

Avaliação de Programas e Avaliação da Aprendizagem

Léa Depresbiteris¹

A finalidade deste ensaio não é a de esgotar quantitativamente os estudos de avaliação realizados através dos tempos. Fosse essa a intenção, seria necessário traçar um perfil histórico desde as épocas mais remotas, quando já era possível detectar algumas atividades avaliativas formais. Alguns mil anos atrás, os chineses já administravam testes para a admissão ao serviço civil; também são frequentes os relatos sobre o emprego de exames pelos professores, na Grécia e Roma antigas.

O que será apresentado é o caminho percorrido pela avaliação, desde uma perspectiva de medida, voltada exclusivamente para os resultados de aprendizagem dos alunos, até uma perspectiva mais ampla, através da qual a avaliação visa ao programa como um todo, expandindo-se aos propósitos educacionais, aos currículos e incluindo relações com outros sistemas: social, econômico e político. A avaliação, nesta última perspectiva, visa os projetos em larga escala, particularmente os que assumem o caráter de projetos nacionais, regionais ou estaduais e, não a avaliação tal como é concebida e praticada no contexto restrito de um aluno, uma sala ou uma só escola.

A. AVALIAÇÃO DA APRENDIZAGEM

As primeiras idéias sobre avaliação da aprendizagem estavam ligadas à idéia de medir.

O uso da avaliação como medida vem de longa data. Através de Ebel, tem-se o relato de Kuo sobre a presença de exames, já em 2205 A.C. Nessa época, o Grande "Shun", imperador chinês, examinava seus oficiais a cada três anos, com o fim de os promover ou demitir. O regime competitivo nos exames na China antiga tinha então como propósito principal prover o Estado com homens capacitados (Ebel e Damrin, 1960).

Avançando na história, Ebel relata que no século XIX, nos Estados Unidos da América, Horace Mann criou um sistema de testagem, sendo um dos pioneiros nesta área. Uma controvérsia entre Mann e os comitês das escolas americanas sobre a qualidade da educa-

¹ Da Equipe de Avaliação, da Divisão de Currículos e Programas, da Diretoria de Tecnologia Educacional, da Coordenadoria de Educação do Departamento Regional do Senai - São Paulo. Doutora em Educação.

ção, fez com que ele propusesse a experimentação de um sistema uniforme de exames, em uma amostra selecionada de estudantes das escolas públicas de Boston. Os resultados dessa experiência reforçaram muitas das críticas feitas por Mann com relação à qualidade da educação e indicou a possibilidade de testar os programas em larga escala, com a finalidade de sugerir melhorias nos padrões educacionais. As principais sugestões de Mann foram: (a) substituir os exames orais pelos escritos, (b) utilizar, ao invés de poucas questões gerais, uma quantidade maior de questões específicas, e (c) buscar padrões mais objetivos do alcance escolar.

Ebel destaca, também, a contribuição de J.M. Rice, que desenvolveu os primeiros testes objetivos para uso em pesquisas de alcance escolar. Seus estudos eram basicamente relacionados: (1) à construção, ao uso e à correção de testes objetivos, (2) à construção, ao uso e à correção de testes dissertativos e (3) à busca de diferenças entre esses dois tipos de teste, na função de medir, avaliar, prever e classificar.

O desenvolvimento de testes objetivos tornou possível o estabelecimento de programas de exames estaduais e regionais. Nos E.U.A. criaram-se associações, comitês, bureaus para o desenvolvimento de testes padronizados. O Conselho Americano de Educação apoiou, através do Serviço Cooperativo de Testes, a produção de testes de rendimento escolar para a escola secundária e universidades. Estabeleceram-se o sistema de testagem para medir competências dos professores e o Bureau de Arquivos Educacionais para prover programas de testagem para escolas públicas.

Uma revisão histórica mais específica, em países como a França e Portugal, aponta para o desenvolvimento de uma ciência chamada Docimologia. A docimologia vem do grego *dokimé*, que quer dizer nota. É, segundo De Landshere (1976), a ciência do estudo sistemático dos exames, em particular do sistema de atribuição de notas e dos comportamentos dos examinadores e examinados.

Tomando por base as obras de Pierón (1969), Bonboir, De Landshere (1976) e Miranda (1982) verifica-se que a investigação docimológica teve início com os estudos de Pierón e Laugier, que evidenciavam a instabilidade das avaliações, no tocante às diferenças inter e intra-individuais e à precisão dos testes. Os primeiros estudos desses autores basearam-se na aplicação de seis testes de aptidão em alunos de escolaridade primária. Os resultados dos alunos, nesses testes, foram comparados com classificações escolares realizadas ao longo e ao final do ano letivo e mostraram estar discrepantes, entre si, levando os autores a colocar em discussão o papel do exame como fator eliminatório decisivo para a classificação do aluno. A docimologia surgiu como crítica à extrema confiança nos métodos tradicionais utilizados, com fins de seleção, nos exames e nos concursos.

Prosseguindo em seu trabalho, Pierón realizou estudos que buscavam analisar as diversas variáveis que influem no ato de medir os conhecimentos, provando, através deles, a fragilidade de métodos considerados infalíveis. Esses estudos, compilados no "*Études docimologiques sur le perfectionnement des examens et des concours*" e publicados em 1935, mostram que a docimologia pautou-se por duas grandes linhas mestras: docimologia clássica, também chamada de negativa, e docimologia experimental, também chamada de positiva.

A docimologia clássica punha ênfase no aperfeiçoamento das técnicas de construção dos instrumentos de avaliação e de análise de resultados. A docimologia experimental centrava-se na avaliação como um comportamento, procurando determinar, experimentalmente, os mecanismos intervenientes na decisão avaliativa, tais como: reação dos aplicadores, reação dos submetidos aos exames, discrepâncias entre situações de exame e entre critérios de aplicadores.

Nos Estados Unidos, a docimologia passou a ter seu lugar de destaque em 1931, quando a Carnegie Corporation pôs à disposição do Instituto Internacional de Educação do Teacher's College da Universidade de Columbia (New York) fundos necessários à realização de um inquérito subordinado ao tema "As concepções, os métodos, a técnica e o alcance pedagógico e social dos exames e concursos".

Uma primeira conferência foi realizada em Eastbourne, reunindo diversas comissões nacionais (norte-americana, inglesa, escocesa, francesa, alemã, finlandesa e suíça). A essa conferência seguiram-se duas outras em Folkstone (em 1935) e em Dinard (em 1938), sendo que nesta última foi criada uma comissão permanente para o estudo das questões levantadas. Dentre as publicações mais importantes deste empreendimento estão: *"The examination of examinations"* (Londres, 1935) e *"La correction des épreuves écrites dans les examens - Enquete experimentale sur le baccalauréat"* (Paris, 1936).

Como se pode verificar, durante as primeiras décadas do século XX, a maior parte da atividade que pode ser caracterizada como avaliação educacional formal estava associada à aplicação de testes, o que imprimia um caráter instrumental ao processo avaliativo.

Para quebrar um pouco essa idéia específica de mensuração, surgiram nos Estados Unidos da América, por volta de 1950, os escritos de Ralph Tyler. Esse autor provocou um grande impacto na literatura com seu "Estudo dos oito anos", realizado com Smith, no qual defendia a inclusão de uma variedade de procedimentos avaliativos, tais como: testes, escalas de atitude, inventários, questionários, fichas de registro de comportamento e outras formas de coletar evidências sobre o rendimento dos alunos em uma perspectiva longitudinal, com relação à consecução de objetivos curriculares (Tyler, 1949).

Segundo Tyler (1982), muitas pessoas consideram a avaliação como sinônimo de aplicação de testes de lápis e papel. Apesar de achá-los extremamente importantes, pois permitem determinar a habilidade dos alunos em alguns assuntos, o autor mostra que existem outros objetivos, como o de ajustamento pessoal, que são avaliáveis através de observação das crianças em situações nas quais estejam envolvidas, por exemplo, em interações sociais. Outros objetivos relacionados a habilidades psicomotoras devem, igualmente, ser avaliados através de observações, de entrevistas, de registros sistemáticos. Assim, de acordo com Tyler, há muitos meios de conseguir evidências sobre mudanças comportamentais. Quando se pensa em avaliação, está se falando em qualquer meio capaz de assegurar evidências válidas sobre os comportamentos em seus diferentes níveis e manifestações.

Tyler defende, também, a idéia de que o processo avaliativo consiste basicamente na determinação de quanto os objetivos educacionais estão sendo atingidos por programas instrucionais. Ele diz que esta concepção de avaliação tem dois aspectos importantes. Em primeiro lugar, implica que a avaliação deve julgar o comportamento dos alunos, pois o que se pretende em educação é justamente modificar comportamentos. Em segundo lugar, pressupõe que a avaliação deve envolver mais do que um único julgamento, em determinada ocasião, e logo outros mais, em instantes subseqüentes, para identificar mudanças que podem estar ocorrendo.

Tyler ainda chama atenção para três variáveis básicas no estudo da avaliação: o estudante, a sociedade e a área de conteúdo a ser desenvolvida, de acordo com os princípios da psicologia da aprendizagem e de uma filosofia de educação bem definida. Na verdade, o autor já defendia a idéia de elaboração e avaliação de currículos, através do modelo reproduzido na figura 1.

A linha tracejada, partindo da avaliação, indica que, para o autor, avaliar é estabelecer uma comparação entre os desempenhos e os objetivos previamente determinados. Os objetivos, para serem definidos, têm como fonte o aluno, a sociedade e os especialistas e, como filtros, a Filosofia e a Psicologia de Educação. De acordo com Vianna (1982a), Tyler criou um modelo de avaliação composto dos seguintes passos:

- (1) formulação e classificação dos objetivos, segundo níveis de generalidade e especificidade,
- (2) definição de cada objetivo em termos comportamentais,
- (3) identificação de situações que demonstrem os comportamentos estabelecidos nos objetivos,
- (4) seleção e experimentação de métodos e instrumentos adequados, ou construção de novos instrumentos, para a coleta de informações sobre cada objetivo,

que se desejar que os alunos alcancassem ou ultrapassassem.

(d) Todo objetivo deve conter as condições, as ações e os critérios explícitos, para permitir uma avaliação que realmente detecte em que o aluno está tendo sucesso ou falhando. As condições são os elementos com os quais o aluno conta para realizar suas ações; as ações são o que se espera que o aluno realize e os critérios são as medidas para determinar se as ações esperadas foram realizadas no nível desejado.

(c) Objetivos são descrições de resultados almejados, isto é, descrições dos padrões do que visa a esclarecer o motivo pelo qual um padrão não foi alcançado.

(b) Um teste é um processo de medir uma determinada característica. Um item de teste é uma amostra de conduta ou de um desempenho. Existem itens de verificação e diagnóstico. O primeiro é aquele que visa a determinar se um padrão foi alcançado, o segundo é o que visa a esclarecer o motivo pelo qual um padrão não foi alcançado.

(a) Medir é um processo de determinar a extensão de uma característica pertencente a um objeto ou a uma pessoa. Avaliar é o ato de comparar uma medida com um padrão e de emitir um julgamento sobre esta comparação.

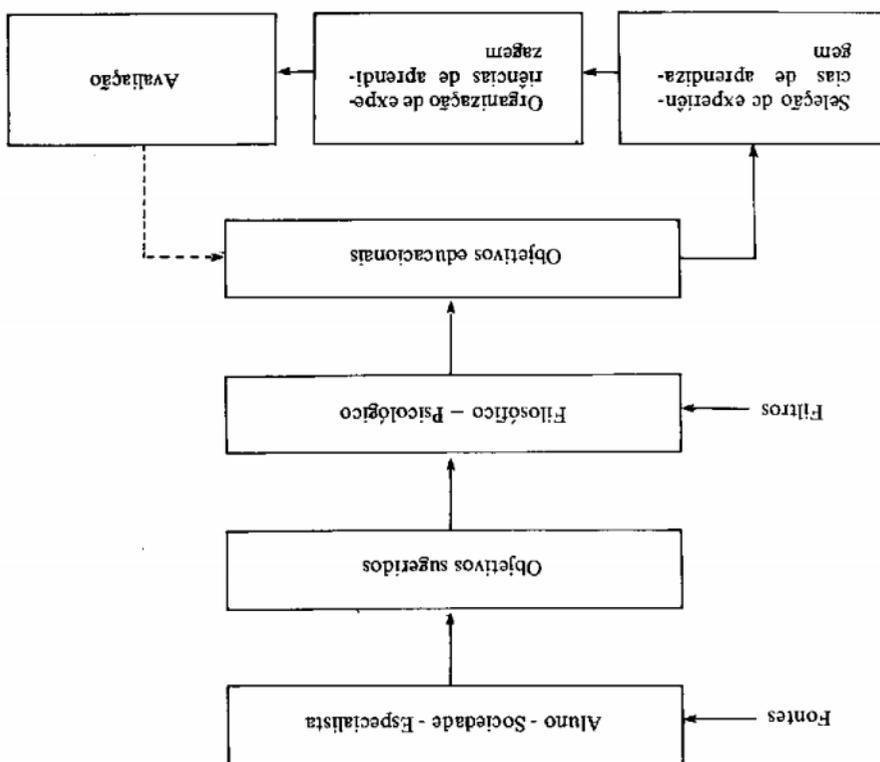
Outro estudo da avaliação foi Mager, cujas principais idéias podem ser assim resumidas (Mager 1977):

Concorrem julgamentos de valor.

de objetivos, sem vinculá-la a um processo contínuo e sistemático, para o qual também a abordagem de Tyler pecava por considerar a avaliação como atividade final de alcance objetivos na avaliação e por conter a idéia de "feedback" para que se efetuassem melhorias, Ainda que inovadora, na época, por atribuir a abordagem de Tyler importância aos objetivos de critérios para a interpretação de resultados.

Fonte: Saul, 1985

Figura 1 – Modelo de avaliação de currículo, segundo Tyler.



O conceito de avaliação foi igualmente enriquecido por Cronbach que, em meados de 1963, foi o primeiro a vincular as atividades de avaliação ao processo de tomada de decisão.

De acordo com Vianna (1982b), Cronbach indicou os seguintes objetivos para a avaliação:

(a) determinar se os métodos de ensino e o material instrucional utilizado no desenvolvimento de um programa são realmente eficientes, com a finalidade básica de aperfeiçoar currículos;

(b) identificar as necessidades dos alunos para possibilitar o planejamento da instrução e julgar o mérito dos estudantes, de modo que estes possam ser incentivados em seus sucessos e auxiliados em suas deficiências;

(c) julgar a eficiência do sistema de ensino e dos professores, de forma a subsidiar decisões de natureza administrativa.

Na linha de tomada de decisão, Cronbach apresenta a idéia de que um teste de aprendizagem deve ser apropriado à decisão a ser tomada e que, para isso, antes de tudo, deve haver definição precisa do propósito do sistema de testagem.

Outra contribuição importante foi a de Bloom, cujas principais idéias a propósito da aprendizagem para o domínio e de uma taxonomia que permitisse um sistema coerente de ensino e avaliação, ganharam ampla divulgação.

Com sua ênfase na aprendizagem para o domínio, Bloom (1963) defende a idéia de que cada vez mais é necessária uma educação contínua, durante toda a vida do indivíduo. Cabe à Escola esforçar-se para assegurar experiências de aprendizagem bem sucedidas, no campo das idéias e do auto-desenvolvimento, a todos os estudantes. Para isso, deve passar por mudanças, primordialmente relacionadas a atitudes dos estudantes, professores e administradores; mudanças que devem ser realizadas, também, nas estratégias de ensino e no papel da avaliação. Uma das mudanças pregadas por Bloom é a do descondicionalismo da idéia de curva normal, pois se o professor é eficaz em sua instrução, a distribuição do rendimento dos alunos deve ser muito diferente daquela apresentada por uma curva normal. De fato, dever-se-ia até considerar que os esforços educacionais fracassam na medida em que a distribuição de rendimento se aproxima desta curva. Diferenças individuais nos estudantes é um fato que pode ser demonstrado de muitas maneiras. O papel da educação é o de descobrir estratégias que levem essas diferenças em consideração, mas que o faça no sentido de promover, ao máximo, o desenvolvimento do indivíduo.

Bloom pressupõe que o domínio da aprendizagem é teoricamente disponível para todos se houver possibilidade de encontrar os meios de ajudar cada estudante. Uma fundamentação para este ponto de vista pode ser encontrada nas normas de avaliação de muitos testes padronizados de rendimento. As aplicações sucessivas demonstram que critérios selecionados, atingidos pelos melhores estudantes em um ano, são atingidos pela maioria dos estudantes em período posterior. Outra fundamentação mais profunda pode ser encontrada em estudos que dizem que os estudantes são capazes de aprender em seu próprio ritmo. Esses estudos mostram que, embora a maioria dos estudantes eventualmente chegue a dominar cada tarefa de aprendizagem, alguns desses estudantes atingem esse domínio muito mais cedo do que outros.

Ligado a este fato, diz Bloom que o grande desafio da aprendizagem para o domínio é encontrar meio de reduzir o tempo requerido pelos estudantes mais lentos, de modo que a tarefa se torne menos longa e cansativa. Para isso, ele defende a idéia de que é necessário utilizar estratégias de instrução muito diferentes para atingir o domínio, pois elas são tentativas de aperfeiçoar a qualidade da instrução em relação à cada habilidade do estudante.

Um dos fatores mais fundamentais na aprendizagem para o domínio de Bloom é o do tempo gasto para aprender. A premissa básica é a de que a aptidão determina o ritmo de aprendizagem e que a maioria dos estudantes pode atingir o domínio, se dispuser da quantidade de tempo requerida para aprender.

Em termos de avaliação, Bloom diz que, para este tipo de aprendizagem, em algum ponto no tempo, os resultados de ensino podem ser detectados ao se avaliar os estudantes. A tônica do autor é a distinção bem marcada entre o processo de ensino-aprendizagem, que tem a intenção de preparar o estudante, e o processo de avaliação final, que tem a intenção de verificar em que extensão o estudante se desenvolveu da maneira esperada. Ambos, professor e aluno, devem ter alguma compreensão de quais são os critérios de rendimento e ambos devem ser capazes de assegurar evidência de progresso em direção a esses critérios, que sempre devem ser absolutos. Dessa maneira, os estudantes devem sentir que estão sendo avaliados em termos de nível de desempenho e não em termos de uma curva normal, ou de outro conjunto de padrões arbitrários e relativos.

Bloom defende a idéia de que testes curtos de diagnóstico podem ser usados para determinar se um estudante dominou ou não unidades de ensino, ou se ele terá que fazer mais alguma coisa para dominá-las. Estes testes não devem receber nota, em hipótese alguma, e sim determinarem o domínio ou a falta de domínio das habilidades, o que permite, tanto ao aluno como ao professor, informações para melhoria de desempenhos ainda não dominados ou incentivo, no caso dos já alcançados.

Para permitir que a informação avaliativa mostre o caminho do estudante no processo ensino-aprendizagem, Bloom e colaboradores criaram em 1956 uma taxonomia de objetivos educacionais para o domínio de operações intelectuais. Por taxonomia, os autores definem uma classificação de comportamentos do aluno que representam resultados esperados do processo educacional (Bloom, 1972).

Essa taxonomia de operações intelectuais compreende seis classes principais, dispostas numa hierarquia de complexidade crescente, onde o nível mais simples é o de conhecimentos, conforme se pode observar no quadro 1.

QUADRO 1

TAXONOMIA DO DOMÍNIO COGNITIVO, SEGUNDO BLOOM.

(Fonte: Bloom, 1972)

CLASSES	TÓPICOS
• Avaliação	<ul style="list-style-type: none"> • juízos em termos de critérios externos • juízos em termos de evidência externa
• Síntese	<ul style="list-style-type: none"> • derivação de um conjunto de relações abstratas • produção de um plano ou conjunto de relações abstratas • produção de uma comunicação singular
• Análise	<ul style="list-style-type: none"> • de princípios de organização • de relações • de elementos
• Aplicação	<ul style="list-style-type: none"> • aplicação
• Compreensão	<ul style="list-style-type: none"> • extrapolação • interpretação • tradução
• Conhecimentos	<ul style="list-style-type: none"> • de princípios universais e das formulações abstratas de um campo • das formas e meios de trabalhar com elementos específicos • de elementos específicos

Em 1957, Bloom e colaboradores criaram a Taxonomia de objetivos para o domínio afetivo. Esta taxonomia compreende níveis, sendo o nível mais elementar o de acolhimento, conforme se pode observar no quadro 2.

QUADRO 2

TAXONOMIA DO DOMÍNIO AFETIVO, SEGUNDO BLOOM.

(Fonte: Bloom, 1977)

NÍVEIS	TÓPICOS
<ul style="list-style-type: none"> • Caracterização por um valor ou complexo de valores 	<ul style="list-style-type: none"> • caracterização • direção generalizada
<ul style="list-style-type: none"> • Organização 	<ul style="list-style-type: none"> • de um sistema de valores • conceituação de um valor
<ul style="list-style-type: none"> • Valorização 	<ul style="list-style-type: none"> • compromisso • preferência por um valor • aceitação de um valor
<ul style="list-style-type: none"> • Resposta 	<ul style="list-style-type: none"> • satisfação na resposta • disposição para responder • aquiescência ou consentimento na resposta
<ul style="list-style-type: none"> • Acolhimento 	<ul style="list-style-type: none"> • atenção controlada ou seletiva • disposição para receber • percepção

Igualmente relevante foi a contribuição de outro estudioso, Glaser, que estabeleceu as diferenças entre testes referentes a normas e testes referentes a critério. O quadro 3 apresenta uma síntese de suas idéias, feita por Depresbiteris (1976).

Outros estudiosos poderiam ser citados, porém para finalizar é interessante lembrar dois autores brasileiros comumente citados na literatura da área, Vianna e Medeiros.

Devemos a Vianna uma vasta produção de artigos de avaliação voltados para a seleção de alunos. Seu livro "Testes em educação" (1982b) apresenta uma síntese de idéias sobre testes de rendimento escolar, voltados mais diretamente a norma.

Algumas idéias principais deste livro são apresentadas a seguir:

(1) A medida do desempenho escolar é fundamental para uma educação eficiente.

(2) O emprego de instrumentos de medida facilita as observações que o professor faz de seus estudantes.

(3) O processo de medida educacional considera diferentes elementos: um conjunto de situações a responder, as respostas às situações e a classificação dos indivíduos segundo as respostas apresentadas.

(4) A construção de um teste não deve ficar sujeita aos azares da informação: exige um planejamento cuidadoso.

(5) Qualquer que seja o teste, seu propósito e sua filosofia, a sua construção se baseia num sistema de valores humanos.

(6) O planejamento de um teste exige que se considerem as diferentes variáveis que afetam o desempenho do estudante: extensão, tempo, tipo de item e o grau de dificuldade.

(7) Um teste não é uma simples seqüência de itens – exige, para sua montagem, a combinação de diversos critérios.

(8) O desempenho do examinado é influenciado por diversas variáveis. Algumas dessas variáveis podem ser controladas tais como: condições ambientais e métodos de ensino; outras, apesar de poderem ser atenuadas, são de difícil controle uma vez que agem diferentemente sobre cada indivíduo como por exemplo: condições emocionais e afetivas.

(9) Interpretações exatras dependem de escores precisos. O tratamento estatístico não melhora a qualidade dos escores obtidos por instrumentos defeituosos.

Este referente a critério	Este referente a norma	
<p>A interpretação da tarefa é feita frente a padrões absolutos, sem referência ao desempenho de outros indivíduos. Nesta medida, não há suposição da distribuição dos escores, segundo uma curva normal de probabilidade.</p>	<p>Interessa-se pela interpretação do desempenho do indivíduo, em comparação a um grupo de referência. A suposição básica é de que a distribuição dos escores segue uma curva normal de probabilidade.</p>	Diferenças conceituais
<p>A elaboração proposta para este tipo de teste é a seguinte:</p> <p>1. analisam-se os resultados comportamentais de um currículo ou área operacionalizada de conteúdo, em termos de objetivos específicos.</p> <p>2. os itens são elaborados para verificar os resultados esperados, isto é, se o indivíduo alcançou o comportamento pretendido pelo objetivo.</p> <p>3. são estabelecidos níveis aceitáveis de desempenho nos itens de teste os quais são escritos para uma forma específica de instrução.</p>	<p>Para a elaboração deste tipo de teste adotam-se os seguintes procedimentos:</p> <p>1. estabelecem-se os conteúdos de uma área em termos de conjuntos comportamentais, a fim de que os itens possam ser medidos.</p> <p>2. os itens de teste são elaborados de forma a:</p> <ul style="list-style-type: none"> - criarem a máxima variância possível nos escores, - serem livres de ambigüidade. 	<p>Implicações para elaboração e uso do teste</p>
<p>Os itens são escolhidos para serem representativos dos comportamentos esperados na instrução. A validade dos itens é de importância vital. As características básicas deste tipo de teste não se identificam com as de testes referentes a normas, pois seu objetivo básico não é a comparação dos indivíduos, nem tampouco a seleção dos mesmos numa posição classificatória.</p>	<p>Os itens são considerados uma amostra lógica de um universo de itens de aprendizagem ou de um construto qualquer, teoricamente definido. Para seleção dos itens de teste é necessário computar os índices de dificuldade, discriminção, correlação ítem-teste total e verificar se os itens estão livres de ambigüidade. Todos estes procedimentos são necessários, pois a característica básica deste tipo de teste é a classificação dos indivíduos a partir da suposição da curva normal de probabilidade.</p>	Seleção de itens
<p>As técnicas clássicas (Nunnally, 1967; Cronbach, 1970) determinam que para estabelecer validade e precisão dos testes é necessária a suposição de variância nos escores. As referidas técnicas devem servir, também, para estabelecer validade e precisão de testes referentes a critério. Cabe, porém, acrescentar que numa situação de instrução individualizada, onde não seria possível pres- supor variância nos escores, o uso de tais técnicas sofreria uma grave restrição.</p>		Estabelecimento de técnicas e técnicas de validação
<p>Não supõe variância nos escores.</p>	<p>Sendo que a ênfase desta medida está na colocação do indivíduo numa curva normal deve supor variância nos escores obtidos.</p>	Interpretação dos escores

RELAÇÕES ENTRE ESTE TESTE REFERENTE A NORMA E TESTE REFERENTE A CRITÉRIO

(Fonte: Depresbiteris, 1976)

(10) O escore de um teste não é um valor absoluto, mas uma estimativa sujeita a erro.

(11) Quanto mais importante forem as decisões a tomar, maior deve ser a fidedignidade do instrumento de medida, para que essas decisões não tenham conseqüências prejudiciais à vida do estudante.

Medeiros também dedicou-se à avaliação como medida. Seu livro "Provas objetivas: técnicas de construção" (1977a) resalta as seguintes idéias:

(1) A medida faz parte da vida diária. A necessidade de medir surgiu tão cedo na vida do homem, que ele foi buscar no próprio corpo as primeiras unidades de comparação — o cúbito (comprimento do cotovelo à ponta do dedo médio esticado), o palmo, o pé e o dígito. Com o correr do tempo, cuidou de refinar o processo, criando instrumentos especiais de mensuração.

(2) A educação é um processo intencional que implica a verificação de resultados. Na escola, o professor também tem interesse na medida. Ele sabe como é importante manter os alunos informados do êxito dos próprios esforços e prestar igual conta do seu avanço a outros interessados, como pais e administradores. Para isto, recorre às provas e lança mão de outros recursos de verificação.

(3) A eficiência do trabalho educativo está intimamente presa à observação sistemática dos resultados alcançados. Na comprovação periódica das mudanças conseguidas nos educandos é que o educador encontra não só uma base realista para a apreciação crítica do próprio trabalho, mas também pontos de referência para a revisão de sua conduta. Esse controle constante das modificações do comportamento dos indivíduos permite ao educador julgar a eficácia e a eficiência da sua atuação; dá-lhe guias para o reexame do que convém ensinar (ou do conteúdo) e da maneira de fazê-lo (ou do método). Constitui, por conseguinte, tarefa básica da atividade docente. Já que a educação sempre se dirige para fins determinados; fazê-la sem verificações periódicas e freqüentes seria o mesmo que treinar tiro-ao-alvo no escuro.

B. AVALIAÇÃO DE PROGRAMAS

A avaliação de programas também é chamada de pesquisa de avaliação. Segundo Seltiz et al (1987), a avaliação de programas "é um tipo especial de pesquisa aplicada, elaborada para avaliar programas, geralmente programas sociais de melhoramentos, tais como: educação remediatória, reformas no bem estar social, métodos de ensino inovadores, sistemas de distribuição de serviços de saúde, programas de treinamento de pessoal e afins. Os resultados desse tipo de pesquisa não se destinam a incrementar nosso corpo de conhecimentos ou a desenvolver teorias. Eles são usados, muitas vezes imediatamente, para decidir se os programas devem parar ou continuar, se as verbas devem ser aumentadas ou diminuídas, se é preciso contratar ou despedir pessoas — tudo isso com base no atendimento do programa àquilo para que fora destinado".

Talmage (1982) diz que a avaliação de programas ou pesquisa de avaliação começou a ganhar corpo já em 1950, quando um grupo de críticos passou a exigir que fossem realizadas avaliações mais amplas, envolvendo inúmeras variáveis de um programa educacional. Essa solicitação, aliada ao incentivo governamental, incrementou os estudos sobre desenvolvimento de currículos e sobre programas de intervenção social.

Com relação a esse assunto, Popham (1983) diz que a iniciativa governamental para a avaliação de currículos teve origem quando o primeiro satélite (Sputnik) foi lançado pela União Soviética em 1957 e não pelos Estados Unidos da América. Os ataques à escola, neste último país, tornaram-se enfáticos, responsabilizando-a pela derrota na corrida espacial, o que resultou num investimento de milhões de dólares em uma série de iniciativas educacionais e, concomitantemente, na sua avaliação.

Retornando a Talmage (1982) tem-se que os estudos sobre o desenvolvimento de currículos, a partir dessa época, serviram para modernizar os conteúdos e métodos de ensino na área de Ciências e Matemática, com a finalidade de tornar os estudantes os melhores em tecnologia.

Quanto aos estudos avaliativos sobre programas de intervenção social, esse autor diz que eles emanaram de uma política denominada "Guerra à Pobreza" (1960-1970) e que tinham por finalidade desenvolver estratégias para concretizar propósitos sociais igualitários.

Os avaliadores de currículo eram oriundos, principalmente, da Pesquisa e Psicologia educacionais, trazendo para a avaliação toda a experiência adquirida nestas disciplinas. Os avaliadores de programas de intervenção social, oriundos das Ciências Políticas e Sociologia, reconheceram de imediato algumas limitações frente às novas exigências avaliativas, como: levantar questões avaliativas pertinentes, desenraizar-se dos "designs" de controle amplamente usados na pesquisa experimental, engajar-se na vida do aluno em sala de aula, enfim, aplicar as teorias no campo naturalístico. Contudo, a integração das múltiplas experiências desses avaliadores possibilitou a melhoria e a inovação de metodologias nessa área.

Talmage apresenta uma distribuição dos avaliadores segundo o domínio de diferentes linhas metodológicas, conforme síntese apresentada no quadro 4.

Aprofundando as idéias apresentadas no quadro 4, sempre com base em Talmage, pode-se verificar o seguinte:

(a) *Com relação à base filosófica*

Os experimentalistas baseiam-se em idéias positivistas, ou seja, de relação causa e efeito. As hipóteses são dedutivas; o pesquisador começa com um referencial teórico, formula uma hipótese e deduz logicamente quais devem ser os resultados do teste da mesma, se ela for correta. Os ecléticos baseiam-se numa filosofia pragmática. Reconhecem o valor do método científico em transportar novos conhecimentos e estabelecer a verdade (até que ela seja negada e outra verdade provada), mas mostram a importância concernente às múltiplas percepções e seus efeitos sobre os dados. Os descritivos rejeitam as premissas positivistas e o pragmatismo dos ecléticos e defendem uma filosofia fenomenológica que se baseia numa linha humanística ao invés de comportamental, mais qualitativa que quantitativa, mais universalista que particularista, mais naturalista do que experimental. Os analisadores de custo-benefício baseiam-se no utilitarismo, pelo qual o programa é julgado pela sua eficiência em atingir o máximo de estudantes, com maior qualidade.

(b) *Focos da metodologia e variáveis*

A principal preocupação dos experimentalistas é a identificação de relações causais e, para isso, baseiam-se em delineamentos experimentais e quase experimentais. As variáveis são predeterminadas para o estudo. Os ecléticos não rejeitam os delineamentos experimentais, mas também acham importante analisar o processo e o contexto dos programas educacionais. Defendem a idéia de estudos de caso e descrições abrangendo não apenas as variáveis predeterminadas, mas também as emergentes da própria situação. Neste ponto, Talmage cita Bryk, para o qual uma síntese da posição dos ecléticos é traduzida nesta frase: "nós deveríamos atentar para traçar os pontos fortes de cada abordagem – a objetividade trazida pelos dados quantitativos e a riqueza da informação qualitativa – para criar uma visão integrada do programa de impacto". Os descritivos defendem o emprego de estudo de caso e observação participante, com descrições do programa no contexto e através do testemunho dos participantes. As variáveis não são pré-selecionadas, mas emergem no processo de descrever a operação do programa no contexto naturalístico. Os analisadores de custo-benefício baseiam-se em abordagens experimentalistas para fundamentar suas análises voltadas para aspectos econômicos dos programas. Estudos de custo-benefício são delineados para determinar como antecipar custos e alternativas de benefício. A metodologia, inspirada no campo econômico, prevê um caminho de converter valores sociais em custo-benefício.

(c) *Interação avaliador e avaliado*

Para os experimentalistas e para os analistas de custo-benefício, o papel dos avaliadores não é ativo, eles funcionam como respondentes. O avaliador, por sua vez, tem um papel independente do programa, funcionando como elemento que subsidia decisões de continuidade ou não deste programa. Para os ecléticos, a relação entre avaliador e avaliado é cooperativa.

Analistas de Custo/Benefício	Descritivos	Parlett e Hamilton (1977) Stake (1975) Patton (1980)	Halperin (1975) Thompson (1980)	Parlett e Hamilton (1977) Stake (1975) Patton (1980)	Descritivos	Analistas de Custo/Benefício
Experimentalistas	Ecléticos	Byrk (1978) Cronbach e associados (1980) R.S. Weiss e Rein (1972)	Ravlin e Jimpane (1974) Riecken e Bonuch dos (1980)	Cook e Campbell (1979) Riecken e Bonuch dos (1980) Ravlin e Jimpane (1975)	Experimentalistas	Analistas de Custo/Benefício
Base Filosófica	Positivista	Modificar o positivismo para o pragmatismo			Base Filosófica	Analistas de Custo/Benefício
Base disciplinar	Psicologia	Psicologia, Sociologia, Ciência Política			Base disciplinar	Analistas de Custo/Benefício
Focos da Metodologia	Identificar relações causais	Ampliação da investigação para relações causais com processo e dados contextuais	Estudos de caso, experimentais, quase-experimentais	Estudos de caso, observação participante, etnografia	Metodologia	Analistas de Custo/Benefício
Variáveis	Predefinidas	Predefinidas e emergentes	Predefinidas	Predefinidas	Variáveis	Analistas de Custo/Benefício
Grupo de comparação ou de controle	Sim	Quase impossível	Sim	Sim	Grupo de comparação ou de controle	Analistas de Custo/Benefício
Papel dos participantes na avaliação	Nenhum	Nenhum para interação	Vários (participantes podem reagir as observações)	Nenhum	Papel dos participantes na avaliação	Analistas de Custo/Benefício
Papel do avaliador	Independente do programa	Cooperativo	Interativo	Independente do programa	Papel do avaliador	Analistas de Custo/Benefício
Pressões políticas (internas e externas)	Controladas em desígnios ou ignoradas	Acomodadas	Descritiva	Ignoradas	Pressões políticas (internas e externas)	Analistas de Custo/Benefício
Foco do relatório de avaliação	Propiciar a decisão de continuar ou não continuar	Interpretar e recomendar melhorias no programa	Representação honesta do processo	Propiciar julgamento	Foco do relatório de avaliação	Analistas de Custo/Benefício

(Fonte: Talmage, 1982)

QUATRO ABORDAGENS METODOLÓGICAS DE AVALIAÇÃO DE PROGRAMAS.

QUADRO 4

rativa, buscando juntos interpretar e recomendar melhorias no programa. Para os descritivos, a relação é interativa e os avaliados, que são chamados de participantes, podem inclusive reagir às observações.

Prosseguindo em sua análise, Talmage apresenta a idéia de modelos de avaliação de programas, os quais se constituem em quadros referenciais que embasam o planejamento da avaliação. O autor diferencia entre os modelos que não explicitam as orientações teóricas sobre as quais se baseiam, tais como o modelo de Stufflebeam e Guba e o modelo de discrepâncias de Provus; e os modelos que explicitam pressupostos tais como: interação entre avaliador e cliente, modos de facilitar decisões, formas de estabelecer relações causais, pressupostos encontrados no modelo de avaliação transacional de Rippey, modelo de avaliação responsiva de Stake e o modelo de avaliação iluminativa de Parlett e Hamilton.

Talmage cita Ross e Cronbach que, em uma crítica ao *"Handbook of evaluation research"*, dividem os modelos em duas categorias: modelos cujas atividades são orientadas pelo avaliador e modelos em que ambos, avaliador e participantes, desempenham papéis ativos no processo avaliativo. De modo similar, Guba usa duas categorias para diferenciar entre os modelos: forma de inquérito convencional e forma de inquérito naturalístico. A primeira é baseada na epistemologia lógica-positivista (causa e efeito) e a segunda na epistemologia fenomenológica.

Outro estudioso citado por Talmage é Popham, que sugere três classificações de modelos: (1) modelos de alcance de objetivos como, por exemplo, os de Tyler e Metfessel-Michael; (2) modelos de julgamentos intrínsecos e extrínsecos, tendo como exemplo do primeiro o estabelecido pelo *"National Study of School Evaluation"* e como exemplos do último os modelos de Scriven e Stake; (3) modelos facilitadores de decisão, dentre os quais os de Stufflebeam e Guba; Stufflebeam, Alkin e Provus.

Ainda nessa perspectiva de diferenciar modelos, Talmage menciona Stufflebeam e Webster, que os classificam em três abordagens: estudos orientados politicamente, que os autores chamam de "pseudo-avaliações"; estudos orientados para questões chamados "quase-avaliações" e estudos orientados para valor ou chamados "verdadeiras avaliações". Para os autores, somente os estudos que permitem julgamento de valor entram na categoria de avaliação verdadeira, que não pode ser confundida com experimentos verdadeiros, como os de Campbell e Stanley.

Certamente, apresentar todos os modelos de avaliação seria uma tarefa inviável neste ensaio, porém é necessário relatar alguns deles, em virtude de sua freqüente citação na literatura da área. Esses modelos são sintetizados a seguir.

Modelos de Stake — Stake propõe inicialmente um modelo racional de avaliação, descrevendo todos os passos que um avaliador deveria executar para poder julgar os resultados obtidos. Depois de alguns anos, Stake apresentou um outro modelo, caracterizado como avaliação responsiva que é aquela que se orienta mais em direção às atividades de um programa educacional do que para suas intenções. A síntese apresentada aqui baseia-se num artigo de Stake (1977) e no quadro traçado por Guba e Lincoln (em Werthein et al., 1985).

No modelo mais racional de Stake, a avaliação é um processo de descrever programas educacionais, através de investigação formal. Formal, porque esse tipo de avaliação pressupõe bases científicas, objetivas e precisas, possibilitando informações imediatas para a tomada de decisão. Contrapõe-se à investigação informal, isto é, investigação baseada em observação casual, que implica metas implícitas, em normas intuitivas, em julgamentos subjetivos. O autor propõe duas matrizes básicas, conforme se verifica na figura 2.

A matriz de descrição tem como entrada as intenções e as observações; a matriz de julgamento, os padrões e os julgamentos. As intenções envolvem o planejamento das condições ambientais da especificação dos objetivos, enfim, de tudo o que pode acontecer no programa educacional. As observações são relativas às atividades que realmente acontecem no decorrer do programa. Os padrões são níveis de julgamento de duas ordens: relativos e absolutos. Os padrões relativos são aqueles que se referem a outros programas; os absolu-



Figura 2 – Matrizes do modelo de Stake.
(Fonte: Bellack e Klebard, 1977)

tos são os padrões referentes aos objetivos do programa que está sendo julgado. Os julgamentos são atribuições de peso, devendo prover uma resposta imediata para a tomada de decisão. Os antecedentes são todas as condições anteriores ao programa e que se relacionam com os resultados, tais como: interesses, recursos e experiências anteriores. As transações são todas as atividades realizadas para o alcance dos objetivos, incluindo relação e interações entre aluno e professor. Os resultados são os produtos da aprendizagem e que podem ser identificados como habilidades, realizações, atitudes, expectativas cumpridas, transferência de aprendizagem. Através da matriz de descrição, o avaliador descreve intenções e observações. A descrição pode ser feita direta e pessoalmente, ou através do uso de instrumentos adequados a cada situação.

No modelo de Stake existem duas maneiras de processar os dados obtidos pela matriz de descrição: contingência e congruência. A figura 3 apresenta o processamento por contingência.

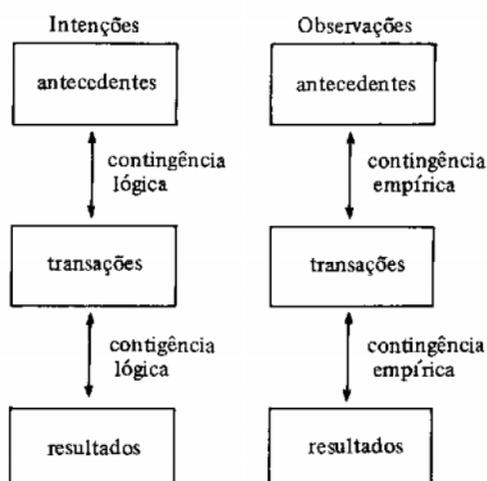


Figura 3 – Processamento dos dados da matriz de descrição de Stake, por contingência.
(Fonte: Bellack e Klebard, 1977)

O processamento dos dados descritivos pode ocorrer por contingência que se refere à ligação vertical entre antecedentes, transações e resultados, isto é, procura verificar até que ponto os resultados foram produtos das transações e dos antecedentes. Na avaliação das intenções, o critério de contingência é lógico, isto é, baseado em experiências prévias, pesquisas, julgamentos de outros especialistas, entre outros. O critério de contingência na avaliação das observações é empírico, isto é, depende da evidência empírica dos fatos. Como já mencionado, o processamento dos dados descritivos pode ocorrer também por congruência, conforme se vê na figura 4.

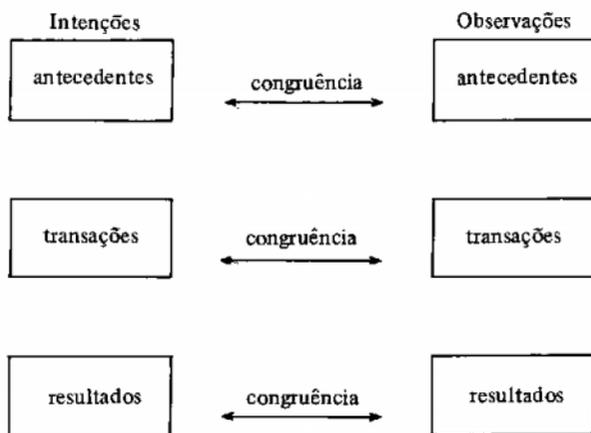


Figura 4 – Processamento dos dados da matriz de descrição de Stake, por congruência. (Fonte: Bellack e Klebard, 1977)

A congruência existe quando o que é pretendido ocorre realmente, ou seja, quando não há discrepância entre antecedentes, transações e resultados pretendidos e observados. O processamento de dados de julgamento pode ser feito através da comparação relativa ou absoluta. A comparação relativa envolve padrões retirados das descrições de outros programas, isto é, programas alternativos, conforme se vê na figura 5.



Figura 5 – Comparação relativa no modelo de Stake. (Fonte: Bellack e Klebard, 1977)

Por comparação absoluta entende-se o conjunto de padrões pelos quais se pode julgar um programa (padrões de excelência). Existem conjuntos de padrões porque há vários grupos de referência ou de pontos de vista (julgamentos pessoais). A figura 6 apresenta a idéia de comparação absoluta.



Figura 6 – Comparação absoluta no modelo de Stake.
(Fonte: Bellack e Klebard, 1977)

O julgamento do programa é feito através dessas duas comparações relativa e absoluta, como mostra a figura 7.

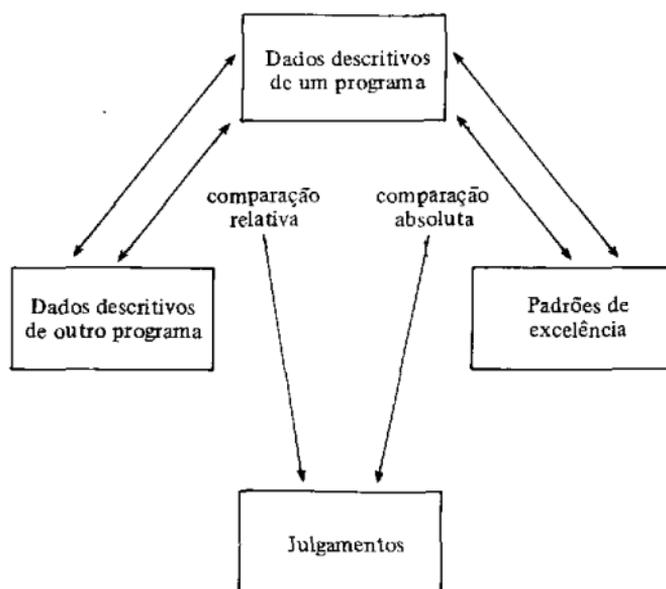


Figura 7 – Julgamento do programa no modelo de Stake.
(Fonte: Bellack e Klebard, 1977)

Através da proposição deste modelo de avaliação racional, Stake quer demonstrar que: (a) o trabalho do avaliador deve ser objetivo ou seja, para descrever e julgar os programas o avaliador deve se basear sempre em procedimentos científicos; (b) as matrizes de descrição e julgamento podem ser aplicadas na sala de aula e (c) os resultados de avaliação devem ser tornados públicos.

O modelo de avaliação responsiva de Stake mostra a idéia de resposta contínua à audiência, numa interação com o avaliador. A avaliação responsiva é aquela que se orienta mais em direção às atividades de um programa educacional de que para suas intenções. Para melhor compreender as características da avaliação sensível de Stake, observe-se o quadro 5, no qual ela é comparada com a avaliação tradicional.

QUADRO 5

COMPARAÇÃO ENTRE AVALIAÇÃO TRADICIONAL E AVALIAÇÃO SENSÍVEL DE STAKE.
(Fonte: Guba e Lincoln, 1981 (Trad. por Werthein et al 1985))

	Tipos de Avaliação	
	Tradicional	Responsiva*
Orientação	Formal	Informal
Valoração	Singular; consensual.	Pluralística; admite a possibilidade de conflito.
Bases do "design"	Intenções do programa, objetivos, metas, hipóteses, idéias do avaliador sobre coisas como desempenho, domínio, habilidade, aptidão, resultados mensuráveis, os valores instrumentais da educação.	As preocupações são as questões da audiência, atividades do programa e de pessoas envolvidas no programa.
Prazo para o "design"	No início da avaliação.	Não há – em contínua evolução.
Papel do Avaliador	Estimulador dos sujeitos para testar desempenho crítico.	Deixa-se estimular pelos sujeitos e pelas atividades.
Métodos	Objetivos: "passar na prova".	Subjetivos, por exemplo, observações e entrevistas.
Comunicação	Formal: relatórios, tipicamente unidirecionais.	Informal; descrições (portrayals) geralmente multidirecional.
"Feedback"	A intervalos discretos, em geral apenas um, ao término	Informal; evoluindo continuamente com as necessidades
Forma de "feedback"	Relatório escrito, identificando variáveis e descrevendo as relações entre elas, interpretação simbólica.	Descrição do tipo narrativo, muitas vezes oral (se a audiência o preferir), gerando imagens do que é o programa, possibilitando experiências simuladas.
Paradigma	Psicologia experimental.	Antropologia, Jornalismo, Poesia.

Modelo de Scriven. As idéias de Scriven apresentadas a seguir foram recolhidas em Anderson (1981) e Scriven (1978).

Para o autor, a avaliação é a determinação sistemática e objetiva do mérito ou valor de alguma coisa. A ênfase está no aspecto comparativo da avaliação, uma vez que avaliar só tem sentido se evidenciar quão bem ou quão mal se saiu o objeto de análise. Scriven divide o processo de avaliação em duas grandes etapas: (a) julgamento dos instrumentos de coleta de dados, da atribuição de pesos aos objetivos e da seleção de critérios e padrões;

(b) julgamento do valor ou do mérito de um programa educacional com base na coleta sistemática de dados de desempenho, na combinação desses dados com um conjunto de metas valoradas e na análise dessa informação. Scriven estabelece uma distinção marcante entre o objetivo e as funções da avaliação. Para o autor, o objetivo da avaliação é julgar o mérito de alguma coisa. As funções da avaliação são de duas ordens: formativa e somativa. A avaliação formativa consiste no fornecimento de informações a serem utilizadas na melhoria de um programa em suas partes ou em seu todo. A avaliação somativa consiste no fornecimento de informações sobre o valor final de um programa instrucional. Scriven ainda relaciona essas funções a dois tipos de julgamento: intrínseco (relativo ao conteúdo, materiais, currículo) e extrínseco (efeitos do programa). O quadro 6 mostra essa relação.

QUADRO 6

RELAÇÃO ENTRE AVALIAÇÃO FORMATIVA E SOMATIVA E JULGAMENTOS INTRÍNSECO E EXTRÍNSECO, NO MODELO DE SCRIVEN (Fonte: Worthen e Sanders, 1973)

	Intrínseco	Extrínseco
Formativa	Julgamento da estrutura e dos conteúdos do programa educacional.	Julgamento dos efeitos intermediários do programa com a finalidade de fornecer realimentação.
Somativa	Julgamento final dos materiais.	Julgamento final dos efeitos dos materiais em alunos, professores, administradores e outros.

Scriven também propôs uma nova idéia de avaliação. Ele, que inicialmente defendia a idéia de que a avaliação deveria estar fortemente atrelada aos objetivos de um programa educacional, passa a encarar a possibilidade de uma avaliação sem referência a objetivos (ASO). A idéia surgiu quando Scriven teve que avaliar um projeto e sentiu dificuldades em separar os fins dos efeitos secundários. Ele concluiu que, evidentemente, a avaliação dos fins é importante, mas não reflete tudo. É necessário, também, detectar os efeitos secundários e avaliar se estes são ou não intencionais, e isso pode ser dificultado se os objetivos direcionam extremamente o avaliador. Para Scriven, a principal ênfase da avaliação sem referência a objetivos é a preocupação com a qualidade. Os dados quantitativos são importantes, porém os dados qualitativos representam a fonte principal. De acordo com este tipo de avaliação, não se considera apenas o aspecto técnico de um projeto, mas também seus efeitos sociais.

Modelo de Stufflebeam. As idéias de Stufflebeam aqui sintetizadas foram colhidas principalmente em seu livro "Educational Evaluation: decision making" (1974).

O autor define a avaliação "como um processo de delinear, obter e fornecer informações úteis ao julgamento das alternativas de decisão". Para ele há oito termos básicos para explicar a avaliação, que são os seguintes:

processo – atividade contínua que inclui muitos métodos e técnicas e envolve várias operações e passos, geralmente sucessivos, e que podem ser repetidos de forma interativa;

delinear – identificar informações solicitadas, através do exame das alternativas de decisão e dos critérios a serem aplicados;

obter – tornar a informação disponível, através da coleta, organização e análise dos dados, utilizando-se de técnicas de medição, processamento de dados e análises estatísticas;

fornecer – organizar e agrupar as informações em sistemas que possam melhor servir aos propósitos da avaliação;

informação – dados descritivos e interpretativos sobre os programas, alunos e materiais instrucionais e seu relacionamento em função de algum propósito;

úteis – as informações devem ser úteis para satisfazer a certos critérios científicos, tais como: validade, confiabilidade e importância;

alternativas de decisão – ações diferentes entre si, que podem ser tomadas em resposta a uma situação que necessita de alteração.

A ênfase que Stufflebeam dá ao processo de julgamento e de tomada de decisão encontra-se refletida na relação entre avaliação, decisão e fases de um projeto, estabelecidas conforme mostra a figura 8.

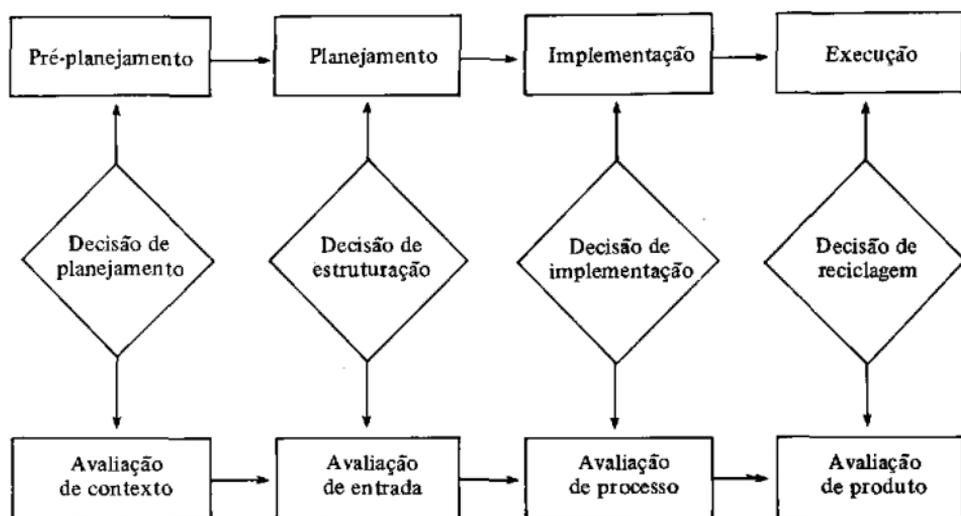


Figura 8 – Representação do modelo de Stufflebeam
(Fonte: Stufflebeam, 1974)

Conforme diz o autor, a avaliação de contexto é o tipo básico de avaliação, cujo propósito é o de delinear, obter e fornecer informações úteis para a determinação dos objetivos de um programa educacional. Através da avaliação de contexto, o ambiente, as condições desejadas e existentes são descritos e são detectadas as necessidades do programa. Esse diagnóstico propicia as decisões de planejamento, através das quais são especificadas as mudanças a serem introduzidas no sistema, caso sejam necessárias. Dessas decisões surge, como consequência, um julgamento com a descrição do que está ocorrendo e do que deveria ocorrer.

A avaliação de entrada deve prover informações para determinar como utilizar recursos no sentido do alcance dos objetivos estabelecidos. O produto dessa avaliação é uma análise de planos de procedimentos em termos de estratégias e custos. As decisões envolvidas nesse tipo de avaliação são as de estruturação que especificam os métodos, o conteúdo, a organização, o pessoal e as facilidades. Essas decisões permitem o planejamento final da ação.

A avaliação de processo, cujo objetivo é fornecer informações periódicas sobre os procedimentos em ação, relações entre os membros da equipe responsável pelo programa, inadequação de recursos, discordância entre meios e fins, permite aos decisores tomar as devi-

das providências durante o processo, evitando o acúmulo de erros futuros. As decisões envolvidas nessa avaliação são as de implementação, tomadas sempre no desenrolar do plano de ação. Essas decisões permitem o estabelecimento da estrutura do programa, a definição de responsabilidades e outros aspectos importantes.

A avaliação de produto mede o alcance dos objetivos estabelecidos pelo programa, não somente ao fim de um ciclo, mas sempre, quando necessário, durante o desenrolar das ações. Esse tipo de avaliação tem, assim como a avaliação de processo, a função de controle do programa, além da função de explicação de seus resultados. A avaliação de processo enfoca os procedimentos, verificando se estes estão sendo efetivados conforme o pretendido; a avaliação de produto enfoca os resultados, ou seja, investiga em que medida os objetivos foram alcançados. As decisões envolvidas nessa última são de reciclagem que determinam a continuação, a modificação ou o término das atividades do programa. Essas condições embasam a fase de execução onde o programa é testado e instalado.

Modelo de Parlett e Hamilton. As idéias desses autores aqui apresentadas baseiam-se em Hamilton e Parlett (1982).

Parlett e Hamilton advogaram em 1971-72 total revisão dos fundamentos e das técnicas de avaliação de programas. Propuseram a avaliação iluminativa, cujo objetivo é estudar intensamente o programa educacional inovador, tal como este funciona: como é influenciado pelas diversas situações nas quais é aplicado; quais são suas operações, vantagens e dificuldades, na opinião de seus principais interessados; quais as condições que mais afetam as tarefas intelectuais e a experiência acadêmica dos alunos.

Os autores salientam a importância de se considerar uma nova concepção de sistema de ensino e de meio de aprendizagem. Por este enfoque, o sistema de ensino não pode ser encarado como um conjunto de manuais preestabelecidos a partir de pressupostos pedagógicos formais, com a definição de técnicas e estratégias das quais os docentes extraem propósitos, objetivos e resultados, verificando se foi ou não conseguido o que era esperado. A avaliação iluminativa pressupõe que um sistema de ensino quando adotado sofre modificações quase sempre muito importantes. O sistema de ensino assume diferentes formas em situações concretas, tendo seus elementos diminuídos ou ampliados, de acordo com as diferentes interpretações que cada situação exige. A formulação "ideal" deixa não apenas de ser precisa, como também de ser importante. Passa-se, então, de uma discussão técnica para uma discussão crítica, o que exige uma nova postura dos envolvidos no processo ensino-aprendizagem.

Nessa discussão, ganha ênfase o meio de aprendizagem, que os autores definem como o contexto sócio-psicológico e material em que docente e alunos trabalham juntos. Reconhecer a diversidade e complexidade dos meios é um pré-requisito essencial para um estudo sério de programas educacionais. Os autores afirmam que qualquer meio de aprendizagem é tão complexo que não se pode lidar com ele através de modelos quantitativos apriorísticos. Os avaliadores devem introduzir-se no meio de aprendizagem, experimentá-lo e estudá-lo, através de um processo de focalização progressiva no qual observadores humanos exploram, julgam, investigam e interpretam, iluminando as perspectivas, muitas vezes conflitantes, dos vários participantes envolvidos. Em termos de técnicas, os autores pregam a diversidade e enfatizam a necessidade de diferentes cruzamentos de dados, sempre levando em conta a amplitude e complexidade das situações. Uma frase desses autores resume bem a idéia metodológica deste tipo de abordagem: "A avaliação iluminativa não é um pacote metodológico, mas uma estratégia geral de pesquisa. Ela visa a ser ágil e eclética". Assim, o que fica patente é a escolha de táticas de pesquisa baseada nas decisões tomadas em cada caso: o problema dita a utilização de certos métodos, sendo que o inverso não é verdadeiro.

A contribuição de Goldbérq

No Brasil, Goldberg apresenta uma contribuição bastante significativa na área de avaliação de programas. Uma contribuição específica é apresentada em seu artigo "*Avaliação e planejamento educacional: problemas conceituais e metodologias*" (1973) que, em síntese, discute os vários aspectos referidos a seguir.

(1) Um programa educacional é uma estrutura de decisões. A mais fundamental delas diz respeito a objetivos educacionais, porque quando se fala em planejamento, refere-se, por excelência, à racionalização da conduta humana.

(2) A avaliação será sempre uma avaliação de objetivos educacionais, isto é, das funções do processo de ensino. A avaliação funciona como controle de qualidade do próprio processo de planejamento educacional.

(3) O controle de qualidade do processo de planejamento educacional é feito verificando-se sua eficácia e eficiência.

(4) A eficácia diz respeito à utilização, pela sociedade, do aluno formado pelo programa educacional. Ela será tanto maior quanto mais alta for a probabilidade de aceitação do aluno formado. Em síntese, a eficácia diz respeito ao valor e à viabilidade dos objetivos educacionais de uma programação. A eficiência diz respeito ao julgamento sobre o impacto de um programa educacional.

(5) Existem quatro tipos de avaliação: (1) diagnóstica, feita antes do planejamento; (2) "ex-ante" posterior ao plano de trabalho, mas antes de sua aprovação; (3) "in-processu", que acompanha a execução do plano e (4) "ex-post", posterior à obtenção dos resultados.

(6) A avaliação diagnóstica e a avaliação "ex-ante" permitem analisar a eficácia das decisões relativas aos objetivos educacionais. A avaliação "in-processu" e avaliação "ex-post" são duas formas de avaliar a eficiência dos objetivos educacionais.

(7) A eficácia diz respeito ao valor (utilidade social), à viabilidade (julgamento do dimensionamento de recursos humanos). Pela eficácia pode-se determinar se os objetivos educacionais estão sendo ou foram atingidos pela maioria dos alunos, da melhor maneira possível. A eficiência diz respeito à produtividade (rapidez, mínimo esforço, baixo custo) e ao rendimento (resultados esperados).

(8) A avaliação diagnóstica oferece ao planejamento critérios pelos quais pode fundamentar suas decisões sobre metas e estratégias educacionais. A avaliação "ex-ante" permite verificar a existência ou não de uma adequação entre objetivos e estratégias e as variáveis relacionadas ao professor e aos alunos. A avaliação "in-processu" determina se a execução está se processando conforme o planejado, se está havendo produtividade na operação do programa. A avaliação "ex-post" determina se os objetivos foram alcançados, se houve rendimento, se este foi aceitável e se pode ser considerado como tendo sido produzido pelo programa.

(9) Em termos de variáveis, existem três grandes classes que permeiam o processo ensino-aprendizagem: as variáveis antecedentes, as variáveis processuais e as variáveis produto. As variáveis antecedentes abrangem todas as condições que preexistem à interação professor-aluno. Dentre as variáveis do professor, destacam-se as que se relacionam a ele enquanto profissional, tais como: conhecimentos e habilidades específicas e cursos de formação. Dentre as variáveis do aluno, incluem-se conhecimentos, habilidades, atitudes e interesses, entre outras. Existem ainda as variáveis contextuais, que representam as condições externas, tais como condições sócio-econômicas e políticas; interação professor e aluno. As variáveis processuais abrangem todas as interações professor-aluno e todas as formas pelas quais o aluno tem que lidar com material didático e de entrar em interação com os demais elementos do grupo instrucional. As variáveis produto compreendem as mudanças comportamentais produzidas no aluno. Essas mudanças podem refletir modificações em conhecimentos, habilidades, interesses e atitudes. A figura 9 mostra a inter-relação dessas variáveis.

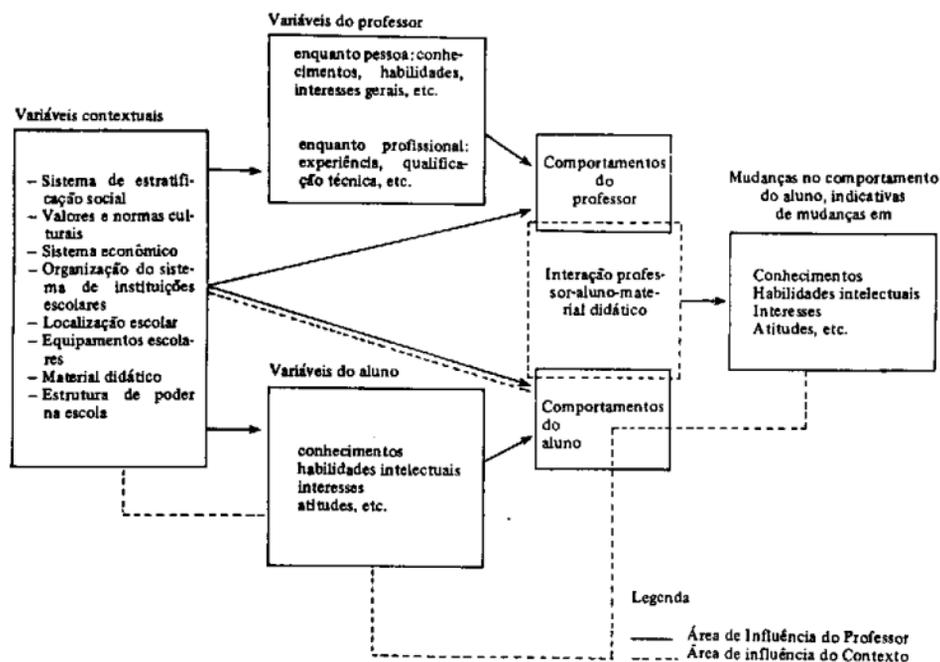


Figura 9 – Variáveis do processo ensino-aprendizagem, segundo Goldberg. (Fonte: Goldberg, 1973).

Embora não pertençam estritamente ao histórico da avaliação aqui apresentado, duas linhas teóricas de pesquisa, que vêm ganhando crescente divulgação entre nós, em anos recentes, serão mencionadas a seguir, porque ambas se ajustam aos propósitos mais amplos de avaliação de programas ou de todo um sistema ou organização complexos. Trata-se da pesquisa-ação e da observação participante.

Pesquisa-ação

A pesquisa-ação descende, em linha direta, das contribuições de Lewin (1890-1947). Conforme assinala Bickman, na introdução que escreveu para *Métodos de pesquisa nas relações sociais* (Sellitz, 1987), "interessado nos assuntos sociais da época, Lewin traduziu suas preocupações em pesquisas sobre tópicos, tais como conflito intergrupar, produtividade do trabalhador e liderança. Ele via esses projetos como pesquisa-ação, combinações de experimentos e aplicações. Lewin sentia que sua pesquisa poderia beneficiar a sociedade, bem como aprimorar o conhecimento científico".

Com relação às bases teóricas, Mitchell (1977) assinala que “boa parte da ciência social decorre do desejo de alterar e melhorar uma situação social, ou ajudar pessoas que necessitam de auxílio. A pesquisa-ação é uma investigação cuja natureza se orienta para essas finalidades, nas quais o objetivo não é apenas coletar informações e chegar a uma melhor compreensão, mas fazer algo prático. Às vezes os expoentes da pesquisa-ação lançam dúvidas sobre a possibilidade de realizar estudos “frios” ou neutros e científicos de assuntos humanos. Eles argumentam, por exemplo, que não há como impedir que um investigador deixe de influenciar o comportamento das pessoas que está estudando, que a experimentação é extremamente difícil, se não impossível, nas ciências sociais, que há intermediação do instrumento humano na medida e que tudo isso prejudica o “status científico da pesquisa social”.

Mitchell prossegue registrando que Argyle, no livro *“The scientific study of social behavior”* (1957), criticou, de modo incisivo, a pesquisa-ação, argumentando que nela a descoberta de resultados científicos é sempre secundária ou de pouca importância, podendo-se, pois, duvidar de sua contribuição para o progresso da ciência.

English (1958) diz que, de modo geral, “a pesquisa-ação é feita por alguns dos próprios participantes no programa de ação. As atividades são realizadas com a esperança de alcançar certos resultados de caráter utilitário ou prático, e a pesquisa é planejada a fim de efetivar um progresso no processo em curso e não meramente em algum processo futuro. No experimento puro, de modo contrastante, as atividades são prescritas primariamente para a obtenção de dados que poderão ser analisados para fins de generalização científica ou tecnológica”.

No contexto educacional, a pesquisa-ação ganhou o interesse dos especialistas a partir dos anos cinquenta, segundo Fleming (1970). De acordo com Schubert (1982), a pesquisa-ação, em termos de currículo, foi implementada de modo prático, no início da década de oitenta, por Stenhouse. Este último autor (1980) vê a pesquisa de currículo como um processo graças ao qual pesquisadores-consultores interagem como professores nos seus ambientes, encorajando a reflexão cooperativa a respeito de situações específicas e o estudo cooperativo dessas situações; o objetivo é melhorar o ensino com a fusão do investigador com o professor.

Em todas essas manifestações de pesquisa-ação em contextos psicológicos e educacionais, é perceptível a influência lewiniana e de seus discípulos. Um destes continuadores da obra de Lewin, Lippitt (1984), escreveu que o termo pesquisa-ação tem três diferentes significados que precisam ser distinguidos. Primeiro, pode-se tratar de uma pesquisa de diagnóstico sobre algum aspecto em marcha de um processo de ação, como anomia comunitária, moral, organizacional ou produtividade de grupo: neste sentido é pesquisa com foco na ação. Segundo, designa um procedimento de coleta de dados por participante de algum sistema de grupo, e a seguir a previsão de um “feedback” sobre o que se encontrou nos dados, como uma intervenção destinada a influenciar, presumivelmente de algum modo útil, o processo de ação em andamento do grupo. Terceiro, é um procedimento no qual os participantes em algum sistema social são envolvidos no processo de coleta de dados sobre si mesmos, e utilizam a seguir os dados gerados para rever os fatos sobre si próprios, a fim de realizar uma ação remediativa ou de desenvolvimento. Lippitt diz que o terceiro tipo de pesquisa-ação é o melhor sucedido para assegurar uma ligação efetiva entre a busca de fatos e sua utilização.

O autor confessa que em relação ao primeiro tipo de pesquisa-ação, a história é desapontadora no que diz respeito às conseqüências práticas. Diz o autor: “usualmente os volumosos relatórios têm sido ignorados ou sucumbem diante da resistência daqueles que os recebem, ou, ainda, são lidos atentamente, mas sem que se use os resultados de modo significativo”.

Quanto aos outros dois tipos de pesquisa-ação, Lippitt lembra que Lewin e colaboradores os conceituaram e operacionalizaram, após a fundação de *“Research Center for Group Dynamics”* no Massachusetts Institute of Technology (MIT), em 1945. A aplicação inicial

deu-se em relação a problemas de preconceitos e mudança de atitudes em relações intergrupais comunitárias (Lewin envolveu-se ativamente nisto, como parte do seu trabalho junto à Comissão de Inter-relações do Congresso judaico norte-americano).

Ainda na década de quarenta, os trabalhos de Cook (1976) trouxeram contribuições relevantes para a pesquisa-ação, sob a forma de levantamento de atitudes na comunidade, seguidos de "feedback" dos habitantes e discussão com estes, das implicações dos resultados e desenvolvimento de planos de ação. Tratava-se, nestes casos, de intervenção deliberadamente planejada para influenciar a mudança de atitudes, por meio de "feedback" de informação. Durante este período, outro desenvolvimento importante em matéria de pesquisa-ação, nesta perspectiva, foi o grupo "T" ou grupo de treinamento, por Lewin e colaboradores. Proporcionava-se ao grupo um observador dos processos nele realizados, com a responsabilidade de fornecer dados aos grupos sobre si mesmos, à medida do transcorrer do trabalho em áreas como liderança, tomada de decisão, capacidade de lidar com "agendas ocultas", abrir linhas de comunicação, melhorar a fixação de alvos e outros fenômenos similares. Um desdobramento posterior foi o programa de intervenções e estudos de "feedback" organizacional na Universidade de Michigan, a que estiveram associados os nomes de Floyd Mann e Rensis Likert.

Lippitt acrescenta que os pesquisadores sociais, mais interessados em criar condições para mudança pessoal e mudança social em virtude da coleta de dados e "feedback", descobriram três vantagens no envolvimento dos grupos nesse processo: (1) quanto mais as pessoas se envolvem no planejamento e coleta de dados, tanto mais adquirem uma "propriedade psicológica" dos dados (sentem-se "donas" deles) e aceitam sua credibilidade e relevância como intervenção para mudança, (2) a coleta de dados por membros do grupo voluntário é muito mais econômica, (3) o treinamento de membros para conduzir suas próprias indagações, que continua sempre como um recurso à disposição do grupo.

Observação participante

A observação participante faz parte de um conjunto mais vasto conhecido como "metodologia qualitativa". De acordo com Thines e Lempereur (1984), "a origem da observação participante como método de pesquisa está na Antropologia e consiste em partilhar, tão intimamente quanto possível, a vida cotidiana do grupo estudado, durante um período suficientemente longo para cobrir o conjunto da vida e perceber progressivamente os seus elementos, e suas estruturas e significações". Observem os autores citados que, "se a identificação do antropólogo com o grupo estudado talvez seja prejudicial ao estudo científico dos fenômenos, a observação objetiva não basta, por si, para alcançar os objetivos procurados".

Situada, de início, no quadro de referência antropológica, a observação participante logo passou a integrar as outras ciências sociais, particularmente a Sociologia. A literatura acumulada a este respeito é extensa e inclui contribuições de Griaule, Vruyn, Williams, McCall e Simons, Filstead e Habenstein, Lofland, Bogden, Delfenhal e, particularmente, Renzin, que reúne vários artigos ilustrativos do método de observação participante. Um sumário de diversas pesquisas nas quais a observação participante foi usada encontra-se em Kidder (1987).

O princípio norteador da observação participante, segundo Sellitz (1987) é o de que os ambientes exercem grande influência sobre as pessoas e, conseqüentemente, a pesquisa deve começar em situações reais e não em circunstâncias arranjadas ou simuladas. Para os observadores participantes, a maneira pela qual as pessoas vêem e explicam seus ambientes é o núcleo da pesquisa, mas esta deve ocorrer no próprio ambiente. Sellitz cita Bogden, ao dizer que a observação participante tem sido definida como um período de intensa interação social entre pesquisadores e sujeitos, no ambiente desses últimos, durante o qual se coletam dados, na forma de notas de campo, de uma maneira sistemática e não reativa.

O autor também afirma que em contraste com as abordagens que utilizam sistemas de categorias, a observação participante envolve procedimentos menos formalizados, embora haja alguns passos em comum. Esses incluem: (1) selecionar um ambiente de pesquisa, (2) obter acesso e desempenhar um papel, (3) tomar notas e (4) formular uma análise. Na observação participante, inicia-se com algumas hipóteses, porém elas não são consideradas como algo definitivo, sendo revisadas à medida que transcorre a pesquisa, através da análise de casos negativos. Essa substitui a análise estatística e permite que o pesquisador procure dados que reflitam a hipótese. Assim, quando um único caso negativo é encontrado, o observador participante revê a hipótese de forma que esta possa abarcar o caso. A análise de casos negativos, com sua revisão e reteste contínuo de hipóteses, é semelhante ao que os experimentalistas fazem quando planejam estudos subsequentes. O observador participante faz, na realidade, uma série de estudos consecutivos, mas os realiza todos na mesma pesquisa, pois não há regra que diga que ele não pode mudar as hipóteses ou questões no decorrer da pesquisa. Além das hipóteses e questões, o observador participante revisa, muitas vezes, o seu tema de estudo.

No Brasil, a pesquisa-ação e observação participante têm encontrado numerosos adeptos e propugnadores. Em seu livro *"Metodologia da pesquisa-ação"*, Thiollent (1985) diz que o principal objetivo deste tipo de pesquisa é oferecer ao pesquisador melhores condições e compreensão das situações analisadas. Resume, também, as principais características da pesquisa-ação:

- (a) ampla e explícita interação entre pesquisadores e pessoas da situação estudada;
- (b) ordem de prioridade feita em conjunto: pesquisador e participantes da pesquisa;
- (c) constante acompanhamento da ação;
- (d) busca maior de esclarecimento dos problemas, do que de sua resolução;
- (e) busca do aumento de conhecimento dos pesquisadores e do nível de consciência das pessoas e dos grupos.

Thiollent ainda cita Zuniga, mostrando que a pesquisa-ação é inovadora do ponto de vista científico, somente quando é inovadora do ponto de vista sócio-político, isto é, quando tenta colocar o controle do saber nas mãos dos grupos que expressam uma aprendizagem coletiva, tanto na sua tomada de consciência como no seu comprometimento com uma ação grupal.

Com relação ao contraste entre a pesquisa-ação e a pesquisa convencional, o autor citado estabelece as comparações apresentadas no quadro 7.

Thiollent indica alguns cuidados que devem ser tomados quanto à pesquisa-ação. Lembra que deve ser preservado um equilíbrio entre conhecimento e participação; além disso, diz ele, é preciso evitar, de um lado, o tecnocratismo e o academicismo e, de outro, o populismo ingênuo e é preciso, ainda, estar atento para o risco de manipulação dos objetivos da pesquisa.

No que respeita à observação participante, há no Brasil notícia de algumas experiências em termos mais amplos de pesquisa participante. Demo (1984) dedicou-se ao estudo deste tipo de pesquisa e refere-se a vários tópicos importantes como os seguintes:

(a) Na pesquisa participante, o problema se origina na comunidade ou no próprio local de trabalho. A finalidade visada, na pesquisa, é a melhoria de vida dos envolvidos. É fundamentalmente uma experiência educativa de caráter coletivo.

(b) A metodologia da pesquisa participante indica três fases principais: (1ª) exploração geral do problema pela população envolvida, (2ª) identificação das necessidades básicas da população e (3ª) elaboração de uma estratégia educativa para a resolução do problema.

(c) A tônica básica do ponto de vista metodológico é a união entre o conhecimento e a ação.

(d) Os pesquisadores participam não apenas da discussão dos resultados da pesquisa, mas sobretudo do processo desta.

QUADRO 7

DIFERENÇAS ENTRE PESQUISA-AÇÃO E PESQUISA CONVENCIONAL

(Fonte: Thiollent, 1985)

Pesquisa-ação	Pesquisa-convencional
– Participação ativa do pesquisador junto às pessoas buscando um trabalho conjunto.	– Não há participação total do pesquisador junto às pessoas da situação observada.
– A ação é imediata.	– Há uma grande distância entre os resultados de uma pesquisa e as possíveis decisões ou ações decorrentes.
– O usuário participa como elemento fundamental de informação e execução.	– O usuário é, ao nível da pesquisa, um meio informante ao nível da ação, mero executor.
– São considerados tanto os aspectos individuais quanto os aspectos globais, que podem ser captados por questionários, mas devem ser enriquecidos com informações trazidas pelo pesquisador.	– São privilegiados os aspectos individuais, tais como opiniões, atitudes, motivações, comportamentos. Esses aspectos, geralmente captados por meio de questionários, não permitem uma visão dinâmica da situação.
– A dinâmica dos problemas, conflitos e ações é um dos principais focos da pesquisa.	– A dinâmica de transformação de uma situação em outra desejada não é o enfoque principal.
– Os pesquisadores afastam-se das preocupações metodológicas relacionadas à formalização, à lógica, às regras estatísticas.	– Os pesquisadores valorizam as estruturas de raciocínio e os critérios estatísticos.
– A pesquisa é realizada numa situação natural, na qual o pesquisador intervém conscientemente.	– A pesquisa é realizada numa situação criada, na qual o pesquisador não intervém.

(e) A escolha dos problemas a serem estudados não se efetua a partir de um conjunto de hipóteses previamente estabelecidas pelos pesquisadores, mas tem sua origem nas situações sociais concretas que os elementos que participam do processo de pesquisa querem estudar e resolver.

Pelo que se pode observar, inúmeras são as abordagens em termos da avaliação de programas e da aprendizagem. Cabe ao avaliador a difícil tarefa de conhecer possibilidades de cada uma delas, verificando sua adequação para a realidade na qual não trabalha. Um fator porém parece ser imprescindível de ser considerado: o fator político.

A essa altura, convém lembrar Weiss (1982), que explicitou bem a necessidade de considerar esse fator na avaliação e especificamente na pesquisa de avaliação. Ela afirma que as considerações políticas nos programas de formação profissional se insinuam de três formas principais e o avaliador que falhar no reconhecimento disso, arrisca-se a uma série de choques e conseqüentes frustrações. A primeira diz que os programas educacionais

tratados pela avaliação são oriundos de decisões políticas. Eles foram propostos, debatidos, aprovados e financiados através de processos políticos e, durante sua implementação, permanecem sujeitos a pressões, tanto de apoio quanto de hostilidade, provenientes do jogo político. A segunda diz que a avaliação é realizada com a finalidade de alimentar a tomada de decisão. Seus relatórios inserem-se no contexto político. Assim, a evidência avaliativa acerca dos resultados de um programa tem que disputar atenção com outros fatores que pesam no processo político. A terceira diz que a avaliação tem um posicionamento político. Por sua própria natureza, ela faz afirmações políticas implícitas sobre temas tais como a natureza problemática de alguns programas, a legitimidade dos objetivos e até mesmo sobre o papel que convém ao cientista social na formação de políticas e programas.

Cabe, portanto, ao avaliador, a difícil tarefa de imprimir ao seu trabalho princípios científicos que produzam resultados válidos e fidedignos e, principalmente, assumir uma postura clara e definida com relação ao seu papel no programa educacional do qual participa.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANDERSON, S.B. et al. (1981). *Encyclopedia of educational evaluation*. San Francisco, Jossey-Bass.
- BELLACK, A.A. & KLIEBARD, H.M. ed. (1977) *Curriculum Evaluation*. Berkeley, McCutchan.
- BLOOM, B.S. (1963) Learning for mastering. *Evaluation Comment*, 1, 2.
- _____ (1972). *Taxionomia de objetivos educacionais; 1, Domínio cognitivo*. Porto Alegre, Globo.
- _____ (1977). *Taxionomia de objetivos educacionais; 2, Domínio afetivo*. Porto Alegre, Globo.
- BONBOIR, A. (1976) *Como avaliar os alunos*. Lisboa, Seara Nova.
- COOK, S.W. et al (1976). *Research methods in social relations*. New York, Holt, Rinehart & Winston.
- DE LANDSHERE, G. (1976). *Avaliação e exames; noções de docimologia*. Coimbra, Almedina.
- DEMO, P. (1984). *Pesquisa participante; mito e realidade*. Rio de Janeiro, SENAC/DN.
- DEPRESBITERIS, L. (1976). *Um estudo de generalizabilidade dos instrumentos do projeto SACI/EXERN*. São José dos Campos, INPE (tese de mestrado).
- EBEL, R.L. & DAMRIN, D.E. (1960). Tests and examinations. In: HARRIS, C.W. ed. *Encyclopedia of educational research*. 3ª ed. New York, Macmillan, pp. 1502-1514.
- ENGLISH, H.B. (1958). *A comprehensive dictionary of psychological and psychoanalytical terms*. New York, Longman.
- FLEMING, R.S., org. (1970). *Curriculo*. Rio - Brasília, Lidador.
- GOLDBERG, M.A.A. (1973). Avaliação e planejamento educacional; problemas conceituais e metodologias. *Cadernos de Pesquisa*, (7): 63-72.
- HAMILTON, D. & PARLETT, M. (1982). Avaliação iluminativa uma nova abordagem no estudo de programas inovadores. In: GOLDBERG, M.A.A. & SOUZA, C.P. orgs. *Avaliação de programas educacionais; vicissitudes, controversias e desafios*. São Paulo, EPU, pp. 38-45.
- KIDDER, L.H. org. (1987). *Métodos de pesquisa nas relações sociais*. São Paulo, EPU.
- LIPPITT, R. (1984). Action research. In: CORSINI, R.J., ed. *Encyclopedia of psychology*. New York, Wiley, vol. 1, pp. 11-13.
- MAGER, R.F. (1977). *Medindo os objetivos educacionais*. Porto Alegre, Globo.

- MEDEIROS, E.B. (1977). *Provas objetivas, técnicas de construção*. Rio de Janeiro, Fundação Getúlio Vargas.
- MIRANDA, M.J. (1982). A docimologia em perspectiva. *Revista da Faculdade de Educação*. (USP), 8, (1): 39-69.
- MITCHELL, G.D. (1977). Action research. In: MITCHELL, G.D. ed. *A dictionary of Sociology*. London, Routledge e Kegan Paul.
- PIERÓN, H. (1969) *Examens et docimologie*. Paris, Presses Universitaires de France.
- POPHAM, W.J. (1983). *Avaliação educacional*. Porto Alegre, Globo.
- SAUL, A.M. (1985). *Avaliação emancipatória: uma proposta democrática para reformulação de um curso de pós-graduação*. São Paulo, PUC. (Tese de doutorado).
- SCHUBERT, W.H. (1982). Curriculum research. In: MITZEL, H.E. ed., *Encyclopedia of educational research*. New York, Macmillan, pp. 420-431.
- SCRIVEN, M. (1978). Perspectivas e procedimentos de avaliação. In: BASTOS, L. & PAIXÃO, L., orgs. *Avaliação Educacional II*. Petrópolis, Vozes.
- SELLTIZ, C. et al. (1987). Introdução. In: Kidder, L.H., org. *Métodos de Pesquisa nas Relações Sociais*. São Paulo, EPU. Vol. 1.
- _____ et al (1987). Observação participante. In: Kidder, L.H., org., *Métodos de Pesquisa nas Relações Sociais*. São Paulo, EPU. Vol. 1.
- STAKE, R.E. (1977). The countenance of educational evaluation. In: BELLACK, A.A. & KLEBARD, H.M. ed. *Curriculum and evaluation*. Berkeley, Mc-Cutchan.
- STUFFLEBEAM, D. et al. (1974). *Educational evaluation and decision making*. Itaska, Peacock.
- TALMAGE, H. (1982). Evaluation of programs. In: MITZEL, H.E. ed., *Encyclopedia of educational research*. New York, Macmillan, pp. 592-611.
- THINES, G. & LEMPEREUR, A. (1984). Observação participante. In: *Dicionário geral das ciências humanas*. Lisboa, Setenta, p. 653.
- THIOLLENT, M. (1985). *Metodologia da pesquisa-ação*. São Paulo, Cortez.
- TYLER, R.W. (1949). *Basic principles of curriculum and instruction*. Chicago, University of Chicago Press.
- _____ (1982). Avaliando experiências de aprendizagem. In: GOLBERG, M.A.A. & SOUZA, C.P. orgs. *Avaliação de programas educacionais; vicissitudes, controvérsias e desafios*. São Paulo, EPU, pp. 35-37.
- VIANNA, H.M. (1982a.) Avaliação educacional; algumas idéias precursoras. *Educação e Seleção*, (6): 61-70.
- _____ (1982b.) *Testes em educação*. São Paulo, Ibrasa.
- WEISS, C. (1982). Pesquisa avaliativa no contexto político. In: GOLDBERG, M.A.A. & SOUZA, C.P., orgs. *Avaliação de programas educacionais; vicissitudes, controvérsias e desafios*. São Paulo, EPU, pp. 23-27.
- WERTHEIN, J. et al (1985). Avaliação educacional: tendências na direção de enfoques mais participativos. In: WERTHEIN, J. & ARGUMEDO, M., orgs. *Educação e Participação*. Rio de Janeiro, Philobiblion.
- WORTHEN, B.R. (1982). Visão geral do mosaico formado pela avaliação e controles educacionais. In: GOLDBERG, M.A.A. & SOUZA, C.P., orgs. *Avaliação de programas educacionais; vicissitudes, controvérsias e desafios*. São Paulo, EPU, pp. 4-9.
- WORTHEN, B.R. & SANDERS, J.R. (1973). *Educational evaluation; theory and practice*. Worthington, Charles A. Jones.

