticam a avaliação e sentem necessidade de uma fundamentação teórica consistente [1].

O entendimento entre especialistas em currículo e elementos da área de avaliação nem sempre é perfeito. Muitas vezes aquele especialista não possui uma idéia perfeita do tipo de colaboração que o avaliador lhe pode proporcionar; por outro lado, conforme Cronbach (1963), muitas técnicas e concepções teóricas da avaliação não se ajustam perfeitamente aos problemas de currículo [2].

A avaliação, no seu sentido mais amplo, pode ser definida como um processo que visa à coleta e ao uso de informações que permitam decisões sobre um programa educacional. A avaliação, portanto, segundo Cronbach (1963), deve ser entendida como uma atividade diversificada, que exige a tomada de vários tipos de decisões e o uso de grande número de diferentes informações. A avaliação com vistas ao aprimoramento de currículos não deve ser confundida, como muitos o fazem, com a construção de instrumentos de medida e a obtenção de escores fidedignos, processos que, eventualmente, podem entrar no contexto da avaliação, mas que não são indispensáveis para que ela possa atingir seus objetivos.

Cronbach (1963) mostra que a avaliação é usada com o objetivo de tomar três tipos de decisões:

- 1) determinar se os métodos de ensino e o material instrucional utilizados no desenvolvimento de um programa são realmente eficientes;
- 2) identificar as necessidades dos alunos para possibilitar o planejamento da instrução; julgar o mérito dos estudantes para fins de seleção e agrupamento; fazer com que os estudantes conheçam seu progresso e suas deficiências; e
- 3) julgar a eficiência do sistema de ensino e dos professores etc.

Assim, de acordo com Cronbach (1963), no primeiro caso, a avaliação permitiria decisões que levariam ao aperfeiçoamento do currículo; no segundo, referir-se-ia aos alunos submetidos a

determinado programa; e, finalmente, no terceiro e último caso, as decisões seriam de natureza administrativa.

### 2.3 Avaliação de curso — problemas

Cronbach (1963), em seu artigo, discute, particularmente, alguns pontos que não podem ser ignorados pelo avaliador educacional. Inicialmente, enfatiza o seguinte: — **quando a avaliação visa ao aprimoramento de cursos, seu principal objetivo é verificar quais os efeitos do curso, ou seja, quais as mudanças que produz no estudante.** O problema, segundo a perspectiva de Cronbach, não está em determinar se um curso é eficiente ou ineficiente apenas. É preciso lembrar que os resultados da instrução são multidimensionais e, desse modo, a avaliação deve promover o mapeamento de todos os efeitos do curso em cada uma de suas dimensões. Um erro freqüente está na concentração em um único escore de diversos desempenhos esperados após a realização de um curso. Isso pode ser enganador, pois o insucesso numa dimensão pode ser compensado pelo sucesso em outra. Os escores compostos englobam e, muitas vezes, ocultam julgamentos sobre a importância de vários resultados; desse modo, para fins de avaliação, é importante que os resultados sejam apresentados separadamente, a fim de que se tenha uma idéia real das mudanças que estariam ocorrendo no estudante como decorrência da influência exercida pelo currículo.

**A avaliação presta um grande serviço quando identifica os aspectos do curso que necessitam de revisão,** no dizer de Cronbach (1963). É evidente que qualquer especialista em currículo gostaria de apresentar evidências sobre a eficiência do seu produto, entretanto, conforme observa Cronbach, costuma haver certa relutância na aceitação de uma **avaliação externa.** O procedimento habitual consiste em submeter o produto à avaliação somente depois de terminado, com vistas a uma confirmação do que foi previamente estabelecido. Esse comportamento não proporcionará bons resultados, além de traduzir um menosprezo pelo importante papel que o avaliador pode efetivamente desempenhar. A fim de apresentar um papel influente, a avaliação deve acompanhar o desenvolvimento do currículo, quando o mesmo ainda se acha em estado fluido, para usar

palavra de Cronbach (1963). A avaliação, assim entendida, possibilita a criação de conhecimentos sobre a natureza das habilidades que constituem os objetivos educacionais do projeto.

**A comparação de cursos não deve ser o objetivo dominante da avaliação**, na concepção de Cronbach (1963). A comparação de resultados de avaliações de cursos deve ser cautelosa para evitar que decisões errôneas sejam tomadas. As diferenças entre os escores médios de diferentes cursos, geralmente, costumam ser pequenas, em virtude da grande diferença que há entre e intragrupos submetidos ao mesmo curso. A impossibilidade de equalizar diferentes grupos prejudica a interpretação dos resultados e representa um problema difícil de superar nos estudos comparativos de cursos. Além do mais, ainda conforme Cronbach (1963), em experimentos educacionais é difícil ocultar dos estudantes o fato de que integram um grupo experimental, sendo igualmente complexo o controle do viés dos professores numa situação experimental. Ocorrem mudanças comportamentais e, assim, nem sempre se pode afirmar com convicção que um determinado resultado decorre, efetivamente, da própria inovação ou se é decorrência do fato de alunos e professores terem sido colocados diante de uma situação nova e experimental.

A comparação entre grupos pode oferecer resultados equívocos; desse modo, Cronbach (1963) propõe que estudos formais sejam planejados sobretudo para determinar o desempenho após o curso por um grupo perfeitamente conhecido, a fim de verificar objetivos importantes e a ocorrência de efeitos colaterais. Cronbach (1963) chama a atenção para o fato de que um experimento em que os tratamentos comparados diferem em inúmeros aspectos, a ocorrência de uma pequena diferença numérica em favor da situação nova não significa grande coisa e não contribui para o aumento de nossos conhecimentos. Estudos analíticos, em escala menor, mas em condições controladas, de versões alternativas de um mesmo curso, oferecem melhores resultados do que pesquisas de campo que aplicam tratamentos dissimilares a grupos diferentes.

## 2.4 Informações quantitativas e qualitativas

Um problema posto em destaque por Cronbach (1963) refere-se aos procedimentos de medida, que variam em função do tipo de decisão a ser tomada. O instrumento usado para uma decisão administrativa deverá ser compreensivo e possuir todas as características psicométricas desejáveis, o mesmo ocorrendo para decisões individuais; contudo, para a avaliação de um curso, o emprego da técnica de amostragem de itens oferece elementos suficientes para um julgamento preciso.

A avaliação de um curso é freqüentemente identificada com a aplicação de um teste; no entanto, conforme acentua Cronbach (1963), existem diferentes maneiras de analisar o desempenho de um estudante, e este desempenho, por sua vez, não constitui o critério único para a avaliação de um curso. A avaliação pode ser realizada por intermédio de levantamentos de opiniões, relatórios de professores usuários do curso (observação de comportamento), observações sistemáticas, medidas ou estudos do processo, medidas de proficiência, medidas de atitudes e estudos longitudinais (**follow-up**), destaca Cronbach em seu artigo.

O levantamento de opiniões de especialistas sobre um determinado curso constitui um procedimento perfeitamente legítimo, ainda que implique diferentes problemas, sobretudo se forem influenciadas por preconceitos, o que, naturalmente, dará margem a controvérsias. O posicionamento em relação a muitas metodologias de ensino pode ser polêmico, sendo necessária a sua experimentação para conhecimento de seus efeitos educativos. A opinião de diferentes especialistas pode servir, segundo Cronbach (1963), para avaliar esses efeitos.

As opiniões manifestas por professores, durante o desenvolvimento de um currículo, são importantes, pois graças aos seus relatórios, que traduzem observações de comportamentos, positivam-se os pontos críticos, as deficiências e todos os demais elementos que estão a exigir maiores cuidados, talvez substituição ou mesmo supressão total da parte comprometida. A observação do professor, ainda que não seja de natureza sistemática, reveste-se de grande importância para a avaliação educacional (Cronbach, 1963).

A observação sistemática é mais imparcial e, freqüentemente, consegue apresentar diferentes aspectos com maior profundidade; contudo, apresenta, também, alguns problemas, como o de custos e o referente à demora que geralmente introduz entre o momento do ensino e a análise dos resultados. Este tipo de observação é particularmente útil durante a fase intermediária de desenvolvimento do currículo, quando as deficiências mais evidentes já foram eliminadas.

O estudo do processo refere-se aos eventos que ocorrem em sala, sendo, pois, de especial valor para o aprimoramento do curso. Nesse estudo, o registro de eventos não se limita àqueles relativos ao estudante durante a fase de instrução, procura, também, enfocar o procedimento do professor em sala de aula. Este conjunto de atividades dos alunos e dos professores possibilita ao construtor de currículos diagnosticar deficiências e propor modificações.

A medida da proficiência, ressalta Cronbach (1963), utiliza de várias técnicas. O uso da amostragem de itens — **itens diferentes para alunos diferentes** — oferece mais vantagens do que o emprego dos mesmos itens para todos os estudantes, na avaliação de cursos. Este tipo de medida pode utilizar-se, ainda, de questões de redação, de questões de resposta aberta curta e, se for o caso, de observação de indivíduos ou de grupos na abordagem de problemas de pesquisa ou na tentativa da solução de um problema complexo<sup>8</sup>.

---

<sup>8</sup> O **item sampling** (amostragem de itens) ou **matrix-sampling** possibilita fragmentar um número elevado de itens e aplicar cada um dos novos conjuntos de forma aleatória a uma amostra aleatória de estudantes, com estimativas bastante aceitáveis da média do teste completo e, na maioria dos casos, de sua variância, excetuando-se, naturalmente, os casos de testes de velocidade (**speed tests**) ou testes heterogêneos, em que a variância será afetada. O **item sampling** é importante na avaliação de programas, em que o interesse é no desempenho de grupos, e no caso de mensuração de mudanças com medidas comparáveis ao longo do tempo. A técnica não se aplica quando medidas de intercorrelações são desejáveis, ainda que possam ser testadas hipóteses sobre correlações se procedimentos adequados forem acautelados. O **item-sampling** é adotado pelo **National Assessment of Educational Progress - NAEP**, e na definição de normas pelo **Educational Testing Service - ETS**. A tecnologia do **item sampling** é apresentada com bastante clareza por K. A. Sirotnik — **Introduction to Matrix Sampling for the Practitioner**, no livro editado por W. J. Popham — **Evaluation in Education - current applications**, Berkeley, California, McCutchan Publishing Corporation, 1974.

Cronbach (1963) destaca o fato de considerar mais importante os dados relativos a um item do que os escores do teste. O escore global pode dar confiança ou não em relação a um determinado currículo, mas pouco informa como aprimorá-lo. Outro aspecto importante em avaliação de currículo, conforme mostra Cronbach (1963), refere-se à *ajustagem do instrumento de medida ao currículo*, que não precisa ser motivo de preocupação para o avaliador. Ainda que isso possa parecer surpreendente, porque em outras situações isso não deve ocorrer, na avaliação de currículo, contudo, e em termos ideais, todos os tipos desejáveis de proficiência devem ser medidos e não apenas os de alguns objetivos selecionados pelo construtor do currículo. Se há um interesse em saber se o currículo alcança os seus objetivos, impõe-se o ajustamento do teste ao currículo, mas se a intenção é a de verificar até que ponto o currículo atende a interesses mais amplos, o ajustamento não é necessário, tendo em vista o desejo de que sejam verificados todos os objetivos possíveis.

Outro ponto enfocado por Cronbach (1963) diz respeito à distinção entre testes factuais e testes para verificar processos mentais complexos, segundo a terminologia de Bloom. A classificação dos itens de acordo com as categorias da Taxonomia é difícil e às vezes impossível. Se uma resposta traduz conhecimento ou compreensão depende de como o aluno foi ensinado e não apenas da questão apresentada. Aplicar um teste apenas para verificar se o aluno "sabe" ou "não sabe" um certo assunto não é inteiramente relevante para fins de avaliação de um curso, de acordo com Cronbach (1963), importando, isto sim, medir o conhecimento em termos de *profundidade e relacionando à capacidade de aplicá-lo a novas situações (habilidade de aplicar)*.

A questão da especificidade das questões é considerada por Cronbach (1963), que discute, detalhadamente, o problema. A especificidade, na maior parte das vezes, concentra-se no uso de uma terminologia própria do curso, que somente é compreendida por aqueles que tiveram a oportunidade de assisti-lo. Ainda que o conhecimento dessa terminologia seja importante, para fins de avaliação, entretanto, é mais significativo a medida de compreensão, de relação e de outras variáveis do curso, que, em princípio, poderiam

ser verificadas, também, em quem não foi diretamente submetido ao curso em questão.

## 2.5 Medidas de atitude e estudos longitudinais

Voltando, novamente, aos diferentes modos de avaliar um curso, Cronbach (1963) discute, ainda, as medidas de atitude e os estudos do tipo **follow-up** (longitudinal)<sup>9</sup>.

A formação de atitudes parece ser uma preocupação constante de todos os responsáveis por desenvolvimento de currículo, pois a expectativa é a de que cada atividade relacionada à aprendizagem contribua para o desenvolvimento de atitudes além da área curricular específica. A mensuração de atitudes pode ser feita de diferentes modos: entrevistas, questionários e outras técnicas semelhantes, sem dúvida valiosas, mas cujos resultados, como adverte Cronbach (1963), devem ser aceitos com as devidas cautelas, porque sujeitos a distorções.

Os estudos de **follow-up** possuem ponto de contacto com as técnicas de observação; no entanto, conforme Cronbach (1963), muitas vezes se mostram de pouca valia para fins de aprimorar um curso ou explicar seus efeitos, em virtude do distanciamento temporal da fase de instrução inicial. Outro problema no **follow-up** é que este avalia os efeitos de um curso como um todo, não sendo possível detectar, por exemplo, desvios da situação ideal, diferenças na eficiência das várias partes do curso, ou, ainda, diferenças de um item em relação a outro item do programa, no dizer de Cronbach (1963). A necessidade, às vezes impossível, de igualar os grupos experimentais e os de controle, constitui outro problema nos estudos de **follow-up**, que, apesar de apresentarem esses pontos críticos, são importantes, especialmente para fins de pesquisas sobre a construção de novos currículos.

---

<sup>9</sup> Os estudos longitudinais, além de exigirem grandes investimentos financeiros, demandam o emprego de complexo tratamento estatístico. Vide, por exemplo, entre outras referências, as seguintes: GOLDSTEIN, H. **The Design and Analysis of Longitudinal Studies**. Their Role and Measurement of Change. London, Academic Press, 1979; PLEWIS, I. **Analysing Change**. Measurement and Explanation Using Longitudinal Data. New York, John Wiley and Sons, 1985; VON EYE, A. **Statistical Methods in Longitudinal Research**. Boston, 2 vols.. Academic Press, Inc., 1990.

As idéias de Cronbach no seu ensaio de 1963 são extremamente provocadoras, no bom sentido, e, pessoalmente, confessamos o nosso débito para com elas, tendo em vista que contribuíram para a formação do nosso pensamento nas atividades de avaliação, justificando-se, assim, a insistência com que temos procurado disseminar esse conjunto bastante coerente de idéias fundamentais para o praticante da avaliação. Assim como Cronbach, acreditamos que as mensurações são importantes em avaliação educacional, mas não se deve ficar limitado ao escore pelo escore, ou apenas para fins de comparação de posições. O importante, no caso, é que saibamos usar esses escores para uma descrição bastante ampla dos resultados obtidos; por outro lado, é preciso, igualmente, que esses escores sejam indicativos das mudanças ocorridas e possibilitem identificar pontos críticos do currículo, do curso, do programa que precisam ser alterados ou submetidos a uma rígida revisão. A avaliação educacional, quando devidamente conduzida, serve para reduzir o mito do valor do escore (bruto) ou de compósitos; na verdade, a grande fonte de informações está na análise das questões/itens e nas diversas respostas apresentadas, pois um item, isoladamente, fornece mais informações do que um simples escore. A amostragem de itens aplicada a uma amostra de alunos, para fins de avaliação, na perspectiva de Cronbach, possibilita a superação de uma série de problemas decorrentes do emprego do mesmo instrumento a todos os sujeitos.

## 2.6 Planejamento geral e interno

Cronbach é uma das raras figuras na área educacional que domina os múltiplos e complexos caminhos da avaliação em suas diferentes perspectivas. Além do importante artigo sobre avaliação de currículo (1963), escreveu livro capital — **Designing Evaluation of Educational and Social Programs** (1982) em que discute, com rara felicidade e grande profundidade, os múltiplos aspectos da avaliação educacional, como mostrou Shrinkfield (In: Madaus et al. (Ed.),

1993)<sup>10</sup>. Alguns dos pontos considerados por Cronbach (1982), envolvendo o planejamento da avaliação, merecem ser destacados, tendo em vista a sua relevância para o contexto geral da avaliação, o que faremos resumindo, conjuntamente, o pensamento de Cronbach (1982) e as observações de Shrinkfield (1993).

Cronbach é de opinião de que a avaliação tem uma **função política**; desse modo, as reações aos dados da avaliação têm uma motivação política. O planejamento de uma avaliação deve ter flexibilidade suficiente para atender à diversidade dos interesses das suas várias audiências, com expectativas as mais diversas possíveis. A avaliação, acreditamos, afeta diferentes áreas de poder (política), devendo, pois, enfrentar desafios.

A avaliação é geralmente realizada por solicitação de um administrador que vê o avaliador como um técnico, ou seja, alguém que domina as técnicas de amostragem, conhece a tecnologia da construção dos instrumentos e é capaz de usar as mais sofisticadas metodologias estatísticas, sem considerar, entretanto, as possíveis implicações daquilo que o avaliador faz ou deixa de fazer, o que dá ao avaliador uma grande responsabilidade. A aprovação de um planejamento de avaliação deve merecer cuidadoso exame, envolvendo diferentes especialistas competentes para a decisão final relacionada à aprovação do projeto.

Nenhum avaliador sozinho, na opinião de Cronbach, possui um conjunto de qualificações, por mais capaz que seja, para tomar todas as decisões relacionadas ao planejamento e à interpretação dos dados. A responsabilidade na avaliação deve ser compartilhada por um grupo — a equipe de avaliação, o que possibilita uma análise e discussão sob diferentes perspectivas. O planejamento de uma avaliação, se bem considerarmos o problema, deve realizar-se, assim, em dois níveis: um **planejamento geral** com vistas a definir prioridades e alocar responsabilidades, e um **planejamento interno**, intra-equipes, que deve basear-se na experiência e na interação dos que formam os vários

---

<sup>10</sup> Sobre avaliação de programas e seu planejamento ver, especificamente, em Cronbach (1982), as pp. 3-4; 13-18; 266; 268; 280. A respeito das fases convergente e divergente da avaliação, para maiores detalhes, consultar, especialmente, as páginas 210-239 e 323 da mesma obra de Cronbach (1982).

grupos ou equipes. O planejamento, desse modo, segundo Cronbach (1982), torna-se uma reflexão sobre o processo, perde o caráter mecânico e padronizado que costuma oferecer em algumas situações. A avaliação, no pensamento de Cronbach, é um caminho para eliminar complexos mecanismos e chegar a decisões sobre ações.

O planejamento de uma avaliação deve fundamentar-se na concepção do que seja efetivamente uma boa avaliação. Um bom planejamento está associado ao aumento de benefícios decorrentes da avaliação e a escolha das alternativas de planejamento deve levar em consideração de que forma a avaliação afetará a educação, o ensino. A avaliação, nos termos apresentados por Cronbach (1982), objetiva, entre outros aspectos, acelerar o processo de aprendizagem das instituições educacionais, por intermédio de informações que, de outra forma, não seriam consideradas ou poderiam ser interpretadas erroneamente.

A avaliação precisa concentrar-se em um conjunto de atividades científicas que garantam a congruência entre as observações apresentadas e o mundo da realidade. Além disso, as atividades científicas devem possibilitar interpretações aprofundadas sobre os dados levantados. Uma avaliação não terá um crédito relevante se tudo aquilo que o avaliador estabeleceu e aprendeu não for incorporado ao conhecimento de diferentes públicos: alunos, professores, pais, administradores, burocratas e a todos os cidadãos interessados em problemas educacionais.

O valor de um avaliador não está apenas em levantar questões, mas em dar respostas aos problemas, comunicando-as de uma forma eficiente: de modo claro, rapidamente, com fidedignidade e validade, oferecendo informações alternativas que satisfaçam às várias audiências. Tudo isso para que a mensagem do avaliador seja compreendida, mereça credibilidade, ofereça respostas significativas, altere prováveis posições preconcebidas e estabeleça um diálogo enriquecedor, conseqüente e, finalmente, permita decisões adequadas [3]. Ainda que nem todos se deixem convencer por seus argumentos, o agir do avaliador deve ser nesse sentido, para que o máximo de pessoas possam usar os dados levantados e as conclusões estabelecidas, mas isto deve fazer parte do seu plano de ações definidas pela equipe de avaliação. A avaliação deve esclarecer

controvérsias, dirimir dúvidas sobre falsos pressupostos e possibilitar ações que resultem da compreensão do objeto avaliado.

Os contrastes entre os aderentes da abordagem científica da avaliação e os entusiastas — às vezes por demais entusiastas — da abordagem holística ou naturalista levam a uma polarização que uma reconciliação parece impossível. Cronbach, assim como nós outros também, acredita que há muitos exageros em ambas as posições e que uma posição conciliatória, uma mistura de estilos, no dizer de Cronbach, é possível, bastando para isso que as mentes estejam abertas e uma boa comunicação se estabeleça entre ambos. A avaliação é um campo próprio para qualquer tipo de investigação, pois, como acentua Cronbach (1982), somente assim a verdade de uma situação pode ser avaliada.

Um dos problemas que o avaliador deve enfrentar consiste na distribuição do esforço investigativo num momento específico, exigindo uma preocupação maior com o equilíbrio de diferentes aspectos do planejamento. Assim, assinala Cronbach, um certo aspecto pode exigir mais ênfase no experimentalismo, enquanto em outros um posicionamento naturalista seria mais adequado. As duas abordagens podem ser usadas em função da natureza da avaliação, ainda que na maioria dos seus trabalhos o planejamento seja do tipo **estudo de caso** ou de **avaliação iluminativa**. A explicação para esse fato decorre de que Cronbach acredita que as abordagens naturalistas ainda não sedimentaram seus fundamentos, o que precisaria ser feito com o aprofundamento das discussões.

Há uma concordância geral, como bem mostra Cronbach (1982), que as instituições sociais, inclusive as educacionais, precisam ser inovativas, o que nem sempre ocorre e, assim, elas acabam ficando muito aquém do que poderiam ser, o que demonstra a necessidade de constantes avaliações.

## 2.7 Humanistas e behavioristas

Cronbach lembra-nos que entre avaliadores humanistas e **behavioristas** há concordância de que a avaliação de um programa deva ser empírica e os eventos examinados nos locais em que ocorrem. O problema estaria possivelmente na idéia de que as conclusões sobre

os fatos podem ser inferências que vão muito além dos dados. As conclusões podem ser falíveis e basearem-se em pressuposições ou meras hipóteses de trabalho; daí precisarmos considerar que as conclusões podem ser afetadas pelo seu maior ou menor grau de plausibilidade. O avaliador, especialmente aqueles que seguem uma metodologia quantitativa, precisam lembrar que as estatísticas são acompanhadas por graus de incerteza, o que nos faz considerar que a avaliação não se pode apoiar inteiramente em métodos estatísticos e nas inferências decorrentes dos resultados. Tudo isso indica que o avaliador, diante dessa situação, deva fazer opções que melhor atendam ao seu caso específico.

A partir do anteriormente colocado, apercebemo-nos que alguns avaliadores, no planejamento e na prática, usam ambas as abordagens, a científica e a humanista. Outros preferirão um estilo uniforme num determinado estudo. Aqueles que favorecem testes somativos possivelmente optarão pela objetividade na observação e análise; ao contrário, a opção por estudo de caso levará a advogar uma interpretação do tipo impressionista. Quando o avaliador, ao planejar, não consegue identificar, o estudo deve prosseguir e a identificação far-se-á no contexto, a partir dos dados coletados e interpretados pelo avaliador com base na sua experiência direta ou na vicária. Um planejamento é considerado forte se identifica apropriadamente, e formula, claramente, as variáveis, permitindo previsões objetivas, reproduzíveis e inquestionáveis, teoricamente fundamentadas.

Cronbach (1982) mostra que não há incompatibilidade entre um controle experimental e o uso de informação qualitativa ou interpretação subjetiva. Um experimento formal pode incorporar entrevistas, por exemplo, e, por outro lado, dados quantitativos podem ser usados na construção de hipóteses para um estudo de caso, sendo possível, ainda, combinar um planejamento forte com interpretação subjetiva. Um observador naturalista, por sua vez, pode introduzir objetividade no seu estudo documentando incidentes à medida que ocorrem e empregando vários observadores para focar um aspecto particular do programa que está sendo avaliado.

A concepção de Cronbach é no sentido de que há um **continuum** entre planejamento científico e humanístico em avaliação e que toda a retórica entre o polarismo de experimentação e estudo de

caso numa visão naturalista é útil por destacar questões críticas para a avaliação educacional. Alguma forma de controle na coleta de dados, na manipulação dos tratamentos e na observação naturalista, sendo importante encontrar um equilíbrio entre os dois estilos que podem ser usados em diferentes momentos da evolução de um programa avaliativo.

## 2.8 Planejamento — fase divergente

Uma avaliação leva, necessariamente, a conclusões que expressam diferentes expectativas de várias audiências em relação às decisões a serem tomadas, inclusive propostas de mudança. Isso significa que o avaliador, na fase de planejamento, se deva preocupar por todos os meios possíveis para que essas conclusões sejam válidas. O que nos leva ao que Cronbach (1982) denominou de **fase divergente e convergente** do planejamento da avaliação. O avaliador não é livre para propor questões a serem respondidas pela avaliação. As questões/problemas decorrem de incertezas dos que constituem a comunidade responsável pela tomada de decisões e a implementação de ações. A fase divergente, quando possíveis questões são levantadas, e a fase convergente, quando prioridades são estabelecidas para essas questões, sendo aconselhável que os responsáveis pela solicitação de uma avaliação considerem essas duas fases **antes** de ser solicitada a colaboração de uma equipe de avaliação.

A **fase divergente** do planejamento é importante porque é nesse momento que o avaliador e administradores formulam o máximo possível de indagações a serem consideradas a respeito da avaliação. Perguntas devem ser feitas — **a quem?**, eis um problema a considerar, tendo em vista que possíveis omissões podem contribuir para que se tenha uma visão restrita do quadro a ser avaliado. Ainda que o avaliador se situe em melhores condições para propor questões, a decisão sobre a sua relevância e inclusão no projeto acaba sendo feita em conjunto pelo avaliador e administradores, considerando suas implicações políticas. É nessa fase que se formam grupos de pressão, pretendendo impor idéias e orientações pessoais, mas os possíveis atritos devem ser superados e as realidades acabam por prevalecer, face, inclusive, às condições políticas. Uma **fonte** de questões, nesta

fase, é representada pelos objetivos propostos para a avaliação. A lista de objetivos deve ser a mais completa possível, abrangendo, inclusive, objetivos não diretamente desejáveis. Cronbach, aliás, ressalta que a relação de objetivos será incompleta se dela não constarem, ou melhor, não for dada suficiente atenção aos **unwanted outcomes**, aos efeitos colaterais da avaliação, que são outro aspecto a considerar, nessa fase, ainda que se saiba que são difíceis de alcançar depois de completada a avaliação, o que exigirá um estudo posterior de **follow-up**. Vemos, assim, que nesta fase o avaliador deve ouvir diferentes segmentos da comunidade profissional e não-profissional a fim de identificar diferentes esperanças e, também, temores dos vários grupos da sociedade.

## 2.9 Planejamento — fase convergente

A **fase convergente** destaca a necessidade de que as questões sejam propostas para uma diversidade de indivíduos e grupos. Deve-se considerar que muitas dessas questões serão abandonadas por razões práticas e não serem prioritárias. O número de variáveis a serem tratadas em uma avaliação muitas vezes precisa ser reduzido por razões de custos, tendo em vista limitações orçamentárias; por sua vez, uma avaliação se torna por demais complicada e difícil de administrar, quando deve considerar uma massa muito grande de informações e, conseqüentemente, muita coisa é perdida no decorrer da redução dos dados e na sua síntese. Por fim, precisa-se considerar, também, que na comunidade responsável pela tomada de decisão poucas são as pessoas que querem conhecer todo o detalhamento da avaliação e nem sempre têm disponibilidade de tempo para manifestar sua opinião a respeito de tudo, ainda que isso fosse desejável.

É nesse momento da fase convergente, mostra Cronbach (1982), que é identificado o que possui valor para ser avaliado; é nesse momento que é decidido o grau de tolerância para a omissão de certas variáveis, considerando-se que o avaliador deve investir seu tempo e esforço naquilo em que tem certeza de que vai obter resposta, há possibilidade de conseguir amplas informações, os custos da avaliação não serão aumentados, e as opções oferecidas pelo avaliador têm alto peso entre as informações necessárias ao processo decisório, ou seja, a

esperança é a de que cada uma das respostas dadas pela avaliação tenha importância nas decisões que serão tomadas.

### 2.10 Perda de informações — as audiências e a equipe

Cronbach (1982) destaca o problema da perda de informações na avaliação. O observador nem sempre comunica o total das suas observações, sendo algumas perdidas no processo de codificação e sumário das estatísticas, enquanto outras informações são omitidas do relatório. Há um processo de seletividade e significação das informações.

Surge, aqui, a questão das várias audiências do avaliador, que são distintas e nem sempre identificadas com a linguagem e o estilo de pensamento do avaliador responsável pelo relatório da avaliação. É necessário que o avaliador indague sobre o que relatar sem dificultar o pensamento do usuário da avaliação, e quais as implicações que isso poderá ter nas primeiras operações de campo da avaliação. O problema da comunicação é, assim, um dos aspectos mais relevantes na disseminação das informações e Cronbach compara o papel do avaliador ao do jornalista, que investiga o que é de interesse público, seleciona o que merece a atenção do público e apresenta as coisas de forma a atrair a atenção do público, o que nem sempre ocorre na avaliação educacional.

Cronbach, Stufflebeam e, anteriormente, Stake acreditam que o trabalho de avaliação deva ser dividido entre todos os membros da equipe, que age, entretanto, como uma unidade solidária. A equipe deve ter uma perspectiva de tudo aquilo que acontece e possa, depois, interpretar corretamente os dados. Ao final, Cronbach (1982) mostra-se otimista quanto à importância da avaliação, na medida em que a sociedade passa a utilizar suas informações para iluminar suas linhas de ação.

### 2.11 Notas e comentários

- [1] Ainda que o assunto — **medidas e avaliação** — já tenha merecido uma certa cobertura, ao fazermos **considerações teóricas e posicionamentos**, voltamos a ele para mais um esclarecimento, tendo em vista que o assunto ainda se acha difuso (e confuso) em nosso contexto educacional. É necessário

não confundir avaliação, **assessment** e medida, palavras que não podem ser usadas de forma intercambiável, ainda que isso ocorra com relativa frequência. Não são sinônimos, diz Choppin, In: Walberg e Haertel (1990), e nem devem ser confundidas com **testing**, sobretudo quando se considera que os tipos de teste usados nos três processos podem apresentar grandes diferenciações.

**Medir** é, na sua definição clássica, anteriormente apresentada, atribuir números a um certo atributo, segundo determinados critérios. As medidas em educação referem-se a atributos, sendo, pois, medidas indiretas. O ato de medir pode ser básico na pesquisa, mas não necessariamente na avaliação. As medidas em educação, realizadas por meio de testes/provas, resultam em escores brutos, que, por sua natureza não-linear e a falta de certeza quanto ao significado do **zero** e do **escore perfeito**, devem ser transformados em escalas, para que se revistam de significado.

**Assessment** é a avaliação aplicada a pessoas, como o caso do rendimento escolar de um estudante, da atitude de um aluno em relação a um curso ou a competência profissional de um professor. A tradução de **assessment** como rendimento escolar, o que ocorre com bastante frequência, parece-nos bastante limitada, face ao exposto. **Assessment** está relacionado a **um** indivíduo. Quando os escores de vários indivíduos são agregados, para fins de análise, interpretação e discussão, atividades que nos permitirão fazer afirmações do tipo — os alunos da região A possuem nível de desempenho melhor que os da região B — no caso estamos diante de uma **avaliação** e não mais de um **assessment**. Exemplo: os resultados do Sistema Nacional de Educação Básica no Brasil (SAEB) traduzem uma avaliação, o mesmo ocorrendo com os escores agregados do **National Assessment of Educational Progress**, nos Estados Unidos.

A palavra **avaliação**, por outro lado, deve ser utilizada em relação a entidades abstratas, como programa, currículo ou variáveis organizacionais. O uso da expressão **avaliação educacional** implica julgamento de valor, conforme o destaque de Scriven (1967), e a palavra avaliação deve ser empregada quando se fazem comparações entre programas, currículos ou entre esquemas organizacionais.

- [2] É importante que se esclareça o seguinte: — quando as pessoas falam em **avaliação de currículo** podem estar significando diferentes coisas, como diz Alkin, In: Madaus et al. (1993): — avaliação de um produto do currículo, ou avaliação de um programa curricular. A avaliação de produto — cursos, livros, materiais etc — emprega critérios externos, que possibilitam julgar em que medida o produto possui adequação curricular, atende aos objetivos estabelecidos, desenvolve as habilidades esperadas e é adequado ao nível de conhecimento, à idade, à etnicidade e ao próprio **background** sócio-econômico do aluno. Os critérios externos antes referidos podem estar relacionados ao nível de leitura (dificuldade), à adequação aos objetivos, ao

conteúdo instrucional, à metodologia instrucional e à validação (impacto). A avaliação formativa ou somativa de um produto visa a alterá-lo durante o seu desenvolvimento ou a validá-lo antes da sua divulgação.

A avaliação de programa, na perspectiva de Alkin (1990), refere-se ao complexo de interações entre um certo programa instrucional e o seu cenário, ou seja, preocupa-se em verificar como um determinado currículo funciona no processo instrucional. Ainda que, sob alguns aspectos, a avaliação de programa se superponha à avaliação de currículo, dela, entretanto, se afasta, pois o seu interesse maior não se concentra no currículo **per se**, mas o currículo como uma parte operacional do programa.

A avaliação de programas curriculares apresenta diferentes subconjuntos, como mostra Alkin (1990), que podem ser caracterizados da seguinte forma: 1) **avaliação orientada pela medida dos resultados** (objetivos), que é o modelo de TYLER e seus seguidores; 2) **avaliação orientada por pesquisa ou metodologia**, em que aspectos particulares da metodologia são considerados, como, por exemplo, na abordagem de GUBA, que procura realçar a riqueza que os dados qualitativos e descritivos podem trazer para a avaliação; 3) **avaliação orientada por valores** na qual os valores são o principal elemento verificado pelos teóricos, casos de SCRIVEN e Eisner, que julgam caber ao avaliador a emissão de juízos de valor, e em menor ênfase no caso de STAKE; e 4) **avaliação orientada para a decisão**, enfoque que procura destacar a preocupação dos responsáveis pela tomada de decisão e dos usuários das informações da avaliação, o que pode ser observado em STUFFLEBEAM, PROVUS e no próprio Alkin.

- [3] A questão do impacto da avaliação foi objeto de algumas considerações bastante pertinentes de Alkin e Dallack, In: Walberg e Haertel (1990). Os autores discutem que as informações coletadas, analisadas e divulgadas, no caso de uma avaliação educacional, destinam-se à tomada de decisões, segundo a colocação de Stufflebeam e, dizem Alkin e Dallack (1990), à expectativa geral. As informações deverão ter influência sobre as **atividades** e as **atitudes** de pessoas ou grupos na área educacional, o que caracterizaria o impacto. As informações, apesar de influenciarem atitudes e valores, nem sempre influenciam as ações (atividades), tendo um impacto meramente conceitual, que, eventualmente, no futuro, pode vir a resultar em ação. Alguns impactos da avaliação educacional podem ser inesperados, mas como identificar os diversos impactos e como medir o possível impacto de uma avaliação? É, na visão de Alkin e Dallack (1990) e de outros especialistas, um problema realmente complexo, considerando que muitos impactos não são diretamente observáveis, além do que outros elementos que não integraram a avaliação podem estar influenciando nos resultados e no seu impacto. A mensuração do impacto, além do mais, é difícil de ser concretizada, o que faz com que muitos aceitem as repercussões da divulgação dos resultados como um reflexo do impacto. O impacto é, também, inferido em relação às ações

subseqüentes dos responsáveis pela tomada de decisões. A questão do impacto das avaliações é um desafio, tendo em vista que há uma multiplicidade de fatores que atuam concomitantemente no processo, maximizando ou minimizando o impacto. A medida do impacto das avaliações é muitas vezes perturbada por fatores políticos e burocráticos. Ainda não se chegou a estabelecer uma metodologia própria para análise do impacto de avaliações, sendo o problema resolvido, na maioria das vezes, pela realização de um *survey* (entrevistas) e de estudos de caso.

### 3. AVALIAÇÃO E SUA LÓGICA: Michael Scriven

#### 3.1 Introdução

Scriven, sem a preocupação de criar um modelo, ao publicar seu ensaio **Methodology of Evaluation** (1967), desenvolveu uma série de idéias fundamentais para a compreensão da lógica da avaliação educacional. Inicialmente, acreditamos que a sua grande contribuição consistiu em estabelecer que a avaliação desempenha muitos papéis (**roles**), mas possui um único objetivo: determinar o valor ou o mérito do que está sendo avaliado. A diferença entre **roles** (papéis) e **goals** (objetivos) foi, assim, uma das muitas contribuições de Scriven para o desenvolvimento de uma possível futura teoria da avaliação. O **objetivo** consistiria em oferecer uma resposta satisfatória aos problemas propostos pelas questões a serem avaliadas; os **papéis** referir-se-iam às maneiras como essas respostas são usadas. Conceitualmente, haveria diferença entre ambos, tendo em vista que os objetivos estão intimamente ligados a questões de valor, mérito [1].

#### 3.2 Metodologia da Avaliação

A **Methodology of Evaluation** (1967) é uma obra extremamente rica em idéias que hoje estão definitivamente incorporadas ao universo da avaliação e fazem parte do patrimônio pessoal de cada avaliador. Assim, ao diferenciar os papéis *formativo* e *somativo* da avaliação, Scriven estava apresentando dois conceitos que iriam influenciar enormemente o futuro e a prática da avaliação. Scriven mostrou que a avaliação *formativa* deve ocorrer ao longo do desenvolvimento de programa, projetos e produtos educacionais, com

vistas a proporcionar informações úteis para que os responsáveis possam promover o aprimoramento do que está sendo objeto de implementação. As discussões entre especialistas ao longo do desenvolvimento de um programa, o exame de seu conteúdo por pessoas que possuem **expertise** na área de seu conteúdo, a aplicação de diferentes instrumentos de mensuração a grupos de crianças e/ou adolescentes etc. seriam, naturalmente, práticas envolvendo a dimensão formativa da avaliação. Scriven mostrou, ainda, que ao longo de um projeto a avaliação formativa deve ser uma constante, a fim de que os responsáveis possam, usando as informações, realizar as modificações que se fizerem necessárias [2] [3].

A avaliação somativa, conduzida ao final de um programa de avaliação, possibilita, ao seu futuro usuário, elementos para julgar a sua importância, o seu valor, o seu mérito. A avaliação somativa de um currículo, por exemplo, pode ser usada para determinar a eficácia desses currículos, a partir da análise dos dados de uma amostra de escolas, professores e estudantes. A conclusão é a de que a avaliação somativa torna-se indispensável para os usuários em potencial de qualquer currículo, programa ou material.

### 3.3 Avaliação formativa e somativa

Ao considerar estes dois tipos de avaliação, dentro do pensamento traçado por Scriven, é necessário atentar, especificamente, para os futuros usuários ou clientes do que está sendo objeto de avaliação, e, ainda, considerar os papéis da avaliação. Estamos diante de coisas completamente diferentes. A avaliação formativa, no caso de um programa, tem a sua clientela concentrada nos elementos responsáveis pelo desenvolvimento do próprio programa. A avaliação somativa desse mesmo programa teria destinatários diferentes: consumidores potenciais (alunos e professores, por exemplo) ou instituições financiadoras (CAPES, FINEP, Fundação Ford etc.) ou, ainda, supervisores educacionais e os próprios responsáveis pelo desenvolvimento do projeto. O primeiro tipo de avaliação — **formativa** — está diretamente relacionada com a decisão de desenvolver um programa, sua modificação ou, ainda, sua revisão, por exemplo. A avaliação do tipo **somativa** está ligada à decisão de

continuar com o programa, encerrá-lo ou adotá-lo, entre outros aspectos.

Os dois tipos de avaliação — formativa e somativa — são fundamentais, considerando que várias decisões devem ser tomadas no decorrer das diversas fases do desenvolvimento do programa, e que, ao término, dever-se-á determinar o mérito desse programa e empreender outras decisões, inclusive em relação ao futuro do programa, se deve continuar ou ser encerrado definitivamente. Apesar da importância desses dois tipos de avaliação, observamos que, habitualmente, os avaliadores concentram seus esforços em avaliações somativas. É preciso que o avaliador tenha cautela, pois, sem a avaliação formativa, o processo de desenvolvimento, seja de um programa, projeto ou material, resultará incompleto e inteiramente ineficiente. A realização de uma avaliação apenas ao final — somativa, portanto — somente tem possibilidade de muitas vezes constatar o fracasso de um projeto.

As dimensões da avaliação formativa e somativa, conforme o destaque de Worthen e Sanders (1987), podem ser combinadas com as dimensões externa e interna da avaliação, dando origem, dessa forma, a uma avaliação (1) — formativa interna, (2) — formativa externa, (3) — somativa interna, e (4) — somativa externa. A primeira — avaliação formativa interna — é conduzida por um avaliador interno, alguém que conhece o programa. A possível falta de objetividade do avaliador não tem as mesmas conseqüências de uma avaliação somativa. Esta é melhor conduzida por um avaliador externo ou uma agência externa, ainda que isso nem sempre seja possível por razões várias, inclusive de ordem financeira ou por falta de **expertise**. O trabalho de um avaliador interno, na avaliação somativa, seguindo a linha de pensamento de Scriven, nem sempre se reveste de objetividade e credibilidade, mas é possível contornar as dificuldades desde que o avaliador selecionado não seja uma pessoa intimamente identificada com o programa ou o objeto que está sendo avaliado. Estudos de avaliação formativa e somativa são indispensáveis como mostra Scriven, na medida em que desejamos aprimorar o desenvolvimento dos sistemas educacionais.

O papel do avaliador formativo externo nem sempre é considerado, apesar do destaque de Scriven (1972). A sua falta de familiaridade com o programa ou com o seu contexto faz com que

possa não se aperceber dos aspectos negativos do programa que deveria considerar, e pode, ainda, ser influenciado por percepções **a priori** do programa e do seu valor. Um avaliador formativo externo tem uma visão fria da questão, e antecipa o avaliador somativo. A sabedoria, ainda conforme Scriven (1972), indica a necessidade de um avaliador formativo externo [4].

### 3.4 Modelos em avaliação

Ao discutir a questão dos modelos (1974), que são numerosos, superiores a quase meia centena, Scriven destaca que o campo da avaliação é cheio de inquietações metodológicas, o que não deixa de ser sinal bastante positivo, mas que gera confusão e dificuldades para o praticante da avaliação. A questão da diversidade das **abordagens** é resultante de posições epistemológicas diferentes, preferências metodológicas diversas e diferentes visões metafóricas da avaliação. Scriven (1967) coloca a questão do julgamento de valor em avaliação, mas muitos discordam desse posicionamento, adotando uma posição prática e associando a avaliação à tomada de decisão, achando que o juízo de valor deva ser feito por quem toma a decisão, o que não é o caso do avaliador. Outros, acham que a avaliação só se completa se o autor do relatório final, responsável pela decisão, apresentar um juízo de valor. Ou seja, a responsabilidade é do avaliador [5].

### 3.5 Avaliação goal-free

A questão da avaliação independente de objetivos (**goal-free**) foi colocada por Scriven (1973). É preciso partir do princípio de que não se podem aceitar tranquilamente, sem maiores análises, os objetivos fixados pelos avaliadores. Os próprios objetivos precisam passar pelo crivo da análise e ser avaliados. Scriven (1973) mostra que muitos objetivos possuem uma alta dose de retórica e em muitos casos se distanciam da realidade dos objetivos definidos em programas, por exemplo, ou, o que é pior, adulteram completamente suas reais finalidades.

A avaliação **goal-free** defendida por Scriven (1973) teria como principal função reduzir os vieses e aumentar a objetividade das

avaliações. O avaliador, diante de uma avaliação orientada por objetivos, tem um conjunto de objetivos que foram na verdade definidos pelo coordenador do desenvolvimento do projeto, o que limita a sua percepção do problema, impedindo-o de verificar os resultados diretamente relacionados aos objetivos que foram anteriormente determinados. Objetivos colaterais existentes no projeto, mas não previstos no programa, nem sempre podem ser identificados pelo avaliador, ficando o mesmo circunscrito aos objetivos que foram pré-definidos.

Worthen e Sanders (1987) vêem as seguintes características em uma avaliação não orientada por objetivos:

- o avaliador, propositadamente, evita tomar conhecimento dos objetivos do programa;
- não se permite que os objetivos pré-estabelecidos provoquem uma limitação na capacidade de análise da avaliação;
- a avaliação **goal-free** concentra todos os seus esforços em identificar os objetivos reais do programa e não os resultados pretendidos do programa;
- o avaliador **goal-free** tem um contacto mínimo com a equipe responsável pelo programa;
- a avaliação **goal-free** aumenta a probabilidade de antecipar efeitos colaterais dos programas.

A avaliação por objetivos e a **goal-free** não são mutuamente exclusivas, reconhece Scriven (1973). Ao contrário, complementam-se. O avaliador **goal-free** vai determinar até que ponto os objetivos pretendidos pelos criadores dos programas estão sendo realmente alcançados, o que constitui uma tarefa especialmente destinada ao avaliador externo.

### 3.6 Critérios para avaliação de produtos

Houve da parte de Scriven (1967) uma preocupação em estabelecer critérios para a avaliação de qualquer produto educacional. Estes primeiros critérios eram bastante limitados. Assim, mais tarde, Scriven voltou ao assunto (1974), estabelecendo padrões para a avaliação de programas educacionais e produtos. Divulgou uma

listagem (**checklist**)<sup>11</sup> para uso de especialistas, identificando, entre outros, os seguintes aspectos a considerar:

1. **Necessidade** — número de pessoas afetadas pelo produto, significação social, ausência de substitutos, efeitos multiplicativos, provas de sua necessidade.
2. **Mercado** — plano de disseminação do produto, tamanho e importância dos mercados potenciais.
3. **Desempenho real — julgamento de campo** — evidência da eficácia da versão oficial, com usuários típicos, com recursos típicos, em contextos típicos, em um tempo típico definido.
4. **Desempenho-consumidor verdadeiro** — testes com todos os “consumidores” importantes — estudantes, professores, diretores, membros de conselhos distritais, estaduais e federais, Congresso e contribuinte.
5. **Desempenho — comparações críticas** — dados comparativos sobre diferentes competidores, competidores existentes, competidores projetados, competidores criados, competidores hipotetizados.
6. **Desempenho — longo prazo** — evidências sobre os efeitos do produto em tempos estabelecidos, tais como de uma semana a um mês, um mês a um ano, um ano e vários anos mais tarde, sobre fases críticas da existência do produto.
7. **Desempenho — efeitos colaterais** — evidências de estudos ou pesquisas independentes relativas a resultados esperados, imediatamente após o uso num longo período de tempo.
8. **Desempenho — processo** — evidência do uso do produto para verificar a descrição do produto, reclamações, e moralidade no uso do produto.
9. **Desempenho — causalidade** — prova da eficácia do produto por intermédio de um estudo experimental, ou quase-experimental, ou **ex post-facto**, ou um estudo correlacional.
10. **Desempenho — significância estatística** — provas estatísticas da eficiência do produto através de análises técnicas, níveis de significância, e interpretações.

---

<sup>11</sup> Existem várias **checklists** elaboradas por diferentes avaliadores, com variadas abordagens (Worthen e Sanders, 1987).

11. *Desempenho — significância educacional* — demonstrada por intermédio de julgamentos independentes, julgamentos de peritos, julgamentos baseados na análise de itens e em escores de testes, efeitos colaterais, efeitos a longo prazo e ganhos comparativos, uso educacional importante.
12. *Custos-benefícios* — análise abrangente dos custos, incluindo custos do julgamento de peritos, custas de julgamentos independentes, comparação de custos de competidores.
13. *Apoio ampliado* — planos para futura coleta de dados do mercado e aprimoramento, treinamento em serviço, atualização de auxílios, estudo de novos usos e dados sobre usuários.

A longa transcrição de Scriven foi intencional, mas para entendermos a sua preocupação com a avaliação de produtos precisamos considerar o contexto norte-americano, a sua atividade intensa na produção competitiva, o oferecimento dos mais diferentes tipos de produtos educacionais, desde pacotes curriculares até os mais variados tipos de serviços. Tudo dentro do espírito capitalista e da prática da empresa privada. A produção de material didático é uma indústria de vários milhões de dólares que tem usuários conscientes dos seus direitos de consumidores que exigem elevados padrões de qualidade. As colocações de Scriven são importantes, ao menos teoricamente; no entanto, cumpre-nos ressaltar que, dificilmente, qualquer produto, mesmo dentro de uma filosofia de alto controle de qualidade, conseguiria satisfazer a totalidade dos critérios apresentados. Uma seleção dos padrões mais importantes, para um determinado caso, pode ser feita na prática, mas para uma avaliação restrita a determinados aspectos.

### 3.7 Questões conceituais — preocupação

A presente exposição é apenas uma síntese de algumas das idéias básicas de Scriven, cuja preocupação com aspectos conceituais da avaliação levou-o a escrever um número bastante significativo de artigos e monografias, dispersos em publicações de alto nível

internacional. Queremos lembrar que, além dos problemas anteriormente apresentados, Scriven abordou temas de grande relevância, podendo ser destacados, entre outros igualmente significativos, os relacionados, por exemplo, com a necessidade da participação em certos projetos de avaliadores que sejam peritos em áreas específicas, dada a impossibilidade de um único avaliador dominar todos os campos do conhecimento envolvidos em um determinado projeto, como seria o caso da avaliação de um currículo de língua nacional ou estrangeira, em que alguém com **expertise** precisaria integrar a equipe de avaliação, no campo da sua especialização. Scriven preocupou-se, também, com o mito do avaliador sem **bias** (viés), que seria imparcial pois realizaria um trabalho supostamente objetivo. O raciocínio é falacioso, porque não podemos esquecer que nos projetos de avaliação existem relações interpessoais e relações interinstitucionais que podem comprometer a suposta imparcialidade dos avaliadores. Ao contrário de Cronbach, Scriven defendeu e justificou plenamente a necessidade da realização de avaliações comparativas, que trariam um maior aporte de informações, permitindo uma tomada de decisão e o estabelecimento de juízos de valor de uma forma mais segura. A sua formação acadêmica e o constante trato com problemas da filosofia da ciência fizeram com que Scriven chamasse a atenção para a necessidade da análise crítica dos diferentes e numerosos modelos surgidos na avaliação, aos quais faltaria consistência lógica e muitas vezes se limitariam a incorporar elementos de outros modelos. Ainda nessa linha de análise das avaliações, Scriven defendeu a idéia de uma **meta-avaliação** (avaliação da avaliação), o que mais tarde foi concretizado sob a orientação de Stufflebean, como será visto posteriormente, na apresentação dos padrões para avaliação de avaliações. A defesa da avaliação somativa externa é um dos pontos destacados por Scriven, que, entretanto, compreende a impossibilidade de algumas vezes realizá-la, aceitando, com restrições, a avaliação somativa interna, mas acentuando que ao avaliador, nessa situação, faltaria perspectiva para examinar adequadamente o trabalho, além de surgirem dúvidas sobre a imparcialidade do seu julgamento. Uma preocupação de Scriven é com os altos custos da avaliação, mas reconhece que é impossível realizar uma avaliação **cost-free**, assim, a compensação estaria nos benefícios

que seriam decorrentes da avaliação, em que o principal beneficiário acabaria sendo o aluno, o currículo e a própria instituição [6].

### 3.8 Notas e comentários

- [1] A presente nota ressalta alguns aspectos da significativa obra de Scriven que não foram mencionados ou explorados devidamente no corpo principal do capítulo a ele dedicado. A obra de Scriven situa-se no conjunto das reações ocorridas ao modelo de Tyler, especialmente face à dificuldade do seu emprego na avaliação em física, química e matemática, entre outras áreas, que foram objeto de avaliação após a surpresa do Sputnik em 1957, para usar a expressão de Guba e Lincoln (1981). Scriven teve um papel destacado ao analisar criticamente a plethora de avaliações posteriores à legislação federal de 1965 (ESEA), escrevendo uma de suas obras metodológicas (1967) mais significativas sobre avaliação, onde de maneira bastante contundente declara de forma expressa que as atuais concepções — referia-se à década de 60 — da avaliação dos instrumentos educacionais (novos currículos, textos de instrução programada etc) eram inadequadas do ponto de vista filosófico e prático. A sua contribuição ao escrever o ensaio **Metodologia da Avaliação** (1967) foi no estabelecimento da diferença entre avaliação formativa e somativa, ou entre aprimorar e julgar o **evaluand**, palavra por ele criada para designar a entidade que está sendo avaliada. Estabeleceu, ainda, distinção entre o avaliador profissional e o amador, atribuindo àquele a responsabilidade de apresentar juízos de valor; discutir a questão dos objetivos na avaliação do rendimento escolar e o valor da sua concretização, além de, entre outros aspectos, considerar a questão da comparabilidade das avaliações, concluindo pela sua utilidade, em oposição à concepção contrária de Cronbach.

Os anos de 1972 a 1974 são de intensa produtividade para Scriven, que discute o problema da subjetividade em avaliação, assunto que será por nós destacado na parte relacionada às posições de Stake; apresenta a idéia de uma avaliação **goal-free**, à qual já nos referimos anteriormente no corpo do capítulo e que será objeto de uma nota especial; discute a possibilidade de maximizar o poder da investigação causal, apresentando o método **Modus Operandi**; e debate o pró e o contra da avaliação **goal-free** por ele próprio apresentada. Scriven, ainda nessa época, levanta a questão dos efeitos intencionais e não-intencionais, dicotomia que considerava arbitrária, sendo igualmente importante o conhecimento dos efeitos colaterais na avaliação de um programa. Uma das críticas apresentadas a Scriven, nessa época, foi com relação ao fato de o seu modelo não indicar como seria a operacionalização de uma avaliação **goal-free** (Guba e Lincoln, 1981), permanecendo apenas no nível conceitual; contudo, apesar disso, o seu modelo teve certo impacto porque demonstrou a possibilidade de realizar uma avaliação

independentemente da fixação de objetivos, o que representou um posicionamento revolucionário à época, porque a avaliação considerava não apenas os efeitos desejáveis (objetivos), mas os efeitos que não tinham sido previamente considerados, o que o afastava bastante do modelo tyleriano. Ainda com relação ao modelo **goal-free** é necessário fixar que, apesar de Scriven associar a avaliação a um julgamento de valor, a proposta apresentada não estabelece como seriam os padrões de julgamento, o que fez com que Scriven, posteriormente, admitisse que a avaliação **goal-free** deveria ser usada como auxiliar e de forma paralela às avaliações orientadas por objetivos.

Um aspecto a esclarecer em relação ao julgamento de valor, proposto por Scriven, é quanto ao seu significado, que muitas vezes é confundido com mérito, palavra usada como sinônimo. Guba e Lincoln (1981) esclarecem o problema, mostrando que o valor associado a um **evaluand** pode ser considerado de duas maneiras bem distintas. Uma é o **valor implícito**, inerente, independente de qualquer aplicação, como é o caso das ciências puras. Nessa situação um **valor intrínseco** deveria ser referido a **mérito**, como o fazem Guba e Lincoln (1981). Por outro lado uma entidade (**evaluand**) pode ter um valor, um contexto de uso ou aplicação, como no caso das ciências aplicadas, em que o valor é determinado pelo contexto, é um **valor extrínseco**. A diferença fica mais clara quando Scriven se refere ao valor intrínseco como mérito e ao valor extrínseco como **pay-off**, também usado por Cronbach, enquanto Guba e Lincoln usam a palavra **worth** (1981), discutindo amplamente a questão em sua obra **Effective Evaluation** (cap. 3).

A obra de Scriven é considerável e sempre produz alguma forma de impacto na área da avaliação, merecendo uma análise que extrapola os limites de uma simples nota de fim de capítulo; entretanto, na presente oportunidade, queremos voltar a um problema que aborda no seu ensaio básico — **Methodology of Evaluation** (1967), ao discutir o papel do avaliador, assunto a que voltaremos no último capítulo — **Avaliação e o avaliador: epílogo** —, mas vendo o problema segundo a perspectiva de Stake. Uma questão relevante é proposta por Scriven (1967) — uma avaliação deve ser conduzida por um avaliador profissional, treinado na tecnologia da avaliação, ou por um avaliador não-profissional, mas que detém um domínio da área substantiva, podendo ser considerado um **expert**? A posição de Scriven (1967) é no sentido da profissionalização, tendo em vista que a avaliação exige pessoas altamente capacitadas para emitirem juízos de valor. Parece-nos que a solução para a questão deva ter um outro encaminhamento. A pesquisa e a avaliação tornaram-se um campo extremamente complexo, diríamos mesmo que enciclopédico<sup>(1)</sup>, sendo humanamente impossível a alguém dominar, com profundidade, todos os campos do saber nessas duas áreas, parecendo-nos, pois, que a melhor solução seja a da formação de uma equipe em que o domínio da tecnologia e o do conhecimento substantivo se completam e, pelo menos teoricamente, anulam suas deficiências.

- (1) Ver **The International Encyclopedia of Education: research and studies** editada por Torsten Husen e T. Neville Postlethwaite. Oxford, Pergamon, 1985, 10 vols. E volumes suplementares publicados pela Pergamon Press (1989 e 1990).
- [2] Antes de voltarmos à questão da avaliação **goal-free** [3], apresentamos importantes aspectos da avaliação formativa e somativa [4], abordarmos as linhas gerais da meta-avaliação na perspectiva de Scriven [5] e finalizarmos, brevemente, com o problema das ideologias em avaliação [6], queremos fazer uma ligeira referência a outra idéia importante explorada por Scriven (1967) relativa à avaliação intrínseca.

Scriven com a expressão **avaliação intrínseca** visou a caracterizar uma avaliação de programas e materiais que se concentra mais intensamente na sua natureza intrínseca (**inerente, peculiar**) do que propriamente nos efeitos que programas e materiais podem provocar. Apesar de oferecer pouca ou quase nenhuma orientação metodológica, como mostra Erault, In: Walberg e Haertel (1990), Scriven procurou mostrar que todas as avaliações se devem concentrar nos objetivos gerais (**goals**), sendo impossível, na sua visão, o empirismo puro.

O componente intrínseco de uma avaliação concentrar-se-ia na **análise da consistência** dos objetivos e em suas divergências; e, ainda, na **análise de seu conteúdo**, abrangendo a sua precisão, a sua amplitude e o seu significado. A avaliação intrínseca, mostra Erault (1990), com base em Scriven (1967), seria de especial importância na avaliação formativa, especialmente se realizada por um avaliador externo. Igualmente, poderia ser útil no início de uma avaliação **somativa**, quando hipóteses são consideradas em função de previsões e da realidade do contexto em que se acha o **evaluand**. Se bem considerarmos, a avaliação intrínseca também poderia ser de utilidade **no final** de uma avaliação somativa, a partir da análise de pressupostos e objetivos gerais, entre outros aspectos, com vistas, ainda, a orientar os responsáveis pelas decisões, os **decision-makers**. Percebe-se imediatamente, como notou Erault (1990), que é neste último momento que a independência do avaliador é importante, crucial, para a credibilidade externa do avaliador. Infelizmente, as tecnologias que seriam adequadas ainda não foram suficientemente desenvolvidas, salvo, naturalmente, a análise de conteúdo, o mesmo não se passando com aspectos da estrutura metodológica da avaliação formativa de produtos educacionais. A questão da avaliação intrínseca está posta à prova e constitui um desafio para teóricos e praticantes da avaliação.

- [3] A perspectiva oferecida por Scriven em seus diferentes trabalhos (1967; 1971; 1972a; 1972b; 1973; 1974, entre outros) — determinação do mérito a partir dos **efeitos** do programa, sem referência aos seus objetivos — representou uma grande contribuição para a teoria da avaliação, porque atacou problema realmente sério que é o viés gerado no avaliador e, conseqüentemente, em seu trabalho, quando parte do conhecimento prévio dos objetivos que acabam por “orientar” suas ações e decisões. Nessa nova visão

(vide, especialmente, os trabalhos de 1972 e 1973), o avaliador não procura fazer uma distinção entre os objetivos antecipadamente propostos e os efeitos colaterais de um programa.

Ao analisar os **prós e contras** da avaliação **goal-free**, Scriven (1972) chama a atenção para o fato de que a retórica do pretendido — os objetivos — é usada para substituir as possíveis demonstrações de evidências do sucesso de um programa educacional, acrescentando, ainda, que na maior parte das vezes os objetivos são limitados e vagos, além de serem superestimados em muitos casos (ou subestimados). A nova visão de Scriven (1971) muda completamente o enfoque da avaliação, que abandona o questionamento sobre o que o programa tenta fazer para um novo posicionamento: — o que o programa está efetivamente fazendo.

É preciso destacar que a avaliação **goal-free** não é um modelo que apresente definições formais, específicas de relações estruturais, orientação para a coleta e análise de dados, e nem procedimentos operacionais. Se bem considerarmos, a **goal-free evaluation** é um posicionamento filosófico em relação à avaliação e que deve servir de orientação para o desenvolvimento de todo o seu processo, quando o avaliador, com base em seus conhecimentos e experiência profissional, procura colocar todo o seu empenho para identificar e documentar os efeitos de um programa. Scriven, por intermédio do método **modus operandi**, mostra como o avaliador deve estabelecer as relações causais entre o programa e seus efeitos, identificando todas as causas potenciais dos efeitos observados, como ocorre, por exemplo, numa investigação policial, em que mesmo os elementos supostamente irrelevantes precisam ser considerado. O avaliador, numa avaliação **goal-free**, deve adotar um posicionamento cauteloso porque sua posição se assemelha à de um pesquisador, por exemplo, em pesquisas científicas, que desconhece quais os doentes submetidos a um **placebo**, e precisa determinar os efeitos de uma **droga** que está sendo utilizada para um determinado fim e deve positivar inclusive suas influências colaterais.

As críticas à avaliação **goal-free** são várias, destacando-se, em particular o fato de que esse tipo de avaliação simplesmente substitui os objetivos do avaliador pelos do responsável pelo desenvolvimento de um programa. O próprio Scriven (1972) reconhece que a proposta apresenta aspectos favoráveis, mas que, em outros momentos, revela elementos negativos. Assim, apesar de seu interesse teórico, a avaliação **goal-free** não usufrui de uma grande popularidade entre os avaliadores educacionais; contudo, não podemos deixar de reconhecer, como aliás o faz Stechter (In: Walberg e Haertel, 1990), que o pensamento de Scriven teve impacto na prática da avaliação, especialmente por destacar a independência do avaliador, que às vezes é bastante cerceado nas suas atividades, por ressaltar a importância dos efeitos colaterais de um programa, o que nem sempre é objeto de consideração; e, também, por submeter os próprios objetivos do programa a uma análise

aprofundada, ainda que não se utilize dos mesmos objetivos para realizar o seu trabalho investigativo.

- [4] As expressões avaliação formativa e avaliação somativa fazem parte do vocabulário dos professores, que muitas vezes ignoram (ou esquecem) que foi graças a Scriven (1967) que ambas surgiram no contexto da avaliação, inicialmente no caso da avaliação de currículo, lembra Lewy, In: Walberg e Haertel, 1990. A sua divulgação somente ocorreu mais tarde, na década de 70, quando Benjamin Bloom e seus associados publicam a obra por muitos considerada como clássica — **Handbook on Formative and Summative Evaluation of Student Learning** (1971) — relacionada com o desempenho escolar, e por Cronbach (1982) com relação a atividades de diferentes programas sociais.

A idéia de avaliação formativa e somativa contribuiu substancialmente para que a concepção de avaliação de programa se alterasse, deixando de ser uma simples comparação entre objetivos de vários programas para a escolha de um único que correspondesse às necessidades do momento. Estudos com esse enfoque nunca permitiram resultados conclusivos, exigindo quase sempre outros estudos mais específicos sobre aspectos particulares dos programas ao longo de seu desenvolvimento (Lewy, 1990), em outras palavras, estudos de avaliação formativa. Estudos comparativos de programas, com a finalidade de continuá-los ou não, passaram a ser um tipo de avaliação no contexto e foram realizados com a intenção de uma avaliação somativa.

Apesar de serem de uso tranqüilo entre professores, o emprego das expressões avaliação somativa e formativa nem sempre é feito com a devida precisão, especialmente quando usadas em relação a diferentes entidades (produtos, processos, pessoal e, sobretudo, a alunos), como um efeito não desejado da grande obra de Bloom et al. (1971), antes referida. Cronbach igualmente, na década de 80, também contribuiu para a divulgação dessas expressões, ao defender a necessidade da reformulação de programas. A partir da publicação da **Metodologia da Avaliação** (1967), de Scriven, a avaliação começa a ser reconhecida como uma atividade fundamental em educação, ganha **status**, digamos, e amplia, com a aplicação dos seus princípios na avaliação de programas, projetos e materiais, o caminho que fora iniciado, na década de 40, por Tyler.

Ante o que foi exposto, podemos dizer que não existe, efetivamente, uma diferença lógica ou metodológica entre avaliação formativa e somativa. Ambas visam a determinar o valor de uma entidade em particular. A diferença existente entre uma e outra está, realmente, no **timing** em que são realizadas, no **tipo de audiência** a que se destinam e na **utilização dos resultados**. Lewy (In: Walberg e Haertel, 1990) mostra que um mesmo tipo de estudo avaliativo pode ser considerado *formativo* ou *somativo*, tudo dependendo da perspectiva em que se situa quem faz a análise do trabalho.

Cronbach na sua longa discussão sobre planejamento e análise de programas (1982) considera que a avaliação formativa tem um maior impacto e, desse modo, é mais significativa do que a somativa, com o que concordamos. Ainda que a avaliação de um programa possa ser considerada fluida, digamos, por ser a avaliação de algo em progresso, a sua contribuição do ponto de vista educacional é bem maior do que a avaliação de um produto acabado, quando alterações não mais podem ser introduzidas. O mesmo raciocínio aplica-se à avaliação do rendimento escolar, com o objetivo de informar o professor e o aluno sobre o processo de ensino/aprendizagem. Assim, na linha de Cronbach, gostaríamos de reiterar que a avaliação formativa possui uma maior carga de contribuição para a melhoria da aprendizagem do que os resultados de instrumentos de medida aplicados ao término de um curso ou até mesmo de uma simples unidade programática. Isso, entretanto, não significa que se deva ignorar a avaliação somativa, que não deixa de ser imperiosa em diferentes situações.

Apesar da importância da avaliação formativa e somativa, não existe, na realidade, definido pelos teóricos, um conjunto de regras de procedimento para a sua concretização. Queremos, entretanto, destacar os posicionamentos de Bloom et al. (1971) e de Stake (In: Hamilton D. (Ed.), 1978). Ao considerar a avaliação do rendimento escolar, Bloom destacou a necessidade, na avaliação formativa, de analisar o material instrucional, mapear e hierarquizar as estruturas das várias tarefas relacionadas à aprendizagem, e de aplicar instrumentos de medida periodicamente, entre curtos intervalos, na orientação da aprendizagem. Uma amostra dos itens da avaliação formativa deve integrar a avaliação somativa, ao fim do curso, com o objetivo de atribuir escores ou conceitos aos diferentes desempenhos. Stake, por sua vez, procura caracterizar as diferentes informações coletadas em situação de avaliação formativa ou somativa. A avaliação formativa seria uma **análise molecular**, pesquisando causas e fixando-se nas experiências dos usuários de um programa; a avaliação somativa seria uma **análise molar**, porque forneceria informações descritivas, estaria interessada em ressaltar a eficiência de programas, e tenderia a dar maior destaque aos seus efeitos locais.

A diferença entre avaliação formativa e somativa teve (e tem) importante papel na avaliação educacional, especialmente em termos de aprimoramento dos planos educacionais, ainda que se façam necessárias pesquisas sobre o assunto para dirimir controvérsias relacionadas ao seu uso na tomada de decisões e no que se refere a seus aspectos metodológicos.

- [5] Outra grande contribuição de Scriven (In: Taylor et al., 1972), assim como do próprio Stufflebeam, entre outros, foi o destaque oferecido à importância de diferentes atividades com o fim específico de determinar a qualidade científica dos estudos avaliativos e de suas conclusões. Ou seja, empreender uma avaliação da avaliação, que pode ser sintetizada em realizar uma **meta-avaliação**.

A meta-avaliação tem como um de seus objetivos identificar vieses de diferentes tipos que possam existir em uma avaliação, usando, para esse fim, diferentes metodologias para, além da importância desses vieses, determinar os seus efeitos sobre a avaliação. A meta-avaliação pode usar, inclusive, metodologias estatísticas, ainda que a elas não deva ficar restrita, pois os vieses muitas vezes podem resultar — e é importante que levemos em consideração esse aspecto — de simplificações ou interpretações errôneas dos resultados, ou até mesmo da falta de compreensão do valor relativo de aspectos da avaliação, como mostram Straw e Cook, In: Walberg e Haertel (1990).

É necessário que tenhamos presente em nosso espírito que nenhuma avaliação é definitiva, assim como nenhum relatório deve ser considerado em termos absolutos. Ainda que, tecnicamente, a avaliação possa ser bem estruturada em termos de metodologia, poderá revelar deficiências ao serem discutidas as suas conclusões e consideradas as suas implicações. Uma avaliação, desse modo, deve ser meta-avaliada sob diferentes pontos de vista e por grupos diferenciados de analistas, oferecendo-se, assim, às várias audiências, uma síntese multi-facetada, a partir de diferentes perspectivas, que apontarão o relevante e o irrelevante em uma avaliação, quando vista a partir de várias óticas.

- [6] As várias áreas do conhecimento, especialmente as de estruturação mais recente, como é o caso da avaliação, estão sujeitas a disputas de natureza ideológica. A avaliação faz uso de diferentes modelos, frutos de concepções as mais diversas, e enfrenta complexos problemas metodológicos difíceis de um consenso. Tudo isso traduz a intensa fermentação que existe em seu campo de atuação, o que é um sintoma bastante positivo, conforme Scriven (1973). As áreas de conhecimento, por força do impacto dessas ideologias, acabam apresentando vieses os mais variados, que terminam por gerar desfigurações, o que ocorre com frequência na avaliação.

Usando a caracterização de Scriven (In: Madaus et al., 1993), vemos que uma delas é a **ideologia separatista**, que reflete uma falsa dicotomia avaliador/avaliado. Talvez, conforme Scriven, e grande é o seu acerto, no seu início histórico, se pudesse pensar separadamente nessas duas figuras, traduzidas por **eu sou o avaliador e você é o avaliado**. Após a definição de novas formas de avaliação, essa falsa dualidade, se alguma vez existiu, não pode ser aceita, porque, na verdade, o avaliador avalia e simultaneamente é avaliado pelo sujeito da avaliação. A separação decorreria de uma reação fóbica, no dizer de Scriven (1993), que estaria associada à questão do valor, do mérito, que é central na conceituação de avaliação por Scriven. Todos temos um certo temor da avaliação, inclusive o próprio avaliador, especialmente quando tem certeza de que ele próprio está sendo avaliado. A tensão social a que é submetido o avaliador acaba por fazer com que ele se incorpore ao grupo avaliado e não se aperceba dos problemas apresentados no

objeto avaliado. O avaliador, que se deveria conservar à margem do objeto avaliado, incorpora-se e passa a ser um nativo integrante do grupo avaliado. O separatismo inicial e a agregação ao grupo acabam por gerar distorções em todo o processo da avaliação, introduzindo vieses nos julgamentos de valor dos programas educacionais. A questão, entretanto, não está apenas na separação e nem apenas na necessidade de afiliar-se ao grupo que elaborou o programa; o problema está em conseguir objetividade, que é um ideal a ser alcançado em avaliação e faz parte da estrutura de um profissional consciente, no dizer de Scriven (1993).

A avaliação surge e se desenvolve tendo como cenário de fundo a discussão filosófica sobre o conflito entre positivismo e idealismo, a controvérsia entre objetividade e subjetividade. As divergências ainda não foram inteiramente superadas em certas áreas, como lembra Scriven (1993), que exemplifica com a possibilidade de objetividade na ética e em outras áreas de pesquisa de valor, como é o caso da avaliação.

Scriven (1993), ao analisar a ideologia do positivismo em avaliação, acha estranho, até mesmo bizarro, para usar sua palavra, que, sendo o trabalho científico altamente avaliativo e objetivo nas suas análises, a ciência no século XX seja apresentada como **value-free**, independa da consideração de valor. O posicionamento de Scriven é perfeitamente justificável tendo em vista que a sua metodologia de avaliação está centrada na determinação do valor. Por isso, Scriven mostra a contradição daqueles que declaram não ser possível nenhum julgamento avaliativo com objetividade científica, e, ao mesmo tempo, avaliam alunos, programas e outros mais. Há, assim, dois posicionamentos conflitivos —um, profissional/científico, que nega a objetividade em avaliação; outro, pedagógico, que aceita a avaliação. A ideologia do positivismo, diz Scriven (1993) com certa e justificável ironia, é o caso típico de uma ideologia gerada para reduzir ansiedades “valorfóbicas”.

A influência da ideologia gerencial é vista por Scriven segundo a perspectiva norte-americana, que oferece um quadro bastante diferente do brasileiro; contudo, reconheçamos, algumas idéias se ajustam perfeitamente ao quadro nacional. Partamos da indagação de Scriven: — quem realiza as grandes avaliações? Geralmente são de iniciativa de órgãos oficiais (federais, estaduais e municipais), que produzem avaliações internas, cujos resultados sofrem a influência do pecado original e, conseqüentemente, procuram apresentar resultados suavizados da sua própria realidade e — é claro — contaminados por esse enviesamento. Outras vezes, as avaliações são externas, no todo ou em parte, e nesse caso também há um certo comprometimento, pois resultados demasiadamente críticos podem comprometer futuras contratações de novos projetos de avaliação, diz Scriven (1993), candidamente. A avaliação sofre muitas vezes pressões bastante acentuadas, ainda que não diretas e ostensivas, mas que acabam por contaminar todo o sistema de avaliação. A avaliação na ótica da ideologia

gerencial é bem sucedida na medida em que os objetivos do programa foram identificados e concretizados, o que nem sempre é do interesse do consumidor da avaliação, mais interessados nos efeitos diretos ou colaterais do programa. A visão gerencial centra-se na obtenção do sucesso em atingir a esse objetivo (plano de aula; plano de curso), muitas vezes via um ensino comprometido. Ambas as ideologias — a positivista e a gerencial — se assemelham, acentua Scriven (1993), na medida em que se recusam, terminantemente, a proferir juízos de valor.

Tudo isso evidencia a extrema complexidade do ato de avaliar.

### REFERÊNCIAS BIBLIOGRÁFICAS

- 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. 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.
- BLOOM, B. S.; HASTINGS, J. J. e MADAUS, G. F. **Handbook on Formative and Summative Evaluation of Student Learning**. New York: McGraw-Hill Book Co., 1971.
- CHOPPIN, B. H. Evaluation, Assessment and Measurement. In: WALBERG, H. L. et al. **The International Encyclopedia of Educational Evaluation**. Oxford, Pergamon Press. 1990.
- 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 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.
- 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.
- 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.

- GLASS, G. V. e WORTHEN, B. R. Evaluation and research: similarities and differences. **Curriculum Theory Network**, Fall. 1971.
- GOLDBERG, M. A. A. e SOUZA, C. P. (Org.) **Avaliação de Programas Educacionais: vicissitudes, controvérsias, desafios**. EPU. São Paulo. 1982.
- 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.
- 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.
- 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.
- HOUSE, E. R. Assumptions underlying evaluation models. **Educational Researcher**, March, 1978, p. 4-12. In: MADAUS, G. F. et al. (Ed.) **Evaluation Models**. Boston. Kluwer-Nijhoff Publishing. 1993.
- KOPPELMAN, K. L. The Explication Model: An Anthropological Approach to Program Evaluation. **Educational Evaluation and Policy Analysis**, nº 3, 1. July-August, 1999, p. 59-64. In: MADAUS, G. F. et al. **Evaluation Models**. Boston. Kluwer-Nijhoff Publishing, 1993.
- 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.
- 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.
- MADAUS, G. F., AIRASIAN, P. W. e KELLAGHAN, T. **School Effectiveness. A Reassessment of the Evidence**. New York, McGraw-Hill Book Company, 1980.
- 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.
- NORRIS, N. **Understanding Educational Evaluation**. Kogan Page. London. 1993. Published in Association with CARE — School of Education, University of East Anglia.
- PARLETT, M. e HAMILTON, D. Evaluation as illumination: a new approach to the study of innovatory programs. In: GLASS, G.V. (Ed.) **Evaluation studies review annual**. Vol. 1. Beverly Hills, California. Sage. 1973.
- PARLETT, M. R. Illuminative Evaluation. In: WALBERG, H.L. et al. **The International Encyclopedia of Educational Evaluation**. Oxford. Pergamon Press. 1990.
- PAYNE, D. A. (Ed.) **Curriculum Evaluation: Commentaries on Purpose, Process, Product**. Lexington, Mass. D.C. Heath and Company. 1974.
- 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.
- POPHAM, W. J. e CARLSON, D. Deep Dark Deficits of the Adversary Evaluation Model. **Educational Researcher**, June, 1977, p. 3-6. In: MADAUS, G. F. **Evaluation Models**. Boston, Kluwer-Nijhoff Publishing. 1993.
- PROVUS, M. M. **Discrepancy Evaluation**. Berkeley, California. McCutchan. 1973.
- PROVUS, M. M. Evaluation of ongoing programs in the public school system. **National Society for the Study of Evaluation - NSSE 68<sup>th</sup> 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.
- SCRIVEN, M. Evaluation perspectives and procedures. In: W. J. POPHAM (Ed.) **Evaluation in Education**. Berkeley, California. McCutchan. 1974.
- SCRIVEN, M. Goal-Free Evaluation. In: HOUSE, E. R. (Ed.) **School Evaluation — The Politics and Process**. Berkeley, California. McCutchan Publishing Corporation. 1973.
- 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. Evaluating Educational Programs. In: CARO, F. G. **Readings in Evaluation Research**. New York. Russell Sage Foundation. 1971.

- SCRIVEN, M. Evaluation as a Paradigm for Educational Research. In: E. R. HOUSE. **New Directions in Educational Evaluation**. London/Philadelphia. The Falmer Press. 1989.
- SCRIVEN, M. Evaluation Ideologies. In: MADAUS, G. F. et al. (Ed.) **Evaluation Models-viewpoints on educational and human services evaluation**. Boston. Kluwer-Nijhoff Publishing. 1993.
- 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. The Concept of Evaluation. In: APPLE, M. W. et al. (Ed.) **Education Evaluation: — Analysis and Responsibility**. Berkely, California. McCutchan Publishing Corporation. 1974.
- SCRIVEN, M. **The Methodology of Evaluation**. Perspectives of Curriculum Evaluation. AERA. Monograph 1, Rand McNally and Co., Chicago. 1967.
- SCRIVEN, M. Pros and Cons about Goal-free Evaluation. **Evaluation Comment**, 3 (4), p.1-7. 1972a.
- SHRINKFIELD, A. J. Designing Evaluations of Educational and Social Progress by Lee J. Cronbach: a Synopsis. In: MADAUS, G. F. et al. (Ed.) **Evaluation Models — viewpoints on educational and human services evaluation**. Boston. Kluwer-Nijhoff Publishing. 1993.
- STAKE, R. E. Case Studies. In: DENZIN, N. e LINCOLN, Y. **Handbook of Qualitative Research**. Sage Publications. Thousand Oaks, California, 1994.
- 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. 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. Responsive Evaluation. 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.
- 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. nº 1, 7, 1976, p. 1-7; nº 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.
- 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 relevance of the CIPP Evaluation Model for Educational Accountability. **Journal of Research and Development in Education**, 5 (1), p. 19-25. 1971.
- TAYLOR, P. A. e COWLEY, D. M. **Readings in Curriculum Evaluation**. W.M.C. Brown Co. Publishing. Dubuque,Iowa. 1972.
- 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.

- TYLER, R. W. Evaluating Learning Experiences. In: HAMILTON, D. **Beyond the Numbers Game**. Mac Millan Ltd. London. 1978.
- 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. 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. General Statement on Evaluation. **Journal of Educational Research**, 35, p. 492-501. 1942.
- VIANNA, H. M. Avaliação Educacional: algumas idéias precursoras. **Educação e Seleção**, jul-dez. nº 6. Fundação Carlos Chagas. São Paulo. 1986.
- WALBERG, H. J. e HAERTEL, G. D. **The International Encyclopedia of Educational Evaluation**. Oxford. Pergamon Press. 1990.
- WORTHEN, B. R.; SANDERS, J. R. **Educational Evaluation: — alternative approaches and practical guidelines**. New York-London. Longman, 1987.
- WORTHEN, B. R.; SANDERS, J.R. **Educational Evaluation: theory and practice**. Belmont, California. Wadsworth. 1973.