AVALIAÇÃO EDUCACIONAL: origens, contradições e desdobramentos no cenário brasileiro

REGILSON MACIEL BORGES [1]
https://orcid.org/0000-0001-6115-364X
regilson.borges@ufla.br

RESUMO
Este artigo apresenta revisão das origens históricas do campo da avaliação educacional e seus desdobramentos no cenário educacional. Identificam-se os principais períodos que marcaram a trajetória internacional da avaliação, destacando suas ênfases predominantes e estabelecendo aproximações com o percurso brasileiro. A justificativa para o estudo apoia-se na afirmação de Stufflebeam e Shinkfield (1987) de que nenhuma introdução aos estudos em avaliação está completa caso não se preste devida atenção à evolução histórica do tema. Trata-se de uma revisão de literatura, conforme Marconi e Lakatos (2003), cujo propósito é construir uma síntese abrangente sobre o tema, organizada de modo lógico e fundamentada nos trabalhos de autores nacionais e internacionais, como Worthen, Sanders e Fitzpatrick (2004), Freitas (2007), Guba e Lincoln (2011), Vianna (1995), entre outros. O percurso histórico abrange desde a ênfase em testes e medidas educacionais até a centralidade da avaliação externa em larga escala e a institucionalização de sistemas nacionais e internacionais, como o SAEB, o ENEM, o TIMSS e o PISA. No caso brasileiro, discute-se como a avaliação foi incorporada à agenda pública, em diálogo com reformas educacionais e sob forte influência de orientações neoliberais. A análise evidencia que a avaliação educacional deixou de ser entendida apenas como medição de desempenho para configurar-se como campo profissional, atravessado por disputas conceituais e políticas, e que hoje cumpre papel estratégico nas políticas educacionais.

Palavras-chave: Avaliação Educacional. Trajetória Histórica. Ênfases da Avaliação.

EDUCATIONAL EVALUATION: ORIGINS, CONTRADICTIONS AND DEVELOPMENTS IN THE BRAZILIAN CONTEXT

ABSTRACT
This article presents a review of the historical origins of the field of educational assessment and its developments in the educational landscape. It identifies the main periods that marked the international trajectory of assessment, highlighting their predominant emphases and establishing connections with the Brazilian trajectory. The justification for the study is based on Stufflebeam and Shinkfield's (1987) assertion that no introduction to assessment studies is complete without due attention to the historical evolution of the topic. This is a literature review, according to Marconi and Lakatos (2003), whose purpose is to construct a comprehensive synthesis on the topic, logically organized and based on the works of national and international authors, such as Worthen, Sanders, and Fitzpatrick (2004), Freitas (2007), Guba and Lincoln (2011), and Vianna (1995), among others. The historical trajectory ranges from the emphasis on educational tests and measures to the centrality of large-scale external

1 Doutor em Educação pela Universidade Federal de São Carlos (UFSCar) com pós-doutorado pela Universidade Estadual de Ponta Grossa (UEPG). Professor do Departamento de Gestão Educacional, Teoria e Práticas de Ensino (DPE) e do Programa de Pós-Graduação em Educação (PPGE) da Universidade Federal de Lavras (UFLA).

assessment and the institutionalization of national and international systems, such as SAEB, ENEM, TIMSS, and PISA. In the Brazilian case, the article discusses how assessment was incorporated into the public agenda, in dialogue with educational reforms and under the strong influence of neoliberal orientations. The analysis highlights that educational assessment has ceased to be understood solely as a measure of performance and has become a professional field, permeated by conceptual and political disputes, and which today plays a strategic role in educational policies.

Keywords: Educational Evaluation. Historical Trajectory. Evaluation Emphases.

EVALUACIÓN EDUCATIVA: ORÍGENES, CONTRADICCIONES Y DESARROLLOS EN EL ESCENARIO BRASILEÑO

RESUMEN
Este artículo presenta una revisión de los orígenes históricos del campo de la evaluación educativa y su evolución en el panorama educativo. Identifica los principales períodos que marcaron la trayectoria internacional de la evaluación, destacando sus énfasis predominantes y estableciendo conexiones con la trayectoria brasileña. La justificación del estudio se basa en la afirmación de Stufflebeam y Shinkfield (1987) de que ninguna introducción a los estudios de evaluación está completa sin la debida atención a la evolución histórica del tema. Se trata de una revisión bibliográfica, según Marconi y Lakatos (2003), cuyo propósito es construir una síntesis exhaustiva sobre el tema, organizada lógicamente y basada en las obras de autores nacionales e internacionales, como Worthen, Sanders y Fitzpatrick (2004), Freitas (2007), Guba y Lincoln (2011) y Vianna (1995), entre otros. La trayectoria histórica abarca desde el énfasis en las pruebas y medidas educativas hasta la centralidad de la evaluación externa a gran escala y la institucionalización de sistemas nacionales e internacionales, como SAEB, ENEM, TIMSS y PISA. En el caso brasileño, el artículo analiza cómo la evaluación se incorporó a la agenda pública, en diálogo con las reformas educativas y bajo la fuerte influencia de las orientaciones neoliberales. El análisis destaca que la evaluación educativa ha dejado de entenderse únicamente como una medida de rendimiento para convertirse en un campo profesional, permeado por disputas conceptuales y políticas, y que hoy desempeña un papel estratégico en las políticas educativas.

Palabras clave: Evaluación Educativa. Trayectoria Histórica. Énfasis de la Evaluación.

1. INTRODUÇÃO

O artigo apresenta revisão das origens históricas do campo da avaliação educacional e seus desdobramentos no cenário educacional brasileiro. Para tanto, partimos da identificação dos principais períodos que marcaram a trajetória internacional do campo, com destaque para suas principais ênfases e ao mesmo tempo fazemos um paralelo com o percurso da avaliação em nosso país.

A justificativa para um trabalho dessa natureza se baseia na afirmação de Stufflebeam e Shinkfield (1987) de que nenhuma introdução aos estudos em avaliação está completa

caso não se preste devida atenção à evolução histórica do tema. Sendo assim, apresentamos neste artigo alguns acontecimentos mais significativos ocorridos no campo da avaliação em diferentes momentos, enfatizando os acontecimentos históricos que se sucederam, bem como a compreensão conceitual de avaliação presente em determinado período.

A distribuição histórica adotada neste artigo está fundamentada na organização proposta por Stufflebeam e Shinkfield (1987), somada às contribuições de outros estudiosos, tais como Ana Maria Saul (2010), Blaine R. Worthen, James R. Sanders e Jody L. Fitzpatrick (2004), Dirce Nei Teixeira de Freitas (2007), Egon G. Guba e Yvonna S. Lincoln (2011), Heraldo Marelim Vianna (1995), entre outros. Assim, do ponto de vista metodológico, a revisão de literatura apresentada compreende uma síntese "referente ao trabalho e aos dados pertinentes ao tema, dentro de uma sequência lógica" (Marconi; Lakatos, 2003, p. 248).

A revisão de literatura adota uma perspectiva histórico-analítica, com a finalidade de compreender a constituição do campo da avaliação educacional a partir de seus principais marcos teóricos e epistemológicos. Nessa direção, alinhou-se ao entendimento de Mattar e Ramos (2021, p. 37), para os quais a revisão tem a intenção de "[...] produzir novos conhecimentos ao tornar explícitas conexões e tensões entre estudos". A seleção do material bibliográfico, não buscou exaurir toda produção existente, mas concentrou-se em fontes representativas, capazes de evidenciar mudanças de orientação, rupturas e permanências do tema pesquisado, em consonância com o caráter seletivo das revisões dessa natureza.

Stufflebeam e Shinkfield (1987, p. 33) organizam a avaliação educacional em cinco grandes períodos: o primeiro deles é conhecido como o "período pré-Tyler", que vai até 1930; o segundo é definido como a "época tyleriana", que vai de 1930 a 1945; o terceiro é denominado de "época da 'inocência", compreendido de 1946 até 1957; o quarto é denominado de "época do realismo", que cobre o período de 1958 a 1978; e o quinto é chamado de a "época do profissionalismo", que inclui a evolução experimentada pela avaliação a partir de 1978.

Além desta periodização proposta por Stufflebeam e Shinkfield (1987), outros autores também demarcam períodos para a avaliação educacional, caso de Daniel L. Stufflebeam, George F. Madaus e Thomas Kellaghan (2000, p. 3) que distribuem a avaliação em sete "idades": a primeira é o período anterior a 1900, chamado de "idade da reforma"; a segunda, de 1900 até 1930, a "idade de eficiência e testes"; a terceira, 1930-1945, é chamada de "idade

tyleriana"; a quarta, de 1946 até 1957, chamam de idade da "inocência"; a quinta, de 1958 a 1972, é a "idade do desenvolvimento"; a sexta, 1973-1983, é a "idade de profissionalização"; e, a sétima 1983 a 2001, chamada de "idade de expansão e integração" .

Guba e Lincoln (2011), por sua vez, apresentam "quatro gerações" para descrever os significados atribuídos à avaliação. A primeira geração é denominada pelos autores de "mensuração", a segunda é chamada de "descrição", a terceira de "juízo de valor" e a quarta geração é a de "negociação". Cada geração apresenta uma acepção de avaliação: na primeira, avaliação e medida são sinônimos; na segunda, a avaliação é vista como descrição dos pontos fortes e fracos; na terceira, a avaliação se caracteriza pela formulação de juízo de valor; e na quarta, a principal dinâmica da avaliação é a negociação.

Cabe destacar que nas propostas de periodizações supracitadas, Ralph W. Tyler aparece como marco para pensar um salto nas discussões sobre a avaliação, isso porque Tyler ampliou as discussões a respeito das medidas educacionais para a questão dos testes educacionais padronizados, antes restritas aos exames e quocientes de inteligência. Tamanha a sua contribuição para a área da avaliação, o faz ser reconhecido na literatura especializada como o "pai da avaliação educacional" (Stufflebeam; Shinkfield, 1987; Fernandes, 2009; Saul, 2010).

Segundo Stufflebeam e Shinkfield (1987, p. 91), Ralph W. Tyler foi o primeiro a desenvolver um "método sistemático de avaliação educacional", resultante de trabalho realizado durante a década de 1930 e 1940 na Universidade Estadual de Ohio. Para Tyler (1974, p. 99) "o processo de avaliação consiste em determinar em que medida os objetivos educacionais estão sendo realmente alcançados pelo programa do currículo e do ensino", trata-se de um "processo mediante o qual se determina o grau em que essas mudanças de comportamento estão realmente ocorrendo".

A avaliação educacional se volta, inicialmente, ao rendimento escolar dos estudantes, quando da aplicação e difusão dos testes padronizados, sendo confundida com medição de desempenho (Vianna, 1995). Com o tempo passa a preocupar-se com os currículos, programas, enfatizando a consecução dos objetivos educacionais (Tyler, 1974). Mais tarde evolui com a incorporação do julgamento de valores, no qual toda informação é levada em conta na emissão de juízos e não somente de especialistas em avaliação (Stake, 1982). Inova com a introdução da dimensão política nos estudos avaliativos (Macdonald, 1982). Assim, de algo muito restrito, o campo da avaliação ampliou-se e hoje constitui-se como "uma indústria tentacular e solidamente profissionalizada" (Dias Sobrinho, 2003, p. 43).

O artigo está organizado em oito seções, além desta introdução e das considerações finais. A segunda seção aborda a ênfase nos testes e medidas educacionais (1840-1930). A terceira seção trata da avaliação com ênfase nos objetivos educacionais (1930-1945). A quarta seção discute a institucionalização da avaliação (1946-1957). A quinta seção expõe a ênfase na avaliação de programas educacionais (1958-1978). A sexta seção apresenta a profissionalização do campo da avaliação (1978-1989). A sétima seção considera a ênfase nos exames internacionais e nacionais (1990-2001).

2. ÊNFASE NOS TESTES E MEDIDAS EDUCACIONAIS (1840-1930)

Este período foi marcado pela ênfase na aplicação de testes em exames escolares que visavam avaliar o desempenho individual dos alunos. Essa prática de avaliação de desempenho individual já era evidente a 2000 a.C., quando, segundo Stufflebeam e Shinkfield (1987), algumas autoridades chinesas realizavam concursos públicos para medir a proficiência dos serviços públicos prestados, e no século V a.C., quando Sócrates e outros mestres gregos se utilizavam de questionários avaliativos verbais como parte de sua metodologia didática.

No entanto, Stufflebeam e Shinkfield (1987) observam que a aplicação de avaliações formais era quase inexistente até meados do século XIX, quando a Inglaterra passou a fazer uso de comissões reais para avaliar os seus serviços públicos. Worthen, Sanders e Fitzpatrick (2004) relatam que a insatisfação com os programas educacionais e sociais na Grã-Bretanha gerou movimentos de reforma baseados nos depoimentos e outros métodos para avaliar as instituições.

Nos Estados Unidos, a partir de 1845, o Comitê Escolar de Boston, liderado por Horace Mann, realizou uma avaliação da educação de Massachusetts baseada no rendimento dos alunos, para averiguar se as escolas estavam ensinando bem os seus alunos. Esta aplicação ficou conhecida como a primeira utilização de testes impressos para a avaliação de desempenho dos alunos. Entre 1887 e 1898, Joseph Rice organizou um programa de avaliação fundamentado nos conhecimentos em ortografia dos alunos de boa parte do território norte-americano (Stufflebeam; Shinkfield, 1987; Worthen; Sanders; Fitzpatrick, 2004).

No final do século XIX também surgiram movimentos para a acreditação de instituições (universidades e escolas de ensino médio) e programas educacionais nos Estados Unidos (Stufflebeam; Shinkfield, 1987). De acordo com Worthen, Sanders e

Fitzpatrick (2004), a acreditação é um processo a serviço da revisão de programas a partir de certos critérios de qualidade com objetivo de habilitar ou credenciar o programa. Os autores citam como exemplo de sistema de acreditação nos Estados Unidos da América, a avaliação de 155 escolas de Medicina dos EUA e do Canadá que foi desenvolvida por Abraham Flexner, em 1908. Desde então, Stufflebeam e Shinkfield (1987) salientam que o processo de acreditação se tornou um dos principais meios para se avaliar os serviços educacionais.

No início de 1900, Edward Lee Thorndike convenceu alguns educadores a investirem na tecnologia da mensuração para determinar as habilidades humanas. Guba e Lincoln (2011) mencionam algumas das aplicações de testes do período, seja para o uso de decisões educacionais e psicológicas, entre as quais se encontram os testes destinados a medir a inteligência, desenvolvidos por Alfred Binet e Théodore Simon, em 1905, na França, e mais tarde aplicados a crianças norte-americanas por Lewis Terman, em 1916; e os testes para propósitos escolares, como os realizados pela Associação Americana de Educação, em 1904, para investigar a classificação e aproveitamento escolar das crianças.

Assim, a década de 1920 presenciou o surgimento dos testes como os principais meios de avaliar as escolas por meio dos níveis individuais de desempenho (Worthen; Sanders; Fitzpatrick, 2004). De acordo com Stufflebeam e Shinkfield (1987), todo esse movimento a favor dos testes estandardizados apontava, por um lado, para esforços que visavam tornar a educação mais eficiente, e por outro, para o desenvolvimento de todo um aparato tecnológico para “a medida de capacidades humanas e a avaliação, em consequência, passa a ter o significado de medida (testing) [...]” (VIANNA, 1995, p. 9).

A ênfase nos testes e medidas educacionais no cenário educacional brasileiro é ressaltada por Sousa (2005), ao analisar estudos sobre avaliação divulgados nos anos 1960, constatou que o significado atribuído à expressão avaliação educacional era de forma dominante de seleção e medida educacional. Igualmente Gatti (1987, p. 34) menciona que “o uso de testes educacionais de modo mais generalizado se deu no Brasil a partir de meados da década de 60” e “seu emprego em nosso meio está muito associado a exames vestibulares”. A autora destaca que foi essa cultura de medição, testes, provas, que se desenvolveu em torno dos processos avaliativos, permanece ainda hoje quando as pessoas se reportam de imediato a esse tipo de avaliação ao se referirem à avaliação educacional (Gatti, 2002).

3. ÊNFASE NOS OBJETIVOS EDUCACIONAIS (1930-1945)

Foi durante a década de 1930 que ocorreu nos Estados Unidos, e em grande parte do mundo, o que ficou conhecido como a Grande Depressão, momento em que as escolas, e demais instituições públicas, presenciaram uma ausência de recursos e também de otimismo (Stufflebeam; Shinkfield, 1987).

Como tentativa de reverter essa situação e salvar a economia norte-americana, foram implementadas diversas medidas adotadas pelo governo do presidente Franklin Delano Roosevelt, dentre as quais estão a criação de órgãos federais "para supervisionar novos programas nacionais de previdência, assistência social, administração de mão-de-obra, desenvolvimento urbano, saúde, educação e inúmeras outras áreas de prestação de serviços [...]" (Worthen; Sanders; Fitzpatrick, 2004, p. 64).

É neste período que, segundo Stufflebeam e Shinkfield (1987), John Dewey e outros educadores contribuíram para se pensar uma educação dinâmica e inovadora, num movimento que ficou conhecido como Educação Progressiva, cuja filosofia refletia o pragmatismo com forte utilização de instrumentos da psicologia behaviorista. Neste mesmo movimento aparece a figura de Ralph W. Tyler, pesquisador e avaliador norte-americano, que liderou um estudo que ficou conhecido como "Estudo de Oito Anos" (*Eight Year Study*), cujo modelo de avaliação educacional apresentava uma metodologia de ligação entre as medidas dos resultados e os resultados desejados (Worthen; Sanders; Fitzpatrick, 2004).

Participaram do estudo liderado por Tyler, 30 escolas secundárias públicas e particulares dos Estados Unidos. O programa teve como propósito demonstrar que os alunos capacitados de acordo com currículos mais responsivos, portanto, não convencionais, seriam capazes de obter bom desempenho na faculdade. A escolha dos oito anos como recorte do estudo possibilitou um acompanhamento durante quatro anos dos alunos concluintes do ensino secundário e quatro anos do ensino superior (Guba; Lincoln, 2011).

A intenção central de Tyler era "aperfeiçoar os currículos que estavam sendo desenvolvidos e confirmar se eles estavam funcionando", para isso era necessário coletar informações a respeito do aproveitamento dos alunos em relação aos objetivos definidos, num processo que analisaria os pontos fortes e fracos dimensionados na coleta, processo esse que "repetia-se sucessivamente a cada curso oferecido, até que se concluísse que o currículo estava produzindo um nível apropriado de aproveitamento" (Guba; Lincoln, 2011, p. 35).

A principal característica do método tyleriano está centrada nos objetivos claramente

fixados, de onde se define, inclusive, a sua noção de avaliação "como algo que determina se certos objetivos foram alcançados", assim, os objetivos referidos ao rendimento proporcionariam a base para elaboração de currículo e de testes de desenvolvimento, diante disso os "[...] avaliadores deviam ajudar aos elaboradores dos currículos a clarificar o rendimento dos alunos, que haviam sido precisamente determinados através da realização de um currículo " (Stufflebeam; Shinkfield, 1987, p. 34, tradução nossa).

A influência e o alcance de Ralph W. Tyler neste período lhe fizeram ficar conhecido como o pai da avaliação educacional (Guba; Lincoln, 2011), justamente por ter cunhado o termo avaliação educacional, expressão essa que, para o autor, designaria o processo pelo qual se determina em que medida os objetivos educacionais estão sendo realmente alcançados pelo programa do currículo e do ensino. No meio acadêmico brasileiro, Saul (2010) destaca que os argumentos de Tyler foram amplamente recebidos, para se ter uma ideia, sua obra "Princípios Básicos do Currículo e Ensino" chegou a ter nove edições no período de 1974 a 1984. Outro fator que, segundo a autora, permitiu a difusão e influência do pensamento de Tyler foi à tradução de trabalhos de seus seguidores Hilda Taba, Robert F. Mager, James Popham e Eva Baker (Saul, 2010).

Freitas (2007) assinala que no Brasil, dos anos de 1930 a 1940, teve início o interesse do Estado pela avaliação, marcada pela institucionalização da pesquisa e pela necessidade do planejamento educacional. Segundo a autora "os motivos para 'medir, avaliar e informar' foram enunciados em termos de necessidade e importância de o Estado conferir e verificar resultados ante os objetivos da educação nacional" (Freitas, 2007, p. 52). Sousa (1995, p. 44) ressalta que "a ênfase dada à avaliação da aprendizagem enquanto mensuração de capacidades e características do homem, por meio de testagem", o que, segundo a autora "reflete o movimento ocorrido nas duas primeiras décadas deste século nos Estados Unidos".

No mesmo sentido, Sousa (1998, p. 162) aponta a preocupação com a importância da avaliação do desempenho escolar na década de 1930, destacando o autor Isaias Alves, que "defendia os testes pedagógicos, argumentando que sua objetividade seria mais adequada que as avaliações subjetivas até então realizadas".

4. ÊNFASE NA INSTITUCIONALIZAÇÃO DA AVALIAÇÃO (1946-1957)

Este período foi designado por Stufflebeam e Shinkfield (1987) como a época da "inocência", por se tratar de um período de expansão e plenitude da sociedade norte-americana, mas envolto a graves problemas sociais, como a pobreza das cidades do interior

e áreas rurais, preconceitos raciais e segregação, excesso de consumo e desperdício de recursos naturais, entre outros.

Na área educacional foi um período de expansão das ofertas, de pessoal e das instalações. Stufflebeam e Shinkfield (1987) relatam que foram construídos novos edifícios escolares, aumentaram as inscrições de professores em programas educacionais, assim como houve aumento no número de matrículas escolares. Foram ainda incluídos novos serviços "[...] para a saúde mental e física, conselhos escolares, cantinas, ensino de música, amplos programas esportivos, educação comercial e técnica e educação social" (Stufflebeam; Shinkfield, 1987, p. 36, tradução nossa).

Em 1949 foi publicada a primeira edição da famosa obra de Ralph W. Tyler, "Princípios Básicos de Currículo e Ensino", na qual o autor apresenta um modo de encarar um programa de ensino como instrumento eficiente de educação. O quarto capítulo da obra trata especificamente da avaliação. Neste capítulo o autor institui a avaliação como "uma operação importante no desenvolvimento do currículo" (Tyler, 1974, p. 97), discute ainda a necessidade da avaliação, suas noções fundamentais, de procedimentos da avaliação, das fases da construção de um instrumento de avaliação e do uso dos resultados da avaliação.

A partir dos anos 1950, Worthen, Sanders e Fitzpatrick (2004, p. 65) mencionam os esforços na tentativa de melhorar a abordagem tyleriana de avaliação, em que se procurava ensinar "os educadores a apresentar objetivos em termos explícitos, mensuráveis [...]". Dentre os esforços empreendidos, os autores destacam os manuais de taxonomia de objetivos educacionais para os domínios cognitivo e afetivo que foram formulados por Benjamin S. Bloom e seus colaboradores.

Em 1957 é lançado pela União Soviética o primeiro satélite artificial, conhecido como Programa Sputnik I. Isso provocou uma rápida reação do governo dos Estados Unidos que buscou averiguar "as supostas deficiências da educação americana que haviam permitido que os russos ganhassem vantagem na exploração do espaço" (Guba; Lincoln, 2011, p. 36). Ampliaram-se rapidamente os investimentos para um ensino mais eficiente, principalmente nas áreas de Matemática e Ciências (Worthen; Sanders; Fitzpatrick, 2004). Tão logo foram criadas equipes para avaliar se a resposta do governo a iniciativa soviética surtia efeito.

No entanto, alguns autores salientam as dificuldades encontradas pelos avaliadores indicados para compor as equipes de avaliação, dentre as quais se encontram a falta de definição de objetivos dos projetos criados e, "[...] além disso, eles não poderiam aceitar uma estratégia de avaliação que não produzisse resultados enquanto o programa não estivesse

totalmente desenvolvido" (Guba; Lincoln, 2011, p. 36). Havia também considerável "empobrecimento conceitual e metodológicos da avaliação [...]", "concepções inadequadas e relatórios irrelevantes", ocasionando projetos insuficientes, dados inválidos, análises imprecisas das avaliações realizadas (Worthen; Sanders; Fitzpatrick, 2004, p. 66).

No Brasil, Sousa (1995) e Freitas (2007) destacam que a ênfase das pesquisas educacionais, no período de 1940 a 1950, se concentravam em discussões na área da psicologia, predominando "temas relativos à avaliação da aprendizagem e do desenvolvimento psicológico" (Freitas, 2007, p. 9). Somente a partir da década de 1950 é que, segundo Sousa (1995, p. 44) "as pesquisas deslocam-se da ênfase psicológica para o estudo da relação entre o sistema escolar e a sociedade".

5. ÊNFASE NA AVALIAÇÃO DE PROGRAMAS EDUCACIONAIS (1958-1978)

No ano de 1958 foi promulgada, pelos Estados Unidos, a Lei Educacional de Defesa Nacional, que entre outras questões, criou programas educacionais e serviços de consultoria e programas de pesquisa em distritos escolares, seguidos por projetos para o desenvolvimento de currículos nas áreas de ciências e matemáticas (Stufflebeam; Shinkfield, 1987).

Como consequência disso, buscava-se avaliar a efetividade dos novos currículos, se "[...] estavam adequadamente financiados e aplicavam cuidadosamente a tecnologia que havia sido desenvolvida durante a década passada ou mais recentemente" (Stufflebeam; Shinkfield, 1987, p. 37, tradução nossa). Contudo, alguns avaliadores como Lee J. Cronbach notaram que os resultados dos estudos avaliativos não contribuíam para a melhoria dos currículos e programas.

Nesse mesmo período, em 1963, Cronbach escreveu um artigo considerado por alguns autores como "seminal" (Worthen; Sanders; Fitzpatrick, 2004) e "histórico" (Stufflebeam; Shinkfield, 1987), no qual critica as avaliações do passado e exige novas diretrizes para se pensar as avaliações futuras. No artigo o autor recomenda que os avaliadores, ao reconceituar a avaliação, não a entendessem como competição, "[...] mas como um meio consistente de coleta e organização de informações que pudessem ajudar aos que elaboram os currículos" (Stufflebeam; Shinkfield, 1987, p. 38, tradução nossa). Stufflebeam e Shinkfield (1987) e Worthen, Sanders e Fitzpatrick (2004) mencionam que as recomendações de Cronbach não tiveram um impacto imediato, pois ficaram reduzidas a pequenos círculos de especialistas em avaliação.

Em 1965, por ocasião dos governos dos presidentes norte-americanos John Kennedy e Lyndon Johnson, houve o desencadeamento da guerra contra a pobreza, com programas de investimentos de bilhões de dólares para as áreas de educação, saúde, moradia, justiça criminal, desemprego etc. (Worthen; Sanders; Fitzpatrick, 2004). Essas reformas pretendiam igualar e aumentar as oportunidades para todos os cidadãos dos Estados Unidos, e só foram possíveis graças à expansão da economia e ao apoio de toda a nação para desenvolver o que o presidente Lyndon Johnson chamou de a Grande Sociedade *(Great Society)* (Stufflebeam; Shinkfield, 1987). Logo surgiram tentativas para avaliar os êxitos desses programas, que se tornaram mais explícitas na Lei do Ensino Fundamental e Médio *(Elementary and Secondary Education Act)*.

Essa Lei incluía requisitos específicos para a avaliação de programas, conforme descrevem Worthen, Sanders e Fitzpatrick (2004). Esse ponto específico foi capitaneado pelo então senador Robert F. Kennedy, que insistia que todos os beneficiários dos programas "preenchessem um questionário de avaliação que mostrasse os resultados dos gastos dos fundos federais. Essa exigência de avaliação por parte do Congresso acabou sendo aprovada para o Título I (ensino compensatório) e para o Título III (programas educacionais inovadores)" (Worthen; Sanders; Fitzpatrick, 2004, p. 69). Era exigido, por exemplo, que cada distrito escolar, que tivesse recebido financiamento, avaliasse anualmente até que ponto os projetos haviam alcançado seus objetivos, utilizando-se para isso de testes estandardizados (Stufflebeam; Shinkfield, 1987).

As dificuldades em decorrência dessas exigências foram enormes, entre as quais se encontravam a inadequação dos instrumentos e estratégias utilizadas na condução das avaliações, visto que os testes estandardizados não respondiam a questões como "[...] as necessidades e níveis de êxito das crianças mais pobres, o que limita os professores na hora de fixar objetivos de comportamento que são significativos para este grupo de alunos" (Stufflebeam; Shinkfield, 1987, p. 39, tradução nossa). Sem contar a pouca condição dos educadores para executar a tarefa de avaliar seu trabalho, assim "professores, administradores e psicólogos escolares estiveram entre os mais sobrecarregados na sua condição de avaliadores [...]" (Worthen; Sanders; Fitzpatrick, 2011, p. 69). Como consequência disso o que se viu foram avaliações inadequadas e com pouca utilidade.

No Brasil, Freitas (2007) indica que, entre 1956 e 1964, as pesquisas deslocaram seu foco dos indivíduos para a própria educação escolar, a qual passou a ser examinada em termos de sua "funcionalidade" social. Nesse período, desenvolveram-se levantamentos,

surveys e inquéritos que permitiram mapear a gestão das redes de ensino, avaliar as condições de funcionamento das escolas, registrar dados quantitativos sobre o rendimento do ensino e compreender as relações entre as instituições escolares e o contexto local. Após 1964, especialmente ao longo dos anos 1970, as pesquisas passaram a privilegiar um enfoque econômico, ganhando destaque os seguintes temas "educação como investimento, os custos educacionais e a relação entre educação escolar e demanda de profissionais" (Freitas, 2007, p.15).

No campo da avaliação de currículo, Saul (2010, p. 41) destaca que, a partir dos anos 1970, esse enfoque ganhou relevância com o aparecimento do projeto de currículo, sendo o modelo de "tomada de decisão", proposto por Stufflebeam, o mais difundido. A autora observa ainda que o primeiro documento oficial do Ministério da Educação (MEC) divulgado com esse teor foi elaborado por João Batista Araújo e Oliveira e Mariza Roda e Oliveira, publicado em 1973, sob o título "A função da avaliação na tomada de decisões educacionais".

6. ÊNFASE NA PROFISSIONALIZAÇÃO DO CAMPO DA AVALIAÇÃO (1978-1989)

A década de 1970 foi particularmente promissora para a avaliação educacional, especialmente para a avaliação de programas, por se tratar de um período em que se organizaram novas associações profissionais para os avaliadores.

Worthen, Sanders e Fitzpatrick (2004, p. 85) mencionam algumas dessas associações profissionais de avaliadores criadas para acolher os avaliadores. Uma das primeiras, segundo os autores, foi a Divisão H da Associação de Pesquisa Americana em Educação, "dedicada a dar um lar aos avaliadores escolares". Em 1976 são fundadas duas outras entidades profissionais, a Rede de Avaliação *(Evaluation Network)* e a Sociedade de Pesquisa em Avaliação *(Evaluation Research Society)*, a primeira reunia avaliadores educacionais e a segunda concentrava cientistas sociais e avaliadores ligados a órgãos do governo. Em 1985 essas duas associações vão se fundir e formar a Associação Americana de Avaliação (*American Evaluation Association)* (Worthen; Sanders; Fitzpatrick, 2004).

Na segunda metade dos anos 1970 e durante toda a década de 1980 ocorreu um considerável desenvolvimento de conteúdo próprio e exclusivo da avaliação, tal como observam Worthen, Sanders e Fitzpatrick (2004, p. 77): "Livros sobre interpretação das avaliações foram publicados. Artigos sobre avaliação começaram a aparecer com frequência cada vez maior em revistas especializadas de diversas áreas". Dentre os periódicos publicados neste período se encontram: *Evaluation and Program Planning; Evaluation*

Practice; Evaluation Review; Evaluation Quarterly; Educational Evaluation and Policy Analysis; Studies in Educational Evaluation; Canadian Journal of Program Evaluation; New Directions for Program Evaluation; Evaluation and the Health Professions; Itea Journal of The and Evaluation, Performance Improvement Quarterly; e Evaluation Studies Review Annual (Worthen; Sanders; Fitzpatrick, 2004).

Entre os livros publicados neste período, Stufflebeam e Shinkfield (1987), Worthen, Sanders e Fitzpatrick (2004) e Guba e Lincoln (2011) destacam as obras de: Blaine R. Worthen e James S. Sanders, *"Educational Evaluation: theory and practice"*, de 1973, e *"Educational evaluation: alternative approaches and practical guidelines"*, de 1987; Elmer L. Struening e Marcia Guttentag, *"Handbook of evaluation research"*, de 1975; James Popham, *"Educational evaluation"*, de 1975; Gene V. Glass, *"Evaluation studies review annual"*, de 1976; Thomas D. Cook e Charles S. Reichardt, *"Qualitative and quantitative methods in evaluation research"*, de 1979; Ernest House, *"Evaluating with validity"*, de 1980; Michael Scriven, *"The logic of evaluation"*, de 1980; Egon G. Guba e Yvonna S. Lincoln, *"Effective evaluation"*, 1981, e *"Fourth generation evaluation"*, de 1989.

Neste período muitas universidades também começaram a oferecer ao menos um curso de metodologia avaliativa, dentre as quais, Stufflebeam e Shinkfield (1987, p. 41) destacam a Universidade de Illinois, Universidade de Stanford, o Boston College, a Universidade da Califórnia, a Universidade de Minnesota e a Universidade de Western Michigan, que, de acordo com os autores, chegaram a "elaborar programas de pós-graduação em avaliação". Em algumas dessas instituições foram fundados centros para a investigação e desenvolvimento da avaliação, entre os quais se encontram a unidade de avaliação da *Northwest Regional Educational Laboratory*, o *Center for the Study of Evaluation da UCLA, or Stanford Evaluation Consortium*, o *Center for Instructional Research and Curriculum Evaluation de la Universidad de Illinois*, o *Evaluation Center da Universidade de Western Michigan e o Center for the Study of Testing and Educational Policy del Boston College* (Stufflebeam; Shinkfield, 1987).

No contexto brasileiro, Sousa (1995) assinala que, ao final da década de 1970, surgem na área educacional estudos que passaram a examinar as relações entre educação e Estado capitalista, enfatizando o papel da escola no processo de reprodução social. Nessa mesma direção, Sousa (1998) ressalta a produção de pesquisas de autores brasileiros que evidenciaram como os estudantes oriundos de camadas populares, marcados pela discriminação socioeconômica, "recebiam das escolas um ensino de pior qualidade e ainda

eram aqueles que respondiam pela maior taxa de evasão e repetência de todo o sistema de ensino" (Sousa, 1998, p. 164).

Saul (2010, p. 40) aponta que, a partir de 1978, surgiram esforços no sentido de "quebrar" como o circuito que até então caracterizava a avaliação no Brasil. Nesse movimento, destacam-se a publicação do artigo de Marli André "A abordagem etnográfica: uma perspectiva na avaliação educacional (1978)", a tradução do texto de Parlett e Hamilton "Avaliação como iluminação: uma nova abordagem no estudo de programas inovadores (1980)", a publicação da revista "Educação e Avaliação (1980,1981)", e a obra organizada por Goldberg e Sousa "Avaliação de programas educacionais: vicissitudes, controvérsias e desafios (1982)". Esses trabalhos ampliaram as perspectivas da avaliação educacional ao criticarem a abordagem quantitativa, proporem alternativas qualitativas e incorporarem a dimensão política ao campo.

Entre 1985 e 1990, Freitas (2007) observa que a avaliação em larga escala e a informação educacional passaram a figurar em pesquisas conduzidas em diferentes áreas do conhecimento. Segundo a autora, esse conjunto de estudos avaliativos, "surgidos ao longo dos anos 1980, revela o crescente interesse pelos temas qualidade do ensino, desempenho escolar, avaliação externa, avaliação de sistemas e informação educacional" (Freitas, 2007, p. 35). Borges e Rothen (2020) ressaltam que nos anos 1980 foram sistematizadas as primeiras experiências de avaliação sistêmica e/ou em larga escala para a educação no país, tanto no nível básico quanto no superior, cuja implementação ocorreu nos anos 1990.

7. ÊNFASE NOS EXAMES INTERNACIONAIS E NACIONAIS (1990-2001)

A partir dos anos 1990, a avaliação educacional ganha espaço com a centralidade da avaliação externa, conforme destaca Fernandes (2009, p. 111), quando aponta que essas avaliações tem a expectativa "de obter informação confiável para melhorar a qualidade do ensino e das escolas". Nessa direção, o autor menciona que a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), "têm gasto, nos últimos anos, milhões e milhões de dólares com a realização de avaliações externas, nomeadamente exames nacionais e todo tipo de provas" (Fernandes, 2009, p. 111-112), caso do Estudo Internacional de Tendências em Matemática e Ciências *(Trends in International Mathematics and Science Study – TIMSS)* e do Programa Internacional de Avaliação de Estudantes *(Programme for International Student Assessment – PISA).*

O TIMSS vem sendo aplicado desde 1995, sob a coordenação da Associação Internacional para a Avaliação do Desempenho Educacional (IEA). A avaliação tem como foco o desempenho dos estudantes do 4º e 8º ano do ensino fundamental nas áreas de Matemática e Ciências. As informações oficiais do site do TIMSS indicam que os dados têm permitido que países em todo o mundo tomem decisões baseadas em evidências para aprimorar as políticas educacionais relacionadas ao ensino e à aprendizagem de matemática e ciências (TIMSS, 2025). Fernandes (2009, p. 141) aponta que as questões propostas pelo TIMSS se preocupam em examinar os currículos, as formas para colocá-los em prática, os contextos de implementação e as aprendizagens pelos alunos ao final de determinado período de escolarização.

De acordo com informações oficiais do Ministério da Educação (MEC) brasileiro, aproximadamente 70 países utilizam os dados de tendências do TIMSS para monitorar a eficácia de seus sistemas educacionais em uma perspectiva comparada. O Brasil passou a integrar a avaliação em junho de 2022, tendo o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) como órgão responsável pelo planejamento e a operacionalização no País. O estudo teve sua aplicação em 2023, por meio de testes digitais realizados em computador, acompanhando a difusão do uso da tecnologia nas escolas e na sociedade (MEC, 2025).

O PISA é promovido pela OCDE desde o ano 2000 e é aplicado a cada três anos. Segundo Fernandes (2009, p. 144), o programa “pretende averiguar em que medida os jovens de quinze anos de cada país estão preparados para enfrentar os desafios das sociedades dos nossos dias”. O programa testa as habilidades e o conhecimento de estudantes em matemática, leitura e ciências, sendo que a cada edição é avaliado um domínio principal. No ano 2000, a ênfase foi na Leitura, “o que significa que dois terços dos testes incidiram nessa área e um terço nas áreas de Matemática e Ciências” (Fernandes, 2009, p. 145). No ano de 2003 a ênfase foi em Matemática e no ano de 2006 se concentrou em Ciências. O Brasil participa do PISA, como país convidado, desde 2000, sendo o primeiro país sul-americano a aderir ao programa.

Nos anos 1990, a avaliação assumiu um papel central na formulação e implementação das políticas educacionais no Brasil. Freitas (2004, p. 149) aponta que “o conceito de avaliação que emergiu privilegiou a avaliação externa em todos os níveis de ensino, sob a lógica da competitividade entre as organizações ou entre os professores”, neste cenário “a qualidade era produto da própria competição e não uma construção coletiva,

a partir de indicadores legitimados socialmente pelos atores". No período foram criados o Sistema de Avaliação da Educação Básica (SAEB), em 1990; o Programa da Avaliação Institucional das Universidades Brasileiras (PAIUB), em 1993; o Exame Nacional dos Cursos (ENC), em 1995; e o Exame Nacional do Ensino Médio (ENEM), em 1998 (Calderón; Borges, 2013).

Além dessas avaliações de âmbito nacional, alguns estados brasileiros criaram seus próprios sistemas de avaliação da educação básica no ano de 1992, Dalben et al. (2011) destacam os casos do Ceará, que criou o Sistema Permanente de Avaliação da Educação Básica do Ceara (SPAECE) e em Minas Gerais, foi criado o Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais, que posteriormente se tornou o sistema atual: o Sistema Mineiro de Avaliação da Educação Pública (SIMAVE). O estado de São Paulo também implantou em 1992, a Avaliação da Jornada Única pela Secretaria de Educação, que mais tarde "culminou com a implantação do Sistema de Avaliação do Rendimento Escolar (SARESP) em 1996".

Freitas (2007) enfatiza que, durante a década de 1990, além da ampliação quantitativa dos sistemas de avaliação, ocorreu também sua crescente sofisticação e a vinculação com programas mais amplos de qualificação da educação e modernização da gestão educacional. Segundo a autora, "esses programas geralmente permitiram aos estados conjugarem 'avaliação-medida-informação' educacional na gestão de seus sistemas de educação e ensino" (Freitas, 2007, p. 156). Nessa mesma linha, Calderón e Borges (2013) destacam estudos desse período que examinam os instrumentos de avaliação, suas intenções e resultados, ressaltando a forte influência do neoliberalismo na formulação das políticas de avaliação no Brasil.

Quanto aos modelos das avaliações externas em larga escala no Brasil, Barretto et al. (2001) destacam que as produções buscam apresentar os propósitos desses modelos em delinear o perfil cognitivo da população com base em informações censitárias, estudos sobre o problema da repetência nos sistemas escolares, esclarecimentos sobre a Teoria da Resposta ao Item (TRI) e a questão da *accountability*. A respeito da avaliação de sistemas, as autoras mencionam que o centro das atenções é a discussão sobre a qualidade da educação, "o enfoque, no entanto recai fundamentalmente sobre a possibilidade de estabelecer um sistema de medição para aferi-la" (Barretto et al., 2001, p. 65).

Calderón e Borges (2013) afirmam que na década de 1990 a avaliação entra na agenda pública, com destaque para o monitoramento a partir da avaliação padronizada do

rendimento escolar. Nesse sentido, os estudos publicados naquele período enfatizavam aspectos relacionados ao desempenho escolar dos alunos, às formas de avaliação de rendimento, aos índices de repetência e as influências que as práticas avaliativas exerciam na relação que se estabelece entre professores e estudantes (Calderón; Borges, 2013).

8. CONSIDERAÇÕES FINAIS

A periodização histórica aqui apresentada possibilita compreender as sucessivas transformações pelas quais passou a avaliação educacional nas últimas décadas. Suas ênfases, abordagens e conceituações encontram-se inseridas a contextos históricos específicos, que orientam práticas avaliativas em diferentes momentos e também influenciam os rumos teóricos do campo.

As contribuições analisadas evidenciam o desenvolvimento da avaliação enquanto campo profissional claramente definido, como assinalam Stufflebeam e Shinkfield (1987), expresso em resultados como: o aprimoramento da comunicação entre pesquisadores e instituições; a formação de profissionais especializados; a cooperação entre organizações dedicadas à avaliação educacional; as divisões conceituais entre correntes metodológicas; e o incremento do diálogo entre defensores de abordagens quantitativas/positivistas e qualitativas/fenomenológicas.

No caso brasileiro, observou-se que a avaliação foi gradualmente incorporada à agenda pública, sobretudo a partir dos anos 1990, como instrumento estratégico de regulação e gestão educacional, em consonância com orientações de cunho neoliberal. Tal processo trouxe ganhos em termos de institucionalização e sistematização, mas também contradições, como a tensão entre concepções tecnocráticas e perspectivas críticas e emancipadoras.

Assim, a avaliação educacional deixou de ser compreendida apenas como medição de desempenho, configurando-se como um campo em constante disputa conceitual e política, com desdobramentos significativos para a formulação, implementação e monitoramento das políticas educacionais contemporâneas.

9. REFERÊNCIAS

BARRETTO, Elba Siqueira de Sá; PINTO, Regina Pahim; MARTINS, Angela Maria; DURAN, Marília Claret Geraes. Avaliação na educação básica nos anos 90 segundo os periódicos acadêmicos. **Cadernos de Pesquisa**, São Paulo, n. 114, p. 49-88, nov. 2001. Disponível em: https://www.scielo.br/j/cp/a/DPqL5s4DdsbDJQTsXNPt8PC/?lang=pt. Acesso em: 26 out. 2025.

BORGES, Regilson Maciel; ROTHEN, José Carlos. Configuração do campo da avaliação educacional na década de 1980 no Brasil. **Revista Diálogo Educacional,** Curitiba, v. 20, p. 318-351, 2020. Disponível em:
https://periodicos.pucpr.br/dialogoeducacional/article/view/25416. Acesso em: 26 out. 2025.

CALDERÓN, Adolfo Ignacio; BORGES, Regilson Maciel. Avaliação educacional no Brasil: da transferência cultural à avaliação emancipatória. **Brazilian Geographical Journal,** Ituiutaba, v. 4, p. 259-274, 2013. Disponível em:
https://seer.ufu.br/index.php/braziliangeojournal/article/view/23189. Acesso em: 25 out. 2025.

DALBEN, Ângela Imaculada Loureiro de Freitas; BROOKE, Nigel; CUNHA, Maria Amália de A.; FALEIROS, Matheus. **A avaliação externa como instrumento da gestão educacional nos estados.** São Paulo: Fundação Victor Civita, 2011.

DIAS SOBRINHO, José. **Avaliação:** Políticas Educacionais e Reformas da Educação Superior. São Paulo: Cortez, 2003.

FERNANDES, Domingos. **Avaliar para aprender:** fundamentos, práticas e políticas. São Paulo: Editora UNESP, 2009.

FREITAS, Dirce Nei Teixeira de. **Avaliação da educação básica no Brasil:** dimensão normativa, pedagógica e educativa. Campinas, SP: Autores Associados, 2007.

FREITAS, Luiz Carlos de. A avaliação e as reformas dos anos de 1990: novas formas de exclusão, velhas formas de subordinação. **Educação & Sociedade**, Campinas, v. 25, n. 86, p. 133-170, abr. 2004. Disponível em:
https://www.scielo.br/j/es/a/Qqvf3NWCXfkmxrgNVYKmCZm/abstract/?lang=pt. Acesso em: 25 out. 2025.

GATTI, Bernardete A. Testes e Avaliações do Ensino no Brasil. **Educação e Seleção**, São Paulo, n.16, p. 12-21, 1987. Disponível em:
https://publicacoes.fcc.org.br/edusel/article/view/2623. Acesso em: 25 out. 2025.

GATTI, Bernardete A. Avaliação educacional no Brasil: pontuando uma história de ações. Eccos - **Revista Científica**, São Paulo, v. 4, n.1, p. 17-41, jun. 2002. Disponível em: https://periodicos.uninove.br/eccos/article/view/291. Acesso em: 25 out. 2025.

GUBA, Egon G.; LINCOLN, Yvonna S. **Avaliação de quarta geração**. Trad. Beth Honorato. Campinas: Editora da Unicamp, 2011.

MACDONALD, Barry. Uma classificação política dos estudos avaliativos. In: GOLDBERG, Maria Amélia Azevedo; SOUSA, Clarilza Prado de (orgs.). **Avaliação de programas educacionais:** vicissitudes, controvérsias, desafios. São Paulo: EPU, 1982. p. 16-17.

MATTAR, João; RAMOS, Daniela Karine. **Metodologia da pesquisa em educação:** abordagens qualitativas, quantitativas e mistas. São Paulo: Edições 70, 2021.

MINISTÉRIO DA EDUCAÇÃO. **Avaliação e Exames Educacionais. Ações Internacionais. TIMSS.** Disponível em: https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/timms. Acesso em: 26 out. 2025.

SAUL, Ana Maria. **Avaliação Emancipatória**: desafio à teoria e à prática de avaliação e reformulação de currículo. 8. ed. São Paulo: Cortez, 2010.

SOUSA, Sandra Zakia Lian. Avaliação da Aprendizagem nas pesquisas do Brasil de 1930 a 1980. **Cadernos de Pesquisa**, São Paulo, n. 94, p. 43-49, 1995. Disponível em: https://publicacoes.fcc.org.br/cp/article/view/837. Acesso em: 24 out. 2025.

SOUSA, Clarilza Prado de. Descrição de uma Trajetória na/da Avaliação Educacional. **Idéias**, São Paulo, n. 30, p. 161-174, 1998. Disponível em: http://www.crmariocovas.sp.gov.br/ext_a.php?t=004. Acesso em: 24 out. 2025.

STAKE, Robert. Conversando sobre avaliação. In: GOLDBERG, Maria Amélia Azevedo; SOUSA, Clarilza Prado (orgs.). **Avaliação de programas educacionais:** vicissitudes, controvérsias, desafios. São Paulo: EPU, 1982, p. 57-62.

STUFFLEBEAM, Daniel L.; SHINKFIELD, Anthony L. **Evaluación sistemática:** guía teórica y práctica. Trad. Carlos Losilla. Barcelona: Paidós, 1987.

STUFFLEBEAM, Daniel L.; MADAUS, George F.; KELLAGHAN, Thomas (Ed.). **Evaluation models:** viewpoints on educational and human services evaluation. Boston/Dordrecht/London: Kluwer Academic Publishers, 2000.

TYLER, Ralph. **Princípios Básicos do Currículo e Ensino.** Trad. Leonel Vallandro. Porto Alegre: Editora Globo, 1974.

TRENDS IN INTERNATIONAL MATHEMATICS AND SCIENCE STUDY. **TIMSS**. Disponível em: https://www.iea.nl/studies/iea/timss. Acesso em: 27 out. 2025.

VIANNA, Heraldo Marelim. Avaliação Educacional: uma perspectiva histórica. **Estudos em Avaliação Educacional**, São Paulo, n. 12, p. 7-24, 1995. Disponível em: https://publicacoes.fcc.org.br/eae/article/view/2297. Acesso em: 24 out. 2025.

WORTHEN, Blaine R. SANDERS, James S.; FITZPATRICK, Jody L. **Avaliação de Programas:** concepções e práticas. Trad. Dinah de Abreu Azevedo. São Paulo: Editora Gente, 2004.