UNIVERSIDADE DE SÃO PAULO
FACULDADE DE EDUCAÇÃO

MICHELLY FRANCINI BRASSAROTO DO AMARAL

Avaliações externas e formação docente em avaliação educacional: limites e perspectivas

São Paulo
2022

MICHELLY FRANCINI BRASSAROTO DO AMARAL

Avaliações externas e formação docente em avaliação educacional: limites e perspectivas

Dissertação apresentada ao Programa de Pós-Graduação em Educação da Faculdade de Educação da Universidade de São Paulo como requisito para obtenção do título de Mestre em Educação.

Área de Concentração: Estado, Sociedade e Educação

Orientador: Prof. Dr. Ocimar Munhoz Alavarse

São Paulo
2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catalogação da Publicação

Ficha elaborada pelo Sistema de Geração Automática a partir de dados fornecidos pelo(a) autor(a)

Bibliotecária da FE/USP: Nicolly Soares Leite – CRB-8/8204

Aa

Amaral, Michelly Francini Brassaroto do
Avaliações externas e formação docente em avaliação educacional: limites e perspectivas / Michelly Francini Brassaroto do Amaral; orientador Ocimar Munhoz Alavarse. -- São Paulo, 2022.
75 p.

Dissertação (Mestrado - Programa de Pós-Graduação Estado, Sociedade e Educação) -- Faculdade de Educação, Universidade de São Paulo, 2022.

1. Avaliação da aprendizagem. 2. Avaliação externa. 3. Política de avaliação. 4. Formação docente em avaliação educacional. I. Alavarse, Ocimar Munhoz , orient. II. Título.

AMARAL, Michelly Francini Brassaroto do

Título: Avaliações externas e formação docente em avaliação educacional: limites e perspectivas.

Dissertação apresentada ao Programa de Pós-Graduação em Educação da Faculdade de Educação da Universidade de São Paulo como requisito para obtenção do título de Mestre em Educação.

Aprovado em: _____

Banca examinadora

Prof. Dr.: _____ Instituição: _____
Julgamento: _____ Assinatura: _____

Prof. Dr.: _____ Instituição: _____
Julgamento: _____ Assinatura: _____

Prof. Dr.: _____ Instituição: _____
Julgamento: _____ Assinatura: _____

DEDICATÓRIA

Ao Everton, meu companheiro, pelo suporte, compreensão e apoio incondicional.
Aos meus avós, Damiana e Anésio (*in memorian*), pelo cuidado, afeto, incentivo nos estudos e por sempre acreditarem em mim.

AGRADECIMENTOS

Ao Prof. Dr. Ocimar Munhoz Alavarse, meu orientador, agradeço o apoio, a partilha dos saberes e todo suporte durante o percurso que foram fundamentais para o desenvolvimento desta pesquisa. Acima de tudo, por estimular meu interesse pelo conhecimento e pelo acolhimento que tornou possível conciliar formação acadêmica, desenvolvimento profissional e vida familiar.

À Profª. Drª. Gladys Rocha e à Profª. Drª Érica Maria Toledo Catalani, pelas leituras críticas e recomendações em meu exame de qualificação e pela composição da mesa de minha defesa. Suas contribuições possibilitaram um rico debate para consolidação da pesquisa apresentada nesta dissertação.

Aos amigos e amigas do Grupos de Estudos e Pesquisas em Avaliação Educacional (Gepave), pela parceria e compartilhamento de saberes e experiências tornando a pesquisa cientifica um trabalho verdadeiramente coletivo.

Ao Everton, meu companheiro, pelo cuidado, afeto, compreensão na ausência e todo o suporte necessário nos estudos e na vida.

RESUMO

AMARAL, Michelly Francini Brassaroto do. **Avaliações externas e formação docente em avaliação educacional**: limites e perspectivas. 73 p. Dissertação (Mestrado em Educação) – Faculdade de Educação da Universidade de São Paulo. São Paulo, 2022.

Nesta dissertação, as avaliações externas em larga escala e os usos de seus resultados são problematizados em face à disseminação dessas avaliações nas últimas décadas, apontadas como capazes de auxiliar na tomada de decisões com vistas à melhoria da qualidade da educação, considerando que esta qualidade se expressa, total ou parcialmente, no desempenho dos estudantes em testes padronizados. Longe de um consenso, tal movimento foi 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. Considerando a sua complexidade, no escopo de políticas educacionais, com seus múltiplos determinantes e consequências, apresentamos considerações que, sem exaurir o debate, podem ser alternativas às características encontradas no quadro atual dessas avaliações, especialmente diante do desafio de incorporar professores e, ao mesmo tempo, constituir-se num processo de formação em avaliação educacional como forma de minimizar ou superar determinados questionamentos presentes na literatura. Para isso recorreu-se ao trabalho de pesquisa bibliográfica, recuperando as contribuições de ordem histórica, sociológica, pedagógica e de política educacional, para traçar um panorama geral das avaliações em larga escala, cotejando com experiências do Grupo de Estudos e Pesquisas em Avaliação Educacional (Gepave) que buscaram romper com a estrutura tradicionalmente utilizada sem a aproximação e participação dos professores que estão em sala de aula. Aponta-se possibilidades para favorecer o uso dos resultados mediante o envolvimento de docentes em várias etapas da organização dos exames, configurando-se, também, como formação na área de avaliação educacional.

Palavras-chave: Avaliação da aprendizagem, Avaliação externa, Avaliação em larga escala, Formação docente em avaliação educacional.

ABSTRACT

AMARAL, Michelly Francini Brassaroto do. External assessment and teacher education in educational assessment. 73 p. Thesis (Masters) – Faculdade de Educação da Universidade de São Paulo. São Paulo, 2022.

In this study, the external evaluations in large-scale and the uses and their results are problematized in view of the dissemination of these evaluations in recent decades, identified as capable of assisting in decision-making with a view to improving the quality of education, considering that this quality expressed, in whole or in part, in the performance of students on standardized tests. Far from a consensus, this movement was accompanied by debates and questions about the pertinence of these assessments to the educational process, in addition to their political and pedagogical consequences, especially for pedagogical work in educational units. Considering its complexity and its connection with educational policies, with its multiple determinants and consequences, we present considerations that, without exhausting the debate, can be alternatives to the characteristics found in the current framework of these evaluations, especially in the face of the challenge of incorporating teachers and, at the same time, to constitute a process of formation in educational evaluation as a way to minimize or overcome certain questions present in the literature. For this purpose, bibliographic research was used, recovering the historical, sociological, pedagogical and educational policy contributions to draw an overview of large-scale assessments, comparing with the experiences of the Group of Studies and Research in Educational Assessment (Gepave) who sought to break with the structure traditionally used without the approach and participation of teachers who are in the classroom. It points to possibilities to favor the use of the results through the involvement of professors in several stages of the organization of the exams, configuring, also, as formation in the area of educational evaluation.

Keyword: Learning assessment, External assessment, Large scale assessment, Teacher education in educational evaluation.

LISTA DE SIGLAS

ANA	Avaliação Nacional da Alfabetização
Aneb	Avaliação Nacional da Educação Básica
Anresc	Avaliação Nacional do Rendimento Escolar
BIB	Blocos Incompletos Balanceados
BNCC	Base Nacional Comum Curricular
CEFAI	Centro de Formação e Acompanhamento à Inclusão
DNFPS	Direção Nacional de Formação de Profissionais da Saúde
DOT-P	Diretorias de Orientação Técnico- Pedagógica
DRE	Diretoria Regional de Educação
ENEM	Exame Nacional do Ensino Médio
Feusp	Faculdade de Educação da Universidade de São Paulo
Gepave	Grupo de Estudos e Pesquisas em Avaliação Educacional
Ideb	Índice de Desenvolvimento da Educação Básica
IdF	Institutos de Formação
Inep	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
Jica	Japan International Agency for Cooperation
Misau	Ministério da Saúde de Moçambique
NAE	Núcleo de Avaliação Educacional
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
Pisa	Programa Internacional de Avaliação de Estudantes
PNDRH	Plano Nacional de Desenvolvimento de Recursos Humanos
Proforsa	Projecto para o Fortalecimento das Capacidades Pedagógicas e Técnicas dos Professores dos Institutos de Formação em Saúde
RAQ	Repartição de Monitoria e Avaliação da Qualidade
RME-SP	Rede Municipal de Ensino de São Paulo
SAAI	Sala de Apoio e Acompanhamento à Inclusão
Saeb	Sistema de Avaliação da Educação Básica
Saresp	Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo
SME-SP	Secretaria Municipal Educação de São Paulo
TCT	Teoria Clássica dos Testes
TIMSS	Tendências no Estudo Internacional sobre Matemática e Ciências
TRI	Teoria da Resposta ao Item

SUMÁRIO

1. INTRODUÇÃO 11
2. AVALIAÇÃO EDUCACIONAL: DEFINIÇÕES 18
3. AVALIAÇÃO DA APRENDIZAGEM: FUNÇÕES 22
4. AVALIAÇÃO EXTERNA EM LARGA ESCALA 25
4.1 Avaliações externas: características do PISA e do Saeb 35
4.1.1 Programa Internacional de Avaliação de Estudantes (Pisa) 35
4.1.2 Sistema de Avaliação da Educação Básica (Saeb) 38
5. FORMAÇÃO DOCENTE EM AVALIAÇÃO EDUCACIONAL 40
6. EXPERIÊNCIAS DE AVALIAÇÕES EXTERNAS COM A PARTICIPAÇÃO DE PROFESSORES 43
6.1 A experiência de Moçambique..... 43
6.1.1 Elaboração e validação das matrizes 45
6.1.2 A elaboração e revisão de itens 47
6.1.3 A análise dos resultados de aplicação dos itens e dos instrumentos 48
6.1.4 Polêmica sobre certificação 50
6.2 A experiência na Rede Municipal de Ensino de São Paulo: Prova Mais Educação 54
7. INDICAÇÕES PARA UMA AVALIAÇÃO EXTERNA 57
7.1 A teoria de Barry MacDonald 58
7.2 A teoria de Samuel Messick 60
8. CONSIDERAÇÕES FINAIS 65
REFERÊNCIAS 69

1. INTRODUÇÃO

Neste capítulo apresentaremos o problema de pesquisa, a escolha da metodologia para amparar as análises e proposições e os objetivos do trabalho tendo como objeto um dos aspectos relevantes no debate em torno das avaliações externas: o envolvimento dos docentes nas avaliações externas, seus instrumentos e usos dos resultados.

Nosso problema é abordar, em linhas gerais, as avaliações externas quanto à formação dos professores em avaliação educacional no escopo do envolvimento dos docentes nesse tipo de avaliação. Considera-se que os professores têm lacunas e limitações quanto à capacidade em realizar a avaliação da aprendizagem de seus alunos, por mais paradoxal que isso possa parecer e tem-se a perspectiva de que as avaliações externas poderiam contribuir nessa formação, o que incluiria o uso dos resultados dessas avaliações e de algumas de suas características no processo pedagógico, até mesmo para problematizar as avaliações internas e criar condições para a superação de algumas limitações encontradas nas escolas no que tange à avaliação da aprendizagem de seus alunos. Assim, apesar de reconhecermos muitas restrições nas avaliações externas, como se demonstrará nesta dissertação, sobretudo quanto às crenças que gestores de redes de ensino têm em relação a elas para determinadas indicações e decisões a partir de seus resultados e alguns efeitos no cotidiano escolar, admitimos que possuem um potencial, para além dos usos de seus resultados, em contribuir na formação docente em avaliação educacional, especialmente se forem incorporadas algumas características em seu delineamento, sobretudo, nas etapas preparatórias, e na maneira como os resultados são disponibilizados e discutidos com os docentes que, direta ou indiretamente, têm relação com eles.

A centralidade que a avaliação educacional, notadamente pelo emprego de testes padronizados aplicados em larga escala, vem assumindo no contexto de formulação e implementação de políticas públicas para a educação, tanto no Brasil quanto em vários países, faz com que a avaliação, gradualmente, venha se configurando como recurso estratégico justificado tanto para a produção de evidências a respeito das aprendizagens dos alunos, de escolas e de redes de ensino, quanto para a possível indução e consolidação de novos objetivos educacionais no âmbito de políticas educacionais.

Especialmente, a implementação de avaliações externas e em larga escala, apontadas como capazes de auxiliar na tomada de decisões com vistas à melhoria da qualidade da educação, considerando que esta qualidade se expressa, total ou parcialmente, no desempenho dos estudantes em testes padronizados, longe de um consenso, é um movimento acompanhado

de debates e questionamentos sobre a própria pertinência dessas avaliações ao processo educativo, ademais de suas consequências políticas e pedagógicas, para o trabalho pedagógico nas unidades educacionais, inclusive pelo fato de contrastarem, como se demonstrará neste trabalho, com aquelas avaliações conduzidas por professores no interior de salas de aula e escolas e com uma tradição de mais de cem anos, avaliações denominadas, usualmente, de avaliações internas.

Nas últimas décadas foram feitos esforços na investigação das avaliações externas, considerando, sobretudo, que essas avaliações foram apresentadas por instâncias governamentais e agências de fomento como uma condição para a melhoria da qualidade da educação escolar. No entanto, tem-se debatido sobre a dificuldade em estabelecer relações diretas e inequívocas entre as avaliações externas e a melhoria da qualidade da educação escolar, como alguns gestores de redes defendem para a implementação dessas avaliações externas. Ainda assim, em muitos países e por uma diversidade de razões, políticas, econômicas, culturais e até ideológicas, as avaliações externas ocupam um papel significativo nas políticas públicas de educação, incluindo aqueles que envolvem vários países, como é o caso mais conhecido do Programa Internacional de Avaliação de Estudantes (Pisa), patrocinado pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Possivelmente porque os dados produzidos e os índices dela independentes, como é o caso do Índice de Desenvolvimento da Educação Básica (Ideb) no Brasil, são utilizados para justificar reformas nos sistemas educativos e, em geral, há uma predisposição social para associar as avaliações externas à exigência, ao rigor e, em última análise, à garantia de qualidade do ensino.

Observa-se que a expansão desse tipo de avaliação vem acompanhada da ideia de que boas avaliações significam ou produzem bons diagnósticos da educação que poderiam, entre outros objetivos: 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. Esse quadro é realçado na literatura da área da avaliação educacional, como encontramos, entre outros autores, em Afonso (2000), Bonamino e Sousa (2012), Brooke e Cunha (2011), Sousa e Oliveira (2003, 2010), Vianna (2003) e Werle (2011). Quadro que talvez se justifique porque, em diversos segmentos sociais, inclusive no acadêmico, se considera que as avaliações externas apresentam 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 negativas, 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, aspecto controverso, como apontam Alavarse, Bravo e Machado (2013) e Machado e Alavarse (2014).

Especialmente no Brasil, iniciativas próprias de avaliações externas por estados e municípios foram ganhando força, constituindo um movimento crescente em face daquele de responsabilidade do governo federal. De acordo com Brooke, Cunha e Faleiros (2011, p. 8), essa expansão subnacional das avaliações refletiria a necessidade dos entes federados em controlar, com maior propriedade, as etapas que envolvem o processo de avaliação e, por consequência, sentirem-se instrumentalizados para construir um diagnóstico mais preciso sobre a aprendizagem de seus alunos e elaborar estratégias pedagógicas e de gestão potencialmente mais eficazes. Podemos adicionar a esse quadro a possível resistência dos entes federativos em terem suas redes de ensino avaliadas pelo governo federal sem que esses tivessem acúmulo teórico e/ou competência técnica para apropriarem-se das avaliações centrais e de seus resultados, ou sem que pudessem ter sua própria avaliação para fazer comparações, ademais de buscarem agilizar a produção de resultados, pois, notadamente tendo-se o Sistema de Avaliação da Educação Básica (Saeb) como referência, há um intervalo muito grande entre a aplicação de provas e a divulgação do resultados.

Como uma das justificativas para a existência de uma avaliação externa, pode-se considerar, a despeito de algumas restrições, no contexto da formação profissional docente, a possibilidade de se estabelecer alguns referenciais para essa formação, como são, por exemplo, os resultados de aprendizagem dos alunos nas provas empregadas usualmente nesse tipo de avaliação. Isso, em nossa visão, não autoriza, por exemplo, que resultados daí advindos possam ser utilizados para fins de certificação, sobretudo na Educação Básica, que deve continuar a cargo das equipes docentes nas instituições.

Se a avaliação externa é uma realidade em diversos países, contudo ainda são questionáveis os usos de seus resultados, dado que, em muitas circunstâncias, ou não são efetivamente utilizados, quer por autoridades educacionais, quer por equipes escolares, ou são empregados para estabelecer hierarquização de escolas sem a devida consistência ou sem necessidade política para tal comparação, pois esses resultados, ainda que importantes, não

são capazes de indicar toda a qualidade do trabalho escolar e tampouco devem ser utilizados com o objetivo de hierarquizar escolas. Mesmo quando permitem sólidos diagnósticos, isso por si só não é suficiente para mudar uma determinada situação, inclusive porque os usos de resultados de avaliação, rigorosamente, extrapolam a avaliação em si mesma, sendo um desdobramento do processo avaliativo. Embora tenha-se avançado nas metodologias de coleta e tratamento de dados das avaliações externas, 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.

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 finalidade, abrangência e possibilidade de análise da realidade escolar (CASASSUS, 2009; HORTA NETO, 2010; MACHADO, 2012; MADAUS; RUSSEL; HIGGINS, 2009), na gestão (AFONSO, 1998; BARRERA, 2000; CASTRO, 2009; KLEIN; FONTANIVE, 1995; VIANNA, 2003), no currículo (CHAPPAZ, 2015; MADAUS; RUSSELL; HIGGINS, 2009), no trabalho do professor (EVERS; WALBERG, 2002; MARCHI, 2012; OLIVEIRA, 2015), efeitos colaterais das políticas (NEAL; SCHANZENBACH, 2010; NICHOLS; BERLINDER, 20070; RAVITCH, 2010) e na formação docente (ANDRADE, 2012; BAUER, 2012), além de pesquisas que procuraram compreender os possíveis contrapontos das avaliações externas sobre as políticas públicas educacionais (ALAVARSE, 2013; BAUER, 2010, 2012; FREITAS, 2013; OLIVEIRA, 2013; OLIVEIRA, 2015).

Chama a atenção o fato de que foram encontradas poucas investigações que tivessem as próprias avaliações como objeto de estudo, isto é, que se debruçaram sobre suas características técnicas, incluindo suas limitações nesse tocante.

Se a avaliação externa é considerada como estratégia que permitiria o alcance dos objetivos de melhoria, da eficiência e da qualidade da educação, pois, por meio da interpretação de seus resultados, seria possível o desenvolvimento de ações de intervenção em aspectos específicos dos processos de ensino e de aprendizagem, como, por exemplo, o currículo, com consequências diretas na melhoria da qualidade da oferta da educação. Isso não elimina questionamentos sobre o uso dos resultados das avaliações para expressarem completamente a qualidade dos processos aos quais estariam associados a produção do aprendizado que são objeto de avaliação. Tais questionamentos nos remetem ao debate da

representação do termo “qualidade” e sua relação com as avaliações. Gatti (2007, p. 3) problematiza, por exemplo, a divulgação e uso dos resultados como tradução da qualidade indicando que

> [...] cabe perguntar: qualidade educacional se traduz apenas por esses resultados? Evidentemente que não. A qualidade da educação passa por questões como a existência de uma filosofia educacional e, pela consciência do papel social da educação – não só seu papel instrumental, de utilidade, por exemplo, para o trabalho, mas seu papel para a civilização humana, para a constituição de valores de vida e convivência, seu papel no desenvolvimento de sensibilidades ao outro, ao meio ambiente, às expressões humanas de cultura. Portanto, passa por elementos formativos que transcendem, embora não dispensem de modo algum, a aquisição de conhecimentos apenas.

Casassus (2007, 2009) também discute o papel das avaliações externas por meio de testes, o aumento de pressão sobre os docentes mediante a oferta de bônus ou penalidades com base nos resultados e aponta o uso extensivo dessas avaliações para finalidades políticas e não educacionais. Ao retomar o debate dos anos 1980, momento em que a qualidade da educação passou a fazer parte da agenda dos governos na América Latina, após um período de expansão precária da oferta dos sistemas de ensino, o autor lembra que o discurso acerca da qualidade educacional emergiu no âmbito político sem que houvesse um debate profundo sobre o significado da “qualidade educacional”. Como consequência, a qualidade passou a ser interpretada como “o equivalente a uma pontuação numa prova estandardizada” (CASASSUS, 2009, p. 72).

Para Sousa e Oliveira (2010), as possibilidades de interferência da avaliação na promoção da melhoria do ensino possuem duas ordens de argumentação. A primeira está mais voltada para a lógica interna dos processos avaliativos, nos quais, por meio dos princípios e procedimentos envolvidos na avaliação, é possível explicitar o enfoque e os critérios adotados que podem ser referências para a discussão de potencialidades e limites das práticas implementadas. Na segunda, as contribuições das avaliações estariam em seus resultados, que podem ser utilizados na condução de políticas educacionais, “examinando-se a validade, a relevância e a oportunidade das iniciativas adotadas como consequência do conhecimento obtido mediante o processo de avaliação” (Op. cit., p. 794). Tal potencialidade das avaliações vem sendo declarada em planos e propostas de governo de diversas instâncias e instituições de ensino (SOUSA, 2003).

Neste trabalho abordamos, sobretudo de uma perspectiva teórica, as avaliações externas, com testes aplicados em larga escala, considerando que se constituem em avaliações da aprendizagem, embora em alguns contextos seus resultados são utilizados e apresentados

para avaliar, em última instância, outras dimensões que não exatamente, a aprendizagem. Exploramos, especialmente, como essas avaliações podem contribuir com a formação dos professores em avaliação educacional.

Aportamos, também, sem a pretensão de exaurir as análises, dados empíricos e relatos de algumas experiências de redes que buscaram estratégias que, de certa forma, romperam com características mais comuns de construção de avaliações externas; nesse sentido recuperamos experiências de avaliações externas conduzidas em Moçambique e na Rede Municipal de Ensino de São Paulo (RME-SP) que indicam alternativas em contraste com várias iniciativas de avaliação externa.

Para fundamentar nosso trabalho, discutimos aspectos sobre esse objeto de avaliação – a aprendizagem –, em face, inclusive das várias definições de avaliação educacional na literatura especializada.

Caso o objetivo da educação esteja voltado mais ao desempenho dos alunos nas provas, o ensino se desloca, tendencialmente, para os conteúdos passíveis de comporem essas provas, sejam elas produzidas por professores ou por agentes externos às escolas, podendo ser negligenciados outros conteúdos do trabalho escolar, mesmo que se possa constatar que os conteúdos sejam relevantes. Uma implicação dessa ponderação é o sentido, o significado, que o ensino passa a ter para alunos e professores, podendo, ainda, fazer com que os resultados nas provas assumam um lugar privilegiado, como se fosse uma "palavra final", em detrimento de retomadas e revisões para o sucesso escolar. Com isso, não se está negando o potencial e a necessidade de provas, mas aventando-se sobre as consequências adversas que seu emprego, sobretudo pelo uso dos resultados, pode ocupar no âmbito escolar. Assim, os testes padronizados nas avaliações externas podem, como um risco para o processo educacional, acabar se constituindo, por identificar e classificar as posições cognitivas dos sujeitos, no que foi o uso dos testes de inteligência desenvolvidos no início do século por Binet-Simon para legitimar cada um numa posição considerada "natural", reforçando a função de controle, segmentação social da educação e da exacerbação da "disputa" por melhores posições.

Casassus, ao analisar casos na América Latina, salienta que a ênfase na avaliação e nas políticas de responsabilização dos profissionais da educação significam um obstáculo aos sistemas de ensino, pois produzem o aumento da desigualdade educacional e por consequência social. Essa afirmação, sem negar sua razoabilidade, demanda maiores evidências, dado que não identificamos nesses trabalhos e na literatura dados que correlacionem o uso dos testes e o suposto aumento das desigualdades educacional e social, o que seria de grande contribuição para a área. Ademais, historicamente a oferta de ensino

público é marcada pela falta de recursos e investimentos e se configura como instrumento de estratificação social e controle de parte significativa da população. Se analisarmos por essa perspectiva, a avaliação nas políticas educacionais pode ser concebida como um processo que evidencia as desigualdades educacionais que podem ser produzidas ou reproduzidas pelos sistemas educacionais e não serem consideradas como instrumento de promoção de tais desigualdades.

Assim, partindo das polêmicas e tensões que envolvem as avaliações externas, nossa investigação buscou analisar as avaliações externas em seus delineamentos e no uso de seus resultados ou procedimentos, pelos profissionais de educação que atuam nas escolas, a fim de melhor evidenciar suas características.

Deste modo, colocou-se como questão de pesquisa: **É necessário e possível envolver os professores na construção de uma avaliação externa se constituindo em um processo formativo em avaliação educacional?**

Esse problema de pesquisa tem origem no contexto de um esforço coletivo no escopo das iniciativas do Grupo de Estudos e Pesquisas em Avaliação Educacional (Gepave), da Faculdade de Educação da Universidade de São Paulo (Feusp), coordenado pelo professor Dr. Ocimar Munhoz Alavarse, que congrega pesquisadores e pesquisadoras com disposição e interesse, acadêmico e prático, de pensar suas atividades de pesquisa de forma articulada e em conjunto, promovendo maior possibilidade de abrangência e avanço ao conhecimento. Tomamos como referência duas experiências delineadas no Gepave, o Projeto Proforsa II, desenvolvido em Moçambique, e a Prova Mais Educação, desenvolvida na Rede Municipal de Ensino de São Paulo, que buscaram o envolvimento de professores em suas avaliações externas e em larga escala da aprendizagem.

Considerando a complexidade da construção de uma avaliação externa da aprendizagem em larga escala com a participação de professores enquanto processo de formação e sua vinculação com uma política de avaliação, com seus múltiplos determinantes e consequências, adotamos uma abordagem qualitativa que permita cotejá-lo com outras avaliações externas para delimitar se constituem uma experiência efetivamente diferenciada no cenário internacional.

2. AVALIAÇÃO EDUCACIONAL: DEFINIÇÕES

A compreensão das avaliações externas, quer teórica, quer empiricamente, nos parece exigir, inicialmente, uma delimitação da própria definição de avaliação, que neste trabalho é 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 empreendermos 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 é 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, pode-se verificar que um conjunto de termos ou são tomados como sinônimos ou como componentes intrínsecos ao processo avaliativo. Assim, nos deparamos com associações ao 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 – ou o que poderia ser avaliar – o fazem em contraste com esses verbos, ou pelo menos em relação a 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.

Para Nérici (1977), a avaliação é uma etapa de um processo que incluiria uma verificação prévia sobre algo que se pretende avaliar na realidade, envolvendo a

aprendizagem dos alunos, e a avaliação propriamente dita, seria o processo de ajuizamento, apreciação, julgamento ou valorização do que o educando revelou ter aprendido durante um período de estudo ou de desenvolvimento do processo ensino e aprendizagem.

Para Vianna (1989, p. 10), "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, mas esta, 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”.

Haydt (1997, p. 10), nesse escopo conceitual, destaca 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.

Nesse sentido, realizando um percurso de 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.

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 poderão utilizar 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 Educação? Com efeito, a definição dos usuários, entre outras providências e características, e sem se constituir no núcleo conceitual da avaliação da aprendizagem, realça a importância das formas e conteúdos da divulgação dos resultados.

Na perspectiva de definir, 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.

Em algumas definições de avaliação educacional a emissão de juízo ou não aparece com tanta ênfase ou abarca um leque mais amplo de elementos, como identificamos em Bloom, Hastings e Madaus (1975, p. 177), que consideram a avaliação como um método de adquirir e processar evidências necessárias para melhorar o ensino e a aprendizagem, incluindo uma grande variedade de evidências que vão além do exame usual de “papel e lápis“. É 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 nessa definição, pode-se constatar a determinação de uma medida adequada de aprendizagem.

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, planificada, 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.

3. AVALIAÇÃO DA APRENDIZAGEM: FUNÇÕES

No âmbito da avaliação educacional, a avaliação da aprendizagem, além de ser uma expressão bastante conhecida, é a mais utilizada, pois, praticamente, se confunde com a difusão da escolarização no século XX. Embora uma avaliação externa também possa ser uma avaliação da aprendizagem, historicamente é por meio da avaliação interna que se dá o acompanhamento das aprendizagens dos alunos.

De maneira geral o processo de avaliação está relacionado à produção de informações sobre determinado objeto e é algo que está bastante presente no cotidiano escolar. É um processo que envolve concepções, valores, princípios, teorias, conceitos, metas, percursos e, por sua complexidade, possui desafios para sua legitimação. Esses elementos ganham maior densidade quando se pensa no desenvolvimento de uma avaliação da aprendizagem organizada de modo externo e em larga escala. Ademais, a avaliação da aprendizagem sempre esteve 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.

Uns 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, além das profundas divergências entre os muitos psicólogos que procura(ra)m defini-la.

Como não se pode avaliar diretamente a capacidade dos alunos, o que se pode, materialmente, avaliar, são os desempenhos ou os resultados que os alunos apresentam na realização de tarefas. Com isso não se está desconsiderando todo um volume gigantesco aportado por várias vertentes da Psicologia voltada para a aprendizagem, apenas delimitando aquilo que, inclusive por razões pragmáticas, pode-se, efetivamente, avaliar. Além disso, as hipóteses psicológicas sobre a aprendizagem podem ser invocadas para explicar, eventualmente, o porquê de determinados resultados ou produções dos alunos.

Assim, para que tenhamos uma avaliação da aprendizagem consistente, é imperioso, ainda que isso não esgote um amplo conjunto de exigências, que: 1- tarefas avaliativas sejam bem elaboradas; 2- os instrumentos, compostos de tarefas, sejam bem construídos; 3- as condições de aplicação dos instrumentos não se convertam em obstáculos cognitivos; 4- os alunos acreditem em um processo avaliativo justo, fidedigno, válido e que seus resultados serão utilizados em seu favor.

Em algumas definições, nos deparamos com as finalidades ou usos dos resultados da avaliação que se condensam em torno de sua função, um aspecto relevante se ponderarmos que no âmbito da educação escolar a avaliação, independentemente de sua finalidade, é um elemento do processo pedagógico, não existindo, portanto, isoladamente, sobretudo porque os conteúdos curriculares são, via de regra, mesmo que com matizações e seleções, a base para a delimitação daquilo que será avaliado enquanto as apropriações, condensadas na aprendizagem, de cada aluno.

Encontramos, assim, na literatura mais do que o emprego da expressão "função", mas, sobretudo, a adjetivação da avaliação como decorrência da função que a ela se atribui. Disso decorrem as expressões "avaliação diagnóstica", "avaliação formativa", "avaliação somativa", entre outras. Também encontramos, é preciso frisar, 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), busca a 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 serão propostas e as aprendizagens anteriores que servem de base àquelas, no sentido de evitar as dificuldades futuras.

A função formativa, conforme Haydt (1995, 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 função da avaliação, tanto ao estudo do aluno como ao trabalho do professor, principalmente através de mecanismos de *feedback*. Estes mecanismos permitem que o professor detecte e identifique limitações/barreiras no processo de ensino, possibilitando reformulações no seu trabalho didático, visando aperfeiçoá-lo. Bloom, Hastings e Madaus (1975) empregam a expressão avaliação formativa, mas, na verdade, ressaltam os desdobramentos e defendem que visaria 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.

A função somativa tem como objetivo, segundo Miras e Solé (1996, p. 378), determinar o grau de domínio do aluno em uma área de aprendizagem ao final de um período

de aprendizagem, de acordo com os níveis de aproveitamento, o que permite outorgar uma qualificação que, por sua vez, pode ser utilizada como um sinal de credibilidade da aprendizagem realizada, podendo ser denominada de função creditativa. A avaliação somativa pode contribuir na observação do progresso realizado pelo aluno no final de uma unidade de aprendizagem, caso existam dados, comparáveis longitudinalmente, levantados anteriormente, o que, por sua vez, podem indicar como o ensino influenciou na aprendizagem, ainda que esta não dependa exclusivamente do ensino, pois existem outros fatores que nela incidem. Corresponde a um balanço final, a uma visão de conjunto relativamente a um todo sobre o qual, até então, só haviam sido feitos juízos preliminares.

Se, como definiram Lukas Mujika e Santiago Etxberria (2009), a finalidade da avaliação é a otimização do objeto de avaliação, isto é, a proficiência do aluno, ganha relevância a função formativa da avaliação, de tal modo que seus resultados possam ser utilizados para essa otimização.

Scriven (1967) indicou a importância da função formativa que, em seus desdobramentos, alimentou proposições como as de Crahay (2012), apontando para o sucesso de todos como a tarefa primordial do processo de escolarização, em contraste com a visão subjacente à avaliação para certificação. Sucesso que depende muito fortemente das condições pelas quais os cursos são oferecidos.

A função somativa pode converter-se à função formativa quando se toma os resultados finais de uma turma num ano letivo como um elemento para se repensar o acompanhamento nos anos subsequentes; podendo-se ainda sustentar a importância para o desenvolvimento curricular com vista às próximas turmas. Em tal quadro pode-se considerar a possibilidade de a avaliação educacional ser desenvolvida para melhorar seu objeto de avaliação, como assinalam Lukas Mujika e Santiago Etxeberria (2009), ao destacarem a relação da avaliação com o processo educacional mais amplo no interior do qual se encontra, facultando a tomada de decisões mais inclusivas e propositivas com vista a esse processo. Tal perspectiva apresenta um contrapeso para o olhar reducionista para a avaliação somativa, muitas vezes pouco valorizada por ser aplicada em finais de processos, como se seus dados não pudessem ser utilizados para intervenções que promovam avanços nas aprendizagens limitando seus resultados a constatações, classificações e seleções.

4. AVALIAÇÕES EXTERNAS DA APRENDIZAGEM EM LARGA ESCALA

Embora as avaliações externas e seus resultados estejam no cotidiano escolar e já façam parte da sua cultura, mesmo que de maneira questionável, muitos autores apresentam críticas a elas por sua finalidade, abrangência e possibilidade de análise da realidade escolar (CASASSUS, 2009; HORTA NETO, 2010; MACHADO, 2012; MADAUS; RUSSEL; HIGGINS, 2009). De outra maneira, os resultados das avaliações vêm sendo apropriados pelas políticas educacionais como se seus efeitos fossem inquestionáveis e as intervenções a partir destes fossem sempre benéficas. O que, por outra perspectiva, desconsidera os diferentes tipos de erro e efeitos destas políticas (NEAL; SCHANZENBACH, 2010; NICHOLS; BERLINDER, 2007; RAVITCH, 2010).

Presentemente, a avaliação educacional apresenta-se como um campo complexo onde coexistem diversas dimensões de avaliação como, "a avaliação dos alunos, a avaliação dos profissionais (educadores e professores), a avaliação institucional das escolas, a avaliação dos sistemas ou subsistemas educativos, a avaliação de projetos e programas e a própria avaliação de políticas (AFONSO, 2014, p. 488-489).

Esteban (2014, p. 467) aponta que

> [...] a avaliação educacional – seja da aprendizagem ou do sistema – tem, indiscutivelmente, importante papel na articulação de um projeto de educação de melhor qualidade do que o atual. Nossa educação ainda é altamente excludente e guarda mecanismos que fazem renascer, sob novas configurações, o fracasso das classes populares, ainda que se reduzam os níveis de evasão e de repetência e se ampliem o acesso à escola e o tempo médio de escolarização.

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 outro tipo de 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, por isso denominada avaliação externa. Assim, 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 pelas denominadas avaliações externas. 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.

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 o desenvolvimento da pesquisa realizada. 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. Entre esses aspectos, podemos destacar a certificação, o credenciamento, o diagnóstico e a rendição de contas. Essas avaliações são, em geral, organizadas a partir de um sistema de avaliação cognitiva dos alunos e são aplicadas de forma padronizada para um grande número de pessoas, entre os quais estão alunos, professores, diretores, coordenadores, por meio de testes compostos por itens de múltipla escolha.

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 ou de sistema. Deste modo, uma avaliação pode ser externa, mas não necessariamente em larga escala.

Portanto, a avaliação na educação deve ser entendida como um processo amplo, com desdobramentos individuais, coletivos e institucionais, que possui comprometimento com um processo educativo e pedagógico. Estudos e pesquisas embasados nessa perspectiva, enfrentam o desafio de articular as necessidades sociais à criação e consolidação de políticas públicas na direção da melhoria da qualidade da educação.

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.

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, segundo Bonamino e Sousa (2012), na terceira geração vai ao encontro do apontamento levantado, pois essa geração é caracterizada pela divulgação dos resultados de forma interna e externa, e também com recompensas, pelo pagamento de bônus em decorrência dos resultados dos alunos.

Por analogia, as matrizes das avaliações externas não substituem os currículos, por várias razões, mas podem ser pontos de ancoragem para a organização do ensino e, especialmente, para as avaliações de aprendizagem internas que requerem, forçosamente, parâmetros para sua validade e fidedignidade, para o que as matrizes podem subsidiar alternativas.

Em análise à forma como gestores incorporam as avaliações externas, Sousa ressalta que 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).

Madaus (1988) aponta outros efeitos indesejáveis que a avaliação externa pode ocasionar o impacto no 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 externas 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.

Com relação à divulgação dos resultados, o 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 capazes de organizar a avaliação da 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).

Essas consultorias e empresas são contratadas para prestarem serviços técnicos, relacionados à elaboração dos itens, à construção dos testes e ao tratamento dos dados, e frequentemente não contribuem para articulação da tarefa técnica com a formação dos profissionais das equipes técnicas das secretárias e, menos ainda, dos professores. Essa articulação, que decorre de uma apreensão política da avaliação, da relação entre a avaliação e trabalho dos profissionais da educação a elas, direta ou indiretamente, vinculados, aproximaria esses profissionais 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 salientam que a avaliação educacional não é abordada na formação inicial e continuada dos professores (FREITAS, 2019; GATTI, 2007). Neste sentido, 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 internos e facilitaria a análise e interpretação das tarefas apresentadas nas avaliações externas e seus resultados.

Tudo indica haver 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 – a proficiência dos alunos em determinadas competências – é pouco associado àquilo que é ensinado ou que os alunos deveriam 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.

Nesse padrão, no qual as provas são elaboradas por especialistas e empresas terceirizadas, 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, efetivamente, de análise, discussão e estabelecimento de linhas de ação. São demasiadamente técnicos, empregando uma linguagem muitas vezes desconhecida pelos docentes ou com termos e expressões 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 geoeducacionais. E quando são feitas as identificações, essas padecem da falta de inteligibilidade mencionada anteriormente. 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. Destacaremos, também, que há uma 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 inexistente, em relação ao sistema e a suas unidades.

Outro ponto de tensão é a validade, atualmente definida como “o grau com que a evidência e a teoria apoiam as interpretações dos escores do teste para um propósito específico” (AERA; APA; NCME, 2014; MESSICK, 1989), sendo necessária uma reflexão sobre as avaliações atualmente desenvolvidas nos vários níveis do sistema educacional, especialmente avaliações em larga escala, compreendendo a pluralidade de todo território envolvido na avaliação.

Vianna (2003, p. 43) apresenta indagações sobre as avaliações de competências e habilidades que, embora sejam mencionadas nos debates, apresentam-se de modo muito pouco consensual, de tal modo que

> As questões que se impõem imediatamente, com o objetivo de aprofundar nossas percepções, podem ser propostas da seguinte forma: – são desenvolvidas competências e habilidades em nosso 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 supostas habilidades? Outra pergunta, que também reflete a nossa perplexidade: – se competências e habilidades foram promovidas, houve, efetivamente, preparo adequado dos educadores em relação a esse complexo e controvertido assunto? E quanto a atitudes, interesses e valores? 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.

Com essas tensões, foi evidenciado que a participação dos professores, possível, por meios adequados, e necessária, tanto para efeito de formação em avaliação educacional quanto para o incremento da credibilidade e da legitimidade política, poderia favorecer, no sentido de aproximar as avaliações externas do que foi ou deveria ter sido ensinado e, sobretudo, para que seus resultados e mesmos seus procedimentos, ainda que nem todos eles, sejam integrados ao trabalho docente.

Neste quadro, podemos ainda ponderar alguns aspectos, sem termos positivos ou negativos. Como um aspecto negativo, é possível considerar a hipótese de que em algumas escolas haveria a tendência de ensinar aquilo que é ponto de partida para a elaboração dos itens dos testes padronizados, expressos nas matrizes de referência, com acentuado emprego do formato da prova utilizando-se de itens de múltipla escolha, o que configuraria um reducionismo curricular e didático. Como exemplo dessa tendência, há o registro de atividades escolares de organização de “simulados” de aplicação de provas padronizadas, algo que deveria ser evitado, inclusive porque não produz o aumento de proficiência esperado, pois este depende de processo de ensino e de estudo. Mais grave, ainda, seria abandonar o foco nas avaliações internas, pois estas, mesmo quando revelam algumas restrições técnicas, integram

a tarefa educativa a cargo, incontornavelmente, dos professores, e que devem abarcar um conjunto de conteúdos que nem sempre, por razões variadas, integram os testes padronizados aplicados em larga escala. Se as restrições às avaliações internas podem ser superadas com formação e capacitação, por seu turno, as avaliações externas não podem e não deveriam se converter em orientadoras privilegiadas dos processos formativos, mesmo quando parecem se constituir no mais saliente traço das políticas educacionais. Distinto disso seria a tarefa de articular, por vários meios e objetivos, as avaliações externas com as internas.

Como aspecto potencialmente positivo, podemos considerar que, com as avaliações em larga escala, a gestão de escolas e redes passa a incorporar indicadores de desempenho como mais um elemento para o conhecimento de suas realidades e, assim, pode estabelecer metas mais precisas e elencar prioridades de intervenção parametrizadas numa realidade mais ampla, envolvendo a comparação, dependendo da avaliação externa referenciada, com resultados do país, do estado e do município.

A despeito desse potencial positivo, salientamos, contudo, que as medidas resultantes dessas avaliações – as estimativas de proficiência – não se constituem, por si só, em avaliação, pois uma medida indica o quanto se atingiu numa determinada escala e a avaliação é o julgamento desse resultado em função de critérios, para os quais a interpretação pedagógica é parte insubstituível do processo avaliativo, que, também, deve levar em consideração as condições específicas de cada rede e escola, reforçando a importância da avaliação institucional. Assim, numa situação na qual uma escola receba seus resultados com uma avaliação, pode-se organizar o debate em torno dos critérios utilizados para se definir determinado juízo.

Outro desafio é avaliar as avaliações externas. Como já apontado, poucos trabalhos se destinaram a avaliar os instrumentos de avaliação e seu processo de construção. Isso não deveria passar incólume porque limitações nesse tocante podem comprometer, sobretudo, a validade e a fidedignidade dos resultados, impedindo vários desdobramentos relevantes para o trabalho, por exemplo, em escolas concernidas à avaliação em questão.

Para que uma avaliação atenda aos critérios técnicos mínimos, é necessário garantir determinadas condições em seu processo de construção, sendo elas:

1) Organização de **Matriz de referência** que descreva cada habilidade 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 para uma população;

8) Definição de **escala de medição** e delimitação de **pontos de corte dos níveis**;

9) Definição de itens âncora e redação de **interpretação pedagógica** dos níveis de resultados;

10) Estabelecimento de **critérios** para avaliação/julgamento dos resultados e **proficiências**;

11) Conclusão da **avaliação e análise de fatores associados**; e

12) Definição dos públicos e divulgação de **relatórios pedagógicos** dos resultados.

13) Atualização do Banco de Itens;

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, poderia aproximar os profissionais que atuam nas escolas dos instrumentos, diminuindo a resistência, qualificando-os para a interpretação dos dados divulgados e validando os resultados.

Não bastando as tensões advindas do caráter externo de uma avaliação, temos pela frente outros obstáculos quando se considera uma avaliação externa e em larga escala, isto é, quando essa avaliação abarca um amplo conjunto de escolas e, mais ainda, numa ampla cobertura nacional. Adicionalmente, pode-se aventar que os resultados de uma avaliação podem ser utilizados para apoiar a tomada de decisões que seriam mais bem embasadas, pois sustentada em processos mais bem organizados, com resultados válidos e fidedignos, especialmente no caso educacional, para guiar indicadores de qualidade.

Nevo (1995), mesmo reconhecendo que muitas das avaliações externas estão inseridas em políticas educacionais que ignoram em suas formulações as necessidades das escolas, trabalha com a necessidade e a possibilidade de diálogo entre as avaliações externas e as avaliações internas – estas conduzidas pelos professores ou pelos profissionais da educação –, pois cada uma das avaliações possui potencialidades e limitações, ressaltando, entretanto, que esse diálogo deve ser baseado na escola, indicando com isso que o ponto de partida e de chegada dos processos avaliativos deve ser a escola e suas necessidades, o que, por seu turno, deve ser mais do que as aprendizagens, embora essas sejam um dos aspectos mais relevantes

de um projeto educativo. Para Nevo (1995), no entanto, as políticas educacionais precisariam favorecer e criar condições para esse diálogo. Deveriam começar pelo reconhecimento de que se as avaliações externas podem contar com pessoal mais especializado tecnicamente e trazer resultados decorrentes de um olhar menos condicionado pelas rotinas da escola e, portanto, de maior amplitude, podendo envolver comparações com um conjunto de escolas, essas avaliações não gozam, via de regra, de reconhecimento acentuado no processo de trabalho de cada uma das escolas, algo que só uma avaliação interna dispõe, ainda que à custa de menor precisão sobre alguns objetos de avaliação. Nesses termos, até mesmo para que a ideia de avaliação formativa possa ser posta em prática, o autor defende que ambas as avaliações se complementam, superando suas limitações e adicionando suas potencialidades.

Nesse sentido, pode-se apontar que as avaliações externas, em si mesmas, não são boas ou más. Como qualquer avaliação, compreendida como um processo, as avaliações externas podem e devem ser avaliadas e disso pode decorrer uma série de questionamentos que, no limite, podem nos levar a uma apreciação negativa a ponto de considerar a hipótese de que seja descontinuada. É na utilização delas, assim como os seus propósitos e as suas relações com as avaliações internas e com o currículo, que se pode determinar o seu real significado e o seu real impacto no que se refere à melhoria dos sistemas educativos. Ou seja, no que se refere às suas relações com a melhoria da qualidade do ensino dos professores e das aprendizagens dos alunos, havendo também problemas bem identificados que podem questionar a credibilidade que, em geral, lhes é atribuída pela sociedade.

As avaliações em larga escala usam como instrumentos os questionários e os testes de proficiência. Os questionários permitem o levantamento de fatores associados intra e extraescolares que podem eventualmente explicar os resultados do desempenho. Porém, os dados levantados por esse instrumento não são diretamente os objetos de avaliação, mas que podem contribuir no entendimento dos resultados.

Os testes de proficiência são elaborados a partir das Matrizes de Referência, que estabelecem as habilidades de uma área do conhecimento, em consonância com os conteúdos de ensino expressos nos currículos, descrita, usualmente, pelo que faz ou deveria fazer um respondente nessa habilidade. A partir das matrizes são elaborados os itens – as tarefas avaliativas – que comporão os testes cujas respostas dos respondentes – seu desempenho – indicarão, numa determinada escala de proficiência, o quanto esse respondente domina em relação àquela competência. Ainda, essas matrizes podem ser segmentadas para cada etapa de escolaridade e serem descritas em termos de habilidades, mas que não esgotam o conteúdo a ser trabalhado em sala de aula.

Desta maneira, podemos considerar que a avaliação da aprendizagem é feita mediante tarefas – itens – a serem resolvidas pelos alunos. As respostas aos itens configuram o desempenho no teste que, devidamente tratado, produzirá um resultado expresso na escala de proficiência a ser comparado com o critério de avaliação, isto é, com aquilo que naquele momento merece o julgamento correspondente. Esse desempenho embora esteja relacionado à aprendizagem não pode ser confundido com ela. Não necessariamente o aprendido "revela-se" nos resultados, pois alguns fatores podem comprometer essa "revelação", podendo estes fatores estarem vinculados ou ao aluno ou aos instrumentos ou aos procedimentos. O que se pode garantir é que o desempenho nunca será maior do que a aprendizagem. Assim, o desempenho cumpre o papel de levantamento de informações que, dados os cuidados na construção do instrumento e em sua aplicação, sejam válidas, que permitam interpretações referentes ao aprendizado, aquilo que o processo de aprendizagem consolidou.

Os instrumentos de avaliação são recursos, portanto, para esse levantamento de informações ao serem processados gerarão resultados que, de acordo com critérios, possibilitam juízos sobre as aprendizagens dos alunos, visto que expressam o que o aluno teria aprendido, deixado de aprender ou ainda precisa aprender sobre os conteúdos avaliados. Os instrumentos apresentam registros de diferentes naturezas: expresso pelo próprio aluno (provas, cadernos, textos e outros) ou expresso pelo professor (pareceres, registro de observação, fichas e outros).

Assim, ao definir os objetos de avaliação, o professor necessita selecionar o instrumento mais adequado, aquele que fornece as melhores informações sobre esse objeto. Para escolher o instrumento de avaliação, é necessário conhecer as possibilidades, aplicação e limitações dos instrumentos, que precisam ser refletidos quanto à sua elaboração, à adequação aos objetivos das decisões que se pretende tomar com base nos resultados da avaliação, ao conteúdo e à metodologia adotada no ensino, completando-se com sua aplicabilidade e o processamento das respostas. Cabe ainda destaque a preocupação com a devolução dos resultados aos vários públicos interessados.

Segundo Vasconcellos (2003, p. 125), a reflexão crítica dos instrumentos de avaliação remete o professor a alguns questionamentos voltados a como os instrumentos são preparados, como são analisados, pontuadas as respostas e como é feita a comunicação dos resultados da avaliação, seus juízos, que podem estar associados a interpretação pedagógica das medidas. Todos esses aspectos necessitam ser amadurecidos pelo professor. Porém, a elaboração do instrumento é um ponto crucial nessa reflexão. O que se faz com os resultados obtidos é elemento do ensino e embora possa se pautar na avaliação não pode ser confundido com ela.

Ainda pouco se sabe, efetivamente, sobre os desdobramentos e efeitos das avaliações externas sobre a organização do trabalho escolar, e, em particular, sobre a ação pedagógica. No entanto, alguns estudos recentes (ESQUINSANI, 2010; PIMENTA, 2012; SILVA, 2011; WELTER; WERLE, 2010) indicam já ser possível afirmar a existência de movimentos de intensificação do uso das avaliações externas e seus resultados por redes educacionais e escolas, no Brasil, na articulação e organização do trabalho pedagógico. Entretanto, cabe perguntar: esse esforço em utilizar os dados das avaliações tem contribuído, de fato, para a garantia da aprendizagem para todos? A interpretação dos resultados provenientes das provas pode contribuir para o cumprimento dessa finalidade educacional? (ALAVARSE, 2012). O desafio consiste, de um lado, na ampliação da experimentação de uso dos resultados com essa perspectiva; de outro lado, na investigação dos seus efeitos sobre o trabalho pedagógico e a aprendizagem dos estudantes. Também convém explorar mais o uso dos resultados na articulação e organização da gestão do ensino.

Para Amaral, Alavarse e Silva (2020), no que tange mais especificamente à relação da gestão com os resultados das avaliações externas, estas ganham maior relevância pelo fato desses profissionais serem os responsáveis pela intermediação entre gestores de rede, usualmente os sujeitos dessas avaliações, e os professores, que prioritariamente devem se apropriar dos resultados para incidir nas salas de aula, local proeminente das aprendizagens.

4.1 Avaliações externas: características do Pisa e do Saeb

Neste tópico são apresentadas as características que nos parecem mais relevante do Pisa e do Saeb, avalições externas que servem como referência internacional e nacional, respectivamente, tanto por suas estruturas e processos de construção como o uso de seus resultados.

4.1.1 Programa Internacional de Avaliação de Estudantes (Pisa)

O Programa Internacional de Avaliação de Estudantes (Pisa), tradução de *Programme for International Student Assessment,* é uma avaliação internacional que mede o nível educacional de estudantes por meio de provas de Leitura, Matemática e Ciências e é realizada a cada três anos pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Seu público-alvo é uma amostra representativa de estudantes na faixa etária de 15 anos, idade

que representa o término da escolaridade básica obrigatória na maior parte dos países. Seus resultados permitem que cada país avalie seu desempenho nas áreas elencadas e possa compará-los com outros países, bem como formular políticas e programas educacionais.

A construção dos instrumentos avaliativos do Pisa é organizada por uma comissão de especialistas, a assembleia de diretores do Pisa e pesquisadores dos países participantes. A validação dos saberes e capacidades julgados essenciais, nos respectivos campos avaliados, foi coordenada por um grupo de pesquisadores internacionais em cooperação com os países integrantes. Os campos de avaliação, bem como sua organização, foram apresentados e pré-testados em alguns países, produzindo resultados para uma breve análise para avaliação. Esta testagem inicial para organização da avaliação teve como foco:

- Elaborar um significado de trabalho para o domínio das pressuposições que apresentam como base essa definição;
- Avaliar de maneira a estruturar os elementos de trabalhos a serem elaborados de modo a fornecer aos pesquisadores informações a respeito das diretrizes políticas e o progresso dos estudantes em cada esfera da avaliação nos países participantes;
- Identificar um grupo de fatores básicos a serem considerados quando a avaliação for estabelecida no nível internacional;
- Instrumentalizar o grupo de elementos básicos a serem empregados na elaboração de testes, tendo como base a literatura atual e no conhecimento adquirido na prática decorrente de outras avaliações em larga escala;
- Examinar minuciosamente as variantes e a avaliação da colaboração de cada um para a abrangência dos obstáculos do trabalho nos países participantes, e
- Elaborar um modelo para a interpretação dos resultados.

A organização foi aprovada tanto no meio científico como também no político e, posteriormente, viabilizou dados para o desenvolvimento dos instrumentos avaliativos. A organização é exposta no relatório da estrutura de avaliação do Pisa 2003 - Conhecimentos e Habilidades em Matemática, Leitura, Ciências e Resolução de Problemas (OECD, 2003). Proporcionaram uma comunicação de fácil compreensão para que todos os países integrantes desenvolvessem uma conformidade em relação aos parâmetros de mensuração do Pisa.

Para retratar o vasto campo de diversidade nacional, cultural e linguística entre os países da OCDE, adotou-se bastante cautela. Como parte desta dedicação, a organização do Pisa contou com apoio de pesquisadores que desenvolveram itens de experimento nos países, bem como Austrália, Reino Unido, Holanda e Japão. Ademais, em relação aos itens

construídos e testados pela organização do Pisa, os países colaboraram com material de avaliação. O grupo de trabalho globalizado composto por profissionais que elaboram o teste julgou apropriada uma quantidade considerável do material exposto em decorrência das condições impostas pelas estruturas da avaliação do programa.

Como produto, o banco de itens da fase inicial de testes contou com o apoio de países, como: Alemanha, Argentina, Austrália, Áustria, Canadá, Coréia do Sul, Dinamarca, Estados Unidos, Finlândia, França, Grécia, Holanda, Irlanda, Itália, Japão, Noruega, Nova Zelândia, Portugal, República Checa, Suécia e Suíça. Os países integrantes do Pisa colaboraram com cerca de um terço dos itens construídos para a introdução no experimento de campo e 37% desses itens foram utilizados na pesquisa principal.

Aproximadamente, 232 unidades, abrangendo cerca de 530 questões, foram utilizadas nos grupos de questões para verificação e possíveis correções nacionais no domínio da matemática, resolução de problemas e ciências. Em seguida ao processo de investigações, o experimento de campo introduziu 115 unidades de matemática com 217 questões. Desse quantitativo de unidades de matemática, o material de estímulo para 53 unidades foi elaborado pelos países e 80 desenvolvidas pela organização do Pisa, e uma unidade foi retirada do relatório Tendências no Estudo Internacional sobre Matemática e Ciências (TIMSS).

Cada questão incluída na combinação de avaliação foi então analisada pelos países particularmente: por seu impacto cultural, gênero ou outra característica; sua relevância para o público de estudantes de 15 anos de idade, tanto no cenário educacional quanto fora; e por sua relevância e nível de interesse. Em um primeiro momento, desenvolveu-se uma primeira análise com os países no que diz respeito à combinação de questões como etapa do processo do experimento de campo. Posteriormente ao experimento de campo, realizou-se uma pesquisa como suporte na escolha das questões a serem utilizadas no estudo principal.

Na continuidade do experimento de campo, no qual todas as questões foram experimentadas nos países integrantes, o grupo de desenvolvimento de questões e especialistas examinou sobre o olhar de diversos elementos a seleção de questões para o estudo principal, como: os resultados do experimento de campo; as decorrências do exame de verificação das questões em cada país; e as incertezas surgidas no decorrer do processo de atribuição de notas do experimento de campo. No decorrer do mês de outubro de 2002, os grupos de elaboração de questões, juntamente com seus especialistas, escolheram um agrupamento final de questões que, no decorrer de um tempo de acordos, foi escolhido pelos países participantes, tanto na competência científica quanto política.

4.1.2 Sistema de Avaliação da Educação Básica (Saeb)

O Sistema de Avaliação da Educação Básica (Saeb) é um conjunto de aplicações de testes padronizados em larga escala realizados bienalmente pelo Inep desde os anos 1990, permitindo que as redes municipais e estaduais de ensino e escolas, estas desde 2005, tenham informações sobre o desempenho de seus estudantes. Oficialmente, o resultado dessas aplicações seria um indicativo da qualidade do ensino oferecido por essas unidades como subsídios para a elaboração, o monitoramento e o aprimoramento de políticas educacionais com base em evidências (Inep, 2021).

São aplicados testes padronizados e questionários contextuais a cada dois anos, censitariamente desde 2005, nas escolas das redes públicas municipais e estaduais e também para uma amostra de escolas privadas. Desde sua implementação, o Saeb passou por várias reformulações. Inicialmente avaliou uma amostra de escolas representativas das 1ª, 3ª, 5ª e 7ª séries do ensino fundamental com ênfase em Língua Portuguesa (LP), Matemática, Ciências Naturais e Redação. Em 1995, a 3ª série do ensino médio foi incluída na amostra e os resultados passaram a ser tratados pela Teoria de Resposta ao Item (TRI). Em 2005, o Saeb foi reformulado mais uma vez, sendo dividido em duas avaliações, a Avaliação Nacional da Educação Básica (Aneb), de caráter amostral, e a Avaliação Nacional do Rendimento Escolar (Anresc), mais conhecida como Prova Brasil. Aplicada de forma censitária aos estudantes do 5º ano e do 9º ano do ensino fundamental avaliando a proficiência em leitura e resolução de problemas, seus resultados passaram a compor, a partir de 2007, o Índice de Desenvolvimento da Educação Básica (Ideb), um indicador de qualidade.

Em 2013 foi a vez dos estudantes do 3º ano do ensino fundamental participarem do Saeb por meio da Avaliação Nacional da Alfabetização (ANA) aplicada de forma censitária para as escolas públicas e amostral para instituições privadas. Em 2017 as provas foram censitárias também para estudantes da 3ª série do ensino médio e, em 2019 as avaliações que compunham o Saeb passaram a se chamar “Novo Saeb”, concentrando todas as provas como um sistema único. A edição de 2019 trouxe ainda outras mudanças, como a inclusão da educação infantil aplicada a uma amostra de escolas, além da adequação das matrizes à Base Nacional Comum Curricular (BNCC); houve ainda a inclusão de Ciências Naturais para estudantes do 9º ano aplicada para uma amostra de escolas públicas e privadas. Atualmente, o

Novo Saeb é composto por avaliações voltadas à educação infantil, 2º, 5º e 9º ano do ensino fundamental e 3ª série do ensino médio.

De acordo com Fernandes e Gremaud (2020, p. 1115), os ciclos do Saeb podem ser divididos em três fases: "os dois primeiros (1990 e 1993) compondo a fase de implementação do SAEB. Depois, a partir do terceiro ciclo de 1995, temos a segunda fase, de consolidação do SAEB como instrumento de diagnóstico e monitoramento da educação básica brasileira e, a partir de 2005, uma terceira fase marcada por acoplar ao SAEB elementos de *accountability*".

> O Brasil possui hoje um sistema de avaliação da educação básica que, sem exagero, poderia ser classificado entre os melhores do mundo. Mas, evidentemente, pode ser aprimorado. Além de retratar o passado, as avaliações têm o papel de sinalizar para o sistema o que se espera das escolas. Então, ao menos para a segunda fase do ensino fundamental e para o ensino médio, as avaliações deveriam ir além de leitura e matemática e incluir ciências da natureza e humanidades. No ensino médio, isso era feito com o ENEM, mas foi revertido com a decisão do INEP de não mais divulgar os resultados do ENEM por escolas, e universalizar o SAEB para as escolas públicas de ensino médio.

Os itens que compõem as avaliações do Saeb são elaborados por profissionais experientes, com formação sólida, acadêmica e prática, que são selecionados via chamamento público e selecionados pelo Inep. Os elaboradores credenciados recebem formação para elaboração de itens oferecida pelo órgão organizador das avaliações participando de todo o período de elaboração e revisão dos itens sob critérios rigorosos de conduta sobretudo no que diz respeito ao sigilo.

Tanto no Saeb quanto no Pisa não há o envolvimento de professores das redes de ensino avaliadas de forma sistemática, tampouco processos formativos organizados amplamente. Sendo os profissionais selecionados por alta especialização e mesmo que atuem nas redes de ensino o trabalho realizado com cunho restritamente técnico não se articula com sua prática docente e estes estão envolvidos apenas na etapa de elaboração dos itens, não tendo contato com a construção dos instrumentos ou ao processo de tratamento e interpretação dos resultados.

5. FORMAÇÃO DOCENTE EM AVALIAÇÃO EDUCACIONAL

A avaliação é uma prática escolar e uma atividade intrínseca à docência, cuja dimensão de avaliador compõe o rol de atribuições dos professores previstas na legislação educacional e fortemente evidenciada nos editais de processos seletivos.

A competência avaliadora, da mesma forma, está presente nos Referenciais Profissionais Docentes, documentos que definem os conhecimentos e habilidades inerentes à prática docente que expressam o que se espera que o professor em termos de conhecimento e práticas profissionais. Temos como exemplo de uso de referenciais para o trabalho docente as províncias canadenses de Ontário e Quebec, Estados Unidos da América do Norte (EUA), Inglaterra, Alemanha, Argentina, Austrália, Chile, Cingapura, Colômbia, Equador, Escócia, Finlândia, Líbano, México, Peru, Polônia, Portugal, Nova Zelândia, recentemente o Brasil, entre outros países.

No entanto, a literatura especializada, indica que a avaliação é pouco abordada ou inexistente nos cursos de licenciatura (GATTI et al., 2010; ROJAS, 2007; SILVA et al., 2016; VILLAS BOAS; SOARES, 2016). Não obstante, segundo Tardif (2013), a formação de professores deveria integrar saberes do conhecimento profissional do cotidiano do trabalho docente. Alavarse (2013) nomeia esta lacuna formativa de "paradoxo docente", uma vez que se requer uma formação que garanta contato prévio com conhecimentos e tarefas da prática profissional e ao mesmo tempo constata-se que o professor raramente recebe formação técnica e teórica para avaliar seus alunos, o que o leva a reproduzir os modelos de avaliação aos quais teve contato enquanto aluno. Tal prática interfere na trajetória escolar dos alunos, pois os expõe a instrumentos de avaliação improvisados e a uma prática avaliativa frágil em conceitos e técnicas, podendo incidir sobre seu futuro acadêmico de forma a favorecer ou até mesmo ampliando as desigualdades sociais, transformando-as em desigualdades escolares, impossibilitando um tratamento equitativo como corroborado por Crahay (2002) e Dubet (2008).

Siqueira (2018, p. 151), destaca que

> Se são muitas as lacunas formativas dos professores quanto à avaliação externa, isso parece assemelhar no que tange à avaliação praticada em sala de aula, alimentada, sobremaneira, por concepções cristalizadas com equívocos históricos quanto a questões básicas da avaliação.

A autora (Op. cit., p. 151) problematiza, ainda, os critérios utilizados pelos docentes a partir dos registros dos Conselhos de Classe, salientando que

> Em muitos Conselhos de Classe é possível verificar critérios e objetos de avaliação flutuantes, comportamento e participação dos alunos como objetos preferenciais de avaliação, além da utilização de uma espécie de "escala imaginária" principalmente nos anos finais para emissão de juízos relativos à aprovação ou à reprovação, conforme o atendimento ou não ao perfil disciplinar desejável. Nessas ocasiões, havia uma tendência à aprovação dos alunos localizados nos extremos dessa escala imaginária, e de reprovação daqueles que se alocavam no meio dessa métrica, a partir de uma combinação entre bom comportamento e nível de aprendizagem que possibilitasse a retomada de um trabalho docente no ano seguinte, uma prática recorrente apoiada em percepções subjetivas, intuitivas dos professores, corroborando a tese de Airasian e Abrams (2003) sobre a diversidade de fontes que alimentam a classificação que o professor faz de seus alunos.

Siqueira, Freitas e Alavarse (2021, p. 12) apontam a necessidade de formação em avaliação educacional na formação docente como forma de superar equívocos e subjetividades nos processos avaliativos com impacto no percurso acadêmico dos alunos, argumentando que,

> Se comportamento e empenho fizerem parte da avaliação, é preciso que se crie uma escala para tal, preservando o aluno de julgamentos equivocados que anunciem tratar-se da dimensão cognitiva, quando na verdade enfatizam outras dimensões do desenvolvimento, confirmando a tese de Merle (2018) quanto à irrelevância pedagógica que as práticas avaliativas dos professores adquirem diante de tamanha amplitude de fatores que são considerados para a atribuição de notas. Desse modo, reiteramos a necessidade de formação docente em avaliação como forma de superar equívocos na ação avaliativa no que se refere a instrumentos, procedimentos, critérios e objetos flutuantes.

Os autores ainda ressaltam que tais ponderações não pretendem deslegitimar as informações consideradas pelos professores ao avaliarem seus alunos, mas sim chamar a atenção para o seu uso e, principalmente, para o impacto que isso pode ter quanto à tomada de decisão de modo a não permitir que impressões ou convicções em relação aos alunos predominem e impactem na sua trajetória acadêmica, seja na forma como os professores os veem, seja no modo como o próprio aluno se enxerga, capaz ou incapaz, a depender do tipo de relação que é estabelecida com o docente.

Compreendemos que a avaliação educacional representa um importante elemento da prática docente, constituindo-se em ferramenta indissociável do acompanhamento da aprendizagem dos alunos, com efeito direto sobre seu percurso escolar. Sua concretização se reflete sobre a qualidade da educação e de sua efetivação definem-se percursos escolares com forte influência, inclusive, sobre a permanência ou a evasão escolar do aluno, como revela Bowers (2019) em relação a marcas que a avaliação deixa nos estudantes e que incidem, inclusive, na definição pela continuidade ou não dos estudos. Bowers (2019) afirma que o

juízo do professor é o que, de fato, acompanha o aluno em sua história escolar, haja vista que o fluxo escolar é definido, via de regra, pelas apreciações que o professor realiza sobre as produções de seus alunos, que se expressam, por sua vez, em notas ou conceitos a partir do que se decidem aprovações e reprovações. Desta forma, consideramos que a formação em avaliação educacional atenuaria ou remiria os impactos das práticas de avaliação instituídas pela ausência da temática da avaliação na formação inicial e continuada.

6. EXPERIÊNCIAS DE AVALIAÇÕES EXTERNAS COM A PARTICIPAÇÃO DE PROFESSORES

Com o objetivo de exemplificar e trazer luz às possibilidades da participação dos professores, coordenadores e gestores no processo de construção de avaliações externas descrevemos a seguir duas experiências: Avaliação Nacional para o diagnóstico da qualidade de formação em Saúde em Moçambique e a Prova Mais Educação de Rede Municipal de Ensino de São Paulo, que embora não se constituam como "modelos" de avaliações externas com a participação e formação docente, apontam possibilidades para viabilidade e contribuições positivas da participação desses profissionais nas etapas de construção de suas avalições externas promovendo um processo formativo.

6.1 A experiência de Moçambique

A equipe do Ministério da Saúde (Misau) de Moçambique, no âmbito da implementação do Plano Nacional de Desenvolvimento de Recursos Humanos 2016-2025 (PNDRH), a Direção Nacional de Formação de Profissionais da Saúde (DNFPS), em alinhamento com o Plano Estratégico para a Garantia da Qualidade de Formação de Técnicos de Saúde 2016-2020, definiram a avaliação educacional, sobretudo a avaliação externa, como um componente para o diagnóstico da qualidade da formação.

A demanda de formação em avaliação e do desenvolvimento de uma avaliação externa em larga escala 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.

É neste contexto que, desde 2012, a Jica através do ProFORSA I e II apoiou a DNFPS na capacitação de técnicos do Misau e docentes das IdF, em Moçambique e no Brasil, em conteúdos sobre a avaliação educacional e assistência técnica especializada por via de assessores, até a realização, pioneira, em dezembro de 2018 da Avaliação Nacional para o diagnóstico da qualidade de formação em saúde em Moçambique.

Na contramão da tendência de realização de avaliações externas em 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 pelo desenvolvimento de cursos e oficinas, conduzidas por membros do Grupo de Estudos e Pesquisas em Avaliação Educacional (Gepave) da Feusp, que visam capacitar essa participação.

Esse envolvimento se deu tanto na participação de professores na construção da avaliação externa, quanto na formação em avaliação educacional que poderá ainda aperfeiçoar a avaliação conduzida pelos professores em suas escolas, mediante o que denominamos diálogos entre avaliações internas e externas.

A tensão em torno da certificação evidenciou uma polêmica sobre as finalidades de uma avaliação externa, bem como sobre possíveis consequências deletérias para um processo de formação profissional cujos dirigentes das IdF devem assumir a responsabilidade dessa formação, sem transferi-la para os estudantes e sem torná-la apenas uma preparação para o exame.

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.

Essa avaliação externa contou com a participação de especialistas em avaliação do Gepave, equipe da RAQ e docentes e técnicos das IdF. Os docentes das IdF colaboram desde o início dessa construção, elaborando e revisando itens, discutindo a Matriz de Referência

para Avaliação e participando da aplicação e da análise dos resultados dos pré-testes desses itens.

6.1.1 Elaboração e validação das matrizes

Para que a produção de itens ocorresse foi necessário o processo de construção e validação de uma matriz de referência compreendeu um conjunto de oficinas envolvendo professores das etapas de ensino que se pretendia avaliar realizando a formação de introdução à avaliação educacional. Os professores selecionaram os principais conteúdos que deveriam ser avaliados ao término da etapa de ensino cujos alunos participariam da avaliação, tendo como referência o currículo oficial. Esse processo de seleção de conteúdos foi muito rico e se constitui num exemplo da importância e da possibilidade de envolvimento de professores numa avaliação em larga escala, destacando-se discussões e debates entre eles, pois muitas vezes não havia um acordo inicial, dado que uns consideravam determinados conteúdos sem a concordância dos demais. Isso demandou esforços de convencimento do que seria prioridade, considerando a realidade e necessidade do país, estado ou município para os estudantes serem avaliados, mas, sobretudo, demandando argumentos tendo em vista uma avaliação da aprendizagem ao final de uma etapa de formação profissional, com potenciais repercussões para o processo de ensino de novas turmas.

Adicionalmente, esse debate destacava a existência de um limite de conteúdos que podem ser contemplados tendo em vista a quantidade de itens que podem ser inseridos numa avaliação da aprendizagem ao final de uma etapa, algo que vale tanto para uma avaliação interna quanto externa, considerando inclusive o ‘efeito cansaço’ dos estudantes, que ocorreria se tivessem de responder a uma quantidade excessiva de itens.

Após o levantamento dos conteúdos, os professores agruparam os mesmos de acordo com sua natureza, formando as competências, e elaboraram a descrição de cada agrupamento. Os agrupamentos elaborados foram apresentados em assembleia para os participantes da formação para contribuições e discussões coletivas, ocorrendo a validação da proposta pelo do grupo.

Nos Exames Nacionais dos Cursos Técnicos de Moçambique (todos os cursos) a Matriz de Referência para avaliação das aprendizagens foi formulada com base na Taxonomia de Bloom revisada (ANDERSON; KRATHWOHL, 2001), que possibilita a articulação dos tipos de conhecimento com os grupos de operações mentais, distribuindo os objetivos em diferentes níveis de aprendizagem. Essa taxonomia organiza a elaboração de itens,

delimitando as competências, os tipos de conhecimentos e operações mentais que serão aplicados aos conteúdos considerados relevantes ao curso; subsidiando a elaboração da matriz que confere transparência ao processo de avaliação, apoiada em sua legitimidade, que por seu turno, decorre de um amplo processo de legitimação, envolvendo a grande maioria dos professores do curso. Igualmente, essa taxonomia favorece o estabelecimento dos conteúdos essenciais indicados na matriz curricular. É importante destacar que a Matriz de Referência não deve ser confundida com o currículo, que é muito mais amplo.

Num segundo momento, a lista de competências proposta e seus respectivos conteúdos para cada curso foram submetidos à consulta aos professores das unidades educacionais que não participaram do grupo de trabalho, como forma de validação. Esta etapa se configura então em um processo rico e muitas vezes desconsiderado: o processo de validação pelos professores.

Nesse instante se propõe uma nova discussão das necessidades: quais seriam os conteúdos prioritários. O que se destaca neste segundo momento são dois aspectos: o primeiro, relacionado à oportunidade dos professores de refletirem sobre o curso e o papel de cada disciplina na formação do estudante e o segundo, de compreenderem melhor o papel da avaliação externa em selecionar os conteúdos mais relevantes.

Durante o processo formativo em Moçambique, uma das situações que mais se destacaram foi quando os professores, depois de algum tempo em discussão, concluíram que, considerando o problema crônico de AIDS e, particularmente no caso do curso Enfermagem em Saúde Materno-Infantil, muitos habitantes que, por preceitos culturais, não externalizavam a gravidez, o que dificultava o acompanhamento necessário para o sucesso da natalidade. Então, neste sentido, todos os conteúdos que, de algum modo, envolviam estes âmbitos não poderiam ser rejeitados. Empiricamente este processo se mostrou exitoso uma vez que 95% dos conteúdos selecionados foram mantidos e não houve propostas de modificação para a organização das competências.

Fez-se necessária a realização de um seminário com palestras e atividades para aprofundamento das bases teóricas utilizadas no processo desenvolvido e validado em momento anterior.

A matriz elaborada subsidiou a elaboração dos itens que constituíram um pré-teste que forneceu elementos para análise dos itens e dados que auxiliaram no aperfeiçoamento e validação da matriz proposta.

Acreditamos que a participação dos professores no processo de construção da matriz de avaliação seja de grande contribuição para validação e apropriação dos resultados uma vez

que esses, além de contribuírem para a aproximação entre o instrumento de avaliação e o currículo, se apropriam dos conhecimentos técnicos da avaliação externa das aprendizagens, antes apropriados somente por equipes técnicas e assessorias.

6.1.2 A elaboração e revisão de itens

O item consiste na unidade básica de um instrumento de coleta de dados que visa "medir" um traço latente. O item é uma tarefa que pode ser na forma de item de múltipla escolha, de item de resposta construída ou item de atuação, que deve estimular uma reação do respondente, cuja capacidade de enfrentá-lo com sucesso depende de sua proficiência.

Para compor o banco é necessária uma quantidade expressiva de itens com comprovada qualidade técnico-pedagógica e psicométrica, incluindo seus pré-testes, os quais irão compor os testes da avaliação externa e em larga escala.

A elaboração de itens de múltipla escolha requer que o elaborador tenha domínio tanto da área de conhecimento a ser avaliada quanto de procedimentos técnicos. O processo de composição do item desenvolve-se artesanalmente e a versão final será alcançada após várias revisões, no que diz respeito ao conteúdo, estrutura, linguística, gramática etc. Por esta razão a elaboração e revisão de itens foi adotada como ponto de partida para participação dos professores no processo de elaboração dos Exames Nacionais e formação inicial em avaliação educacional no formato de oficinas.

Durante a construção de uma avaliação externa é necessário desenvolver oficinas de elaboração e revisão de itens precedidas de uma formação inicial em avaliação educacional onde são apresentados os conceitos e princípios que seriam adotados no processo.

Na experiência de Moçambique a realização das oficinas de elaboração e revisão de itens contaram com a participação de membros do Misau, dos IdF e dos docentes dos cursos envolvidos oriundos de diferentes IdF, sendo mediadas pelos peritos do Gepave.

Por meio das oficinas os professores tiveram contato com as teorias que balizam a elaboração de itens e puderam coletivamente praticar a construção e revisão para aperfeiçoamento dos itens. Essas ações de formação visam a um efeito multiplicador, tanto para o próprio docente que participava das oficinas, quanto para seus colegas de unidade educacional por meio da socialização das informações e experiências.

A prática de interpretação de descritores e elaboração de itens em oficinas com os pares demonstrou ser uma estratégia de aproximação dos docentes com as avaliações externas

produtivas, uma vez que no processo de elaboração os professores puderam refletir sobre a construção de enunciados e aprofundar a reflexão acerca da qualidade dos itens e instrumentos de avaliação.

Para garantir que o item vá a pré-teste é indispensável que ele passe pelo menos por um revisor técnico-pedagógico, para atender a todas as características descritas e um revisor de linguagem. O revisor técnico-pedagógico do item, na primeira leitura, deve devolver o item ao elaborador sempre que identificar um ou mais problemas de acordo com os critérios previamente estabelecidos. Para orientar a análise, registrar as observações e fundamentar a revisão técnica-pedagógica final do item, é utilizada uma ficha de revisão. Esta ficha é composta por critérios, divididos em blocos, relacionados a todos os aspectos do item, possibilitando uma análise detalhada.

O revisor técnico-pedagógico possui autonomia de fazer pequenas modificações no item necessárias para a sua aprovação, na primeira ou na segunda leitura, desde que não sejam muitos problemas.

Cabe ressaltar que o revisor técnico-pedagógico se constitui como parceiro do elaborador no processo de construção do item. A revisão funciona como momento de aprimoramento e refinamento do item para sua adequação aos critérios e normas preestabelecidas. Os itens que necessitam de revisão referente ao conteúdo devem ser encaminhados para revisão de um especialista.

6.1.3 A análise dos resultados de aplicação dos itens e dos instrumentos

A realização do pré-teste é fundamental para avaliar a qualidade dos itens elaborados nas oficinas e com isso verificar uma série de padrões que auxiliam na revisão e até mesmo no descarte de itens.

Os itens nas duas experiências utilizadas como base foram analisados pela Teoria Clássica dos Testes e o curso Técnico de Enfermagem de Saúde Materno e Infantil teve seus dados tratados pela Teoria de Resposta ao Item (TRI), metodologia estatística usada nas provas visando calibrar os itens e posicioná-los em uma escala. A TRI considera três parâmetros diferentes para sua formação: o grau de dificuldade das questões, de acerto casual e de discriminação dos itens, que assegura que participantes com conhecimentos distintos tenham graus diferentes de acerto. Ao expor os itens a uma população semelhante àquela que irá responder de fato à avaliação, ocorre a definição desses elementos.

As etapas de constituição da base de dados e o processamento foram realizadas na Universidade de São Paulo pelos membros do Gepave, uma vez que o Misau não conta com uma equipe técnica em tratamento estatístico e não houve a possibilidade dessas etapas serem realizadas em Moçambique com membros do Gepave ou no Brasil com membros do Misau. Caracterizando uma lacuna no processo, pois visando à formação e autonomia da equipe de Moçambique era de suma importância a capacitação de membros locais nessas etapas.

Após os processamentos, as análises e revisões foram desenvolvidas em oficinas contando com a participação de docentes e técnicos para interpretação pedagógica de resultados como referência para os níveis de proficiência aos quais estão associados os resultados e a definição dos pontos de corte dos critérios de avaliação e dos próprios termos e expressões da avaliação – do julgamento de seus objetos.

Essa etapa consistiu em uma oficina com o objetivo de construir uma escala e elaborar uma interpretação pedagógica dos itens, na qual os professores analisaram os itens aplicados no exame e realizam a escrita coletiva das sentenças descritoras. A sentença descritora explicita a tarefa que o respondente deverá desenvolver para responder ao item. O conjunto de tarefas desenvolvidas pelo estudante, caracterizadas pelo acerto na resolução dos itens, possibilitam posicioná-lo em uma escala.

Após elaboração das sentenças descritoras, as fichas foram expostas sobre uma mesa posicionadas de acordo com os blocos formados no tratamento estatístico. Os professores individualmente analisam os conjuntos de itens e apontam o bloco que consideravam o mínimo de tarefas que todos os alunos deveriam ser capazes de responder. As respostas de cada um dos professores são registradas e a partir do conjunto de respostas dos professores é demarcado o ponto de corte da escala onde nenhum aluno deveria estar abaixo.

Como apontado anteriormente, os professores são responsáveis pela avaliação das aprendizagens, entretanto não recebem formação inicial ou continuada em avaliação. Esse cenário não é diferente entre Moçambique e a Rede Municipal de Ensino de São Paulo e associa-se a ele o fato dos professores não possuírem formação de nível superior e/ou em Educação em Moçambique. Em sua maioria os docentes possuem formação técnica e por se destacarem e/ou possuírem o acúmulo de experiência na prática em saúde tornam-se professores nos institutos e ensinam por meio da socialização dos seus saberes e da reprodução de técnicas.

Diante deste quadro, faz-se necessária uma formação em Avaliação Educacional, tratando dos conceitos básicos em avaliação educacional, seus princípios, instrumentos e

técnicas, subsidiando os docentes para reflexão e articulação entre as teorias apresentadas e suas práticas.

Esse processo revelou-se potente uma vez que não somente apresentou teorias, mas proporcionou a escuta e o diálogo entre os participantes revelando concepções e práticas e contribuindo para formação inicial em avaliação educacional dos envolvidos.

A participação dos docentes no processo de discussão e elaboração dos itens, das matrizes e da análise dos resultados, constitui-se em um processo formativo de grande potencial uma vez que forma a equipe técnica local, com o intuito de promover a autonomia em relação à elaboração dos exames por terceiros, e promove a reflexão dos processos relacionados ao percurso formativo dos docentes envolvidos e do desenvolvimento do currículo proposto para os cursos avaliados. Assim, o processo em construção em Moçambique se contrapõe à tendência de organização de avaliações externas que, de maneira geral, são implementadas por técnicos/empresas especializadas, não havendo a participação dos docentes no processo.

6.1.4 Polêmica sobre certificação

Durante o desenvolvimento do Proforsa, nos deparamos com uma polêmica que, guardadas as proporções, se assemelha às políticas de *accountability*, pois o Misau havia decidido, em meados de 2017, desenvolver os Exames Nacionais dos Cursos Técnicos de Enfermagem Geral, Enfermagem em Saúde Materno-Infantil e Medicina Geral como avaliação para certificação, provavelmente extensível a outros cursos, e isso tornou necessário um posicionamento e a exposição da concepção de avaliação educacional que norteava o trabalho do Gepave, até então. Esta exposição teve o objetivo primordial de evidenciar as consequências negativas da certificação e a oposição nas condições propostas da formação profissional em saúde e seu antagonismo com as finalidades do Proforsa II, especialmente da proposta de avaliação externa em curso.

Na oportunidade destacou-se que, desde 2012, se enfatizava a concepção de que as avaliações externas, no que consistem os Exames Nacionais, deveriam ser organizadas com a finalidade de que seus resultados fossem incorporados pelos professores e pelas equipes gestoras nas Instituições de Formação (IdF), como parte da avaliação institucional – a avaliação que tem por objeto a IdF como um todo, seu projeto educacional e seus resultados –, tal como preconizado por Nevo (1998). Isso reforçava a noção de que a avaliação educacional,

embora disseminada mundialmente como um mecanismo para a excelência – para encontrar os "melhores" – que se traduz muito mais na exclusão de alunos do que em sua inclusão, como salientou Perrenoud (1999), pode e deve ter outras lógicas e finalidades.

Foi salientado, também, com base em com vários argumentos, a necessidade de avaliações mais democráticas e, igualmente, mais densas pedagogicamente, sobretudo, pela adoção da avaliação formativa, aquela que transforma os resultados da avaliação em referência para eventuais alterações no transcurso das disciplinas e cursos. Isto é, a avaliação que deve servir, especialmente, aos formadores para que possam rever as estratégias e procedimentos de ensino, o que, de certa forma, foi garantido nas oficinas de elaboração e revisão de itens e de provas.

Ponderou-se que a característica mais marcante de uma avaliação para a certificação está em sua finalidade de certificar – ou não – um concluinte de Curso Técnico de Saúde com base nos resultados de uma prova, normalmente, composta por itens de múltipla escolha. Em oposição a isso, apresentou-se que esses resultados devem ter outra finalidade. Mais ainda, que esses resultados não teriam condições de garantir quem deveria – ou não – ser certificado. Seria exigir algo que uma prova com itens de múltipla escolha não pode garantir.

Assim, a principal tarefa com a adoção de avaliações externas, sobretudo pelas resistências e impactos negativos constatados em vários países, como evidenciados por Bauer, Alavarse e Oliveira (2015), é integrá-las na construção de um verdadeiro sistema de avaliação que abarque vários elementos visando tirar o foco exclusivo do aluno, como tradicionalmente tem ocorrido, para incorporar as condições de ensino, como se encontra nas formulações e ponderações de Alavarse (2013). Nestes termos, mediante a análise e a interpretação dos resultados seria possível avaliar – julgar – a formação profissional de qualidade, com base, entre outros critérios, na medição do conhecimento dos alunos. Adicionalmente, permitiria práticas constantes, nas IdF, de remediação da aprendizagem dos alunos em função de suas necessidades, notadamente, tendo em vista a adoção de novos currículos.

Foi apresentado um conjunto de argumentos que justificavam a adoção dos referidos Exames Nacionais, sem transformá-los em instrumentos de avaliação para certificação, pois a adoção da certificação teria, entre outras, as seguintes implicações:

1- Retiraria das IdF a condição de ser o lugar privilegiado e autônomo para a certificação, pois são os professores e gestores das IdF quem devem desenvolver processos de acompanhamento dos alunos para, em diálogo com resultados de avaliações externas, garantir-lhes as condições para o incremento de suas aprendizagens e garantir uma formação nos mais elevados patamares de qualidade.

2- Aumentaria as perdas nesses Cursos, indo, portanto, no sentido contrário do que as condições de saúde pública de Moçambique exigiam, um dos motivos, aliás, para o desenvolvimento do Proforsa. Os registros de processos de certificação têm fatalmente levado a esses resultados. Deixar de efetuar a certificação nos Exames Nacionais não significa, em hipótese alguma, abandonar a preocupação com a formação de qualidade, mas significa considerar os prejuízos sociais obtidos com a ausência de profissionais da saúde que a adoção da certificação acarretaria. As justas preocupações com as limitações de capacidade técnica na conclusão de vários cursos deveriam conduzir à preocupação com suas origens ao que, tudo indica, estariam concentradas nas IdF.

3- Concentraria a atenção de professores e alunos na aplicação da prova padronizada, transformando os cursos em preparação para mesma, em detrimento de aspectos não abarcados por uma prova. Típica situação na qual a Matriz de Avaliação, que embasa a prova, se transformaria em referencial para o currículo, invertendo os processos, pois o mais indicado é que a avaliação seja uma decorrência dos processos de ensino baseados num currículo, estes, sim, a principal referência para a formação profissional. Ademais, desconsideraria aquilo que na literatura configurou-se denominar de corrupção dos processos, ao atribuir maior peso aos resultados de determinados instrumentos de avaliação, sempre limitados por suas configurações e condições de aplicação, do que para os procedimentos e condições de ensino.

4- Acentuaria a crença na capacidade de uma única prova poder dimensionar e avaliar as competências que os egressos de um curso deveriam dominar. Mesmo que bem elaborada, uma prova não tem essa propriedade, inclusive porque a avaliação em larga escala não contemplaria itens de atuação e sequer itens de resposta construída, dada à grande dificuldade de padronizar a aplicação e a pontuação numa avaliação em larga escala como ocorre com avaliações externas administradas nacionalmente. Para que uma avaliação seja válida é preciso que a mesma efetivamente avalie o que se propõe a avaliar. No caso da avaliação para a certificação isso significa que a prova pudesse medir um amplo leque de competências que um profissional precisa dominar para ser certificado. Contudo, isso demandaria um conjunto tão grande de itens na prova, incluindo formato de itens não apenas de múltipla escolha, que não há como garantir nas condições que se tem nos Exames Nacionais. Aliás, a experiência de algumas aplicações de exames de certificação comprovou cabalmente isso.

5- Impossibilitaria a ação supervisora no âmbito das IdF que se está procurando construir. A presença regular de supervisores deveria ter o propósito de oferecer apoio para a condução das atividades de ensino nos patamares exigidos para uma formação profissional de alto nível,

com certeza, muito mais do que preparar alunos para uma prova, o que, tudo indicaria, passaria a ser o principal objetivo da ação supervisora.

6- Enfatizaria ainda mais a responsabilidade dos alunos pelos resultados. Mesmo reconhecendo que os alunos devam ser chamados a assumir um papel ativo em sua própria formação, seguramente, aumentar a responsabilidade dos alunos não favorece a criação de um clima para práticas avaliativas formativas que sejam aceitas e implementadas nas IdF. Aliás, reforçariam práticas internas de avaliação que, supõe-se, já incidam nessa responsabilização, acentuando o caráter de exclusão das mesmas.

7- Reforçaria a noção meritocrática de que alguns terão sucesso e outros não. As condições particulares de Moçambique impõem que o sucesso escolar nos Cursos Técnicos seja para todos os seus alunos, o que, por decorrência, acarreta que professores e gestores assumiriam um lugar de protagonismo com vistas a esse sucesso de todos.

8- Criaria barreiras para que um frutuoso diálogo se estabeleça entre as avaliações externas, como são os Exames Nacionais com seus dados nacionais, e as práticas pedagógicas e didáticas aí incluídas aquelas denominadas de avaliação interna, as avaliações conduzidas pelos professores. Aliás, sobre as avaliações internas deveria recair a atenção no intuito de que fossem aperfeiçoadas para que não incorressem na imprecisão e na subjetividade, traços que parece lhes marcar. A pressão pelos resultados, própria da certificação, tenderia a se propagar para o dia a dia das IdF, reiterando a concepção de avaliação educacional como um processo excludente.

9- Exigiria a demanda imediata por qualificações técnicas em avaliação educacional para a elaboração de provas para certificação quando se constata que se está muito distante disso. Contudo, essa eventual qualificação na perspectiva da certificação, ainda que fosse colocada como uma tarefa, não transformaria a realização de uma avaliação externa como ponto de apoio para o trabalho pedagógico nas IdF. A validação de uma matriz, a construção de um banco de itens que eliminasse o viés de exposição e garantisse fidedignidade das medidas etc. são elementos que, no caso de certificação, demandaria um tempo e esforços que, pelas dificuldades vislumbradas, faria dessas provas instrumentos desprovidos das características indispensáveis para uma avaliação voltada para o sucesso de todos os alunos.

10- Intensificaria o processo de injustiça social, pois as condições sociais diferenciadas em Moçambique, principais fatores de diferenças educacionais, encontradas entre as Províncias e entre as escolas moçambicanas, tenderão a favorecer uns e prejudicar outros. Na perspectiva de justiça social, essas diferenças precisam ser atenuadas e para isso a avaliação educacional

deve ser marcada pela preocupação de se ter um horizonte, tanto material quanto político-educacional, de atenuá-las.

Assim, para que os objetivos do Proforsa II, no que tange à avaliação educacional, pudessem ser atingidos, os dirigentes do Misau reviram a posição sobre a finalidade dos Exames Nacionais enquanto avaliação para a certificação, compreendendo que o processo é constituído de etapas para construção de um instrumento qualificado, que a proposta não fosse adversa aos objetivos do Proforsa II e não impedissem que peritos associados à concepção de avaliação educacional esboçada se integrassem na próxima nova fase do projeto.

Durante a formação em avaliação educacional e todo o processo foram expostas e discutidas as etapas para a elaboração de uma avaliação externa, de acordo com o modelo proposto, ainda em construção, e mesmo havendo uma mudança de concepção no projeto, a avaliação ainda encontrava-se em um processo de inicial a intermediário, faltando muitas etapas para que o instrumento pudesse atender as características que justificassem seu uso para avaliar os alunos e subsidiar a certificação, uma vez que o próprio instrumento estava sob avaliação e construção.

A experiência narrada aponta para a necessidade de clareza das possibilidades e limites dos instrumentos de avaliação para as decisões políticas no âmbito educacional.

6.2 A experiência na Rede Municipal de Ensino de São Paulo: Prova Mais Educação

A Prova Mais Educação na RME-SP foi instituída pela portaria SME nº 3.611 de 29 de maio de 2015, sendo uma avaliação censitária, de larga escala, a ser aplicada bimestralmente por todas as Unidades Educacionais que mantinham o Ensino Fundamental da Rede Municipal de Ensino, a partir do 3º ano do Ciclo de Alfabetização e em todos os anos dos Ciclos Interdisciplinar e Autoral. Foram adotados procedimentos metodológicos formais e científicos para coletar e sistematizar dados, inclusive sobre as condições intraescolares que incidem sobre o processo de ensino e aprendizagem bem como produzir indicadores sobre as aprendizagens no Ensino Fundamental. Sendo composta por instrumentos para os componentes curriculares de Língua Portuguesa, Matemática e Ciências, com questões de múltipla escolha e de resposta construída (conhecidas como questões dissertativas ou abertas), para cada ano/turma.

Esta prova tinha por objetivos:

- contribuir para o desenvolvimento, em todos os níveis educativos, de uma cultura avaliativa que estimule a melhoria dos padrões de qualidade e equidade da educação paulistana e acompanhamento de seus resultados pelas Unidades Educacionais;

- contribuir para a melhoria da qualidade do ensino, redução das desigualdades e democratização da gestão do ensino público nos estabelecimentos de Ensino Fundamental da Rede Municipal de Ensino, em consonância com as metas e políticas estabelecidas pelas diretrizes curriculares;

- ser uma avaliação prevista em calendário definido oficialmente;

- oportunizar aos profissionais de educação da Rede Municipal de Ensino o acompanhamento do processo de ensino e aprendizagem por meio das informações sistemáticas sobre as Unidades Educacionais.

Importante destacar que não havia a intencionalidade da Prova Mais Educação substituir o Saeb, sob a responsabilidade do governo federal.

Os itens e os instrumentos da prova foram elaborados pela equipe da SME/DOT Núcleo de Avaliação Educacional, observados os padrões estabelecidos para avaliações padronizadas, com participação das equipes de DOT-P das DRE e de professores da RME-SP indicados por essas equipes.

Os itens elaborados por professores da Rede participantes do Curso de Elaboração de Itens promovido pela SME foram integrados ao banco de itens da Secretaria Municipal de Educação e, posteriormente, também seriam disponibilizadas no banco de itens para o Professor.

Foram disponibilizados materiais de apoio ao professor e às equipes gestoras das Unidades Educacionais em versão on-line, em links específicos da plataforma de avaliação, tais como: I- o Guia de Acesso da Prova Mais Educação: que constitui documento informativo sobre a realização de download dos instrumentos para reprodução; II- o Guia de Digitação da Prova Mais Educação: que discorre sobre os procedimentos de digitação e contém os gabaritos das questões de múltipla escolha e a grade de correção das questões de resposta construída (questões abertas); III- os relatórios de desempenho, de frequência e níveis percentuais que eram obtidos no prazo de 24 horas após a digitação dos resultados da Prova Mais Educação, que eram disponibilizados em dois formatos: por turma com informações individuais dos educandos e por Unidade Educacional; IV- o Caderno de Fichas Técnicas dos Itens, contendo a ficha detalhada de cada item, o descritor que deu origem ao item, outros descritores correspondentes e a justificativa de cada alternativa, disponibilizado a cada prova, 10 (dez) dias após o término da digitação dos resultados; V- o Caderno de

Análises Pedagógicas, contendo material de apoio para intervenções do professor e da equipe gestora das Unidades Educacionais, disponibilizado a cada prova, 10 (dez) dias após o término da digitação dos resultados na plataforma de avaliação.

Houve sob responsabilidade das DOT-P das DRE, a partir das orientações da SME/DOT Núcleo de Avaliação Educacional, de acordo com o Programa Avaliar para Aprender, um processo de formação para análise, interpretação e encaminhamentos referentes às estratégias para disseminação dos resultados obtidos na Prova Mais Educação para respectivas Unidades Educacionais, considerando o caráter de complementariedade das diferentes dimensões avaliativas dispostas nas legislações em vigor, sobretudo a avaliação para a aprendizagem.

A Secretaria Municipal de Educação em conjunto com as Diretorias Regionais de Educação - DRE, por meio da Supervisão Escolar e das Diretorias de Orientação Técnico-Pedagógica - DOT-P, ofereceram suporte técnico-pedagógico e administrativo às Unidades Educacionais, de modo a garantir a plena realização das atividades da Prova Mais Educação, inclusive no que se refere ao apoio técnico e aos recursos específicos que viabilizassem a participação dos educandos público-alvo da Educação Especial. Constituindo-se uma exceção em avaliações externas em larga escala com adequações para estudantes com diferentes barreiras pois foi ofertada a prova em Libras, Braile, ampliada, ledor, escriba e guia-intérprete, oportunizando a inclusão dos estudantes público-alvo da educação especial da RME-SP.

O direito à avaliação para aprendizagem dos educandos, público-alvo da Educação Especial, considerou a elaboração de diferentes formas e instrumentos, respeitadas as condições próprias de cada educando. A aplicação da Prova Mais Educação pressupôs decisão conjunta, respeitadas as discussões e orientações dos profissionais da Unidade Educacional e do Supervisor Escolar, podendo contar com o apoio do Centro de Formação e Acompanhamento à Inclusão - CEFAI e do Professor Regente da Sala de Apoio e Acompanhamento à Inclusão - SAAI, em concordância com a opção do aluno e da família, não sendo a deficiência motivo de impedimento para a participação dos educandos.

Foram realizadas oficinas com representantes das equipes gestoras, coordenadores pedagógicos e professores de todas as unidades educacionais para análise e interpretação pedagógica e estatística dos itens e dos resultados.

O processo de elaboração e análise dos itens para constituição da prova, bem como as oficinas para análise e interpretação estatística e pedagógica posteriores a aplicação, constituíram-se em um rico processo formativo em avaliação educacional aproximando os

professores, coordenadores pedagógicos e gestores escolares do que foi tratado no programa como “caixa preta da avaliação”, pois até então o banco de itens e a elaboração das provas da rede e seus itens eram tratados com sigilo absoluto sem acesso aos profissionais da educação da RME-SP.

7. INDICAÇÕES PARA UMA AVALIAÇÃO EXTERNA

Como apresentado até aqui, grandes são os desafios oriundos da avaliação externa, sobretudo considerando sua concepção e sua implementação para que seja, de fato, reconhecida e apropriada pelos professores, ainda mais quando se vislumbra constituir-se como ponto de apoio para a formação docente em avaliação educacional.

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 aumentar as 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.

Para Afonso (2014, p. 495),

> [...] a introdução de bonificações (que é uma forma parcelar de uma conce[p]ção conservadora de *accountability*) ilude e protela o enfrentamento das questões anteriores e de outras questões fundamentais. Diversos estudos revelam que os seus efeitos ou não são os esperados, ou aparecem referidos como não sendo conclusivos relativamente aos objetivos pretendidos, ou são vistos como tendo consequências contraproducentes. Não critico, porém, os professores ou as escolas que, perante salários muito baixos e escassez de recursos públicos disponíveis, aderem a estas bonificações. O que digo é que elas são um paliativo que não resolve (nem sequer ameniza) os problemas estruturais. Neste sentido, partilho do ponto de vista de todos aqueles que referem ser muito melhor, mais urgente e mais justo, aumentar significativamente os salários dos professores, adotar modelos de avaliação amplamente discutidos e democraticamente consensualizados, e dignificar de vez o seu estatuto socioprofissional.

Como afirma William Smith (2016, p. 7), "a cultura avaliativa global tem uma qualidade confirmatória que dá origem a um ambiente em que a avaliação se torna sinônimo de responsabilidade. Os antecedentes históricos da cultura avaliativa global no campo da educação e prestação de contas que se torna sinônimo de qualidade educacional. Em outras

palavras, a noção de uma cultura global avaliativa reflete a constatação de que práticas que envolvem o uso de rankings, indicadores de desempenho e responsabilidade e prestação de contas com base nos resultados de diversas avaliações são evidentes em todo o mundo (Rizvi; Lingard, 2010). Além disso, segundo Morgan (2011, p. 56), o Programa Internacional de Avaliação de Estudantes (Pisa) funciona como uma "máquina de apontar responsabilidades e exigir prestação de contas". Assim, o que pode ser chamado de cultura global avaliativa está intimamente relacionado ao que Pasi Sahlberg chama de Movimento Global de Reforma da Educação (GERM). Trata-se de uma abordagem de reforma educacional que, em linhas gerais, segue os princípios do neoliberalismo e da nova gestão pública (NPM). Está estruturado em torno de um conjunto compartilhado de ideias em relação às políticas públicas que incluem gestão baseada em padrões, avaliação de desempenho e responsabilidade e prestação de contas (SAHLBERG, 2016). A conclusão dessas observações é que os dados sobre os resultados e padrões de aprendizagem dos alunos são o foco dos formuladores de políticas (ADDEY et al., 2017)

Afonso (2014, p. 494) ainda destaca que

> [...] passará ainda muito tempo até que se perceba que não é com políticas culpabilizantes e formas autoritárias de controle, nem mesmo com bonificações conjunturalmente compensatórias, que se pode dignificar a profissão docente e conseguir que a escola pública passe a estar ao serviço, não de uma discriminação classista meritocrática (encoberta por uma suposta neutralidade ou igualdade formal de oportunidades), mas sim ao serviço de uma "pluralidade de excelências".

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.

7.1 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. 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. Em 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 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. Quatro, 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 houvesse mais explicações em relação à 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, dando mais relevância

ao 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 à falta de participação de mais agentes educacionais nas discussões de seleção das prioridades a serem avaliadas, entre outras partes dos processos que, se investidos de forma criteriosa, poderiam contribuir para diminuir o distanciamento dos professores da avaliação externa, seus resultados e potencialidades.

7.2 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 àquilo 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 à validade e até mesmo quanto às 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 *característica de um teste*: a medida em que mede aquilo que pretende medir.	Validade é considerada um *argumento* 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 *condição necessária para a validade*.	Confiabilidade pode ser vista como *um tipo de evidência de validade*.
A validade era frequentemente estabelecida através de *correlações* de um teste com outros.	Validade é argumentada com base em um número de tipos de *justificativas* e evidências, incluindo as consequências da avaliação.
Validade de construto era vista como um dos *três tipos de validade* (conteúdo, relacionada a critério e construto).	Validade é um *conceito unitário*, 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 *pesquisadores da avaliação*, responsáveis pelo desenvolvimento de testes de grande escala e alta relevância.	A justificativa de validade de um teste é de responsabilidade de *todos os usuários de um teste*.

FONTE: CHAPELLE (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 pode ser simplificado. 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 elaboram os critérios.

Enfim, acreditamos que estas propostas se contrapõem 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.

Estudos apontam que nos modelos usuais de avaliação externa, no qual as provas são elaboradas por especialistas e empresas terceirizadas, 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. Por outro lado, esses mesmos resultados são apresentados em termos globais, sem identificação, como seria desejável, das unidades educacionais, referindo-se muitas vezes de forma genérica às aprendizagens dos alunos, limitando encaminhamentos que os colocassem a serviço do trabalho docente. 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, como afirma Stecher (2002), mesmo no quadro de políticas de excessiva responsabilização dos docentes.

Ravitch (2010) analisa como as novas reformas se distanciaram da preocupação fundamental com o currículo e passaram a acreditar que mudanças na gestão e na estrutura do sistema seriam prontamente a solução para os problemas da educação. Os princípios empresariais de gestão e contratação de profissionais, escolha, recompensas e punições para incentivar a força de trabalho, decisões e metas baseadas em um bom sistema de dados, passam a ser soluções por si mesmas, negligenciando as dimensões pedagógica e política da educação. Por fim, a autora endossa que o sistema educacional defina claramente seus objetivos, através de um currículo rico e coerente que englobe uma formação abrangente, além das habilidades básicas, que têm sido o alvo restrito dos testes atuais.

Com vistas a esse diálogo, devemos ponderar que a avaliação da aprendizagem interna é aquela que ocorre nos espaços escolares tendo educadoras/educadores como sujeitos, isto é, como aqueles que têm um controle quase absoluto sobre essa avaliação, inclusive dando-lhes uma aura de poder sobre os alunos, mesmo que sendo anunciadas perspectivas diagnóstica e formativa, pois tudo indica que a principal característica dessa avaliação é ser somativa para fins de aprovação ou reprovação desses alunos. Ademais, várias pesquisas indicam que são escassas as ações pedagógicas orientadas pelos resultados das avaliações internas. Alavarse (2013) salienta que professores, apesar desse "poder" exercido por meio da avaliação, raramente possuem formação para a organização e implementação de avaliações que sejam mais consistentes. Dessa constatação, Alavarse (Op. cit.) não propõe a substituição dos professores como avaliadores, mas, sim, a tomada de consciência de várias, e eventualmente severas, restrições, como é o caso da constatação frequente de forte subjetividade na emissão de juízos dos professores sobre as aprendizagens de seus alunos.

Este trabalho não visa a superação dessas restrições, porém se inscreve num esforço sistemático e articulado de envolver professores em etapas críticas de uma avaliação externa como forma, inclusive, de formação em avaliação educacional, algo que pode vir a ter influência nas avaliações internas. Nestes termos, a construção de uma avaliação externa da aprendizagem, em larga escala, deve propiciar oportunidades, ainda que não necessariamente para todos os professores, de contato com procedimentos, conhecimentos, técnicas etc., relativos à organização de uma avaliação que, efetivamente, possa subsidiar de modo mais consistente o processo de ensino.

Ainda que as avaliações externas não devam pautar o currículo e, consequentemente, o ensino, estas devem integrar o diálogo, destacado por Nevo (1998), quer por meio das discussões sobre suas matrizes na medida em que estas, se devidamente validadas, demarcam pontos de partida para a elaboração de itens – as tarefas avaliativas – que são submetidos aos alunos, quer pela interpretação pedagógica dos resultados das avaliações externas. No que tange as matrizes, um aspecto relevante diz respeito às conexões com os conteúdos desenvolvidos pelo currículo. Igualmente, devemos considerar os resultados das avaliações externas como a possibilidade de as equipes das instituições possuírem um panorama mais amplo de avaliação do trabalho na Unidade Educacional ao disporem de dados de outras escolas e isso não pode ser confundido com uma simples hierarquização de uma escola em relação a outras.

Acreditamos que o envolvimento dos docentes e equipes técnicas em todas as etapas do desenvolvimento das avaliações, desde a constituição das matrizes, banco de itens e provas, até o tratamento dos dados e análises dos resultados, represente um rico processo formativo que contribui para o desenvolvimento de uma avaliação que se aproxima o máximo possível daquilo que em termos de currículo é desenvolvido durante o percurso formativo e que o envolvimento de todos no processo tenha importante papel na formação docente, colaborando para a apropriação dos resultados pelos institutos, favorecendo seu uso pedagógico e possibilitando a autonomia das equipes técnicas em relação à terceirização de serviços e contratações externas. Sobretudo, aproxima os professores das discussões que apontam as vantagens e limitações das avaliações externas em larga escala de forma a ressignificar o papel das avaliações externas e consequentemente as internas.

Não obstante, esses resultados podem produzir sinalizações para que essas equipes tenham os citados referenciais como um dos balizadores para a organização da avaliação da aprendizagem interna às escolas.

8. CONSIDERAÇÕES FINAIS

Consideramos que a pouca participação dos professores no processo de construção das avaliações externas constitui uma limitação identificada como uma tendência nessas avaliações. Isso, por sua vez, tem restringido tanto a utilização de seus resultados quanto tem apresentado restrições no tocante à aceitação por parte de docentes. No que diz respeito à possibilidade tivemos dois desafios. O primeiro se refere à própria formulação de um modelo. O segundo, quanto à sua viabilização da aplicabilidade desse modelo em escolas e redes de escolas, o que nos restringiu em sugerir possibilidades práticas de envolvimento dos professores nas etapas de construção das avaliações externas.

Inclusive como um fundamento empírico, as experiências de Moçambique e da RME-SP demonstraram ser possível e necessário que iniciativas de envolvimento de professores na construção de uma avaliação externa se constituam em importante ponto de apoio para as avaliações internas, pois as oficinas, como uma das atividades, possibilitaram reflexões acerca dos resultados com interpretação pedagógica, dos instrumentos padronizados, dos formatos de itens, do cotejamento desses resultados com as notas atribuídas pelos docentes e das discussões das matrizes e interpretação pedagógica de resultados com os planos de ensino. Ademais de fortalecer a legitimidade política das avaliações externas, mesmo com a ponderação de que essas experiências padecessem de limitações de várias ordens. Trata-se, isso sim, de demarcar como se pode engajar professores nesse tipo de avaliação e, simultaneamente, desenvolver formação em avaliação educacional; uma resposta, ainda que não absolutamente conclusiva, para nosso problema de pesquisa.

Como um movimento inicial de elaboração teórico-prescritiva, apontamos que uma 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- constituir-se 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 e o currículo.

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.

Por não conhecer ou participar da elaboração e demais processos que envolvem as avaliações externas, articuladas aos usos nocivos de seus resultados, observa-se uma tendência da não utilização de seus dados/resultados pedagogicamente.

O envolvimento de professores na elaboração de etapas importantes de uma avaliação externa seria uma possível saída para este conflito podendo ser um fator que os potencialize como usuários, sobretudo, pelo fato de que começam a entender os meandros desse processo e, simultaneamente, no entendimento dos conteúdos desses resultados. Adicionalmente, ainda que não altere a característica da avaliação como externa, essa participação cria laços e diminui a rejeição que, muitas vezes, espontaneamente surge simplesmente por se tratar de uma avaliação externa. Isso também pode colaborar para qualificação dos instrumentos por eles elaborados nas avaliações internas, uma vez que propicia a formação em avaliação educacional, sobretudo quanto as técnicas de elaboração e revisão de itens. O que possibilitaria o diálogo entre as avaliações internas e externas como proposto por Nevo (1998); reiterando, igualmente, que não se propõe a substituição dos professores como avaliadores, mas, sim, a tomada de consciência de várias, e eventualmente severas, restrições, como é o

caso da constatação frequente de forte subjetividade na emissão de juízos dos professores sobre as aprendizagens de seus alunos.

Diante desse quadro, um importante passo é a discussão sobre o sentido do que está sendo avaliado. E o vínculo entre esses aspectos é, em geral, explicitado através da construção do que usualmente se denomina de Matriz de Referência. A explicitação dessa matriz e da forma como ela é utilizada nas avaliações permite acrescentar sentidos mais pedagógicos ao processo, além de significados aos índices e percentuais que normalmente expressam o resultado dessas avaliações. E de maneira geral, essa Matriz de Referência é fruto de um currículo, cujo papel principal é ser um balizador das expectativas educacionais e que orientam o processo de ensino e aprendizagem das instituições de ensino.

Tendo por foco específico o sentido das matrizes, Costa (1995) compara a necessidade e viabilidade de uma matriz como o suporte fornecido por um mapa para um viajante. Quando não está munido de um mapa, o viajante pode ter sua viagem comprometida, menos segura e proveitosa.

Portanto, podemos concluir, a partir dessas considerações, que a matriz de referência possui papel preponderante no processo de construção de um sistema de avaliação, especificamente na construção dos itens. Além disso, analisar uma matriz significa, então, buscar identificar sua relação com o currículo prescrito, e também, sua relação com os objetivos e finalidades das diretrizes curriculares mais gerais. Essa relação, por sua vez, se expressa na própria organização ou estruturação das matrizes, na medida em que nela se explicitam a forma pela qual estão sendo propostas articulações de conhecimentos/conteúdos e aspectos cognitivo-formativos.

Como exposto por Ferrer (2006), devem existir metas nacionais, portanto, um compromisso assumido e conhecido por todos: professores, diretores e sociedade em geral, sendo interessante aproximar a equipe técnica de avaliação e os profissionais de educação.

A proposta de incorporação dos professores, coordenadores e gestores nas etapas das avaliações externas tem, também, como preocupação essencial a formação dos professores em avaliação educacional, em geral, e em avaliação da aprendizagem, em particular, na compreensão de que os modelos em voga revelam uma concepção pela qual bastariam a divulgação dos dados ou, como encontramos, entres outras fontes, em Becher e Justo (2020), 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.

Tomamos como concepção que as avaliações externas devem ser organizadas com a finalidade de que seus resultados sejam incorporados pelos professores e pelas equipes gestoras das instituições de ensino, como parte da avaliação institucional – a avaliação que tem por objeto a instituição de ensino como um todo, seu projeto educacional e seus resultados –, tal como preconizado por Nevo (1998). Isso reforça a noção de que a avaliação educacional, embora disseminada mundialmente como um mecanismo para a excelência – para encontrar os "melhores" – que se traduz muito mais na exclusão de alunos do que em sua inclusão, como salientou Perrenoud (1999), pode ter outras lógicas e finalidades.

Realizando-se a defesa, com vários argumentos, da necessidade de avaliações mais democráticas e, igualmente, mais densas pedagogicamente, sobretudo, pela adoção da avaliação formativa, aquela que transforma os resultados da avaliação em referência para eventuais alterações no transcurso das disciplinas e cursos. Isto é, a avaliação que deve servir, especialmente, aos formadores para que possam rever as estratégias e procedimentos de ensino, o que, de certa forma, propõe-se garantir por meio da participação dos docentes nas oficinas de elaboração e revisão de itens e de provas.

O envolvimento de professores na elaboração de etapas importantes de uma avaliação externa pode ser um fator que os potencialize como usuários, sobretudo, pelo fato de que começam a entender os meandros desse processo e, simultaneamente, no entendimento dos conteúdos desses resultados. Adicionalmente, ainda que não altere a característica da avaliação como externa, essa participação oportuniza a criação de laços, o que potencialmente diminuiria a rejeição que, muitas vezes, espontaneamente surge simplesmente por se tratar de uma avaliação externa. Ou seja, parece que essa exterioridade é atenuada por esse envolvimento. Esses elementos devem potencializar o uso dos resultados pelas equipes pedagógicas e docentes, na perspectiva daquilo que Nevo (1998) denominou como “um diálogo entre suas avaliações e as externas”, especialmente para fins de melhorar a qualidade dos cursos. Assim, uma maior participação poderia fazer com que essa exterioridade não fosse um obstáculo para o reconhecimento da avaliação externa como um expediente que pode contribuir com o processo pedagógico.

Nestes termos, a construção de uma avaliação externa da aprendizagem, em larga escala, deve propiciar oportunidades, ainda que não necessariamente para todos os professores, de contato com procedimentos, conhecimentos, técnicas etc., relativos à organização de uma avaliação que, efetivamente, possa subsidiar de modo mais consistente o processo de ensino.

REFERÊNCIAS

ADDEY, Camilla et al. The rise of international large-scale assessments and rationales for participation. **Compare: A Journal of Comparative and International Education**, p. 1-19, April 7th, 2017.

AFONSO, Almerindo Janela. **Políticas educativas e avaliação educacional**. Braga: Universidade do Minho, 1998.

AFONSO, Almerindo Janela. **Avaliação educacional**: regulação e emancipação: para uma sociologia das políticas avaliativas contemporâneas. 4. ed. São Paulo: Cortez, 2009. [Original 2000]

AFONSO, Almerindo Janela. Questões, objetos e perspectivas em avaliação. **Avaliação**, Campinas, Sorocaba, v. 19, n. 2, p. 487-507, jul. 2014.

ALAVARSE, O. M. **Avaliações externas**: perspectivas para o trabalho pedagógico. São Paulo: Centro de Estudos e Pesquisas em Educação, Cultura e Ação Comunitária – Cenpec, 2012.

ALAVARSE, Ocimar Munhoz. 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, Ocimar Munhoz; BRAVO, Maria Helena; MACHADO, Cristiane. 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.

AMARAL, M. F. B.; ALAVARSE, O. M.; SILVA, F. A. 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. **Olhar de Professor**, Ponta Grossa, v. 23, p. 1-18, 2020.

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION – AERA; AMERICAN PSYCHOLOGICAL ASSOCIATION – APA; NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION – NCME. **Standards for educational and psychological testing**. Washington, DC: American Educational Research Association, 2014.

ANDERSON, Lorin W.; KRATHWOHL, David R. (Ed.). **A taxonomy for learning, teaching and assessing**: a revision of Bloom's taxonomy of educational objectives. Complete ed. New York: Longman, 2001.

ANDRADE, E. L. M. **A formação continuada dos diretores escolares no contexto da política pública das avaliações externas**. Dissertação (Mestrado em Educação Profissionalizante em Gestão e Avaliação da Educação Pública) – Universidade Federal de Juiz de Fora. Juiz de Fora, 2012.

BAUER, A. 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; ALAVARSE, Ocimar Munhoz; OLIVEIRA, Romualdo Portela de. Avaliações em larga escala: uma sistematização do debate. **Educação e Pesquisa**, São Paulo, v. 41, n. especial, p. 1367-1382, dez. 2015.

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.

BECKER, Fernanda da Rosa. Avaliação educacional em larga escala: a experiência brasileira. **Revista Iberoamericana de Educación**, v. 53, p. 1-11, 25 jun. 2010.

BECHER, Ednei Luís; JUSTO, Jutta Cornelia Reuwsaat. 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.

BLACK, Paul; WILIAM, David. Assessment for learning in the classroom. In: GARDNER, J. (Ed.). **Assessment and learning**. London: Sage, 2006. p. 9-26.

BLOOM, Benjamin S., HASTINGS, J. Thomas; MADAUS, George F. **Evaluación del aprendizaje**. Buenos Aires: Troquel, 1975.

BLOOM, Benjamin S.; HASTINGS, J. Thomas; MADAUS, George F. **Manual de avaliação formativa e somativa do aprendizado escolar**. Tradução de Lilian Rochlitz, Maria Cristina Fioratti Florez e Maria Eugênia Vanzolini. São Paulo: Pioneira, 1983. (Biblioteca Pioneira de Ciências Sociais). [Original 1971]

BONAMINO, Alicia; SOUSA, Sandra Maria Zákia Lian. 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, p. 373-388, abr./jun. 2012.

BROOKE, Nigel; CUNHA, Maria Amália de A. Cunha. A avaliação externa como instrumento da gestão educacional nos estados. **Estudos & Pesquisas Educacionais**, São Paulo, n. 2, p. 17-79, nov. 2011.

BROOKE, Nigel P.; CUNHA, M. A.; FALEIROS, M. **A avaliação externa como instrumento da gestão educacional nos estados**: relatório final. Belo Horizonte: Game/UFMG; Fundação Victor Civita, 2011.

CASASSUS, Juan. El precio de la evaluación: la perdida de calidad y la segmentación social. **Revista Brasileira de Política e Administração da Educação**, Porto Alegre, v. 23, n. 1, p. 71-79, jan./abr. 2007.

CASASSUS, Juan. Uma nota crítica sobre a avaliação estandarizada: a perda de qualidade e a segmentação social. Tradução de Gabriela Lourenço. **Sísifo: Revista de Ciências da Educação**, n. 9, p. 71-78, maio/ago. 2009.

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

CASTRO, Maria 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.

CHAPPAZ, R.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.

CRAHAY, Marcel. **Poderá a escola ser justa e eficaz?**: da igualdade das oportunidades à igualdade dos conhecimentos. Tradução de Vasco Farinha. Lisboa: Instituto Piaget, 2002. (Horizontes Pedagógicos, 92). [Original 2000].

CRAHAY, Marcel. Como a escola pode ser mais justa e mais eficaz? Tradução de Fernando Santos. **Cadernos Cenpec**, São Paulo, v. 3, n. 1, p. 9-40, jun. 2013.

DEMO, Pedro. **Metodologia científica em ciências sociais**. São Paulo: Atlas, 1981.

ESQUINSANI, R. S. S. Tá lá, em cima da mesa: os dados das avaliações em larga escala e a mediação do coordenador pedagógico. In: WERLE, Flávia Obino Corrêa (Org.). **Avaliação em larga escala**: foco na escola. São Leopoldo: Oikos; Brasília: Liber Livro, 2010, p. 134-146.

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

ESTEBAN, Maria Teresa. A negação do direito à diferença no cotidiano escolar. **Avaliação**, Campinas, Sorocaba, v. 19, n. 2, p. 463-486, jul. 2014.

FERNANDES, Reynaldo; GREMAUD, Amaury Patrick. Avaliação educacional em larga escala e *accountability*: uma breve análise da experiência brasileira. **Revista Pesquisa e Debate em Educação**, Juiz de Fora, v. 10, n. 1, p. 1103-1137, jan./jun. 2020.

FERRER, Guillermo. **Sistemas de evaluación de aprendizajes en América Latina**: balance y desafíos. Santiago: Preal, marzo 2006. (Documentos, n. 43).

FRANCO, Creso. 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. Caminhos da avaliação de sistemas educacionais no Brasil: o embate entre a cultura da auditoria e a cultura da avaliação. In: BAUER, Adriana; GATTI, Bernardete A.; TAVARES, Marialva R. (Org.). **Vinte e cinco anos de avaliação de sistemas educacionais no Brasil**: origens e pressupostos. Florianópolis: Insular, 2013. p. 147-176.

FREITAS, Pâmela Félix. **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, Bernardete. **Avaliação e qualidade da educação**. Trabalho apresentado no Simpósio Brasileiro de Política e Administração da Educação. Porto Alegre, 2007.

_____ et al. Formação de professores para o ensino fundamental: instituições formadoras e seus currículos. **Estudos & Pesquisas Educacionais**, São Paulo, n. 1, p. 95-138, maio 2010.

GONÇALVES, A. L.; LARCHERT, J. M. **Avaliação da aprendizagem**: Pedagogia, módulo 4, v. 6. Ilhéus, BA: Editus, 2011.

GONZAGA, A. M. A pesquisa em educação: um desenho metodológico centrado na abordagem qualitativa. In: PIMENTA, Selma Garrido (Org.). **Pesquisa em educação**: alternativas investigativas com objetivos complexos. São Paulo: Loyola, 2006. p. 65-92.

HAYDT, R. C. **Avaliação do processo ensino-aprendizagem**. São Paulo: Ática, 1995.

HAYDT, R. C. **Curso de Didática Geral**. 4. ed. São Paulo: Ática, 1997.

HORTA NETO, João Luiz. **Limites para a utilização dos resultados de avaliações externas estandardizadas**: o caso da utilização do Saeb por um ente federado. Trabalho apresentado no I Congresso Ibero-Brasileiro "Espaço Público da Educação: emergência de políticas e práticas de gestão local, regional e nacional". Elvas, 2010.

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

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

LAFOURCADE, Pedro Dionisio. **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, Cipriano Carlos. **Avaliação da aprendizagem escolar**. São Paulo: Cortez, 2005.

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

LUKAS MUJIKA, Jose Francisco; SANTIAGO ETXEBERRIA, Karlos. Concepto, componentes y fases de la evaluación. In: _____. **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, Cristiane. Avaliação externa e gestão escolar: reflexões sobre usos dos resultados. **Revista Ambiente e Educação**, v. 5, n. 1, p. 70-82, jan./jun. 2012.

MACHADO, Cristiane; ALAVARSE, Ocimar Munhoz. Qualidade das escolas: tensões e potencialidades das avaliações externas. **Educação & Realidade**, Porto Alegre, v. 39, n. 2, p. 413-436, abr./jun. 2014.

MADAUS, George F. The influence of testing on the curriculum. In: TANNER, Laurel 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, George F.; RUSSEL, M.; HIGGINS, J. **The paradoxes of high stakes testing**: how they aff ect students, their parents, teachers, principals, schools, and society. Charlott e, NC: Information Age Publishing, 2009.

MARCHI, S. M. **Concepções de professores acerca da avaliação externa e suas decorrências nas práticas pedagógicas de uma escola de Ensino Fundamental**. Dissertação (Mestrado em Educação) Centro Universitário Moura Lacerda. Ribeirão Preto, 2012.

MESSICK, Samuel. 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, Samuel. **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, Mariana; SOLÉ, Isabel. 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]

NEAL, D.; SCHANZENBACH, D.W. Left behind by design: proficiency counts and test-based accountability. **The Review of Economics and Statistics**, p. 263-283, May 2010.

NÉRICI, I. G. **Metodologia do ensino**: uma introdução. São Paulo: Atlas, 1977.

NEVO, David. **School-based evaluation**: a dialogue for school improvement. Oxford: Pergamon, 1995.

NEVO, David. Avaliação por diálogos: uma contribuição possível para o aprimoramento escolar. In: TIANA, Alejandro (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.

NEVO, D. Evaluation in education. In: SHAW, I. F.; GREENE, J. C.; MARK, M. **Handbook of evaluation**: policies, programs and practices. London: Sage, 2005. p. 441-460.

NICHOLS, S. L.; BERLINDER, D. C. **Collateral damage**: how high-stakes testing corrupts America's schools. Cambridge, MA: Harvard Education Press, 2007.

OECD. The PISA 2003 Assessment Framework – Mathematics, Reading, Science and Problem Solving Knowledge and Skills. Paris: OECD, 2003.

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.

OLIVEIRA, Romualdo Portela de et al. Análise das desigualdades intraescolares no Brasil. **Estudos e Pesquisas Educacionais**, v. 4, p. 19-112, nov. 2013.

PERRENOUD, Philippe. **Avaliação**: da excelência à regulação das aprendizagens: entre duas lógicas. Tradução de Patrícia Chittoni Ramos. Porto Alegre: Artes Médicas Sul, 1999.

PIMENTA, Cláudia de Oliveira. **As avaliações externas e o trabalho de coordenadores pedagógicos**: estudo em uma rede municipal paulista. Dissertação (Mestrado em Educação) – Faculdade de Educação, Universidade de São Paulo. São Paulo, 2012.

RAVITCH, Diane. **The death and life of the great American school system**: how testing and choice are undermining education. New York: Basic Books, 2010.

RIZVI, F.; LINGARD, B. **Globalizing education policy**. London: Routledge, 2010.

ROJAS, Hugo de los Santos. Formação do professor do ensino básico e a avaliação educacional. **Estudos em Avaliação Educacional**, São Paulo, v. 18, n. 37, p. 7-40, maio/ago. 2007.

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, Michael. The methodology of evaluation. In: TYLER, Ralph Winfred; GAGNE, Robert; SCRIVEN, Michael. **Perspectives of curriculum evaluation**. Chicago, IL: Rand McNally, 1967. (AERA Monograph Series on Curriculum Evaluation). p. 39-83.

SAHLBERG, Pasi. The global educational reform movement and its impact on schooling. **The handbook of global education policy**, v. 12, n. 4, p. 128-144, 2016.

SILVA, M. O. E. **A análise de necessidades de formação na formação contínua de professores**: um caminho para a integração escolar. Tese (Doutorado em Educação) – Faculdade de Educação da Universidade de São Paulo. São Paulo, 2000.

SILVA, M. J. A. **Regulação educativa: o uso de resultados de proficiência de avaliações do Proeb por diretores escolares em Minas Gerais**. Tese (Doutorado em Educação) – Faculdade de Educação, Universidade Federal de Minas Gerais. Belo Horizonte, 2011.

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.

SILVA, Marcela Mara dos Santos et al. A abordagem da avaliação educacional em larga escala nos cursos de graduação em Pedagogia. **Revista Brasileira de Estudos Pedagógicos**, Brasília, v. 97, n. 245, p. 46-67, jan./abr. 2016.

SIQUEIRA, Valéria Aparecida de Souza; FREITAS, Pâmela Félix; ALAVARSE, Ocimar Munhoz. Professores e lacunas formativas em avaliação da aprendizagem: evidências e problematizações. **Educação e Pesquisa**, São Paulo, v. 47, e241339, p. 1-17, 2021.

SMITH, W. C. (Ed.). **The global testing culture**: shaping education policy, perceptions, and practice. Oxford: Symposium Books, 2016.

SOUSA, Sandra M. Z. L. Possíveis impactos das políticas de avaliação no currículo escolar. **Cadernos de Pesquisa**, São Paulo, n. 119, p. 175-190, jul. 2003.

SOUSA, Sandra 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, Sandra M. Z. L.; OLIVEIRA, Romualdo Portela de. 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.

SOUSA, Sandra M. Z. L.; OLIVEIRA, Romualdo Portela de. **Sistemas de avaliação educacional no Brasil**: característica, tendências e uso dos resultados: relatório final. São Paulo: Faculdade de Educação da USP, 2007.

SOUSA, Sandra Maria Zákia Lian; OLIVEIRA, Romualdo Portela de. 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.

STECHER, Brian M. Consequences of large-scale, high-stakes testing on school and classroom practice. In: HAMILTON, Laura; STECHER, Brian M.; KLEIN, Stephen P. (Ed.). **Making sense of test-based accountability in education.** Santa Monica, CA: Rand, 2002. p. 79-100.

TARDIF, Maurice. A profissionalização do ensino passados trinta anos: dois passos para a frente, três para trás. Tradução Marisa Rossetto. **Educação & Sociedade**, Campinas, v. 34, n. 123, p. 551-571, jun. 2013.

VASCONCELLOS, Celso dos Santos. **Avaliação da aprendizagem**: práticas de mudança – por uma práxis transformadora. São Paulo: Libertad, 2003.

VIANNA, Heraldo Marelim. **Introdução à avaliação educacional**. São Paulo: Ibrasa, 1989. (Biblioteca Educação, 40). p. 19-20.

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

VILLAS BOAS, Benigna Maria Freitas; SOARES, Sílvia Lúcia. O lugar da avaliação na formação de professores. **Cadernos Cedes**, Campinas, v. 36, n. 99, p. 239-254, maio/ago. 2016.

WELTER, C. B.; WERLE, S. M. M. Avaliação externa em Picada Café: um relato de experiência. In: WERLE, Flávia Obino Corrêa (Org.). **Avaliação em larga escala**: foco na escola. São Leopoldo: Oikos; Brasília: Liber Livro, 2010. p. 98-109.

WERLE, Flávia Obino Corrêa. 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.