

AVALIAÇÃO DE POLÍTICAS PÚBLICAS COMO PESQUISA SOCIAL: QUESTÕES CIENTÍFICAS, POLÍTICAS E IDEOLÓGICAS

L. M. de SOUZA

Universidade Federal do Rio Grande do Norte, CCHLA
linconlmoraes@ufrnet.br

Submetido 07/03/2018 - Aceito 02/11/2018

DOI: 10.15628/rbept.2018.7066

RESUMO

A avaliação de políticas públicas, como pesquisa social e atividade estatal, tem enfrentado obstáculos no plano científico e político e ideológico. Além do núcleo comum construído pelos seus autores clássicos, e as diferenças internas que marcam qualquer disciplina científica, defronta-se com dois usos pela ideologia neoliberal que a afetam negativamente. O primeiro, ou avaliação gerencialista, fala-se de avaliação para justificar a privatização do Estado. O segundo, são os testes atuais

visando perseguir o funcionalismo. No Brasil, além da educação, encontramos a adoção por uma espécie que tem proliferado na sua fauna e no legislativo em particular. Referimo-nos a alguns políticos profissionais de direita que resolveram abraçar ardorosamente a ideologia neoliberal e falar em nome da avaliação. Daí, o (mal) uso de expressões como avaliação de desempenho e outras do gênero.

PALAVRAS-CHAVE: Avaliação de políticas públicas, avaliação gerencialista, testes.

EVALUATION OF PUBLIC POLICIES AS SOCIAL RESEARCH: SCIENTIFIC, POLITICAL AND IDEOLOGICAL ISSUES

ABSTRACT

The evaluation of public policies, such as social research and state activity, has faced scientific, political and ideological obstacles. Besides the common nucleus constructed by its classic authors and the internal differences that mark any scientific discipline, it is faced with two uses of the neoliberal ideology that affect it negatively. The first, or managerial evaluation, talks of evaluation to justify the privatization of the State. The second is the current tests aimed at pursuing public

functionalism. In Brazil, besides education, we find adoption of a species that has proliferated in its fauna and in the legislature in particular, that is some right-wing professional politicians who have resolved to embrace neoliberal ideology ardently like form of evaluation. For this reason, the (bad) use of expressions such as performance evaluation and others of the genre in this academic work.

KEYWORDS: Evaluation of public policies, managerial evaluation, tests.

1 INTRODUÇÃO

Por mais dedicação e esforço que façam os predadores e as predadoras para exterminar com a fauna brasileira, ela sempre ressurgirá repaginada ou fantasiada de qualquer coisa, especialmente de cordeiro ou ovelha. É o que acontece, por exemplo, com uma espécie de políticos e políticas profissionais de direita no país, como a senadora Maria do Carmo Alves do DEM de Sergipe, séria candidata a representante de uma nova espécie no *patropi*. Quando se imaginava que todas as atividades surrealistas e pós-modernas no Brasil teriam sido esgotadas, eis que a eminente parlamentar apresenta no Senado o PL 116/2017, defendendo que os chamados servidores municipais, estaduais e nacionais passem pelo que chama de *avaliação de desempenho*.

O que a ilustre senadora intitula de avaliação, consiste em testes e provas que os funcionários teriam que se submeter semestralmente e passando, dentre outros, pelos crivos dos chamados rendimento, responsabilidade, qualidade e relacionamento profissional. E, caso recebam notas abaixo de 30% do máximo por quatro vezes seguidas, seriam exonerados. Também seriam postos para fora, quem tiver menos que 50% em cinco das denominadas avaliações.

A julgar pela nobre intenção da parlamentar, incluindo, segundo ela, sua preocupação em melhorar a qualidade do serviço público, os testes estariam na moda e entrando sem cerimônia na (mal) intitulada casa do povo. Além do mais, para quem pensava que esta espécie da fauna era infértil enganou-se redondamente, pois mais recentemente ainda (setembro de 2014) tivemos notícia de um rebento denominado por alguns de Enem dos professores. Seu nome de batismo oficial é Exame Nacional Avaliação do Magistério da Educação Básica ou simplesmente ENAMEB e atinge professores da educação básica das redes públicas e privadas. Seu genitor inicial é o ex-senador Wilson Matos do PSDB DO Paraná e teve o parecer favorável do deputado Valtenir Pereira do PMDB do Mato Grosso. E tramita na Câmara com o número de 6144/2009. E propõe no Art. 3, que o ENAMEB passe a aferir o *desempenho, competências* etc. dos professores. Para completar, a deputada Professora Dorinha Seabra, do DEM de Tocantins, considerou a atividade muito bondosa para com os docente e mudou para pior a redação. Como? Os professores terão que pagar esta suposta avaliação e os testes, ao invés de cinco anos, foram reduzidos para dois. E que os exames, pasmem, seriam ligados à qualificação e valorização dos professores.

Mas, além desses testes e sua presença como elemento da ideologia neoliberal aplicada à avaliação, a denominada avaliação gerencialista também tem ocupado um espaço muito grande como justificativa de privatização do Estado etc., uma vez que ambos os componentes são faces diferentes e interligadas desta ideologia.

De um lado, portanto, temos a ideologia neoliberal vigente na avaliação de políticas públicas e materializada no gerencialismo e na maioria dos testes atuais e suas aplicações, especialmente utilizados na educação. De outro, a avaliação de políticas públicas enquanto pesquisa social e a utilização do método científico. E vai ser partir desta última perspectiva, que

iremos trabalhar nos planos científico e político-ideológico tomando a educação como principal ilustração e elemento comparativo.

2 A AVALIAÇÃO DE POLÍTICAS PÚBLICAS COMO PESQUISA SOCIAL

Em contraste com a avaliação gerencialista e os atuais testes e as aplicações, a avaliação das políticas públicas foi produzida, já desde seu início, como pesquisa social ou algo do gênero. Não obstante algumas diferenças, os autores clássicos da teoria e teorias da avaliação, no caso principalmente Campbell, Cronbach, Scriven, Weiss, Rossi, Stake e Wholey, mesmo com variações ou com uso de termos diversos a consideravam dessa maneira. Isto foi mostrado claramente, dentre outros, por estudiosos como Shadish Júnior., Cook e Leviton (1991) e Alkin e Christie (1984) e, inclusive, deixado claro pelos autores nucleares da avaliação, como os já referidos.

Em outras palavras, a avaliação de políticas públicas, tal como foi construída pelos clássicos da literatura do tema, nunca foi reduzida a ferramenta de privatização do Estado e nem entendida como simples testes visando estabelecer escores e reproduzir hierarquias. Daí, se entender o motivo de textos clássicos, especialmente os Weiss (1978) e Rossi, Lipsey e Freeman (2004), ressaltarem e insistirem mais de uma vez na utilização de metodologia científica na avaliação, bem como alguns mais atuais como Chen (1990), que também expressaram a mesma concepção. Ou, assim compreendermos como um autor clássico, no caso Stake (1971) ainda no século XX, criticarem clara e detalhadamente esses mesmos testes aplicados à educação, como veremos depois.

Mesmo de forma bastante sintética, podemos mencionar alguns elementos importantes e demarcadores. Como pesquisa social, por exemplo, a avaliação de políticas públicas tem suas particularidades quando comparada às outras pesquisas em ciências sociais. De acordo com a concepção de Scriven (1967) e Weiss (1978) e outros, a avaliação, para se ter uma ideia, tem um significado mais valorativo que as outras pesquisas e é mais diretamente aplicada e aplicável. Tem uma espécie de núcleo comum envolvendo, de modo amplo, elementos como a efetividade, eficácia, eficiência e conteúdo da política, assim como os processos de implementação, os efeitos e impacto. Diferencia-se em alguns casos, das avaliações das chamadas políticas setoriais ou de áreas como a educação, saúde, assistência outras. Na educação, por exemplo, além da pesquisa, o monitoramento, a auditoria e os testes também, em muitos casos, são considerados avaliação.

E também como decorrência da sua característica de pesquisa social, a avaliação de políticas públicas distingue-se da análise de políticas públicas, monitoramento, auditoria e testes. No tocante às diferenças com a análise, esta é mais geral e engloba a agenda, formulação, implementação e a própria avaliação. Além do mais, é menos direcionada e valorativa que a avaliação. Comparando ao monitoramento, o mesmo diz respeito muito mais a uma atividade regular visando observar o que está ocorrendo em alguma atividade, tem um caráter mais restrito e, portanto, não chega a constituir uma pesquisa. A auditoria, por sua vez, tem suas atividades direcionadas para o cumprimento de normas, especialmente as contábeis, procura saber até que ponto estão sendo praticadas e envolve possíveis correções e sanções. Já os testes,

finalmente, são bem mais específicos e materializam uma situação criada com o fim de se observar os resultados.

Dessa forma, a avaliação de políticas públicas constitui um subcampo do campo análise de políticas públicas. E é nesta condição, que envolve várias disciplinas, especialmente a ciência política, a teoria das organizações, a sociologia e a economia. Ultimamente, através da abordagem qualitativa e de estudos de caso a antropologia começou também a marcar presença na avaliação.

Quanto aos tipos de avaliação de políticas públicas, os principais são a de impacto, de processo e avaliação política, como já chamaram a atenção Figueiredo e Figueiredo (1986) e procuramos desenvolver num livro anterior (Souza, 2014). Mais recentemente, tem crescido a avaliação institucional e a meta-avaliação ou avaliação da avaliação. Todas essas modalidades de avaliação, contudo, terminam, de um modo ou de outro, excluindo, no fundamental, a avaliação gerencialista e os testes como avaliações de políticas públicas. O que não acontece com uma parte da educação, pois a mesma tem usado diferentes conceitos de avaliação, como já analisamos (Souza, 2017).

3 ALGUMAS CLASSIFICAÇÕES DIFERENTES SOBRE A AVALIAÇÃO EM EDUCAÇÃO

Como já foi mencionado, a avaliação de políticas públicas, enquanto subcampo mantém relações com diferentes áreas, setores ou domínios. Ao mesmo tempo, esses domínios, como a própria educação, podem apresentar uma diversidade de visão e de tipos de avaliação. É o que explica, assim, que podemos encontrar o entendimento da avaliação como pesquisa, monitoramento, auditoria e teste em educação como se fossem também avaliação.

Essa diversidade de significados deriva, em grande parte, do fato da educação constituir, igualmente, um aparelho ideológico susceptível de grandes disputas de significados e de conteúdos científico, político e ideológico. A ideia dos testes atuais e suas aplicações como avaliação, por exemplo, expressam no momento, principalmente, a aplicação da ideologia neoliberal na própria avaliação.

Igualmente, fazem parte da diversidade existente na educação, as variadas classificações. Como ressaltam Alba, Barriga e Viesca (1984), a avaliação em educação teve seu início voltada para a aprendizagem dos alunos, depois foi direcionada para os planos de estudo e currículo e no final ficou centrada nas instituições e relações com a sociedade.

Para Alves e Acevedo (2002) e ligando à aplicação, poder-se-ia falar em avaliação institucional (engloba diversos componentes e processos), avaliação curricular (inclui principalmente os planos educativos), avaliação de programa (conjunto de atividades num contexto e ligado à consecução de objetivos), avaliação do ambiente físico (condições matérias vinculadas ao processo educacional), avaliação do ensino e avaliação de aprendizagem (envolvendo a aprendizagem, desempenho do docente e do programa pedagógico).

De acordo com Sousa (2005), por sua vez, existiriam a avaliação de rendimento escolar em diversos níveis de ensino, avaliação de currículo, avaliação de material didático, avaliação de instituições, avaliação de cursos, avaliação de professores e avaliação de sistemas.

Contrastando com as concepções referidas antes e sem nos alongarmos, podemos apontar para um tipo de visão que estamos criticando que é a de Castro (2005), para quem a avaliação educacional mais importante é ligada aos *testes*. De acordo com o autor, existem testes que avaliam os *sistemas educativos* através de provas aplicadas aos alunos visando conhecer o ensino como um todo e a escola, como o Saeb no Brasil, mas pode-se também medir o *aluno* procurando saber o nível de aprendizado ligado ao estado ou país, como o Pisa. Dessa forma, existiriam avaliações que buscam os objetivos de diagnosticar os *sistemas educativos* como o Provão e quando determinam trajetória dos *alunos* como o Vestibular e o Enem. Ainda refere-se, brevemente, à *avaliação institucional* do ensino superior centrada na descrição da instituição, insumos e processo de ensino. E aponta o que seria um outro tipo de avaliação institucional, no caso a *auto-avaliação*.

Vejam agora, como a chamada avaliação gerencialista no geral e depois os testes atuais e suas respectivas aplicações são utilizados e alguns de seus desdobramentos, novamente com a ênfase na educação.

4 A AVALIAÇÃO GERENCIALISTA COMO FERRAMENTA DE PRIVATIZAÇÃO GERAL DO ESTADO E DA EDUCAÇÃO

Ao olharmos a literatura clássica sobre avaliação de política públicas, além da atribuição de pesquisa à atividade, como falamos, não encontramos nenhuma defesa ardorosa do papel da avaliação como instrumento de privatização do Estado. E mesmo que um dos autores clássicos, no caso Whole (2010), estabeleça uma ligação direta entre a avaliação e sua importância na gestão governamental, não atribui à mesma uma ferramenta de privatização como atualmente se faz. A presença da avaliação como instrumento de privatização, direta ou indiretamente é um traço da ideologia neoliberal e atua como um dos elementos habituais a partir da década de 1970.

Como a chamou Derlien (2001), a avaliação gerencialista ou gerencialismo tem permeado, ressaltamos, várias atividades do Estado capitalista e, evidentemente, com mais intensidade na gestão dos governos de direita, seja os que assumem o neoliberalismo ou os que arranjam designações mais palatáveis como social-liberalismo, terceira via ou outras. Mesmo de modo sintético, podemos apresentar algumas de suas características centrais em seguida, além do que, como temos repetido, constitui uma das principais expressões teóricas da ideologia neoliberal (Souza, 2014). E, de uma maneira ou de outra, funciona também como base para tentar justificar as privatizações, a perseguição ao funcionalismo, o ranqueamento das instituições de educação, a aplicação dos testes etc., pois apresenta-se, embora mal vestida em alguns momentos, muitas vezes com trajes respeitosos e elegantes da ciência. E é neste sentido, portanto, que se materializa, igualmente, como instrumento de gestão, controle e mecanismo de legitimação como aponta Derlien (2001).

Como acentuou esse autor e outros, no que se convencionou chamar de funções da avaliação e que era produzir conhecimentos e fornecer informações, a avaliação gerencialista enfatiza a chamada reforma do Estado, a alocação de recursos no processo orçamentário e adota elementos de auditoria e controle de gestão, bem como busca mais a legitimação. De base na administração e gestão dos programas, o gerencialismo remete a atividade principalmente para o ministério da fazenda. E a ênfase na avaliação dos programas é redirecionada para a redução dos mesmos. Daí, entender-se que os programas de educação pública e estatal passem a ser esvaziados e os recursos sofram transformações visando reduzir seu papel.

No fundamental, podemos dizer que existem duas concepções básicas que fundamentam e nucleiam a avaliação gerencialista, ou seja, as chamadas teorias *Public Choice* ou Teoria da Escola Pública e a *New Public Management* ou Nova Administração Pública. No geral, a teoria mais vinculada à ideologia neoliberal e que respinga bastante nas políticas públicas é a *Teoria da Escolha Pública*, mas é a *Nova Administração Pública New* que se projeta mais direta e especificamente para a gestão e a avaliação.

Iniciemos, então, pela *Teoria da Escolha Pública*, até porque vai exercer uma forte influência sobre a segunda concepção e teve um respaldo de peso nos anos de 1970, bem como na década de 1980 no governo Reagan nos Estados Unidos, ou seja, uma base legal e institucional considerável. E no seu discurso, os elementos são bem expressivos e apontam com clareza seus traços aplicativos no geral e em particular na educação, como veremos depois a partir da análise de Ravitch (2011).

Como mostramos em outra ocasião (Souza, 2014), podemos apontar alguns elementos típicos a esse respeito. Os bens públicos, por exemplo, deveriam, semelhantes aos bens privados, ser sujeitos à *medição* dos seus custos, *rendimentos* dos serviços públicos etc., no sentido de introduzir os mecanismos do mercado capitalista. Em outras palavras, a ênfase na medição através dos testes e a defesa da inserção da escola no circuito do mercado capitalista adquiriram um peso enorme nos Estados Unidos e, diríamos, também na maioria dos países. O chamado *rendimento*, como assinalamos antes, parece ser uma das coisas relevantes para senadora brasileira ou, pelo menos, o que ela entende por isto em relação ao funcionalismo. E completa-se com o rebento intitulado Enameb como já falamos, cujos zelosos pais mais imediatos, no caso, Valtenir Pereira do PMDB do Mato Grosso e Professora Dorinha Seabra, do DEM do Tocantins, voltam sua caridosa atenção para os professores.

Nas atividades do Estado, o gerencialismo defende com ênfase o discurso da concorrência, pois esta, de acordo com a *Teoria*, levaria à *eficiência* e à melhoria dos serviços e na sua entrega. E os problemas decorreriam, principalmente, do uso ineficiente dos recursos uma vez que o Estado não estaria seguindo o caminho da empresa privada, no caso a portadora desta suposta eficiência, competitividade e eficácia. Está claro, evidentemente, que para a ideologia neoliberal a escola pública, assim como outras atividades governamentais, precisariam ter uma postura de empresa privada.

Mas prossigamos. A *Teoria* também postula no seu discurso, que deveria haver o que chama de livre escolha e as preferências individuais, isto é, as pessoas necessitariam ter a liberdade de escolher diversos bens e serviços no mercado capitalista e não dispor apenas das

atividades governamentais, como no caso das escolas. Para tanto, o cidadão teria que ser tratado como *cliente* para usufruir das benesses do mercado. A burocracia pública, por sua vez, teria uma postura egoísta e é considerada culpada pelos aumentos dos gastos e falência dos serviços públicos. E daí, a defesa ardorosa do discurso da privatização das atividades do Estado, bem como da descentralização das atividades, da contratação de empresas externas para reduzir o desperdício etc.

E a *Teoria*, como assinalamos, projeta-se claramente em outra vertente que é a intitulada *Nova Gestão Pública* ou *Nova Administração Pública* como também tem sido chamada no Brasil. E esta, como se sabe, é a base do chamado empreendedorismo e claramente expressa no conhecido livro de Osborne e Gaebler (1998), intitulado *Reinventando o governo*, onde encontramos o costumeiro discurso de governo eficiente, governo voltado para os consumidores etc. (Souza, 2014) e não surpreende que Ravitch (2016) chame a atenção para isto na educação.

Dessa forma, junto e articulada com a *Teoria*, a *Nova Administração Pública* tem exercido considerável influência nas chamadas reformas administrativas na OCDE e em outros países. Emergiu, não por acaso, no final dos anos de 1970 e anos de 1980 no governo neoliberal de Thatcher, em governos municipais dos Estados Unidos, na Nova Zelândia e Austrália.

Alguns traços centrais são bem presentes no seu discurso. De acordo com isto, o tamanho do governo deveria ser reduzido para poder ser *eficiente* e, assim, dever-se-ia, dentre outras coisas, efetuar cortes no orçamento público e adotar uma administração mais ligada aos *custos*. Também seria necessário a *privatização* dos serviços públicos, a prioridade atribuída ao *mercado*, a *contratação* de empresas privadas e a adoção de *gestão privada* nas organizações empresariais e a utilização das denominadas *parcerias*.

Além do mais, a *Nova Administração* não se furta de defender claramente que o governo deveria, igualmente, buscar o *lucro* e procurar a *competição* na área pública e não somente gastar. A atenção deveria ser direcionada para o *cliente* e *consumidor*, buscando-se o que a Nova Administração denomina de busca de *resultados/inputs* e não a procura de programas. A tudo isto, ainda seria importante agregar o discurso da prestação de contas ou *accountability*, a defesa do vínculo das agências com os *contratos* e da auditoria de *desempenho*. E, como uma espécie de presente antecipado ou premonição para a senadora e os genitores imediatos da Enabeb, é defendida uma espécie de auditoria que é (mal) denominada de avaliação de *desempenho* e *avaliação ligada à melhoria de qualidade*.

No caso da *Nova Administração* e os vínculos com a educação, alguns estudos, especialmente os de Afonso (1999, 2000, 2008, 2013), Pasqual (2012) e Lima (1997, 2009, 2011, 2012) têm analisado de modo claro o que está ocorrendo, principalmente em Portugal. Como apontamos anteriormente (Souza, 2014), além do estabelecimento dos testes atuais, a avaliação passou, dentre outras coisas, a funcionar também no controle das despesas públicas, na mudança da compreensão do setor público e no esmaecimento da fronteira entre o espaço público e privado. Daí as *políticas avaliativas neoliberais* como as intitulou Afonso (1999), assemelhadas diretamente com as chamadas eficiência, eficácia e custos no sentido estreito e economicista do neoliberalismo.

Nessa *educação contábil*, como a chama Lima (1996), a avaliação é apresentada como uma espécie de técnica inovadora da gestão e da qualidade mais ligadas, na verdade, a uma função ou papel de controle (Lima, 2011). O gerencialismo, como enfatiza o autor, é centrado na competitividade entre as escolas, criação de mercado, visão de *desempenho* etc. (Lima, 2000).

O que é óbvio e envolve também a educação, é que a mesma passa a ser encarada como uma simples mercadoria e sujeita às intempéries do mercado capitalista ou, como diz a ampla maioria dos noticiários da TV no Brasil, aos humores do mercado. A avaliação gerencialista, corporificada nas duas visões indicadas, funciona como alicerce fundamental para legitimar a privatização das escolas e de outras atividades públicas e estatais.

5 A MODA E OS PROBLEMAS DOS TESTES ATUAIS APLICADOS EM EDUCAÇÃO

Como analisarmos anteriormente (Souza, 2017), até meados da década de 1930 os testes eram considerados o principal instrumento de avaliação e esta era mais centrada na educação. Com um maior aprofundamento e sistematização da avaliação de políticas públicas a partir da década de 1970, observamos uma separação relativa dos testes e da avaliação de políticas públicas no sentido mais amplo. Vejamos, então, como a literatura de avaliação de políticas públicas entende os testes ou não os leva em conta.

5.1. Os testes e a literatura de avaliação de políticas públicas

Na ampla maioria dos trabalhos mais conhecidos dos autores clássicos, como os de Campbell e Stanley (1979), Scriven (1967), Cronbach (1963), Weiss (1979), Rossi, Lipsey e Freeman (2004), Stake (1967) e Wholey (2010), a regra é que, mesmo quando o foco está voltado para uma abordagem mais quantitativa e de modo predominante para a educação, o foco da avaliação é o currículo. Os testes atuais, tais como são conhecidos e aplicados, geralmente são criticados ou praticamente não aparecem e nem são considerados como, propriamente, avaliação. E direta ou indiretamente, como regra são vistos como insuficiente para uma avaliação ou mesmo inseridos na discussão dos desenhos experimentais e quase-experimentais muito mais no plano do procedimento estatístico.

Em outro texto e mais direcionado de Stake (1971), já naquela época o autor indicava questões importantes que deveriam ser levadas em conta, tais como: nem tudo pode ser medido via testes, como as emoções etc.; mesmo falando-se que os testes são objetivos, não se dar a devida importância aos erros, inclusive os que levam a decisões errada; os testes podem medir vocabulário e habilidade com destreza, mas não podem medir adequadamente a compreensão da habilidade para analisar os lados opostos de um argumento; medem coisas muito específicas, mas não medem princípios e nem conseguem sintetizar um argumento político.

E o mesmo Stake (1971) acentuava o que considerava os quatro principais erros. Quanto à *Escolha dos objetivos*, nem todos eles são mensuráveis facilmente e nem sempre são exatos. Os *Crítérios de procedimentos dos testes*, por sua vez, apresentam problemas de correlacionar os escores dos testes para um critério verdadeiro, pois entre o que o teste mede e o que é sempre

conhecido e o que o grupo que é suporte. Qualquer teste padronizado somente será válido para um ou poucos critérios variáveis. E como medir performances complexas com simples testes?

Os terceiro erro, diz respeito à *Análise dos escores de ganho* ao se observar somente o curto prazo, as questões sobre o calendário escolar (observa-se somente o final do ano etc.) e olha-se mais as habilidades e menos para o que uma pessoa possa fazer de tarefas, podendo ocorrer que os escores apresentem apenas a aparência do que não ocorre, como falsos ganhos e falsas perdas. O último erro remete para *O processo social*, onde deve-se contemplar a interação aluno-professor, dentre outras coisas.

Em alguns autores importantes após os clássicos, como Chen (1990), Stufflebeam (2001), House (2000), Guba e Lincoln (2011) e outros, os testes, como regra, são relativizados, mas nem por isto são considerados como elementos muito positivos de um modelo de avaliação de políticas públicas.

Stufflebeam (2001), por exemplo, ao apresentar seus vinte e dois modelos de avaliação, refere-se a um deles e que denomina de *Objective Testint Programs* ou programas de teste de objetivo e cita, inclusive, trabalhos já escritos na década de 1950. E lembra alguns programas que aplicaram testes nos Estados Unidos, como o *American College Testing Program* e o *National Merit Scholarship Testing Program*. E chega, inicialmente e de modo rápido, a referir-se ao que considera *vantagens* dos programas de teste padronizados: a principal, é que seria eficiente em produzir informação confiável e válida sobre a performance dos estudantes em muitas áreas do currículo da escola e que seriam uma estratégia familiar de cada nível do programa escolar nas escolas dos Estados Unidos.

Entretanto, o mesmo Stufflebeam (2001) alude também a *desvantagens* dos programas de teste padronizados e tece mais considerações: as principais *limitações*, é que trazem dados apenas sobre os resultados dos estudantes; reforçam o teste de comportamento de múltipla escolha, mais do que seus escritos e comportamento de fala; tendem a direcionar somente os objetivos de aprendizagem de pouca instrução; em muitos casos talvez sejam um indicador melhor dos níveis socioeconômicos dos estudantes de um dado programa.

E o autor, apoiando-se no mesmo texto já citado de Stake (1971), afirma que os testes teriam uma pobre aproximação do que os professores estariam ensinando. Frequentemente, os conteúdos dos testes não correspondem aos objetivos do programa. Cita estudo em que alguns dados dos testes foram perdidos e não se conseguiu determinar se alguns estudantes foram ou não excluídos do teste ou se ocorreram problemas com as taxas. Segundo o autor, foi possível avaliar o mérito dos programas das escolas porque existiam apenas dados da uma faceta da qualidade dos programas. Existiam poucos dados sobre os currículos, qualidade dos professores, materiais, envolvimento dos pais ou responsáveis etc.

No caso de House (2000), não obstante suas críticas aos enfoques dominantes de avaliação, chega a considerar, mesmo de modo crítico, a utilização dos testes como um dos enfoques da mesma. Para ele, existiriam oito enfoques de avaliação, ou seja, análise de sistema, objetivos comportamentais, de decisão, sem objetivos definidos, crítica de arte, revisão profissional e estudo de caso.

Dos oito enfoques considerados pelo autor, dois deles dizem respeito diretamente à utilização de testes. O primeiro ou *análises de sistemas*, trabalha-se com as pontuações de *testes em educação* envolvendo as diferenças nos programas e as variações nos indicadores. Um de seus antecedentes vem desde 1965, ocorreu com a análise de sistema no Departamento de Defesa dos Estados Unidos e a partir de 1965 com a ampliação do *Planning, Programming and Budgeting System* ou PPBS. Posteriormente, foi aplicado na educação, saúde e bem-estar. E já na década de 1970, foram encontrados escritos a respeito dos vínculos entre os testes e os processos decisórios.

O modelo *análise de sistemas*, segundo o autor, deriva da microeconomia e traz como suposto básico que os indivíduos e também as organizações apresentam comportamento que eleve ao máximo os resultados e os testes são vistos como a única medida de êxito. Mas outros aspectos estão em jogo, como a análise dos custos dos programas, a medição dos programas governamentais, o vínculo com questões do impacto e eficiência econômica etc. E adotaria, uma metodologia positivista, considerada não recomendável por ele.

O outro enfoque que guarda relações com os testes e também vínculos com a educação, é o *enfoque de objetivos comportamentais* e que tem suas raízes nas propostas de Tyler e a avaliação do currículo e dos objetivos. A conduta dos alunos deveria ser medida através da aplicação de *testes de rendimento* e está ligada à ideia também de competência ou teste de competência.

Contudo, do ponto de vista mais geral, House (2000) considera, criticamente, que os testes seguiriam uma *epistemologia objetivista liberal* e enfoques gerenciais. Seriam orientados para a gestão visando supervisionar os programas e, no seu entendimento, melhorá-los. Teriam como preocupação central, a responsabilidade, eficiência e controle de qualidade e nesta epistemologia as informações são consideradas cientificamente objetivas e conseguidas através de *testes* ou questionários e que os resultados seriam reproduzíveis.

5.2. As metamorfoses dos testes e algumas observações

Mas os testes e suas aplicações atuais, é importante destacar, vão materializar de modo mais intenso os elementos da ideologia neoliberal só mais recentemente. No início mesmo, na escola medieval, o exame vai surgir visando determinar o rendimento dos estudantes e, ao mesmo tempo, selecionar para alguma coisa como a educação superior ou outros aspectos da sociedade. Buscava-se o *domínio de um saber* e não como hoje, onde se visa definir o lugar do indivíduo na produção, bem como investigar a *idoneidade dos candidatos* e uma prova de suas realizações escolásticas como assinalam Alba, Barriga e Viesca (1984). Depois, os exames vão surgir nas universidades de Oxford e Cambridge no século XVIII. Dessa forma, a realização dos exames nas propostas didáticas de antigamente, até o final do século XIX, era *promover a aprendizagem* do aluno e sua acreditação escolar como apontam os autores.

No final do século XIX e início do século XX, com a chamada psicologia científica vão sendo desenvolvidos os *testes de inteligência*. E já na primeira guerra mundial, estes instrumentos passaram a ser empregados nas forças armadas dos Estados Unidos para escolher os seus membros. E a aplicação desta psicologia considerada científica, foi sendo expandida e legitimada.

Depois, passou-se a considerar, através dos testes, que as prostitutas eram débeis mentais, os negros cretinos etc. e sua aplicação na indústria, escola e ambiente penal expandiu-se.

No fundamental, podemos dizer que as concepções vigentes *atuais* de avaliação tiveram lugar, inicialmente, nos Estados Unidos, até porque, como lembra Schwartzman (2005), neste país a importância do ensino privado vem de longe. Mais especificamente, ligadas ao seu processo de industrialização e ao conceito de manejo científico do trabalho de Taylor ou, lembraríamos, administração científica do trabalho como se traduz no Brasil. E já continuam, como lembram Alba, Barriga e Viesca A. (1984) e Barriga (2008), ligação com alguns elementos bastante expressivos da ideologia dominante atual: noção de *eficiência* do capital, como tempos e movimentos frente à produtividade; ideia de maior *rendimento* através de movimentos mais rápidos visando aumentar a produção da empresa; a chamada *seleção científica* do trabalhador para colocar os mais apropriados para cada posto, como se ver claramente na seleção de pessoal para candidatos a posto; significado latente de controle visando assegurar a *eficácia*.

Posteriormente com Fayol, o autor vai introduzir a noção de *controle* na administração visando verificar se ocorre o que estava programado visando, assim, mostrar falhas e erros para reparar e evitar repetições.

Guba e Lincoln (2011), por sua vez, associam os testes ao que intitularam de avaliação de primeira geração. Caracterizada, como acentuam, pela mensuração que era assimilada ao processo avaliativo na educação. Inicialmente os testes eram direcionados na mensuração de crianças e jovens em idade escolar, tentando verificar se os alunos estavam dominando o conteúdo ministrado, sejam de Aristóteles ou outros. E estes testes, no fundamental, examinavam mais a *memória* dos alunos. Lembram também, que nos Estados Unidos, no final do século XIX, pensava-se, também, em ampliar o *currículo* no sentido de *incluir arte, música* e outros temas. Posteriormente, a partir do século XX passou-se a utilizar o chamado QI.

Chamaram também a atenção, para o que poderíamos denominar de outros aspectos gerais e destacam dois. O primeiro, foi uma crescente legitimidade da ciência e o segundo o movimento intitulado de administração científica e que ganhou muito terreno nos setores empresariais e industriais. Nesse último sentido, alguns elementos são bem expressivos da assimilação: o administrador deve tornar o trabalho dos seres humanos mais *eficaz* e *eficiente*; crescimento dos estudos sobre o tempo e o movimento visando métodos de trabalho mais produtivos; procura da motivação do trabalhador.

E essas concepções da chamada administração pública, ressaltam, já tinham entrado na escola na década de 1930: alunos vistos como matéria-prima a ser processada e a escola como fábrica; necessidade de um superintendente escolar; *testes* visando especificar se os alunos correspondiam às especificações estabelecidas pela escola; preparação para a faculdade; papel do avaliador para conhecer bem o instrumento visando a mensuração. Podemos acrescentar, ainda, que não era exatamente típico que estes testes fossem muito aplicados por instituições externas à escola e nem em grande escala como vemos atualmente.

Os testes na educação com vistas à classificação, como aponta House (2000), foram baseados em Thorndike e este também influenciou bastante a chamada gestão científica da

educação com a ideia de *eficiência*. Até a década de 1930, como já analisamos (SOUZA, 2017), os testes eram vistos como a avaliação em educação. Sob a influência de Tyler, entretanto, a partir da década citada a avaliação passou a ser mais associada ao estudos dos currículos e dos objetivos.

E na década de 1960 e de 1970, de acordo com House (2000), os testes em educação voltam a ter uma grande importância e vão ganhando maior espaço, até porque, como vimos no *enfoque de sistema*, algumas ideias como eficiência, custo-produto e outras ganham terreno.

Nos anos de 1960 e 1970, vamos assistir um grande crescimento da utilização dos testes em educação nos Estados Unidos e os vínculos com uma concepção de avaliação. Como sintetiza House (2000), em 1965 foi incluída uma cláusula sobre avaliação no domínio da educação, nos Estados Unidos pensando-se, dentre outras coisas, em informar aos pais sobre como trabalhavam as escolas dos seus filhos. Ao mesmo tempo, os educadores seriam os responsáveis perante aos usuários. E ainda no mesmo ano, foi criado no governo federal um organismo intitulado *Assistant Secretary for Program Evaluation* ou ASPE, que tinha como uma das tarefas elaborar programas que deveriam ser *medidos* no sentido custo-benefício e identificar os mais eficazes para usar os recursos de modo mais *eficiente*. O enfoque do ASPE, procurava relacionar como o dinheiro e programas estariam vinculados a resultados como as pontuações dos *testes*.

A partir de 1968, diz o autor, o ASPE levou a cabo enquetes nacionais e os resultados foram considerados decepcionantes nas pontuações dos *testes*. E daí foi decidido que o governo deveria controlar a avaliação e as variáveis iniciais dos programas e passou-se a trabalhar com experimentos e o programa *Follow Through* foi inserido nesta perspectiva e procurou-se pontuações mais elevadas nos *testes*. A avaliação ligada a testes, assim, passou a ter como fim a chamada *eficiência* da produção dos serviços sociais.

Em 1968, na avaliação do *Follow Through*, pensava-se em levar em conta critérios amplos de avaliação como o desenvolvimento cognitivo e afetivo, grau de participação dos pais, grau de atenção médica etc. No final, entretanto, apesar de vários protestos, como os dos pais, predominou a aplicação de *testes cognitivos*.

Posteriormente, ainda segundo House (2000), foram usados mais testes e ocorreram também novos protestos dos pais devido à sua exclusão nas decisões mas não houve grandes transformações até 1974. Também passou-se neste ano a vincular as pontuações dos *testes* com avaliação de impacto, no sentido dos efeitos mostrados. No final, terminou priorizando-se a aplicação restrita de *testes* como medida de êxito e excluiu-se a participação dos pais e professores no processo avaliativo.

Com o avanço crescente da ideologia neoliberal também na educação, um dos elementos bastante expressivo foi o aumento da aplicação dos atuais testes e suas respectivas características e consequências. Seguindo agora, especialmente a análise e denúncia de Ravitch (2011) sobre os Estados Unidos, podemos apontar exemplos bastante significativos a este respeito como veremos a seguir, até porque vários elementos da análise da autora, não obstante as diferenças, podem ser encontrados também em outros países, como o Brasil.

Em 1994 o governo dos Estados Unidos acabou com as referências nacionais, como em história, e posteriormente foram deixadas a cabo de cada estado. Em 2002, ao assumir e introduzir o programa *Nenhuma Criança Fica para Trás*, o governo Busch passou a adotar os testes em leitura e matemática, aplicar a chamada *responsabilização* e a intitulada *escolha escolar*. Ambas seriam, de acordo com o raciocínio governamental, critério de sucesso. Com isso, outras disciplinas fora a leitura e a matemática passaram a ser excluídas ou perder sua importância anterior, ou seja, os conteúdos e temas tratados nas escolas ficaram cada vez mais reduzidos e, evidentemente, o universo da aprendizagem das crianças. Dessa forma, nas palavras de Ravitch (2011), a *responsabilização* baseada nos testes, além de não elevar os padrões, terminou imbecilizando as escolas ao se agir em torno de metas não realistas.

E os altos escores de acordo com Ravitch (2011), no entendimento da ideóloga Michelle Rhee, dependiam exclusivamente da *qualidade* dos professores e não se considerava, dessa forma, a família ou a pobreza dos estudantes, até porque os professores não deveriam precisar de formação pedagógica. Neste sentido, como fica claro, a aprendizagem era *reduzida aos professores* e não se levava na devida conta um conjunto de outras variáveis, como mencionaremos depois, e o contexto mais amplo. Além de que, também segundo a autora, não se pensava que os próprios professores e os pais achavam sobre os resultados dos testes.

5.3. Problemas teórico-metodológicos com os atuais testes e sua utilização

Além de algumas questões de ordem político-ideológicas que apontaremos depois, é necessário mostrar problemas de ordem científica ligados aos atuais testes e suas aplicações.

Inicialmente, devemos recorrer a Alba, Barriga e Viesga (1984) e a Barriga (2008) quando associam o uso de testes ao que chamam de *enfoques administrativos* na avaliação das instituições de educação e não, diríamos, com base na pedagogia e na pesquisa. Desta forma, não se explicita, por exemplo, a *referência conceitual* e a dimensão teórica e neste enfoque criticado existiriam noções explícitas que o constituiriam, como medição, teste, objetividade, neutralidade valorativa, cientificidade e elementos implícitos como controle de pessoal e social, caráter axiológico, legitimação e que determinam o *lugar oculto*.

Afirma-se no discurso dominante que os testes iriam garantir de forma clara *resultados quantificáveis e irrefutáveis*. Na verdade, seguindo os autores, somente captam o fenômeno educativo via *aspectos formais*, como os indicadores de *eficiência* final da instituição, grau acadêmico dos professores, relação *número* de mestres e de alunos, relação de *orçamento* da instituição e número de alunos e outros. Não estão ligados a uma passagem da qualidade à quantidade como dizem os apologistas neoliberais, mas exprimem uma passagem do *formal-aparente para o quantitativo* ou ideologia tecnocrática.

Na concepção dos autores, portanto, podemos considerar que os testes e suas aplicações não teriam, propriamente, um caráter científico e não captariam o real como dizem seus defensores, mas apenas o aparente. Mas é com Ravitch (2011) que os problemas de ordem científica são expressos de modo contundente. E mesmo que sua análise crítica seja focada no que viveu e analisou nos Estados Unidos, não perde de maneira alguma sua possibilidade de

generalização. Sem uma hierarquização, podemos alinhar por tópicos os mais significativos para facilitar a visualização.

. *A realidade reduzida à dimensão quantitativa.* Como enfatiza a autora, o que não pudesse ser mensurado no plano quantitativo não contava e o sucesso era pensado em termos dos escores dos testes padronizados. Não obstante, ressalta, nem tudo pudesse ser quantificado.

. *Monopólio equivocado do espaço.* Os testes passaram a ser vistos como o *único espaço comum e possível na educação*, embora os mesmos não refletissem somente o que ocorre na escola e expressassem muito mais as características dos estudantes testados, como sua motivação e engajamento dos pais. E terminou ocorrendo a *indústria da testagem*, para usar uma expressão da autora ao se referir também à chamada *responsabilização*. Com o predomínio desta, foi introduzido o *pagamento por mérito* ligado ao resultados dos testes.

. *Variáveis importantes excluídas dos testes.* Por diversas vezes, Ravitch (2011) chama a atenção para inúmeros elementos que precisavam ser levados na devida conta para se ter uma apreensão melhor da aprendizagem e de seus resultados. Lembra, por exemplo: não se atribuía a importância devida ao perfil demográfico dos lugares e as transformações na sua riqueza; desprezava-se o conhecimento sobre a classe social dos estudantes e suas experiências enquanto membros de uma família; excluía-se a importância do desempenho anterior dos estudantes e que também poderia influenciar no resultado dos testes; não se pensava na emoção e na distração dos alunos fora de aula ou dentro da mesma, e que também poderiam afetar os estudantes; nunca se incluía a importância do tempo do professor na escola e seu desempenho em anos diferentes.

. *A imprecisão dos testes.* Ainda sem entrar em detalhes, vejamos algumas observações da autora. Segundo ela, os testes não são precisos como se fossem um termômetro e objetivos, pois não têm precisão de uma balança ou régua. Existe, lembra, fatores que não se controla e exercem influência nos testes. Assim, é difícil saber qual foi o fato mais importante ou se houve uma combinação de fatores, uma vez que as mudanças são incrementais. E cita alguns pontos centrais: os testes variam em qualidade; algumas vezes as questões são mal escritas; as respostas podem ser pontuadas de forma errada; as respostas consideradas certas são erradas ou ambíguas; outras vezes, podem ocorrer duas respostas certas; todos os testes têm margem de erro; um estudante pode passar num teste num dia e no seguinte ser reprovado ou o contrário; podem ocorrer melhores resultados apenas em algumas séries e não em outras.

. *Utilização isolada dos testes.* Os testes deveriam ser utilizados exclusivamente para o que foi projetado e não extrapolar. Em outras palavras, é errado usá-los de maneira isolada para tomar decisões sobre os estudantes, pois o processo requer outras avaliações como a participação em sala de aula e outras.

. *Prioridade para o curto prazo.* A utilização dos testes também levou a que fosse dada grande importância simplesmente ao momento na educação. O desempenho anterior dos estudantes, por exemplo, pode influenciar no resultado dos testes, além das suas experiências enquanto membro de famílias pobres e isto não era contemplado.

. *Período não recomendado de aplicação.* A aplicação dos testes era anual e, desta forma, sujeito a erros aleatórios e não confiáveis. A possibilidade de variações aleatórias, aumenta e deveria haver testes no início e no fim do ano escolar visando conhecer a efetividade da escola.

. *Consequências nas escolas.* Na concepção dominante, deveria haver uma pontuação alta nos testes de habilidades básicas, mas não era exigido nenhum currículo e não se elevava o padrão pois, com isto, o próprio currículo tornava-se muito estreito. Devido à busca por escores mais altos, as escolas deixaram de levar na devida conta a aprendizagem dos estudantes.

. *Efeitos nas escolas públicas.* Redução do número de alunos em escolas públicas em Milwaukee em 1998, pois os alunos passaram para as escolas com isenção de impostos e para as chamadas escolas autônomas. Muito tempo depois, contudo, não foi apresentada a melhora prometida.

. *Estreitamento do currículo e redução da qualidade da educação.* Para se ter uma ideia, o programa *Nenhuma Criança Fica para Trás*, do governo Busch, ignorava estudos importantes como os de história, educação, literatura, ciências, educação cívica, artes e geografia, como se viu em Nova York a partir de 2003 e assim, ao centrar em leitura e matemática, excluía-se outros elementos essenciais da educação. Em San Diego, o ensino ficou reduzido apenas a letramento e matemática e, com isto, as escolas eximiram-se de dar a devida atenção à saúde dos estudantes e de ministrar educação física, artes, conhecimento cívico etc. No Texas, como o *currículo* foi enxugado, os estudantes passaram a ficar atrelados apenas a testes estaduais. E em San Diego, o currículo foi reduzido a letramento e matemática.

. *Efeitos negativos sobre os estudantes.* Alguns foram bastante claros: preparavam-se para os testes e não para a aprendizagem; ao ficarem concentrados nos testes, terminaram não sabendo mais comentar um texto ou escrever uma resposta refletida; como eram ensinados para responder aos testes deixaram de responder, por exemplo, questões de história. Em alguns distritos, os alunos foram preparados semanas e meses para os testes, ou seja, para responder às questões semelhantes e específicas que iriam aparecer nos testes estaduais. Com isto, os estudantes poderiam passar nestes testes que foram treinados mas não em outros e nem no *mundo real* e no mundo do trabalho. É neste sentido, então, que se pode compreender que a autora afirme que os escores podem subir mas os jovens ficarem *ignorantes* sobre os eventos contemporâneos.

. *Abandono dos estudos.* No Texas, para se ter uma ideia, os estudantes deixaram a escola ao não passar nos testes, principalmente os afro-americanos e hispânicos.

. *Questões de generalização.* Considerava-se que o sucesso de uma escola poderia ser facilmente transferido para as outras, como se fosse uma peça de maquinário.

. *Conclusões erradas a partir de comparações equivocadas.* A partir de algumas vantagens materiais das escolas particulares chamadas *autônomas* (recursos financeiros adicionais dos empresários patrocinadores, melhores alunos matriculados, turmas menores, computadores para todos (as) os (as) estudantes etc.), comparava-se às escolas públicas com seus vários problemas (superlotação, baixa frequência, baixas taxas de graduação, condições físicas deterioradas etc.), e atribuía-se a estas a incompetência, ineficácia ou algo do gênero. E mesmo

que nem sempre as *escolas autônomas* apresentassem os melhores resultados nos testes, as comparações tendenciosas, como podemos depreender, prosseguiram.

. *Problemas com a amostra e os resultados.* Estudantes com escores mais altos que tinham origem rica e eram de um distrito rico, apresentavam mais facilidades que outros situados na pobreza. Em Washington e outros lugares, um estudo em 2008 e outros mostraram que as chamadas escolas autônomas não estavam acolhendo crianças com grandes necessidades e que só permaneciam nas escolas de alta performance os melhores e mais motivados. Para tanto, usavam diferentes procedimentos como cartas de recomendação e outros.

. *Informações falsas.* Em algumas ocasiões, professores, estudantes e pais responderam como se estivessem bastante satisfeitos com a escolas autônomas, mas isto foi porque não queriam que a nota de sua escola fosse baixa e afetasse a sobrevivência da mesma.

. *Diferenças nos resultados.* Ocorreu diferenças entre as notas da prefeitura de Nova York e as do Departamento Estadual de Educação e do Governo Federal. E os (as) estudantes preparavam-se para um tipo de teste e não se saíam bem em outros diferentes. No Texas, estudantes passavam nos testes estaduais mas não ficavam bem em outras mensurações acadêmicas como o SAT e NAEP ou Avaliação Nacional de Progresso Educacional.

. *Elevação artificial dos resultados.* Os estados tinham a prática de inflacionar os resultados ao reduzir os pontos necessários visando se dar bem nos testes. Em Nova York em 2006, foi reduzido o limiar para que estudantes passassem de um nível a outro superior (precisavam antes de 41% de pontos e depois 17,9% etc.), já que os estados norte-americanos podiam também reduzir o ponto de corte dos testes (limiar da aprovação) e testar somente uma parte das referências e, assim, os testes ficaram previsíveis anualmente. Em Chicago, por sua vez, foi contratada uma empresa que reduziu o corte dos testes.

. *Falsificação planejada dos resultados.* Ocorreu também, ressalta a autora, que determinados estudantes eram estimulados a não comparecer à escola no dia dos testes. Já em Houston e Dallas, devido à pressão pela responsabilização, a *trapaça* mais encontrada foi entre as escolas autônomas, envolvendo o fato de que diretores e professores vazavam as questões antes das provas.

Não se viu, entretanto, as diferenciações internas e a distância entre os (as) alunos, como brancos, negros etc. e os efeitos disto na escola. Muitas vezes, os estudos mostraram, mesmo nos testes, que não haviam diferenças significativas entre escolas públicas e escolas privadas. Além do mais, frisa Ravitch (2011), ainda *não se tem muito claro* o que produz os resultados apontados nos testes. Além do mais, países como Japão e Finlândia, que ultrapassaram os Estados Unidos em termos dos escores ligados aos testes em educação, não ficaram concentrados com excesso em matemática e ciências pois têm currículo mais amplo.

5.4. Os testes e as questões político-ideológicas e outros efeitos

Embora Ravitch (2011) não fale explicitamente de ideologia neoliberal e seus vínculos com os testes aplicados em educação, sua análise é extremamente clara a este respeito. Para se ter uma ideia, cita autores tipicamente neoliberais como Friedman (1955) e Osborne e Gaebler

(1998) como referências importantes das ações dos governantes dos Estados Unidos na educação. Usa expressões como reforma de mercado, indústria de testagem e outras para designar a aplicação dos testes e os seus significados. Aponta grandes empresas capitalistas como mandantes das mudanças neoliberais e os vínculos com o mercado capitalista. Refere-se, em várias momentos, ao embate e a pressão sobre os professores das escolas e seu sindicato, no caso uma das principais vítimas dos neoliberais.

Também é importante ressaltar que, mesmo quando a autora não desenvolve sua crítica e não a nomeie claramente, nem por isto deixa, igualmente, de assinalar alguns elementos centrais da ideologia neoliberal. E mesmo que parte dos exemplos sejam os mesmo já apontados, nem por isto deixam de ser bastante expressivos também nos planos político e ideológico ou mesmo econômico. Vejamos, então, alguns bem significativos e sem fazer nenhuma hierarquização.

. *Reprodução e ampliação da desigualdade.* Como a autora assinala, as escolas públicas nos Estados Unidos são obrigadas por lei a acolher todos os tipos de estudantes e independentes de suas origens econômicas, de suas notas anteriores etc. E como os chamados estudantes pobres matriculam-se preferencialmente nas mesmas, suas dificuldades de aprendizagem ou seus escores nos testes muitas vezes são considerados baixos e apresentados, igualmente, como uma característica intrínseca da escola pública. As escolas privadas, por seu turno, muitas vezes fazem uma triagem anterior visando matricular somente os alunos que trazem notas altas anteriormente ou que possam conseguir isto pois, desta forma, a instituição passaria a ser apresentada como eficiente, eficaz e, inclusive, com grande probabilidade de ter grande financiamento. Com isto, como fica claro, reproduz-se e se amplia a desigualdade, uma vez que fica estabelecida uma clivagem separando as famílias que têm uma renda maior e as demais com rendas menores e seus respectivos filhos e filhas nos diferentes tipos de escolas.

. *Desvalorização do espaço público.* Como decorrência das notas de grande parte dos alunos pobres e especialmente da imagem criada pela ideologia neoliberal, as escolas públicas passaram a ser divulgadas e vistas por parte da população como algo inferior diante das escolas particulares. Como se o espaço público, exatamente por ser público, fosse por natureza ruim por não ter o perfil das empresas privadas e não apresentar o suposto perfil com que as mesmas são consideradas pela ideologia dominante.

. *Redução do espaço público.* Além das campanhas regulares e gerais contra o espaço público no caso da educação nos Estados Unidos, como foram também criticadas por Ravitch em entrevista (Iwasso, 2010) e com mais detalhes em seu próprio livro (Ravitch, 2011), a autora cita ainda o fechamento de escolas. Neste caso, lembra, as escolas públicas que não estivessem atingindo as metas ligadas aos testes eram fechadas. E como parte significativa destas escolas eram públicas, isto levava a que a redução do espaço público passasse também por isto.

. *A competição como discurso ideológico dominante.* Como podemos depreender a partir de House (2000), os testes, de uma maneira ou de outra, terminam fortalecendo uma concepção capitalista e de mercado, uma vez que estabelecem *competição* entre as partes. Além da pressão sobre a competição entre os próprios alunos e alunas numa mesma escola, a aplicação dos testes e a comparação entre as escolas amplia bastante as atividades sujeitas a esta norma tipicamente capitalista. A educação, com isto, não é vista como um processo pedagógico e que possa

estabelecer ligações com os intitulados direitos e cidadania e passa a ser preferencialmente tratada como mais uma mercadoria a ser comprada pelos clientes e consumidores.

. *A ideologia da competência e desempenho.* Diretamente ligados à questão anterior, podemos apontar também outros elementos ideológicos. Em Washington e outros lugares, por exemplo, um estudo em 2008 e outros mostraram que as chamadas escolas autônomas (particulares) não estavam acolhendo crianças com grandes necessidades e que só permaneciam nas escolas de alta performance os considerados melhores e mais motivados. Para tanto, usavam diferentes procedimentos como cartas de recomendação e outros. Dessa forma, só tinham direito de entrar nas escolas particulares aqueles alunos e alunas que eram considerados (as) competentes, que tivessem um chamado desempenho compatível com os testes ou algo do gênero e que pudessem apresentar uma performance compatível com o mercado capitalista.

. *Imagem negativa das classes populares.* No Texas, os estudantes deixaram a escola ao não passar nos testes, principalmente os afro-americanos e hispânicos. Com isso, diríamos, vai sendo criada e ampliada uma imagem de incompetência dos alunos e alunas ligadas à estas etnias. Assim, de uma maneira ou de outra, termina-se favorecendo o preconceito de classe no sentido mais econômico e a ideia do vencedor e do perdedor. E simultaneamente, respingando em algo racista o que, aliás, como se sabe e é óbvio, é algo bem presente e contínuo nos Estados Unidos.

. *Aumento da repressão.* A repressão aos professores e diretores de escolas, especialmente os docentes, também foi um traço apontado por Ravitch (2011) no seu livro e também em entrevista (Iwasso, 2010). A autora denominou de *abordagem punitiva* a concepção que materializava a pressão sobre os diretores e professores para que atinjissem os escores através da aplicação dos testes. E o termo para isto foi a chamada *responsabilização*, no sentido de que seria o corpo docente o principal responsável pelos resultados obtidos pelos estudantes.

. *Ampliação do mercado capitalista da educação.* Ravitch (2011) cita várias empresas privadas, intituladas de fundação, que passaram a exercer influência na educação através de investimentos e outras atividades. Em Chicago, chegou-se ao ponto de contratar uma empresa privada para ser a responsável pela aplicação dos testes e esta, para se ter uma ideia, reduziu o corte dos testes visando atingir o resultado que era desejado. No Brasil, como apontam Oliveira e Menegão (2012), a presença de empresas privadas em São Paulo têm atuado com o governo estadual através das chamadas parcerias como em “Todos pela Educação” e “Parceiros da Educação”. Com isto, foi aberto um mercado para as consultorias para definir os projetos governamentais, a elaboração de testes de larga escala etc.

6 ALGUMAS ASSOCIAÇÕES ÓBVIAS COM O BRASIL

Mesmo que seja pouco recomendável generalizar para o Brasil de forma direta e imediata as críticas de Ravitch (2011) e de outros autores e autoras à avaliação gerencialista e aos testes atuais e sua aplicação, isto não nega a importância de algumas associações e generalizações provisórias.

A presença crescente e progressiva das empresas privadas na educação, como ressaltaram Oliveira e Menegão (2012) e outros, não deixa de ser expressivo. A adoção do currículo único, igualmente, bem como as chamadas avaliações padronizadas não deixa de ser preocupante.

Também muito significativo foi o Edital da seleção para doutorado em direito de uma universidade privada em Marília, São Paulo. Está bastante claro em http://unimar.br/cursos/posgraduacao/strictosensu/direito/Edital_Mestrado_Direito_2017_02 que haverá descontos na mensalidade do aluno que publicar em periódicos classificados pela Capes como Qualis A1, A2 e B1 ou livro integral com apenas três autores. E prossegue: publicação em periódico A1 terá 10% de desconto, em A2 7,5% de desconto e 5% em B1. Observa-se, assim, que o score da Capes, que trabalha bastante com o ranqueamento do ensino superior e a mesma lógica do gerencialismo e dos testes, funciona como uma luva na chamada mão invisível do mercado capitalista.

No tocante especificamente aos testes como o Ideb e semelhantes, por exemplo, seu significado, a aplicação em larga escala, o ranqueamento e outros aspectos são claramente identificáveis e associáveis com o que tem havido nos Estados Unidos, como foram mostrados em textos como os de Rocha (2017) na sua avaliação no Rio Grande do Norte.

Finalmente, a projeção dos testes e o ranqueamento para outras áreas fora da educação, como criticamos na Introdução deste texto, começaram a atingir o legislativo brasileiro como uma forma de perseguir o funcionalismo, mesmo que isto seja apresentado como uma preocupação com a qualidade dos serviços públicos e outras fraseologias neoliberais. E por mais que sejam dadas desculpas esfarrapadas, na verdade, estão ligadas ao contexto geral do sistema capitalista que exprimem, ao mesmo tempo, vínculos do capital e do Estado com a ideologia neoliberal como indicamos em outro texto (2015).

7 REFERÊNCIAS

- Afonso, A. J. (2000). *Avaliação educacional: regulação e emancipação*. São Paulo: Cortez Editora.
- _____. (2008). Escola pública, comunidade e avaliação: resgatando a avaliação formativa como instrumento de emancipação. In M. T. Esteban (Org.), *Avaliação: uma prática em busca de novos sentidos* (5a ed.). Petrópolis: DP et alii.
- _____. (1999). Estado, mercado, comunidade e avaliação: esboço para uma rearticulação crítica. *Educação & Sociedade*, 20(69), 139-164.
- _____. (2013). Mudanças no Estado avaliador: comparativismo internacional e teoria da modernização revisitada. *Revista Brasileira de Educação*, 18 (53), 267-284.
- Alba, A., Barriga, A. D., & Viesca A. M. (1984). Evaluación: análisis de una noción. *Revista Mexicana de Sociología*, 46(1), 175-204.
- Alkin, M. C., & Christie, C. A. An evaluation theory tree. In M. C. Alkin (Ed.), *Evaluation roots: tracing theorist' views and influences*. Los Angeles: University.

- Alves, E., & Acevedo, R. (2002). *La evaluación cualitativa: reflexión para la transformación de la realidad educativa* (2a ed.). Colombia: Petroglifo Producciones.
- Barriga, Á. D. (2008). Uma polêmica em relação ao exame. In M. T. Esteban (Org.), *Avaliação: uma prática em busca de novos sentidos* (5a ed.). Petrópolis: DP et Alii.
- Campbell, D. T., & Stanley, J. C. (1979). *Delineamentos experimentais e quase-experimentais de pesquisa*. São Paulo: Ed. Da Universidade de São Paulo.
- Cronbach, L. J. (1936). *Course Improvement through Evaluation*. Columbia: Teachers College Record, 68.
- Castro, C. M. (2005). Avaliação não é para amadores. In A. M. SOUZA. *Dimensões da avaliação educacional* (Org.), Petrópolis: Vozes.
- Derlien, H. U. (2001). Una comparación internacional en la evaluación de las políticas públicas. *Revista Do Serviço Público*, 52 (1), 105-124.
- Friedman, M. (1995). The role of government in education. In R. SOLO (Ed.), *Economics and the public interest*. New Jersey: Rutgers College.
- Guba, E. G., & Lincoln, Y. S. (2011). *Avaliação de quarta geração*. Campinas: Editora da Unicamp.
- Iwasso, S. (2010). *Nota mais alta não é educação melhor*. São Paulo: Estado de São Paulo.
- Lima, L. C. (2009). A democratização do governo das escolas públicas em Portugal. *Sociologia*, 19, 227-253.
- _____. (2012). Elementos de hiperburocratização da administração educacional. In J. R. Silva Júnior (Orgs.), *Trabalho e educação no século XX: experiências internacionais*. São Paulo: Xamã.
- _____. (1997). O paradigma da educação contábil: políticas educativas e perspectivas gerencialistas no ensino superior em Portugal. *Revista Brasileira em Educação*, 4, 43-59.
- _____. (2011). Políticas educacionais, organização escolar e trabalho dos professores. *Educação: Teoria e Prática*, 21 (38), 1-18.
- House, E. R. (2000). *Evaluación, ética y poder* (3a ed.). Madrid: Ediciones Morata.
- Oliveira, S. B., & Menegão, R. C. S. G. (2012). Vida e morte do grande sistema escolar americano: como os testes padronizados e o modelo de mercado ameaçam a educação. *Educação & Sociedade*, 33(119), 647-660.
- Osborne, D., & Gaebler, T. (1998). *Reinventando o governo: como o espírito empreendedor está transformando o setor público* (10a ed.). Brasília: MH Comunicações.
- Pasqual, M. (2012). Dialogando sobre políticas de regulação e avaliação no campo da educação. *Roteiro*, 37(2), 185-190.
- Ravitch, D. (2011). *Vida e morte do grande sistema escolar americano: como os testes padronizados e o modelo de mercado ameaçam a educação*. Porto Alegre: Sulina.
- Rocha, M. V. F. (2017). *Avaliação do Índice de desenvolvimento da educação básica (IDEB): um estudo nos municípios do Rio Grande do Norte (2011-2015)* (Tese de Doutorado). Universidade Federal do Rio Grande do Norte, Natal, Brasil.

- Rossi, P., Lipsey, M. W., & Freeman, H. (2004). *Evaluation: a systematic approach* (7a ed.). Thousand Oaks: Sage Publications.
- Schwartzman, S. (2005). As avaliações de nova geração. In A. M. Souza (Org.), *Dimensões da avaliação educacional*. Petrópolis: Vozes.
- Scriven, M. (1967). The Methodology Of Evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven. (Eds.), *Perspectives of Curriculum Evaluation* (Serie Monograph Series on Curriculum Evaluation, n. 1). Chicago: Rand McNally & Co.
- Shadish Júnior., W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: theories of practice*. California: Sage Publications.
- Sousa, S. Z. (2005). 40 Anos de Contribuição à Avaliação Educacional. *Estudos em Avaliação Educacional*, 16 (31).
- Souza, L. M. (2017) Avaliação de políticas públicas e avaliação em educação: relacionamentos antigos, autonomia relativa e carreira solo. In: _____, A. S. Q. Sousa (Orgs.), *Avaliação de políticas públicas e avaliação de educação*. Curitiba: Editora CRV.
- _____. (2014). *Três ensaios sobre avaliação de políticas públicas*. Natal: EDFURN.
- Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record*, 68, 1 (7), 523-540.
- _____. (1971). *Measuring what learners learn*. Urbana: Center for Instructional Research and Curriculum Evaluation.
- Stufflebeam, D. L. (2001). Evaluation Models. *New Directions for Evaluation*. 89, 7-89.
- Weiss, C. H. (1978) *Investigación evaluativa: métodos para determinar la eficiencia de los programas de acción*. México: Editorial Trillas.
- Wholey, J. S. (2010). Exploratory evaluation. In J. S. Wholey, H. P. Hatr, K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (3a ed.). San Francisco: Published by Jossey-Bass.