

**O DESAFIO DA CONSTRUÇÃO DE UM MODELO ALTERNATIVO DE AVALIAÇÃO EXTERNA EM LARGA ESCALA DA APRENDIZAGEM: UM DESAFIO, SIMULTANEAMENTE, CONCEITUAL, TÉCNICO E POLÍTICO COMO PONTO DE APOIO PARA O SUCESSO DE TODOS OS ALUNOS**

**THE CHALLENGE OF BUILDING AN ALTERNATIVE MODEL FOR LARGE-SCALE EXTERNAL LEARNING ASSESSMENT: A CONCEPTUAL, TECHNICAL AND POLITICAL CHALLENGE AS A SUPPORT POINT FOR THE SUCCESS OF ALL STUDENTS**

**EL DESAFÍO DE CONSTRUIR UN MODELO ALTERNATIVO PARA LA EVALUACIÓN DEL APRENDIZAJE EXTERNO A GRAN ESCALA: UN DESAFÍO CONCEPTUAL, TÉCNICO Y POLÍTICO COMO PUNTO DE APOYO PARA EL ÉXITO DE TODOS LOS ESTUDIANTES**

**MICHELLY FRANCINI BRASSAROTO DO AMARAL\***

**OCIMAR MUNHOZ ALAVARSE\*\***

**FERNANDO AUGUSTO SILVA\*\*\***

**Resumo:** Neste artigo, parte-se do reconhecimento da centralidade que as avaliações externas em larga escala da aprendizagem assumiram nas políticas educacionais nas últimas décadas e problematiza-se a relação com a denominação genérica de avaliação externa, indicando que em alguns casos essa expressão sequer significa avaliação e, em outros, não se trata de avaliar aprendizagem. Depois, são apontadas limitações em relação ao modelo vigente, sobretudo no que tange ao envolvimento de professores e elementos que poderiam compor um modelo de avaliação externa da aprendizagem associado à formação dos professores em avaliação educacional, destacando as teorias em torno da validade e da avaliação democrática. Finalizando são apresentados relatos de uma experiência em Moçambique similar ao modelo alternativo proposto, com a valorização da conceituação, das condições e das características que uma avaliação da aprendizagem, externa ou interna, deve possuir para se colocar como ponto de apoio para o sucesso de todos os alunos.

**Palavras-chave:** avaliação da aprendizagem; avaliação externa; política educacional

**Abstract:** In this article, we start from the recognition of centrality as the large-scale external assessments of learning assumed in educational policies in the last decades and are problematized in a relationship with a generic denomination of external evaluation, which are used in some cases and, in others, it's not about evaluating learning. Then, limitations in the relationship with the current model are pointed out, mainly with regard to the involvement of teachers and elements that monitor or compose an external evaluation model of learning associated with teacher training in educational evaluation, highlighting as theories in evaluation validity and democratic evaluation.

\* Mestranda em Educação pela Faculdade de Educação da Universidade de São Paulo (FEUSP). Diretora de Escola na Rede Municipal de Ensino de São Paulo. Membro do Grupo de Estudos e Pesquisas em Avaliação Educacional (GEPAVE). E-mail: mfrancini@hotmail.com

\*\* Doutor em Educação pela Universidade de São Paulo. Professor em curso de Graduação e Pós-Graduação, da Faculdade de Educação da Universidade de São Paulo (FEUSP). Coordenador do Grupo de Estudos e Pesquisas em Avaliação Educacional (GEPAVE). E-mail: ocimar@usp.br

\*\*\* Doutor em Ensino de Ciências (Modalidade: Física) pelo Instituto de Física da Universidade de São Paulo (IFUSP). Professor em Rede Particular de Ensino. Membro do Grupo de Estudos e Pesquisas em Avaliação Educacional (GEPAVE). E-mail: fernandofas1985@gmail.com

Finally, reports are presented of an experience in Mozambique similar to the proposed alternative model, with an appreciation of the conception, conditions and resources that an evaluation of learning, external or internal, should use as a support point for the success of all students.

**Keywords:** learning assessment; external evaluation; educational politics

**Resumen:** En este artículo, partimos del reconocimiento de la centralidad como las evaluaciones externas de aprendizaje a gran escala asumidas en las políticas educativas en las últimas décadas y se problematizan en una relación con una denominación genérica de evaluación externa, que se utilizan en algunos casos y, en otros, no se trata de evaluar el aprendizaje. Luego, se señalan las limitaciones en la relación con el modelo actual, principalmente con respecto a la participación de los docentes y los elementos que monitorean o componen un modelo externo de evaluación del aprendizaje asociado con la capacitación docente en la evaluación educativa, destacando como teorías en la validez de la evaluación y evaluación democrática. Finalmente, se presentan informes de una experiencia en Mozambique similar al modelo alternativo propuesto, con una apreciación de la concepción, las condiciones y los recursos que una evaluación de aprendizaje, externa o interna, debe usar como punto de apoyo para el éxito de todos los estudiantes.

**Palabras clave:** evaluación del aprendizaje; evaluación externa; política educativa

## INTRODUÇÃO

As avaliações de aprendizagem externas e em larga escala se destacam pela centralidade que vêm assumindo no contexto de formulação e implementação de políticas públicas de educação, no Brasil e em vários países, configurando-se como instrumento estratégico justificado tanto para a produção de evidências a respeito de alunos, escolas e redes de ensino, quanto para a possível indução e consolidação de objetivos educacionais, especialmente como condição para a melhoria da qualidade da educação escolar, inclusive para associar seus resultados a indicadores, como é o caso do Índice de Desenvolvimento da Educação Básica (Ideb).

Da perspectiva da investigação científica, sobretudo nas últimas décadas e com diferentes abordagens, questiona-se os argumentos que indicam relações diretas e inequívocas entre essas avaliações e a melhoria da qualidade da educação escolar, embora, por uma diversidade de razões – políticas, econômicas, ideológicas, culturais e até simbólicas –, isso não tem impedido o peso das mesmas no âmbito das referidas políticas. Essa “força” argumentativa pode ser proveniente de uma predisposição social para considerar tais avaliações como expedientes metodologicamente rigorosos e com resultados inquestionáveis de tal forma que a qualidade da educação escolar ser expressaria, total ou parcialmente, no desempenho dos estudantes em testes padronizados, outra marca dessas avaliações.

Observa-se que a expansão das avaliações vem acompanhada da ideia de que boas avaliações e bons diagnósticos da educação poderiam reverter o drama do fracasso escolar; melhorar o fluxo escolar, com impactos nos recursos financeiros; aumentar a participação das famílias e das comunidades nas escolas; melhorar a gestão dos sistemas e das escolas; organizar a agenda política da educação; promover políticas de equidade; e melhorar a qualidade do ensino, conforme encontramos na literatura da área da avaliação educacional (AFONSO, 2000; BONAMINO; SOUSA, 2012; BROOKE; CUNHA, 2011; SOUSA; OLIVEIRA, 2003, 2010; VIANNA, 2003; WERLE, 2011). Esse quadro talvez se justifique porque, em diversos segmentos sociais, inclusive no acadêmico, se considera que as avaliações externas apresentem vantagens que podem compensar suas restrições, como se constata na sistematização do debate organizada por Bauer, Alavarse e Oliveira (2015).

Contudo, esse processo está longe de um consenso absoluto, é um movimento acompanhado de debates e questionamentos sobre a própria pertinência dessas avaliações ao processo educativo, ademais de suas consequências políticas e pedagógicas, especialmente para o trabalho pedagógico nas unidades educacionais, como salientou Lopes Júnior (2016) em trabalho no qual sintetiza aspectos da literatura sobre a temática. Dentre as muitas ressalvas, destaca-se o uso de seus resultados como se fossem, praticamente, os únicos indicadores da qualidade do trabalho escolar, está longe de ser um consenso nesses segmentos como apontam Alavarse, Bravo e Machado (2012) e Machado e Alavarse (2013).

Como parte da compreensão desse quadro complexo da disseminação das avaliações externas, procedemos a um levantamento da produção acadêmica sobre o tema, no qual foram identificados alguns agrupamentos dos estudos referentes aos efeitos e aos usos das avaliações nas escolas ou nos sistemas de ensino. Encontramos abordagens com foco na gestão (AFONSO, 1998; BARRERA, 2000; CASTRO, 2009; KLEIN; FONTANIVE, 1995; VIANNA, 2003), no currículo (CHAPPAZ, 2015; MADDAUS; RUSSELL; HIGGINS, 2009), no trabalho do professor (EVERS; WALBERG, 2002; OLIVEIRA, 2015) e na formação docente (BAUER, 2012), além de pesquisas que procuraram compreender os possíveis reatamentos das avaliações externas sobre as políticas públicas educacionais (ALAVARSE, 2014; BAUER, 2010, 2012, 2013; SOUSA; OLIVEIRA, 2010; OLIVEIRA *et al.*, 2013). Chama atenção o fato de que foram encontradas poucas investigações que tivessem as próprias avaliações como objeto de estudo, isto é, que tivessem se debruçado sobre suas características técnicas, incluindo suas limitações nesse tocante.

Entretanto, apesar da amplitude e profundidade na coleta de dados em várias avaliações externas, que abrangem testes padronizados e questionários contextuais, ainda é preciso construir mecanismos para que os resultados sejam utilizados por gestores e professores de modo a colaborarem para tomadas de decisão que influenciem na melhora da qualidade do ensino oferecido e, por conseguinte, no incremento da aprendizagem dos alunos. Algo que contesta as afirmações usuais de responsáveis por essas avaliações, como se bastasse a existência das mesmas para a garantia de consecução desses propalados objetivos, e coloca uma questão, entre outras: se mantidas as avaliações externas, seria possível um modelo alternativo para sua organização?

Neste artigo, propomos enfrentar essa questão respondendo-a afirmativamente e buscamos, ainda que sem a pretensão de esgotar as possibilidades, elementos que indiquem um modelo alternativo de avaliação externa em larga escala de aprendizagem que se constitua como um processo de formação envolvendo professores e os aproximando da apropriação e uso dos resultados. Para tanto abordamos, a seguir, aspectos conceituais e procedimentais envolvidos nesse desafio que, a nosso ver, não tem merecidos o devido destaque, curiosamente, tanto por parte daqueles que contestam *in totum* as avaliações externas quanto por aqueles que consideram os modelos vigentes plenamente satisfatórios.

## **AValiação EDUCACIONAL E AValiação DA APRENDIZAGEM**

3

A compreensão das avaliações externas de aprendizagem, quer teórica, quer empiricamente, nos parece exigir, inicialmente, uma delimitação da própria definição de avaliação, aqui assumida no âmbito escolar e, por isso, a consideramos como avaliação educacional. Ainda que, como se demonstrará adiante, haja, na literatura da área, um grande consenso em torno de uma definição, ou pelo menos de um de seus elementos, constata-se uma profusão, nessa mesma literatura, de definições e determinadas condições e desdobramentos da avaliação educacional.

Sem necessariamente emprendermos um estudo cronológico, pode-se localizar que um aporte extremamente relevante no sentido de definir avaliação, sobretudo pelas repercussões junto a outros autores, encontra-se na produção de Scriven (1967), para quem avaliar significa a determinação do valor ou mérito de alguma coisa, que encontra sustentação na própria etimologia do termo, pois a palavra “avaliar” contém a ideia de valor, de modo que se evidencia uma primeira natureza dela: atribuir valor a algum objeto.

Adicionalmente, essas considerações iniciais apontam, dentro de um amplo universo de questões, que a própria conceituação de avaliação educacional deve ser merecedora de atenção por conta de suas implicações políticas e investigativas, com repercussões práticas. Ou seja, a definição adotada condiciona uma série de desdobramentos práticos na realização da avaliação.

Nas muitas definições encontradas, pôde-se verificar que um conjunto de termos ou são tomados como sinônimos ou como componentes intrínsecos ao processo avaliativo. Assim, nos deparamos com associações com o verbo avaliar como sendo: medir, testar, observar, registrar, verificar, refletir, compreender, decidir, planejar, intervir, acompanhar, promover, transformar etc. Por decorrência, alguns teóricos, ao definirem o que seja avaliar o fazem em contraste com esses verbos, ou pelo menos em relação alguns. Para Lafourcade (1969), por exemplo, avaliação é a interpretação de medidas ou descrições

qualitativas, consideradas em relação a uma norma anteriormente estabelecida e considera que o termo “avaliação” é, por vezes, empregado de forma imprópria, como sinônimo de medida. No entanto, para o autor, a avaliação envolve, além de mensurações, descrições qualitativas e requer, ainda, interpretação, o que lhe é essencial. Haydt (2004, p. 10), nesse escopo conceitual, esclarece as principais diferenças entre testar, medir e avaliar:

**Testar** – Verificar um desempenho através de situações previamente organizadas, chamadas testes.

**Medir** – Descrever um fenômeno do ponto de vista quantitativo.

**Avaliar** – Interpretar dados quantitativos e qualitativos para obter um parecer ou julgamento de valor, tendo por base padrões ou critérios.

Para Viana (1989), “avaliação” e “medida do rendimento” são frequentemente usadas como expressões intercambiáveis, e refletem imprecisões no emprego das palavras medir e avaliar. O autor ainda destaca que, eventualmente, a medida pode levar à avaliação, que, entretanto, só se realiza quando são expressos julgamentos de valor. Conclui que avaliar é, assim, emitir um julgamento de valor sobre a característica focalizada, podendo esse valor basear-se, parcial, mas não exclusivamente, em dados quantitativos. Na visão de Miras e Solé (1996, p. 375), os objetivos da avaliação são traçados em torno de duas possibilidades: emissão de “um juízo sobre uma pessoa, um fenômeno, uma situação ou um objeto, em função de distintos critérios”, e “obtenção de informações úteis para tomar alguma decisão”.

Realizando um levantamento de definições, Nevo (2006, p. 442-443) aponta que a avaliação envolveria a coleta sistemática de informações para atribuição de valor relativo à natureza e qualidade do objeto educacional, de tal modo que

Essa definição combina descrição e julgamento, mas distingue entre eles por causa de sua natureza diferente. Descrição pode se basear na coleta de dados e isso resulta em informação altamente objetiva. Julgamento é baseado em critérios, que na maior parte dos casos são determinados por valores, normas sociais e preferências pessoais das partes interessadas com a avaliação. O julgamento pode ter uma natureza bastante subjetiva. Descrição e julgamento, embora coexistam na maioria das avaliações, são usados em diferentes proporções para diferentes propósitos e por diferentes tipos de avaliadores.

4

Por sua vez, Castillo Arredondo e Cabrerizo Diago (2009, p. 39) consideram a avaliação como um processo dinâmico, aberto e contextualizado, a ser desenvolvido em determinado período de tempo. Apresentam como três características básicas da avaliação:

1º Obter informação: Aplicação de procedimentos válidos e confiáveis para conseguir dados e informação sistemática, rigorosa, relevante e apropriada que fundamente a consistência e a segurança dos resultados da avaliação.

2º Formular juízos de valor. Os dados obtidos devem permitir fundamentar a análise e a avaliação dos fatos que se pretende avaliar, para que se possa formular um juízo de valor o mais acertado possível.

3º Tomar decisões: De acordo com os juízos emitidos sobre a informação relevante disponível, será possível tomar as decisões cabíveis a cada caso.

Podemos, ainda, de acordo com Alavarse (2013, p. 147), considerar que o processo avaliativo deve, entre outros aspectos, levar em conta os usuários dos resultados da avaliação, pois serão, ao fim, quem poderá aproveitar esses resultados para que determinadas ações sejam desencadeadas no processo pedagógico ao qual a avaliação esteja vinculada, notadamente, quando a preocupação esteja voltada para a melhoria desse processo, mais ainda quando se traduz na ampliação das aprendizagens dos alunos concernidos a esse processo. Assim, antecipadamente, é preciso estabelecer quais serão os usuários dos resultados: somente os professores? Os alunos? Os pais e responsáveis? A comunidade escolar? A comunidade educacional? O Ministério da Saúde e/ou Educação? Com efeito, a definição dos usuários, entre

outras providências e características, e sem se constituir no núcleo conceitual da avaliação educacional, realça a importância das formas e conteúdos da divulgação dos resultados.

Na perspectiva de conceituar, mais estritamente, a avaliação e, ao mesmo tempo destacar elementos que com ela estabelecem relações, ainda que potenciais, temos a contribuição de Luckesi (2018) que reconhece que a tomada de decisão e a intervenção não fazem parte da avaliação, pois sua característica marcante, para que a avaliação ocorra de fato, é necessário um juízo de valor sobre os dados de uma realidade que se pretende avaliar, que pode estar associada, mas não confundida com a necessidade de tomada de decisão, por exemplo, por parte do professor.

Para efeito de nosso trabalho, entendemos que Lukas Mujika e Santiago Etxeberria (2009, p. 91-92) apresentam uma definição que melhor responde ao desafio de conceituar e apontar seus desdobramentos, referidos anteriormente como uma implicação prática. Para os autores,

A avaliação é o processo de identificação, levantamento e análise de informação relevante de um objeto educacional – que poderá ser quantitativa ou qualitativa –, de forma sistemática, rigorosa, planejada, dirigida, objetiva, crível, fidedigna e válida para emitir juízo de valor baseado em critérios e referências preestabelecidos para determinar o valor e o mérito desse objeto a fim de tomar decisões que ajudem a otimizá-lo.

Desta forma, tendo como referência a definição apresentada por Lukas Mujika e Santiago Etxeberria (2009), compreendemos que para que uma avaliação atenda as características postas, e sobretudo, seja crível e válida, se faz necessário o envolvimento de professores em suas diferentes etapas de construção e consolidação para o uso dos resultados, traço que grande parte das avaliações externas, no Brasil e em outros países, não exibem.

De modo geral, essas definições contribuem efetivamente para a compreensão do que seria avaliar e/ou sua função e, simultaneamente, permitem questionar, particularmente considerando a emissão de um juízo como elemento incontornável de uma avaliação educacional, se determinadas definições seriam mesmo relativas à avaliação. Por exemplo, identificamos em Bloom, Hastings e Madaus (1975, p. 177) a delimitação de avaliação como um método de adquirir e processar evidências necessárias para melhorar o ensino e a aprendizagem, sendo, ainda, um auxílio para classificar os objetivos significativos e as metas educacionais, um processo para determinar em que medida os alunos estão se desenvolvendo dos modos desejados, um sistema de controle da qualidade, pelo qual pode ser determinada etapa por etapa do processo ensino/aprendizagem, a efetividade ou não do processo e, em caso negativo, que mudanças devem ser feitas para garantir sua efetividade. Mesmo que considerássemos que a determinação de uma medida adequada de aprendizagem seria, ao cabo, um julgamento com base num critério, nota-se, como em muitos outros autores, que não seria, rigorosa e explicitamente, uma delimitação conceitual de avaliação educacional, pois nessa formulação nos deparamos muito mais com finalidades e usos dos resultados da avaliação, o que caracterizaria muito mais sua função; certamente, um aspecto relevante se ponderarmos que no âmbito da educação escolar a avaliação é um elemento do processo pedagógico e nela são apoiadas muitas decisões e desdobramentos.

Encontramos, assim, na literatura mais do que o emprego da expressão “função”, mas, sobretudo, a adjetivação da avaliação como decorrência de sua função que a ela se atribui. Disso decorrem as expressões “avaliação diagnóstica”, “avaliação formativa”, “avaliação somativa”, entre outras. No entanto, é preciso frisar a existência de diferenças na conceituação dessas expressões. A função diagnóstica, de acordo com Miras e Solé (1996, p. 381), é representada pela avaliação inicial, que possibilita o levantamento de informações acerca das capacidades do aluno antes de iniciar um processo de ensino/aprendizagem, ou ainda, segundo Bloom, Hastings e Madaus (1975), a busca pela determinação da presença ou ausência de habilidades e pré-requisitos, bem como a identificação das causas de repetidas dificuldades na aprendizagem, averiguando a posição do aluno face a novas aprendizagens que lhe vão ser propostas e a aprendizagens anteriores que servem de base àquelas, no sentido de obviar as dificuldades futuras.

A função formativa, conforme Haydt (2006, p. 17), permite verificar se os alunos estão, de fato, atingindo os objetivos pretendidos, verificando a compatibilidade entre os objetivos e os resultados efetivamente alcançados durante o desenvolvimento das atividades propostas. Outro aspecto destacado pela autora é o da orientação fornecida por este tipo de avaliação, tanto ao estudo do aluno como ao trabalho

do professor, principalmente através de mecanismos de *feedback* que permitem ao professor detectar e identificar deficiências na forma de ensinar, possibilitando reformulações no seu trabalho didático, visando aperfeiçoá-lo. Para Bloom, Hastings e Madaus (1975), a avaliação formativa visa informar o professor e o aluno sobre o rendimento da aprendizagem no decorrer das atividades escolares e a localização das deficiências na organização do ensino para possibilitar correção e recuperação.

## AVALIAÇÃO DA APRENDIZAGEM

No âmbito da avaliação educacional, a avaliação da aprendizagem, além de ser uma expressão bastante conhecida, é a mais difundida, pois, praticamente, se confunde com a difusão da escolarização no século XX e, igualmente, sempre envolvida em polêmicas por conta do uso de seus resultados com a finalidade de decidir o “destino” dos alunos quanto ao fluxo no interior das escolas, isto é, para fins de aprovação ou reprovação. Nas últimas décadas, ganhou maior densidade quando se considera que determinadas políticas de avaliação foram proposta para organizar a avaliação da aprendizagem de modo externo e em larga escala.

Um dos principais desafios da avaliação da aprendizagem diz respeito ao seu objeto. Se consideramos que o foco é a aprendizagem, rigorosamente, estamos nos reportando a um processo interno, intangível, latente aos indivíduos, ademais das profundas divergências entre os muitos psicólogos que procura(ra)m defini-la. Genericamente, pode-se aventar que a aprendizagem estaria fortemente associada à capacidade dos alunos na realização de determinadas tarefas ou, como um equivalente, na capacidade de expressar um domínio ou um conhecimento sobre determinados conteúdos, mas esses desempenhos derivariam de um potencial interno aos alunos. Portanto, se consideramos que avaliar é julgar “algo” e sendo este interno aos indivíduos, tem-se que não se pode avaliar diretamente essa capacidade dos alunos – o que se pretende avaliar. Restaria, então, avaliar os resultados ou os desempenhos que os alunos apresentam na realização das tarefas a eles propostas e, assim, seriam esses resultados o objeto da avaliação.

Com isso não se está desconsiderando todo um gigantesco volume de aportes de várias vertentes da Psicologia voltada para a aprendizagem, mas apenas estamos delimitando aquilo que, inclusive por razões pragmáticas, pode-se, efetivamente, avaliar. Ademais, as hipóteses psicológicas sobre a aprendizagem podem ser invocadas para explicar, eventualmente, o porquê de determinados resultados ou produções dos alunos, mas, rigorosamente, não seriam o objeto de avaliação.

Assim, para que tenhamos uma avaliação consistente, é imperioso, ainda que isso não esgote um amplo conjunto de exigências, que: 1- existam boas tarefas avaliativas; 2- os instrumentos sejam bem construídos; 3- as condições de aplicação dos instrumentos não se convertam em obstáculos cognitivos; 4- os alunos acreditem que o processo avaliativo seja justo, fidedigno, válido e que seus resultados serão utilizados em seu favor e, como consequência, dediquem-se a realizar com afinco essas tarefas avaliativas propostas. Somente com essas características é que se poderia fazer inferências entre a “intensidade” – daí a noção de medida educacional – do que teria aprendido – no contexto escolar, aquilo que tem-se denominado de aprendizagem – e seu desempenho nas tarefas para então, eis o mais importante, de acordo com algum critério julgar esse desempenho, supostamente representando o que o aluno aprendeu. Evidentemente, esse critério é que daria todo sentido ao ato de julgar, pois um resultados não em si mesmo não é “bom” ou “ruim”. Nestes termos, a definição do critério concentraria a problemática da subjetividade da avaliação a depender de quem é o sujeito a defini-lo. Se a avaliação por si mesma não exclui algum aluno do processo pedagógico, mas, sim, o uso ou consequência desse juízo que, como definido por Luckesi (2018), não se confunde com a avaliação ainda que a ela se associe, esse uso parte de um juízo e este só é concluído por conta de um critério.

Não bastando os desafios, teóricos e práticos, de realização da avaliação da aprendizagem em geral, há um adensamento quando introduzimos o problema do sujeito dessa avaliação. Ou seja, as implicações em torno de quem conduz ou determina como se dará a avaliação da aprendizagem. Se a escola de massas instaurou o professor como avaliador, configurando o que mais tardiamente passou a ser denominada de avaliação interna, definindo a escola, ou mais precisamente a sala de aula como o lócus, por excelência, para essa avaliação, essa mesma escola de massas, passou a conhecer uma avaliação, igualmente da

aprendizagem, mas conduzida por quem não se encontra, pelo menos completamente, no interior das escolas e, mais ainda, das salas de aula. Acentuando-se a partir dos anos 1960, especialmente nos Estados Unidos da América do Norte (EUA), o quadro da avaliação educacional passou a ser composto com as denominadas avaliações externas.

Assim, neste trabalho procuramos realçar que uma parte considerável dessas avaliações devem ser denominadas rigorosamente de avaliações externas da aprendizagem, considerando seus objetos, articulados com foco naquilo que os alunos revelam como o principal resultado da escolarização, isto é, o que aprenderam.

Entretanto, destaca-se que no bojo de diferentes avaliações externas, seja a nível municipal, estadual, nacional ou internacional, há exemplos em que a denominação “avaliação”, pelo menos no que tange a discussão apresentada até aqui, não caberia, pois essas iniciativas apesar de se serem apresentadas como “avaliação” não vão além apresentarem aos alunos a elas concernidos um instrumento – uma prova –, pois, independente da qualidade de suas mensurações do aprendido, não apresentam um critério e, por decorrência, muito menos um julgamento do desempenho na realização das tarefas constantes nessa prova. O Exame Nacional do Ensino Médio (Enem) implementado em 1998 pelo Inep/MEC é um exemplo, cuja proposta inicial tinha como mote contribuir para a aferição da qualidade da educação básica e a auto avaliação (BRAVO, 2017), transformou-se, principalmente a partir de 2009 em um exame para acesso ao Ensino Superior, finalizando-se na emissão do desempenho de cada “candidato”. Ou seja, a informação para cada candidato de seu desempenho configura-se, como mencionado anteriormente, como uma medida, mas sem critérios estabelecidos que favoreçam a algum tipo de julgamento.

Prova Brasil, rigorosamente, seria outro exemplo pelo qual tem-se medidas de desempenho, no caso em leitura e resolução de problemas, sem que haja por parte do Inep um julgamento desses desempenho, de tal sorte que, com a divulgação dos resultados, não se possa ir além da alocação das médias de desempenho dos alunos das e da distribuição percentual desses desempenhos nos níveis de sua escala de medida.

## A AVALIAÇÃO EXTERNA DA APRENDIZAGEM

7

Como definiu Madaus (1988, p. 86), uma avaliação é considerada externa quando, em última instância, é conduzida por alguma autoridade que se encontra fora da escola, ainda que mantendo alguma vinculação com esta. Ou seja, a exterioridade de uma avaliação é definida pela “distância” que seu sujeito estabelece com a escola, local privilegiado de aprendizagem, o que assinala sua dimensão política e se constitui como perspectiva importante para se compreender as disputas de entendimento e aceitação dessas avaliações externas. Assim, uma avaliação externa, entre outras características, marca-se pelo desafio de ser legitimada pelos profissionais que se encontram atuando na escola, usualmente docentes, para diminuir conflitos e aumentar seu potencial de melhorar os processos de ensino e de aprendizagem.

Se isso não nos leva, mecânica e automaticamente, a qualquer apreciação de suas qualidades, contudo, coloca o problema de sua legitimação, sobretudo quanto ao reconhecimento de seus resultados, como um aspecto político desse tipo de avaliação, que pode ser para fins de certificação, credenciamento, diagnóstico e prestação de contas. Essas avaliações são, em geral, organizadas a partir da priorização da dimensão cognitiva dos alunos, mediante o recurso de testes padronizados compostos por itens de múltipla escolha, e alcançando, diretamente, um grande número de alunos e, indiretamente, de professores, diretores, coordenadores vinculados a determinados níveis de escolarização.

A expressão “avaliação externa” abrange características gerais de qualquer avaliação educacional, por consistir em um processo de coleta e análise de informações que, mediante a utilização de instrumentos e procedimentos adequados às características do objeto avaliado, permite a emissão de um julgamento de valor baseado em critérios e referências estabelecidos, que potencialmente, pode ser usado para tomada de decisões. A aplicação de instrumentos padronizados se fundamenta nos estudos psicométricos para definir uma medida educacional, sendo que esta possibilita o desenho de testes e a construção de escalas que balizam as inferências sobre a aprendizagem.

A expressão “em larga escala” é usada para designar a abrangência da aplicação de uma avaliação e, por isso, não é sinônimo de avaliação externa, pois se esta marca a posição de um agente avaliador externo, não necessariamente uma avaliação em larga escala é externa. Deste modo, uma avaliação pode ser externa, mas não necessariamente em larga escala; a recíproca também é verdadeira.

Com efeito, embora não seja usual na literatura esse tipo de ponderação, muitas avaliações externas são, ou deveriam ser, avaliações externas da aprendizagem, pois o objeto é a aprendizagem de cada um dos participantes do processo, aprendizagem expressa nos resultados de realização de tarefas educacionais, ainda que seu sujeito seja externo às escolas.

As avaliações externas tanto podem ser um instrumento para que as escolas, gestores e demais profissionais repensem seu trabalho, quanto para se efetuar classificações que pouco contribuem para a melhora da qualidade das escolas. Se a avaliação for concebida apenas como mecanismo de controle ou se for ignorada, seja pelos professores, pelos gestores ou pelos responsáveis pela formação de docentes, pouco contribui para a melhoria do ensino (SOUSA, 2007).

Neste sentido, reforçando a proposta da avaliação externa da aprendizagem, podemos apontar avaliações externas que, a despeito de suas denominações, não priorizam a aprendizagem como objeto, pois, no fundo, seus resultados estão mais associados à avaliação, por exemplo, do trabalho docente e de gestão, como fica patente nas situações em que há bonificação dos profissionais da educação, tornando a aprendizagem, de algum modo secundária para efeito da avaliação. O Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (Saresp) se enquadra neste modelo, em que a bonificação é quem informa a busca por melhores resultados por diferentes atores educacionais (SILVA, 2015). Não obstante, a alocação do Saresp por Bonamino e Sousa (2012) à terceira geração vai ao encontro do apontamento levantado, geração esta caracterizada pela divulgação dos resultados de forma interna e externa, e também sanções ou recompensas (bônus) em decorrência dos resultados dos alunos.

Madaus (1988) aponta que a avaliação externa pode ter impacto com efeitos na progressão dos alunos, em termos de estreitamento do currículo, do desenvolvimento de práticas de “ensino-para-o-teste” e do desenvolvimento de práticas de avaliação de sala de aula que mimetizam, no seu conteúdo e na sua forma, a avaliação externa. Kellaghan e Greaney (2004, *apud* BECKER, 2010, p. 3) apontam uma grande lacuna entre o que as avaliações demandam dos professores e a visão dos professores a respeito do papel deles no processo de avaliação, citando exemplos em países africanos. Os autores ainda destacam a questão da vontade política uma vez que, sem apoio do governo, possíveis recomendações que possam ser feitas, com base nos dados levantados, não serão postas em prática.

Outra crítica com relação à divulgação dos resultados refere-se ao excesso de técnicas complexas na apresentação das informações, considerando que, se os processos de modelagem e as estatísticas utilizadas são complexos, ao transmitir as análises para o público em geral se faz necessário utilizar uma linguagem menos técnica, mais acessível. Franco (2001) propõe a utilização de gráficos simples que possam ser interpretados sem conhecimentos específicos de métodos quantitativos para que se consiga atingir o público alvo.

Como uma tendência em âmbito internacional, ministérios e secretarias, por não possuírem equipes para avaliar a aprendizagem de seus alunos, acabam recorrendo a consultores ou empresas que, muito frequentemente, após realizarem seu trabalho não contribuem para a formação de profissionais dessas instâncias demandantes (KELLAGHAN; GREANEY, 2004, SOUSA; OLIVEIRA, 2007).

No geral, as empresas são contratadas para prestarem serviços técnicos, relacionados a elaboração dos itens, dos testes e do tratamento dos dados, não necessariamente possuem a atribuição da articulação da tarefa técnica com a formação dos profissionais das equipes técnicas das secretarias e dos professores. Essa articulação, que decorre da política de avaliação, aproximaria os profissionais da educação das técnicas e teorias que embasam as avaliações externas e este processo poderia se configurar como uma formação inicial em avaliação educacional, uma vez que, embora sejam apontadas as características de uma avaliação para que essa seja considerada minimamente adequada, estudos apontam que a avaliação educacional não é abordada na formação inicial e continuada dos professores (GATTI, 2007; FREITAS, 2019). Com efeito, a participação dos professores no processo de construção das avaliações externas, sobretudo, das etapas de elaboração e revisão dos itens, subsidiaria uma formação introdutória à avaliação

educacional, bem como, auxiliaria na qualificação dos seus instrumentos, uma vez que, promoveria a reflexão das metodologias adotadas para elaboração dos instrumentos interno e facilitaria a análise e interpretação das tarefas apresentadas nas avaliações externas e seus resultados.

Supondo um padrão no qual instâncias administrativas do Estado, como são os Ministérios, contratam avaliadores externos para a execução da tarefa, com pouca adesão de professores e baixo uso dos resultados, ademais, via de regra, de responsabilizar docentes e gestores de escolas como responsáveis por esses resultados. Ainda, tem-se nesse padrão que o objeto de avaliação pouco leva em consideração aquilo que é ensinado, ao definir que esse objeto representa o que os alunos devem saber ou que deveriam ter aprendido ou, também, o que deveria ter sido ensinado, mesmo afastando-se a hipótese de que se pretende controlar o currículo com essa iniciativa de avaliação externa, pois a preocupação com o desenvolvimento curricular, pertinente aos desafios de ampliar a formação dos alunos, não pode se concentrar na avaliação externa como um expediente para garantir a execução de determinados planos curriculares. Ainda, nesse padrão, com pouca utilização dos resultados para fomentar políticas educacionais. Com ausência de apropriação de conhecimentos sobre o processo de avaliação externa por técnicos dessas instâncias.

Analisando modelos no qual as provas são elaboradas por especialistas e empresas contratadas, observa-se que o impacto dos resultados pode ser considerado mínimo, por várias razões: os relatórios, elaborados para administradores, técnicos e, em geral, para os responsáveis pela definição e implementação de políticas educacionais, não costumam chegar às mãos dos professores para fins de análise, discussão e estabelecimento de linhas de ação. São demasiadamente técnicos, empregando uma linguagem muitas vezes desconhecida pelos docentes e que poderiam ser evitados. Por outro lado, esses mesmos resultados são apresentados em termos globais, sem identificação, como seria desejável, das unidades educacionais, referindo-se, quando muito, a unidades macro, os estados, e, nestes, eventualmente, às regiões geoeeducacionais. Ainda que os resultados dos desempenhos sejam apresentados em escalas elaboradas por intermédio de rigorosos procedimentos estatísticos, e com a especificação dos vários níveis correspondentes de competência, dificilmente os professores têm condições técnicas para interpretar dados que resultam da expertise técnica dos responsáveis pelos relatórios. Destaquemos, também, que há certa resistência, por parte de professores e alunos, aos resultados de avaliações externas, traduzindo, assim, certa dose de incredulidade em relação aos resultados. Tudo isso faz com que importantes avaliações tenham o seu impacto, quando ocorre, bastante restrito, ou até mesmo seja inexistente, em relação ao sistema e a suas unidades.

Outro ponto de tensão é a validade curricular das matrizes de avaliação, sendo necessária uma reflexão sobre as avaliações ora operacionalizadas nos vários níveis do sistema educacional, especialmente avaliações em larga escala, abrangendo a diversidade geográfica e multicultural. Fala-se, e com bastante destaque, ainda que nem sempre de forma consistente, na avaliação de competências e habilidades, mas de modo discutível e muito pouco consensual. As questões que se impõem imediatamente, com o objetivo de aprofundar nossas percepções, podem ser propostas da seguinte forma: são desenvolvidas competências e habilidades no sistema educacional de uma forma sistemática, ou, explicitando, é o nosso ensino orientado para o desenvolvimento de competências? Se for, qual a natureza dessas competências e as supostas habilidades? Se competências e habilidades foram promovidas, houve, efetivamente, preparo adequado dos educadores em relação a esse complexo e controvertido assunto? As indagações partem do princípio de que somente se pode avaliar aquilo que efetivamente foi desenvolvido, além de considerar que não se avalia em abstrato, mas considerando a problemática em que se situam os avaliados. Neste sentido, acreditamos que somente com a participação dos professores é possível aproximar o máximo possível as avaliações do que foi ou deveria ter sido ensinado.

Diante deste quadro, propomos duas discussões teóricas que vem ao encontro de uma avaliação externa da aprendizagem e que busque, de alguma forma, contribuir para que vários dos aspectos levantados em relação aos desafios a serem enfrentados (professores, matriz de referência e equipes de avaliação) possam ser superados.

## PROPOSTAS TEÓRICAS PARA O PROCESSO DE CONSTRUÇÃO DE UMA AVALIAÇÃO EXTERNA

Como apresentado até aqui, grandes são os desafios oriundos da avaliação externa entre a elaboração e a implementação. Para Bauer, Alavarse, Oliveira (2015, p. 1378),

*É importante observar que apesar de muitos pesquisadores se dedicarem ao estudo das avaliações de sistema e aos problemas delas decorrentes, estas reflexões parecem ter pouco impacto entre os gestores das políticas, visto que a incorporação da crítica e uso dos conhecimentos elaborados para a reorientação das avaliações parece estar fora do rol de intenções dos técnicos responsáveis pelo gerenciamento das avaliações.*

Essa desarticulação entre esses dois diferentes campos é reforçada pelos problemas, indagações, e tempo de ação institucional. Para aumentaras tensões nessa realidade, há o debate entre o que seria atribuível à avaliação e seu emprego como parte de políticas de responsabilização, chegando-se, em algumas circunstâncias, à defesa da extinção da avaliação, pois, para muitos, a melhor saída para acabar com o bônus, por exemplo, é deixar de aplicar as avaliações externas. Porém, em muitos contextos, a dependência de uma com a outra, pode fazer com que a avaliação externa seja afetada, comprometendo as análises de informações ao não se dispor de dados fidedignos e válidos da realidade da educação.

Diante dessa situação, apontamos o que poderia ser feito para remodelar as avaliações externas da aprendizagem, como forma de estabelecer um primeiro diálogo. Para tanto, expomos duas teorias que podem auxiliar nessa tarefa.

### A TEORIA DE BARRY MACDONALD

A primeira é de Barry MacDonald, que foi professor emérito na Universidade de East Anglia (Norwich, Inglaterra) e professor Honoris Causa pela Universidade de Valladolid, na Espanha, tendo sido durante vários anos diretor do Center for Applied Research and Education (CARE), onde trabalhou com Lawrence Stenhouse<sup>1</sup>. Em geral, sua teoria no Brasil é mais utilizada para a avaliação de programas e a avaliação institucional, mas pouco concatenada aos estudos de avaliação educacional da aprendizagem (externa e interna). Ele propõe a chamada **avaliação democrática** (MACDONALD, 1974) e, para isso, estabelece uma diferença entre três tipos de avaliação, a saber: avaliação burocrática, avaliação autocrática e avaliação democrática (GREENE, 2006, p. 119-120), sendo que,

Em primeiro lugar, a avaliação burocrática é um serviço incondicional às agências do governo já com poderes para alocar recursos educacionais e determinar as direções políticas. O papel do avaliador burocrático é de consultor de gestão, e seu trabalho não é independente nem está disponível para o escrutínio público. Clientes de avaliação burocrática mantêm controle sobre os produtos deste trabalho. Segundo, a avaliação autocrática é um serviço condicional às mesmas agências governamentais. O avaliador autocrático mantém a independência como consultor especialista externo e, assim, mantém a propriedade dos produtos de avaliação. Seu trabalho é validado pela comunidade de pesquisa científica e, portanto, quando válido, serve para defender as orientações políticas existentes.

Em contraste, o avaliador democrático reconhece o pluralismo de valores a serviço do direito público de saber. Na avaliação democrática, os métodos e resultados devem ser apresentados de maneiras acessíveis a múltiplos públicos não especialistas, em um relatório que aspira ao status de “best-seller”. (Tradução nossa).

Diante destes três tipos de avaliação, MacDonald se aprofunda nas explicações do que seria a avaliação democrática e de acordo com Norris (2015, p. 136-137)

As principais características do modelo democrático são, primeiro, que a avaliação deve ser um serviço independente e desinteressado e que o teste de avaliação do patrocinador não concede

<sup>1</sup> Para saber mais sobre MacDonald: <https://www.uea.ac.uk/education/research/care/news-and-reports/barry-macdonald>. Acesso em: set. 2019.

nenhuma reivindicação especial ao avaliador. Em segundo lugar, os avaliadores devem reconhecer e **valorizar explicitamente o pluralismo** e procurar representar uma gama de interesses e valores em seu trabalho. Em terceiro lugar, as decisões sobre o foco, conteúdo e relatórios de avaliação devem prosseguir com base em **negociações com patrocinadores, participantes e outras partes interessadas**. Quarto, os avaliadores devem oferecer confidencialidade aos informantes e dar a eles o controle sobre as informações de uso que eles fornecem. Quinto, os relatórios de avaliação devem ser não recomendáveis e acessíveis a audiências não especializadas. Em sexto lugar, o critério de sucesso em uma avaliação democrática deve ser o leque de públicos atendidos (Tradução nossa e grifo nosso).

Assim, observamos que seria possível diminuir o descompasso, quando se propõe que os avaliadores, que no caso brasileiro, em geral, são os governos (secretarias educacionais), possam se preocupar também em como será apresentado determinado resultado. Deveria haver um combate diário para que os resultados fossem divulgados com maiores detalhes e que houvessem mais explicações em relação a elaboração da prova e cálculo dos números apresentados. Além disso, particularmente no que tange a segunda e terceira proposta do modelo democrático, que revelam o desafio da pluralidade e das negociações, a participação dos professores das escolas envolvidas contribuiria neste sentido, realizando uma conexão mais fidedigna com aquilo que os professores lecionam e, por vezes, dão mais relevância no processo de ensino. Favorecendo assim, um processo e por conseguinte uma avaliação mais democrática.

Os esclarecimentos de MacDonald podem, por um lado, parecer complexos e até mesmo defasados, porém, por outro, há aspectos relacionados a falta de participação de mais agentes educacionais nas discussões de seleção das prioridades a serem avaliadas, entre outras partes do processos que, se investidos de forma criteriosa, poderia contribuir para diminuir o distanciamento dos professores da avaliação externa, seus resultados e potencialidades.

## A TEORIA DE SAMUEL MESSICK

Messick discutia sobre o conceito de validade, conceito que já vinha sendo muito discutido na área da psicologia devido a sua grande importância em relação a determinados testes, pois é a partir deste elemento que poderia dar um entendimento mais preciso do que estava sendo medido. Em outras palavras, é uma condição *sine qua non* demonstrar a validade dos instrumentos. Pasquali (2003, p. 159), a esse respeito, comenta que

Isto é particularmente o caso nos enfoques que, em Psicologia, trabalham com o conceito de traço latente, onde se deve demonstrar a correspondência (congruência) entre traço latente e sua representação física (o comportamento). Não causa estranheza, portanto, que o problema da validade tenha tido, na história da Psicologia, uma posição central na teoria da medida, constituindo-se, na verdade, no seu parâmetro fundamental e indispensável.

Entendendo, de maneira geral, que traço latente seria a manifestação através de algum procedimento – responder a um item, por exemplo; e a partir desta manifestação inferir ou constatar o que se quer medir, destaca-se a importância e preocupação em buscar compreender se de fato corresponde aquilo que se mede com o que se quer medir através da sua validade, tornando-se imprescindível e incontornável a sua presença no processo.

Existem relevantes aspectos a serem considerados em relação a validade e até mesmo quanto as suas diferentes possibilidades de utilização em um teste. No entanto, esta visão apresentada, por muitos autores considerada tradicional, contempla somente aspectos que podemos generalizar como “psicométricos”.

A principal crítica de Messick (1989) era o fato do conceito de validade ser elaborado por especialista em medidas, focados numa visão psicométrica, ser fragmentado e incompleto, pois do modo como o conceito era estabelecido, não levava em consideração as implicações de valor do significado dos resultados e, principalmente, as consequências sociais oriundas dos usos (ou não) desses resultados, quer dizer, pelo fato da análise e uso dos resultados envolver uma prática social, afirmava que a dimensão social e política deveriam estar presentes. Para Messick (apud TOFFOLI et al., 2016, p. 346), “validade

é um julgamento avaliativo integrado do grau em que as evidências empíricas e teóricas sustentam a adequação e a qualidade das inferências e ações com base nos resultados de testes ou em outros meios de avaliação”. Ou seja, para Messick (1989), a validade pressupõe um julgamento que considera as ações baseadas nos resultados dos testes, nas explicações teóricas e nas evidências empíricas que confirmam uma relação direta com as interpretações realizadas. Portanto, a validação se daria também pelo modo como os resultados são apresentados para os potenciais usuários. Messick (1994, p. 16) comenta ainda que,

O novo conceito unificado de validade inter-relaciona essas questões como aspectos fundamentais de uma teoria mais abrangente da validade de construto, abordando tanto o significado da pontuação quanto os valores sociais na interpretação dos testes e no uso do teste. Ou seja, a validade unificada integra considerações de conteúdo, critérios e consequências em uma estrutura de construção para testar empiricamente hipóteses racionais sobre o significado da pontuação e as relações teoricamente relevantes, incluindo aquelas de natureza aplicada e científica (Tradução nossa).

Assim, o autor reforça a importância da compreensão e apropriação dos resultados, a partir da necessidade de que o mesmo tenha significado, ou seja, faça sentido na realidade de cada usuário, ultrapassando assim as fronteiras da mera responsabilização por melhores resultados ou mais eficientes.

Scaramucci (2011) elucida as modificações e visões mais modernas em relação ao conceito de validade, após as discussões disparadas por Messick, como consta na Tabela 1.

TABELA 1—AS NOVAS VISÕES EM RELAÇÃO AO CONCEITO DE VALIDADE APÓS MESSICK

Validade era considerada uma <i>característica de um teste</i> : a medida em que mede aquilo que pretende medir.	Validade é considerada um <i>argumento</i> relativo à interpretação e uso: a medida em que as interpretações e usos de um teste podem ser justificados.
Confiabilidade era vista como distinta e uma <i>condição necessária para a validade</i> .	Confiabilidade pode ser vista como um <i>tipo de evidência de validade</i> .
A validade era frequentemente estabelecida através de <i>correlações</i> de um teste com outros.	Validade é argumentada com base em um número de tipos de <i>justificativas</i> e evidências, incluindo as consequências da avaliação.
Validade de construto era vista como um dos <i>três tipos de validade</i> (conteúdo, relacionada a critério e construto).	Validade é um <i>conceito unitário</i> , em que a validade de construto ocupa uma posição central (validade de conteúdo relativa a critério podem ser usadas como evidência da validade de construto).
O estabelecimento da validade era uma tarefa de responsabilidade de <i>pesquisadores da avaliação</i> , responsáveis pelo desenvolvimento de testes de grande escala e alta relevância.	A justificativa de validade de um teste é de responsabilidade de <i>todos os usuários de um teste</i> .

FONTE: CHAPPELLE (1999, p. 258, APUD SCARAMUCCI, 2011, p. 111)

Como podemos observar, há um reposicionamento do conceito de validade, que depende de mais pessoas (usuários), tendo como pauta a necessidade de um investimento muito maior para que realmente seja possível realizar interpretações tangíveis para efetivos usos e, por outro lado, parece se distanciar das medidas psicométricas.

Acreditamos que as contribuições de Messick são muito valiosas, considerando a situação atual das avaliações em geral. Entretanto, entendemos que esta nova proposta de validade deve ser concatenada com o conceito anterior, numa visão mais complementar, pois avistamos que o tratamento dos dados não podem ser simplificados. Por outro lado, para que haja a validade, oriunda da confiabilidade por parte dos professores, que de alguma forma subsidiam todo este projeto, mas que, em geral ficam apartados de todo o processo, seria importante que os professores participassem de diferentes partes do processo de construção de um sistema de avaliação, tendo em vista ser nestes momentos em que se consolida grande parte do objeto que se deseja avaliar ao mesmo tempo em que se elabora os critérios.

Enfim, acreditamos que estas propostas contrapõem-se à noção de responsabilização, na busca de maiores aprofundamentos e, ao mesmo tempo, participação de um grupo de professores que possam

validar o processo, tornando possível uma efetiva apropriação e contribuição dos resultados da avaliação externa para a melhoria da qualidade da educação.

## UM MODELO ALTERNATIVO: A PARTICIPAÇÃO DOS PROFESSORES NO PROCESSO

Consideramos que a necessidade desse modelo alternativo decorre de limitações identificadas nos modelos disseminados que, por sua vez, tem restringido tanto a utilização de seus resultados quanto tem apresentado restrições no tocante à aceitação por parte de docentes, entre outras. No que diz respeito à possibilidade temos dois desafios. O primeiro se refere à própria formulação. O segundo, quanto à sua viabilização em escolas e redes de escolas.

Para que um instrumento atenda aos critérios técnicos mínimos é necessário garantir etapas em seu processo de construção, sendo elas: 1) Organização de **Matriz de referência** que descreva cada competência avaliada; 2) Elaboração, revisão e aprovação de itens para constituição de **Banco de Itens**; 3) Realização de **Pré-Teste** com itens aprovados, com relatórios de aplicação; 4) **Análises estatísticas**, pela TCT ou TRI, e **pedagógicas** dos itens do Pré-Teste; 5) **Construção de provas e cadernos de provas** de itens aprovados no Pré-Teste; 6) **Aplicação de provas**, com ou sem Blocos Incompletos Balanceados (BIB); 7) **Análises estatísticas**, pela TCT ou TRI, e **pedagógicas** de itens e respostas; 8) Atualização do Banco de Itens; 9) Definição de **escala de medição** e delimitação de **pontos de corte dos níveis**; 10) Redação de **interpretação pedagógica** dos níveis de resultados; 11) Estabelecimento de **critérios** para avaliação dos resultados e **proficiências**; 12) Conclusão da **avaliação**; e 13) Divulgação de **relatórios pedagógicos** dos resultados.

Tais etapas, mesmo que realizadas primorosamente pelas equipes técnicas, não são de conhecimento dos docentes, coordenadores pedagógicos e gestores, e o desenvolvimento destas etapas com o envolvimento dos profissionais da educação, sobretudo dos professores e coordenadores pedagógicos, além de se constituir como um processo formativo, aproximaria os profissionais que atuam nas escolas dos instrumentos, diminuindo a resistência, qualificando-os para a interpretação dos dados divulgados e validando os resultados.

Portanto, como um movimento inicial de elaboração teórico-prescritiva, apontamos que um modelo alternativo de avaliação externa da aprendizagem e em larga escala deve:

- 1- incorporar professores na:
  - a. discussão das finalidades da avaliação;
  - b. delimitação dos objetos de avaliação;
  - c. elaboração e revisão de itens;
  - d. análise dos resultados de aplicação dos itens e dos instrumentos;
  - e. interpretação pedagógica de resultados como referência para os níveis de proficiência aos quais estão associados os resultados; e
  - f. definição dos pontos de corte dos critérios de avaliação e dos próprios termos e expressões da avaliação – do julgamento de seus objetos.
- 2- se constituir num processo de formação docente em avaliação educacional; e
- 3- ser ponto de apoio para as avaliações internas, mediante:
  - a. resultados com interpretação pedagógica;
  - b. instrumentos padronizados;
  - c. formatos de itens;
  - d. cotejamento desses resultados com as notas atribuídas pelos docentes; e
  - e. discussões das matrizes e interpretação pedagógica de resultados com os planos de ensino.

Isso, sendo consubstanciado, superaria restrições, encontradas em várias iniciativas nacionais ou internacionais, que acabam por dificultar essa incorporação e, entre outras consequências, não ter seus resultados devidamente utilizados como um referencial para os processos de ensino articulando a avaliação interna com a externa ou mesmo serem ignorados ou contestados por equipes docentes.

Este modelo alternativo tem, também, como preocupação essencial a formação dos professores em avaliação educacional, em geral, e em avaliação da aprendizagem, em particular, compreendo que os modelos em voga revelam uma concepção pela qual bastariam a divulgação dos dados (BECHER; JUSTO, 2020, entre outros), ignorando a efetiva participação em outras etapas do processo, indubitavelmente, mais relevantes, como indicados anteriormente. De algum modo, esta estratégia favorece uma avaliação democrática, proposto por MacDonald e com validação dos professores, como propõe Messick, influenciando as expectativas em relação ao papel da avaliação externa da aprendizagem no contexto da sala de aula.

## EXEMPLO DE MOÇAMBIQUE

Na contramão dos modelos utilizados nas avaliações externas de diferentes países, mediante o envolvimento de docentes dos Institutos de Formação (IdF), a Repartição de Monitoria e Avaliação da Qualidade (RAQ) do Misau, no âmbito do “Projecto para o Fortalecimento das Capacidades Pedagógicas e Técnicas dos Professores dos Institutos de Formação em Saúde de Moçambique” (ProForsa-Moçambique), com financiamento da Japan International Agency for Cooperation (Jica) e em parceria com a Faculdade de Educação da Universidade de São Paulo (Feusp), organizou os Exames Nacionais de Cursos Técnicos na Área da Saúde. Para viabilizá-los nessa perspectiva, a participação docente deu-se em várias etapas da organização dos Exames, configurando-se, também, como formação na área de avaliação educacional que visam capacitar essa participação.

Essa formação em avaliação, no desenvolvimento de uma avaliação externa em larga escala da aprendizagem, surge da necessidade de formação dos docentes, da equipe técnica e do levantamento de informações sobre os conhecimentos adquiridos pelos alunos durante o período de formação como forma de subsidiar o Misau e as IdF. Sobretudo, pelo perfil dos docentes dos cursos técnicos na área da saúde em Moçambique, uma vez que a formação, majoritariamente, se dá em nível técnico e na área específica de formação, não tendo os profissionais formação mínima em didática ou educação para atuarem como docentes, e tendo como critério de seleção o tempo de experiência ou destaque na atuação.

Esse envolvimento se deu tanto na participação de professores na construção da avaliação externa, quanto na formação em avaliação educacional na perspectiva de isso poderia ser desdobrado no aperfeiçoamento da avaliação da aprendizagem conduzida pelos professores em suas escolas, mediante o que denominamos diálogos entre avaliações internas e externas.

Simultaneamente à formação dos técnicos RAQ, iniciaram-se, em 2013, as oficinas de elaboração e revisão de itens com professores das IdF, especificamente para os Cursos Técnicos de Farmácia, Laboratório e Medicina Preventiva e Saneamento do Meio. Em 2015 foram incorporados a esse processo os Cursos Técnicos de Enfermagem Geral, Enfermagem de Saúde Materno-Infantil e Medicina Geral. Essas oficinas permitiram dar início ao processo de constituição de uma avaliação externa e em larga escala das aprendizagens, tendo como objetivo oferecer subsídios para que docentes, escolas e sistema de formação profissional em Saúde possam redirecionar práticas, reestruturar projetos pedagógicos e redefinir políticas públicas, voltados para a igualdade de oportunidades educacionais e a qualidade do ensino ofertado.

## CONSIDERAÇÕES FINAIS

As ponderações até aqui elencadas nos permitem sustentar que um movimento teórico se faz necessário no terreno da avaliação educacional para, por mais paradoxal que possa parecer, “livrá-la” de políticas que ou não garantem a avaliação, pois que não vão além da medida, ou a conduzem para avaliar outros objetos que não a aprendizagem. A defesa de uma avaliação externa em larga escala da

aprendizagem, com toda a gama de desafios técnicos e políticos que nisso estão implicados, especialmente a delimitação dos critérios, decorre da necessidade de valorizarmos aquilo que o objetivo central do trabalho escolar: as aprendizagens visadas para seus alunos.

Ao mesmo tempo, indissolúvelmente, isso precisa ser fruto de uma política que, sem contrapor avaliação externa à interna, na perspectiva preconizada por Nevo (1998) do diálogo entre essas modalidades, a impulse fazendo dela um processo de formação dos professores em avaliação educacional. Por sua vez, se a avaliação não substitui o ensino, menos ainda deva condicioná-lo, sem ela os professores, que dão aulas para que os alunos aprendam, não podem saber o quanto os alunos aprenderam e se esse “quanto” é – eis o núcleo conceitual de avaliação –, suficiente, adequado ou qualquer outra expressão de julgamento. Um julgamento que seja válido, por dizer respeito ao que de fato se deva avaliar, e fidedigno, por não estar solidamente ancorado em bom levantamento de informações sobre a aprendizagem de cada aluno. Isso não garante mecanicamente o sucesso de todos os alunos, mas sem isso não se poderá trabalhar para tal objetivo.

## REFERÊNCIAS

- AFONSO, A. J. **Avaliação educacional**: regulação e emancipação: para uma sociologia das políticas avaliativas contemporâneas. 4. ed. São Paulo: Cortez, 2009. [Original 2000]
- ALAVARSE, O. M. Desafios da avaliação educacional: ensino e aprendizagem como objetos de avaliação para a igualdade de resultados. **Cadernos Cenpec**, São Paulo, v. 3, n. 1, p. 135-153, jun. 2013.
- ALAVARSE, O. M.; BRAVO, M. H.; MACHADO, C. Avaliações externas e qualidade na educação básica: articulações e tendências. **Estudos em Avaliação Educacional**, São Paulo, v. 24, n. 54, p. 12-31, jan./abr. 2013.
- BARRERA, S. Políticas educacionais, qualidade da educação na América Latina e aporte da avaliação como ferramenta de gerenciamento social. In: CONFERÊNCIA INTEGRAÇÃO E CONVERGÊNCIA ENTRE O MERCOSUR E A COMUNIDADE ANDINA DE NAÇÕES, 2000, La Paz. Anais... La Paz: Instituto Internacional de Integração (I.I.I.), 2000.
- BAUER, Adriana. Usos dos resultados das avaliações de sistemas educacionais: iniciativas em curso em alguns países da América. **Revista Brasileira de Estudos Pedagógicos**, Brasília, v. 91, n. 228, p. 315-344, maio/ago. 2010.
- BAUER, A. É possível relacionar avaliação discente e formação de professores? **Educação em Revista**, Belo Horizonte, v. 28, n. 2, p. 61-82, jun. 2012.
- BAUER, Adriana. Limites, desafios e possibilidades das avaliações de sistemas educacionais: contribuições do Ciclo de Debates para as políticas de avaliação. In: \_\_\_\_\_; GATTI, Bernadete A.; TAVARES, Marialva R. (Org.). Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: implicações nas redes de ensino, no currículo e na formação de professores. Florianópolis: Insular, 2013. p. 281-294.
- BAUER, A.; ALAVARSE, O. M.; OLIVEIRA, R. P. Avaliações em larga escala: uma sistematização do debate. **Educação e Pesquisa**, São Paulo, v. 41, n. especial, p. 1367-1382, dez. 2015.
- BECHER, E. L.; JUSTO, J. C. R. As avaliações em larga escala na formação de professores de Matemática e supervisores de um município da região metropolitana de Porto Alegre. **REMAT: Revista Eletrônica da Matemática**, Bento Gonçalves, v. 6, n. 1, p. 1-13, abr. 2020.
- BECKER, F. R. Avaliação educacional em larga escala: a experiência brasileira. **Revista Iberoamericana de Educación**, v. 53, p. 1-11, 25 jun. 2010.
- BLOOM, B. S.; HASTINGS, J. T.; MADAUS, G. F. **Evaluación del aprendizaje**. Buenos Aires: Troquel, 1975.

BONAMINO, A.; SOUSA, S. Z. L. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola. **Educação e Pesquisa**, São Paulo, v. 38, n. 2, abr./jun. 2012.

BRAVO, M. H. de A. **Enem e o percurso histórico do conceito de avaliação**: implicações das e para as políticas educacionais. Dissertação (Mestrado em Educação) – Faculdade de Educação, Universidade de São Paulo. São Paulo, 2017.

BROOKE, N.; CUNHA, M. A. A. C. A avaliação externa como instrumento da gestão educacional nos estados. **Estudos & Pesquisas Educacionais**, São Paulo, n. 2, p. 17-79, nov. 2011.

CHAPPAZ, R. de O. A Prova São Paulo e as tensões das avaliações externas: diálogos com o currículo oficial da rede municipal de ensino de São Paulo. Dissertação (Mestrado em Educação) – Faculdade de Educação, Universidade de São Paulo. São Paulo, 2015.

CASTILLO ARREDONDO, S.; CABRERIZO DIAGO, J. **Avaliação educacional e promoção escolar**. Tradução de Sandra Martha Dolinsky. Curitiba: Ibpex; São Paulo: Unesp, 2009.

CASTRO, M. H. G. A consolidação da política de avaliação da educação básica no Brasil. Meta: avaliação, Rio de Janeiro, v. 1, n. 3, p. 271-296, set./dez. 2009.

EVERS, W. M.; WALBERG, H. J. (Ed.). *School accountability*. Stanford, CA: Hoover Institution Press, 2002.

FRANCO, C. O SAEB – Sistema de Avaliação da Educação Básica: potencialidades, problemas e desafios. **Revista Brasileira de Educação**, n. 17, p. 127-133, 2001.

FREITAS, L. C de. Caminhos da avaliação de sistemas educacionais no Brasil: o embate entre a cultura da auditoria e a cultura da avaliação. In: BAUER, A.; GATTI, B. A.; TAVARES, M. R. (Org.). *Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: implicações nas redes de ensino, no currículo e na formação de professores*. Florianópolis: Insular, 2013. p. 147-176.

FREITAS, P. F. **Formação docente em avaliação educacional**: lacunas, consequências e desafios. Tese (Doutorado em Educação) – Faculdade de Educação da Universidade de São Paulo. São Paulo, 2019.

GATTI, B. **Avaliação e qualidade da educação**. Trabalho apresentado no Simpósio Brasileiro de Política e Administração da Educação. Porto Alegre, 2007.

GREENE, J. E. Democracy and social change. In: SHAW, I.; GREENE, J; MARK, M. (Ed.). **The handbook evaluation**. Newcastle; Sage, 2006.

HAYDT, R. C. C. Avaliação do processo ensino-aprendizagem. 6. ed. São Paulo: Ática, 2004. (Série Educação). [Original 1988]

HAYDT, R. C. C. Curso de didática geral. 8. ed. São Paulo: Ática, 2006. 327 p. (Educação).

KELLAGHAN, T.; GREANEY, V. **Assessing student learning in Africa**. Washington, DC: World Bank Documents, 2004.

KLEIN, R.; FONTANIVE, N.S.F. Avaliação em larga escala: uma proposta inovadora. **Em Aberto**, Brasília, ano 15, n. 66, p. 29-34, abr./jun. 1995.

LAFOURCADE, P. D. **Evaluación de los aprendizajes**. Buenos Aires: Kapelusz, 1969.

LOPES JUNIOR, J. Avaliação em larga escala como eixo da pesquisa realizada pela escola e pela Pós-Graduação: sobre as necessárias revisões e as urgentes prospecções. In: CARVALHO, L. M. O.; CARVALHO, W. L. P.; LOPES JUNIOR, J. (Org.). **Formação de professores, questões sociocientíficas e avaliação em larga escala**: aproximando a Pós-Graduação da escola. São Paulo: Escrituras, 2016. (Educação para a Ciência). p. 253-275.

LUCKESI, C. C. **Avaliação em educação**: questões epistemológicas e práticas. São Paulo: Cortez, 2018.

- LUKAS MUJIK, J. F.; SANTIAGO ETXEBERRIA, K. Concepto, componentes y fases de la evaluación. In: LUKAS MUJIK, J. F.; SANTIAGO ETXEBERRIA, K. **Evaluación educativa**. 2. ed. Madrid: Alianza, 2009. p. 87-126.
- MACDONALD, B. Evaluation and control of education. In: MACDONALD, B.; WALKER, R. (Ed.). **Innovation, evaluation, research and problem of control: some interim papers**. Norwich: University of East Anglia, Centre for Applied Research in Education, 1974. p. 36-48.
- MACHADO, C.; ALAVARSE, O. M. Qualidade das escolas: tensões e potencialidades das avaliações externas. **Educação & Realidade**, Porto Alegre, v. 39, n. 2, p. 413-436, abr./jun. 2014.
- MADAUS, G. F. The influence of testing on the curriculum. In: TANNER, L. N. (Ed.). **Critical issues in curriculum: Eighty-seventh Yearbook of the National Society for the Study of Education**. Part I. Chicago, IL: University of Chicago Press, 1988. p. 83-121.
- MADAUS, G. F.; RUSSELL, M.; HIGGINS, J. **The paradoxes of high stakes testing: how they affect students, their parents, teachers, principals, schools and society**. Charlotte, NC: Information Age, 2009.
- MESSICK, S. Validity. In: LINN, R. L. (Ed.). **Educational measurement**. 3rd ed. Washington, DC: National Council on Measurement in Education: American Council on Education, 1989. p. 13-103.
- MESSICK, S. **Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning**. Princeton, NJ: Educational Testing Service, Sept. 1994. 33 p. (Research Report RR-94-45).
- MIRAS, M.; SOLÉ, I. A evolução da aprendizagem e a evolução no processo de ensino e aprendizagem. In: COLL, César; PALACIOS, Jesús; MARCHESI, Álvaro (Org.). **Desenvolvimento psicológico e educação: v. 2: psicologia da educação escolar**. Tradução de Angélica Mello Alves. Porto Alegre: Artes Médicas, 1996. p. 374-385. [Original 1993]
- NEVO, D. Avaliação por diálogos: uma contribuição possível para o aprimoramento escolar. In: TIANA, A. (Coord.). **Anais do Seminário Internacional de Avaliação Educacional, 1 a 3 de dezembro de 1997**. Tradução de John Stephen Morris. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais (Inep), 1998. p. 89-97.
- NORRIS, Nigel. Democratic evaluation: the work and ideas of Barry MacDonald. **Evaluation**, v. 21, n. 2, p. 135-142, 2015.
- OLIVEIRA, R. P. de *et al.* Análise das desigualdades intraescolares no Brasil. **Estudos & Pesquisas Educacionais**, São Paulo, n. 4, p. 19-112, nov. 2013.
- OLIVEIRA, A. A. A política de avaliação externa na Secretaria Municipal de Educação do Rio de Janeiro: implicações para práticas docentes. Dissertação (Mestrado em Educação) - Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2015.
- PASQUALI, L. **Psicometria: teoria dos testes na psicologia e na educação**. Petrópolis, RJ: Vozes, 2003.
- SCARAMUCCI, M. V. R. Validade e consequências sociais das avaliações em contexto de ensino de línguas. **Lingvarvm Arena**, Porto, v. 2, p. 103-120, 2011.
- SCRIVEN, M. The methodology of evaluation. In: TYLER, Ralph Winfred; GAGNE, Robert; SCRIVEN, Michael (Ed.) **Perspectives of curriculum evaluation**. Chicago, IL: Rand McNally, 1967. (AERA Monograph Series on Curriculum Evaluation). p. 39-83.
- SILVA, F. A. **O Saresp e o professor de Física: contribuições, sentidos e perspectivas**. Dissertação (Mestrado) – Programa de Pós-Graduação Interunidades em Ensino de Ciências, Universidade de São Paulo. São Paulo, 2015.
- SOUSA, S. M. Z. L. Avaliação, ciclos e qualidade do ensino fundamental: uma relação a ser construída. **Estudos Avançados**, São Paulo, v. 21, n. 60, p. 27-44, ago. 2007.
- SOUSA, S. M. Z. L.; OLIVEIRA, R. P. de. Políticas de avaliação da educação e quase mercado no Brasil. **Educação & Sociedade**, Campinas, v. 24, n. 84, p. 873-895, set. 2003.

SOUSA, S. M. Z. L.; OLIVEIRA, R. P. **Sistemas de avaliação educacional no Brasil**: característica, tendências e uso dos resultados: relatório final. São Paulo: Faculdade de Educação da Universidade de São Paulo, 2007.

SOUSA, S. M. Z. L.; OLIVEIRA, R. P. Sistemas estaduais de avaliação: uso dos resultados, implicações e tendências. **Cadernos de Pesquisa**, São Paulo, v. 40, n. 141, p. 793-822, set./dez. 2010.

VIANNA, H. M. **Introdução à avaliação educacional**. São Paulo: Ibrasa, 1989. (Biblioteca Educação, 40).

VIANNA, H. M. **Avaliações nacionais em larga escala**: análises e propostas. São Paulo: Fundação Carlos Chagas / DPE, 2003. 41 p. (Textos FCC, 23).

WERLE, F. O. C. Políticas de avaliação em larga escala na educação básica: do controle de resultados à intervenção nos processos de operacionalização do ensino. **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v. 19, n. 73, p. 769-792, out./dez. 2011.