

CRESCIMENTO E DISSEMINAÇÃO DE AVALIAÇÕES EM LARGA ESCALA E DE RESPONSABILIZAÇÕES BASEADAS EM TESTES: UMA SOCIOLOGIA POLÍTICA DAS REFORMAS EDUCACIONAIS GLOBAIS

Antoni Verger*

Lluís Parcerisa**

Clara Fontdevila***

RESUMO

O *Movimento de Reforma Educacional Global (Global Education Reform Movement – GERM)* está se expandindo internacionalmente e alcançando países que aparentavam ser imunes a essa abordagem de reforma educacional até muito recentemente. Assim, os sistemas educacionais do mundo têm sido cada vez mais articulados em torno de três princípios políticos principais, a saber: responsabilização, padrões e descentralização. As avaliações nacionais em larga escala (*National large-scale assessments – NLSAs*) são um componente central do GERM; essas avaliações têm sido cada vez mais usadas para fins de responsabilização, bem como para garantir que as escolas alcancem e promovam padrões de aprendizado centralmente definidos e avaliáveis. Neste artigo, testamos essas premissas com base em um banco de dados novo e original sobre NLSAs, bem como em dados provenientes dos questionários do Programa de Avaliação Internacional de Estudantes (*Programme for International Students Assessment – PISA*). Também discutimos como diferentes teorias sobre a disseminação/globalização de políticas explicam a disseminação internacional das NLSAs e da responsabilização baseada em testes em todo o mundo, além de refletirmos sobre o potencial de uma abordagem de sociologia política para analisar esse fenômeno globalizante.

Palavras-chave: Testes padronizados. Avaliação educacional. Responsabilização baseada em testes. Instrumentos de política. Política educacional.

ABSTRACT

THE GROWTH AND SPREAD OF LARGE-SCALE ASSESSMENTS AND TESTBASED ACCOUNTABILITIES: A POLITICAL SOCIOLOGY OF GLOBAL EDUCATION REFORMS

The Global Education Reform Movement (GERM) is expanding internationally and reaching countries that seemed to be immune to this education reform approach until quite recently. Accordingly, more and more educational systems in the world

* Pesquisador do Departamento de Sociologia da Universitat Autònoma de Barcelona. Doutor em Sociologia (Universitat Autònoma de Barcelona). E-mail: antoni.verger@uab.cat

** Doutorando no Departamento de Sociologia da Universidade Autònoma de Barcelona. É licenciado em Sociologia (2013) e mestre em Política Social, Trabalho e Bem-Estar (2014). Desde 2013 é membro da GEPS (Globalização, Educação e Políticas Sociais). E-mail: lluiparcerisa@hotmail.com

*** Licenciada em Sociologia pela Universidade Autònoma de Barcelona. Mestranda em Estudos do Discurso na Universidade Pompeu Fabra. E-mail: clara.fontdevila@gmail.com

are articulated around three main policy principles: accountability, standards and decentralisation. National large-scale assessments (NLSAs) are a core component of the GERM; these assessments are increasingly used for accountability purposes as well as to ensure that schools achieve and promote centrally defined and evaluable learning standards. In this paper, we test these premises on the basis of a new and original database on NLSAs, as well as on data coming from the Programme for International Students Assessment (PISA) questionnaires. In the paper we also discuss how different theories on policy dissemination/globalisation explain the international spread of NLSAs and test-based accountability worldwide, and reflect on the potential of a political sociology approach to analyse this globalising phenomenon.

Keywords: Standardized tests. Educational assessment. Test-based accountability. Policy instruments. Educational policy.

RESUMEN

CRECIMIENTO Y DIFUSIÓN DE EVALUACIONES A GRAN ESCALA Y RENDICIÓN DE CUENTAS BASADA EN PRUEBAS: UNA SOCIOLOGÍA POLÍTICA DE LAS REFORMAS EDUCATIVAS GLOBALES.

El movimiento de reforma de la Educación Mundial (GERM) se está expandiendo internacionalmente y llegando a países que parecen ser inmunes a este enfoque de reforma educativa hasta hace muy poco. Así, los sistemas educativos del mundo se han articulado cada vez más en torno a tres principios políticos principales, a saber: rendición de cuentas, normas y descentralización. Las evaluaciones nacionales a gran escala (NLSAS) son un componente central de GERM; Estas evaluaciones se han utilizado cada vez más para propósitos de rendición de cuentas, así como para asegurar que las escuelas alcancen y promuevan patrones de aprendizaje definidos y evaluables de forma centralizada. En este artículo, probamos estos supuestos basados en una base de datos nueva y original sobre Nlsas, así como los resultados de los cuestionarios del programa internacional de evaluación de estudiantes (PISA). También discutimos cómo diferentes teorías sobre la diseminación/globalización de políticas explican la diseminación internacional de marcha nlsas y la rendición de cuentas basada en pruebas en todo el mundo, así como reflexionando sobre el potencial de un enfoque para Sociología política para analizar este fenómeno globalizador.

Palabras clave: Pruebas estandarizadas. Evaluación educativa. Rendición de cuentas basada en pruebas. Instrumentos de política. Política educativa.

Introdução

O Movimento de Reforma Educacional Global (GERM) é um conceito metafórico segundo o qual a maioria das reformas educacionais atualmente adotadas no mundo respondem a problemas e prioridades similares e analisam esses problemas a partir de uma abordagem gerencial muito semelhante. Responsabilização, padrões, descentralização e autonomia escolar são os principais princípios políticos do GERM (SAHLBERG, 2016). Conforme

mostrado no presente artigo, as avaliações nacionais em larga escala são o instrumento político que reúne esses princípios em um pacote de reformas coerente.

As avaliações nacionais em larga escala (NLSAs) não são elementos novos nos sistemas educacionais. Elas têm sido historicamente usadas para fins de certificação de estudantes ou para diagnosticar os problemas e desafios enfrentados pelos

sistemas educacionais (JONES, 1996; KAMENS; MCNEELY, 2010). No entanto, as NLSAs têm sido cada vez mais usadas no contexto do movimento de reforma educacional para monitorar a distribuição de currículos padronizados e para garantir a responsabilização de escolas, diretores e professores.

A origem do GERM é geralmente associada às reformas educacionais neoliberais adotadas em países como os Estados Unidos, o Reino Unido e a Austrália desde a década de 1980 (SAHLBERG, 2016). Essas reformas – ou alguns de seus principais componentes – seriam posteriormente exportadas para países de baixa renda e para economias em transição por meio da condicionalidade de empréstimos concedidos por bancos internacionais de desenvolvimento (HARGREAVES, 2001; KLEES, 2008). O GERM é um fenômeno globalizante, embora, segundo Sahlberg (2016), nem todos os sistemas educacionais tenham sido “infectados” pelo GERM ainda – e, de fato, alguns deles expressaram reservas quanto ao potencial e à adequação dessa abordagem. Segundo o referido autor, esse seria o caso dos países do Leste Asiático, tais como Japão e Coreia do Sul, e de países europeus como França, Alemanha, Dinamarca, Noruega, Bélgica e, é claro, Finlândia.

Neste artigo, mostramos empiricamente que, embora os países anglo-saxônicos e os dependentes de ajuda tenham sido os primeiros a adotar o GERM, essa abordagem está atualmente alcançando muitos outros países, incluindo os países europeus e membros da OCDE que pareciam ser imunes a ele até muito recentemente. Testamos essa premissa com base em um banco de dados novo e original sobre as avaliações nacionais do aprendizado estudantil, bem como em dados dos questionários do Programa de Avaliação Internacional de Estudantes (PISA). As avaliações nacionais são o componente principal do, e um instrumento central no GERM. De acordo com os nossos dados, as NLSAs têm sido cada vez mais usadas para fins de responsabilização, bem como para garantir que as escolas, em sistemas educacionais cada vez mais descentralizados, atinjam e cumpram padrões de aprendizagem centralmente definidos. Dedicamos as partes centrais do artigo para testar esses argumentos. Na última seção, discutimos como diferentes teorias de política global (ou de disseminação

de políticas) explicam a disseminação internacional das NLSAs e da responsabilização baseada em testes em todo o mundo, além de refletirmos sobre as possibilidades e potencialidades de adotar uma abordagem de sociologia política (KASSIM; LE GALÈS, 2010; LASCOUMES; LE GALÈS, 2007) para analisar esse fenômeno globalizante.

O Movimento de Reforma Educacional Global: a função principal das NLSAs

Sobre as origens do GERM

A globalização, entendida como a crescente interdependência e competição entre territórios e atores nos domínios econômico, político e cultural, tem contribuído para desterritorializar os processos de políticas educacionais, e para que os atores e espaços nacionais percam sua centralidade na definição das reformas educacionais (ROBERTSON, 2012). Em grande parte, a globalização implica a redefinição da escala, do espaço e da dinâmica através dos quais a política educacional tem sido negociada, formulada e implementada (VERGER; NOVELLI; ALTINYELKEN, 2018). Nesse contexto, as organizações internacionais e as redes transnacionais de especialistas têm apresentado uma capacidade cada vez maior de determinar as agendas educacionais e definir as prioridades dos países por meio de seus recursos materiais e discursivos (DALE, 1999).

A globalização dos espaços de políticas educacionais e a crescente dinâmica competitiva que envolve os sistemas educacionais têm estabelecido as condições para o surgimento e a rápida disseminação do chamado *Movimento de Reforma Educacional Global* (GERM) (SAHLBERG, 2005). O GERM é um pacote de reforma educacional global que visa fortalecer a eficácia dos sistemas educacionais e é essencialmente articulado em torno de três princípios principais de políticas educacionais, a saber: padrões, responsabilização e descentralização (GABLE; LINGARD, 2015; SAHLBERG, 2016). De acordo com Sahlberg (2016), a ideia do GERM foi primeiramente elaborada por Hargreaves e colaboradores no livro *Aprendendo a Mudar: o Ensino para além dos Conteúdos e da Padronização*. Nesse livro, os autores apontam que:

Uma nova ortodoxia oficial de reforma educacional está sendo estabelecida com rapidez em muitas partes do mundo. Isto está ocorrendo sobretudo em países predominantemente anglo-saxônicos; no entanto, através de organizações financiadoras como o Banco Mundial e da distribuição global de estratégias de políticas, elementos dessa ortodoxia também estão sendo exportados, cada vez mais, para várias partes menos desenvolvidas do mundo. (HARGREAVES et al., 2001, apud SAHLBERG, 2016, p. 132).

O GERM surgiu originalmente em países como os Estados Unidos e o Reino Unido, que adotavam o paradigma neoliberal na prestação de serviços públicos (ROBERTSON, 2015). Essa abordagem de reforma seria posteriormente difundida na América Latina, na Europa Oriental e na antiga União Soviética com o apoio financeiro de bancos internacionais de desenvolvimento. No entanto, outros países estão atualmente adotando o GERM por razões não diretamente relacionadas à promoção de reformas pró-mercado na educação. Em grande medida, a expansão do GERM está intimamente associada ao paradigma de aprendizagem centrado no aluno e ao fato de que a melhoria do desempenho educacional se tornou o principal objetivo da maioria dos reformadores educacionais e das organizações internacionais (SAHLBERG, 2005, 2016).

Atualmente, resultados padronizados e mensuráveis de aprendizagem são considerados o indicador mais relevante da qualidade da educação. Muitos formuladores de políticas e estudiosos associaram a qualidade da educação com a ideia mais específica das conquistas dos alunos, não necessariamente por razões ideológicas, mas devido ao fato de que os resultados da aprendizagem são mais concretos, comparáveis e mensuráveis do que outros tipos de indicadores de qualidade da educação mais sensíveis ao contexto (VERGER; PARCERISA, 2018). A fim de responder a essas novas demandas e justificativas para a garantia de qualidade, muitos governos desenvolveram padrões de aprendizagem alinhados e combinados com a implementação das NLSAs, que enfocam os assuntos centrais do currículo nacional (SMITH, 2016). No entanto, o GERM não pode ser isolado de mudanças mais amplas na administração do setor público e do fato de que as reformas desse setor são cada vez mais informadas pelos princípios da Nova

Gestão Pública (*New Public Management – NPM*). A descentralização, a fragmentação dos serviços públicos em provedores mais autônomos, a gestão baseada em resultados e a responsabilização não são princípios que penetraram exclusivamente o setor da educação, mas todos os tipos de domínios políticos – frequentemente através de reformas da NPM (GUNTER et al., 2016).

O papel fundamental das avaliações nacionais no âmbito do GERM

Embora o GERM não seja um modelo monolítico de reforma educacional e possa evoluir através de diferentes manifestações políticas (BALL; JUNEMANN; SANTORI, 2017), ele é geralmente estruturado em torno de três princípios políticos principais, a saber: padrões, descentralização e responsabilização. A *padronização* da educação envolve a definição de padrões de aprendizagem e a prescrição de um currículo nacional estruturado em torno de padrões centrais comuns mensuráveis – pelo menos em relação a temas curriculares essenciais.

A *descentralização* da educação implica uma mudança “nas competências territoriais do governo central para as autoridades regionais, estaduais ou locais” (WEILER, 1990, p. 434). Em determinados contextos, a descentralização da educação também implica a devolução de responsabilidades organizacionais e/ou pedagógicas às escolas (por exemplo, modelos de autonomia escolar e de gestão baseada na escola), assim como a introdução de modelos de liderança gerencial que visam tornar as escolas “mais profissionais e mais semelhantes às empresas” (BALL; JUNEMANN; SANTORI, 2017, p. 3). No entanto, de acordo com Weiler (1990), a descentralização administrativa e a autonomia escolar não andam sempre de mãos dadas, e alguns países podem combinar, por exemplo, um alto grau de descentralização educacional e um baixo grau de autonomia escolar.

O terceiro componente do GERM, *responsabilização*, implica tornar os atores educacionais (incluindo escolas e professores) mais responsivos a, e responsáveis por suas ações e resultados. A suposição por trás dos sistemas de responsabilização performativa é que a vinculação da aprendizagem a

incentivos, sanções e/ou implicações reputacionais levará à melhoria do ensino e da aprendizagem, tanto na sala de aula quanto em nível escolar (SAHLBERG, 2016).

As avaliações nacionais em larga escala são altamente funcionais para a implementação prática dos princípios políticos que dão forma ao GERM. Primeiramente, essas avaliações representam o principal instrumento político nas mãos dos governos para garantir que os atores escolares cumpram as metas nacionais e adotem os padrões de ensino e aprendizagem definidos no currículo nacional. Em segundo lugar, em um contexto de crescente descentralização dos sistemas educacionais, as NLSAs são uma tecnologia política fundamental adotada para manter o poder do Estado e para “conduzir à distância” a complexa rede de provedores e atores que configuram o sistema escolar (OSBORNE; GAEBLER, 1992). Em terceiro lugar, as NLSAs estão no centro das dinâmicas e sistemas de responsabilização consequencial. A forma predominante de responsabilização que está se espalhando no contexto do GERM possui natureza baseada no

desempenho. Essa abordagem de responsabilização se concentra nos resultados de aprendizado dos alunos e envolve a geração de dados por meio das NLSAs, fato que a faz ser geralmente conhecida como responsabilização baseada em testes ou *test-based accountability* (TBA) (HAMILTON; STECHER; KLEIN, 2002). Especificamente, as NLSAs promovem relações de responsabilização de natureza administrativa e de mercado. Nos sistemas de responsabilização administrativa, os resultados obtidos pelas escolas e professores nas NLSAs estão ligados a consequências e incentivos geralmente distribuídos pela administração educacional. As NLSAs também funcionam como um dispositivo de responsabilização de mercado quando os resultados dos testes são publicados e usados para informar as escolhas escolares das famílias. Tabelas e rankings de pontuação e classificação escolares são ferramentas políticas que derivam diretamente das NLSAs e cuja intenção é tornar as escolas mais conscientes das, e sensíveis às pressões competitivas e dinâmicas da demanda escolar (OLMEDO; WILKINS, 2017).

Tabela 1 – O papel das avaliações nacionais no âmbito do GERM

Princípio do GERM	Definição e principais políticas	Papel das avaliações nacionais	
Padrões	Prescrição de um currículo nacional e estabelecimento de padrões de qualidade	Avaliações nacionais usadas para garantir que as escolas atendam e adiram aos padrões de aprendizado avaliados	
Descentralização	Transferência de competências e autoridade do governo central para níveis administrativos mais baixos	Avaliações nacionais usadas para controlar as autoridades estaduais, regionais, provinciais e locais	
	Devolução das responsabilidades gerenciais e/ou pedagógicas aos diretores e às escolas	Avaliações nacionais usadas para governar à distância uma gama de provedores autônomos através dos princípios de gestão baseada em resultados	
Responsabilização	Atores educacionais responsabilizados por suas ações/resultados por meio de alguma forma de avaliação vinculada a consequências.	<i>Responsabilização administrativa</i>	Testes de resultados associados a incentivos ou sanções para escolas, diretores e professores
		<i>Responsabilização de mercado</i>	Resultados de testes usados para informar a escolha da escola e promover a competição escolar

Fonte: Elaborada pelos autores deste artigo.

A adoção internacional das NLSAs e da responsabilização baseada em testes

Abordagem metodológica e considerações conceituais

Um dos principais objetivos deste artigo é melhorar a compreensão da disseminação das NLSAs e da TBA como componentes centrais do GERM. Para tanto, documentamos e examinamos dois fenômenos diferentes (embora interligados) nesta seção: em primeiro lugar, o surgimento e a expansão das NLSAs e, em segundo lugar, os propósitos mutáveis que essas avaliações servem – com ênfase em se eles estão sendo mais frequentemente usados para fins de responsabilização.

O estudo se baseia em diferentes fontes de dados: a) Relatórios comparativos de organizações internacionais ou regionais que fornecem uma visão geral das NLSAs em vigor (BENAVOT; TANNER, 2007; EURYDICE; EDUCATION, AUDIOVISUAL AND CULTURE EXECUTIVE AGENCY, 2009; UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION, 2007, 2015);¹ b) dados autorrelatados recuperados de arquivos e repositórios nacionais, geralmente armazenados nos sites dos ministérios da educação, institutos de avaliação e agências públicas similares; c) fontes acadêmicas e de mídia para fornecer uma compreensão mais completa das características e propósitos das avaliações, especialmente quando há dados administrativos incompletos sobre as características das NLSAs; e d) dados do questionário dos diretores no PISA (relatos de 2003 2006, 2012 e 2015) sobre como as escolas usam os dados das NLSAs.

Para fins de sistematização, os dados coletados dessas fontes foram compilados e organizados em um banco de dados novo e original sobre as NLSAs, abrangendo 34 países ou divisões administrativas autônomas.² Esse exercício de sistematização

fornece uma visão geral histórica das NLSAs, especialmente no que diz respeito aos propósitos, cobertura e escopo, assuntos testados, níveis-alvo e unidade principal de análise desses instrumentos.³ O universo de países considerados em nosso banco de dados é limitado aos países membros⁴ da OCDE – dada a disponibilidade muito limitada de dados sobre práticas de avaliação em países de baixa e média renda,⁵ mas também pelo fato de acharmos especialmente relevante estudar a penetração do GERM no espaço político da OCDE. O período de tempo da nossa análise vai de 1995 a 2014.⁶

Nossa pesquisa enfoca as NLSAs, ou seja, testes padronizados concebidos e administrados externamente às escolas (geralmente, pela administração educacional ou por agências externas de avaliação) a fim de garantir consistência e comparabilidade (EURYDICE; EDUCATION, AUDIOVISUAL AND CULTURE EXECUTIVE AGENCY, 2009; ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2013a; POPHAM, 1991). Também consideramos as chamadas formas híbridas de avaliação, ou seja, exames desenvolvidos externamente, mas implementados internamente e/ou administrados em nível local, uma vez que são avaliações também adequadas para fins comparativos em nível de sistema.⁷

Quando se trata dos *usos/finalidades* das avaliações, uma série de refinamentos conceituais deve ser feita neste ponto. A maior parte da literatura sobre avaliações nacionais distingue entre avaliações formativas e somativas; no entanto, em nosso banco de dados, classificamos as NLSAs de acordo com uma gama mais ampla de categorias, que incluem:

Escócia e Irlanda do Norte no Reino Unido.

- 3 Os *usos* específicos das avaliações nacionais foram explorados através de dados provenientes do questionário das escolas do PISA, dadas as limitações de informações sobre este tópico em particular disponíveis em fontes administrativas.
- 4 As Figuras 6 a 10 incluem também países não-membros da OCDE ou economias que participam do PISA, uma vez que utilizam dados dos principais questionários administrados como parte do programa.
- 5 As edições de 2007 e 2015 do Relatório de Monitoramento Global da Educação constituem uma exceção a isso.
- 6 O ano de 1995 foi selecionado como o ponto de partida, dadas as dificuldades em coletar dados comparáveis e confiáveis para os anos anteriores. No entanto, como relatado em Eurydice e Education, Audiovisual and Culture Executive Agency (2009), os testes nacionais não foram difundidos até a década de 1990.
- 7 Por outro lado, as avaliações internas da escola (ou seja, não padronizadas e não externas) não são consideradas em nosso banco de dados.

1 Essa informação foi, por sua vez, complementada por documentos de base por trás dos relatórios referenciados. Este é especialmente o caso dos Relatórios dos Países preparados no contexto da Revisão sobre os Quadros de Avaliação e Exame para Melhorar os Resultados das Escolas - *Review on Evaluation and Assessment Frameworks for Improving School Outcomes* (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2013a).

2 Este é o caso da comunidade de língua flamenga e da Comunidade Valônia-Bruxelas na Bélgica, e o caso da Inglaterra, País de Gales,

- a) Avaliação para fins de certificação, *streaming* e seleção de alunos: são exames padronizados de alto risco para os alunos, mas não necessariamente para as escolas. Essa categoria incluiria exames de saída ou qualquer tipo de avaliação padronizada capaz de certificar o progresso da aprendizagem e/ou permitir que os alunos tenham acesso a níveis superiores de educação, ou avaliações que tenham uma consequência formal para a carreira dos alunos (decisão sobre *streaming*, progressão de um ano escolar para o próximo etc.).
- b) Avaliação para fins de diagnóstico: avaliações padronizadas que não têm implicações formais para os estudantes ou para as escolas. Possuem um caráter mais formativo e visam fornecer aos alunos e às escolas dados úteis e feedback para que possam melhorar seu desempenho futuro. Essas avaliações também são úteis para identificar as necessidades de aprendizado dos alunos e adaptar o ensino de acordo.
- c) Verificação para fins de monitoramento e avaliação: NLSAs usadas para fins de monitoramento e avaliação (medindo o progresso e o desempenho em relação a diferentes objetivos e em diferentes níveis do sistema educacional). Inclui qualquer teste orientado a comparar o desempenho entre jurisdições, escolas ou provedores, avaliando o sistema ou um programa específico, ou criando os dados necessários para fins de responsabilização.

Essas categorias não são mutuamente exclusivas, já que muitos programas de avaliação são projetados para atender a vários propósitos. Vale ressaltar que as avaliações são categorizadas de acordo com os usos formais e explicitamente relatados pelas autoridades responsáveis e conforme eles são descritos nos relatórios comparativos. Em outras palavras, usos informais, não intencionais ou não oficiais dados às avaliações nacionais não são considerados nesta análise. Além disso, nosso banco de dados sobre a NLSA não captura a intensidade de diferentes programas de teste – a duração e o número de testes estabelecidos por cada programa.

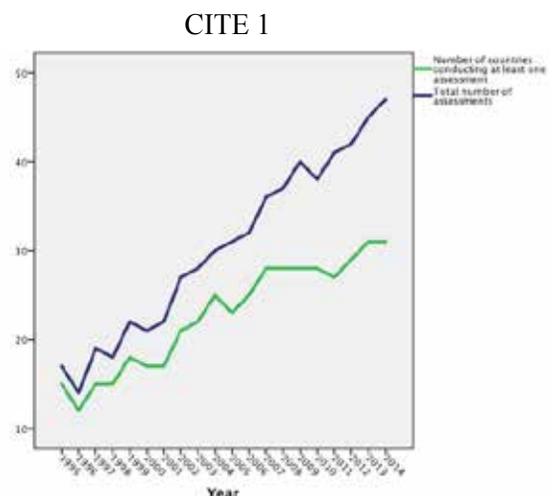
Finalmente, um atributo adicional que levamos em consideração para entender os usos das NLSAs é seu escopo, isto é, a distinção entre avaliações baseadas em censos e avaliações baseadas em amostras administradas, respectivamente, a todo o universo das escolas ou a uma amostra representativa. As avaliações baseadas em censos são as mais apropriadas para fins de responsabilização, pois permitem que a pressão performativa seja colocada em todas as escolas que compõem o sistema educacional.

Nas próximas seções, exploramos como as NLSAs, nos países membros da OCDE, evoluíram em termos quantitativos e qualitativos, além de mostrar que há uma tendência internacional ao uso das NLSAs para fins de responsabilização.

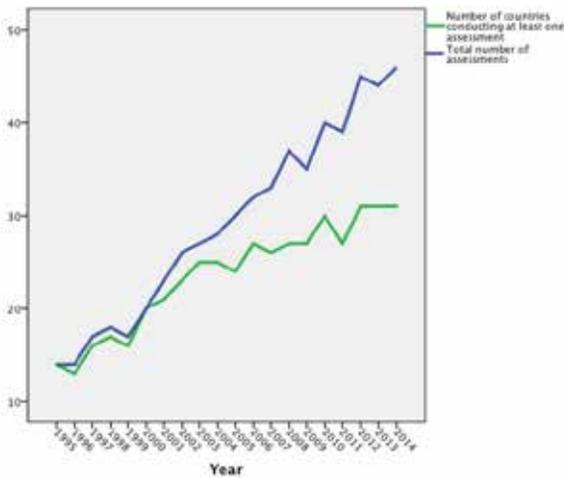
O aumento e a natureza mutante das NLSAs

O número de países que realizam pelo menos uma NLSA aumentou significativamente desde meados dos anos 1990 no espaço da política de educação da OCDE. Esse número quase duplicou desde 1995 tanto no ensino primário como no secundário inferior (Figura 1). No entanto, o número total de programas de avaliação em vigor aumentou a um ritmo maior do que o número de países que adotam as NLSAs, o que significa que um número crescente de países está administrando mais de um programa nacional de avaliação.

Figura 1 – Número de países que realizam pelo menos uma avaliação e número total de Avaliações.



CITE 2



Fonte: Elaborado pelos autores.

Nota: CITE 1 e CITE 2 correspondem, respectivamente, ao ensino primário (ou primeira fase do ensino básico) e ao ensino secundário inferior (ou segunda fase do ensino básico, que geralmente vai até aos 15/16 anos).

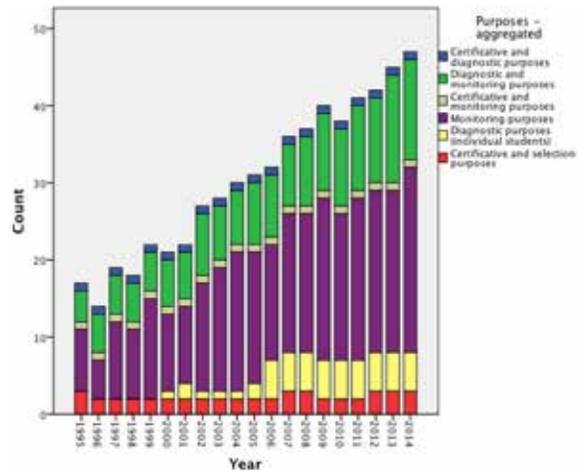
Embora o número de países com pelo menos um programa de avaliação pareça estar se estabilizando, esse não é o caso do número total de avaliações sendo promovidas. Avaliações nacionais em larga escala parecem estar em um estágio de crescimento exponencial ou o que o modelo epidemiológico de disseminação global denomina uma “fase explosiva” (WATTS, 2003 apud STEINER-KHAMSI, 2016). De acordo com esse modelo, a fase explosiva geralmente é seguida por uma fase de esgotamento, na qual a tendência de adoção de políticas diminui. No entanto, essa fase de desaceleração ainda não foi alcançada no caso estudado e o número de NLSAs adotadas no campo de políticas da OCDE provavelmente continuará aumentando.

Ao olharmos para os objetivos específicos buscados pelas NLSAs, observamos que o seu crescimento é, em grande medida, explicado pelo surgimento de uma variedade específica de avaliações, aquelas com propósitos de monitoramento e avaliação (Figura 2). Esse é o caso do ensino primário e do ensino secundário inferior, embora, no ensino primário também seja possível discernir um aumento substancial das avaliações com fins de diagnóstico. Essas mudanças são indicativas de uma transformação muito significativa da natureza

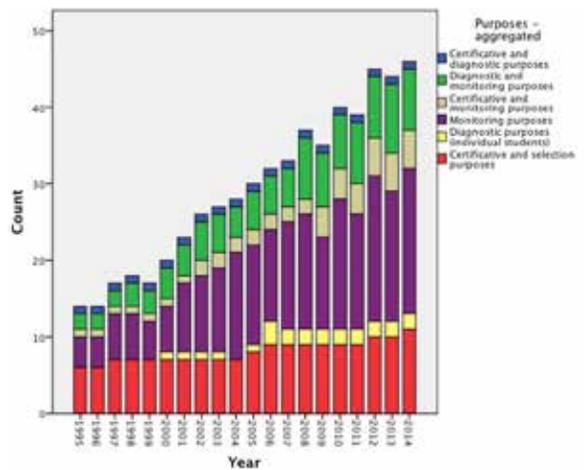
das avaliações. Assim, os testes parecem estar cada vez mais orientados para medir realizações com outras finalidades que não a tomada de decisões sobre a carreira dos alunos, entre as quais os usos de responsabilização em nível escolar são suscetíveis de se destacar.

Figura 2 – Número de avaliações realizadas por ano – distribuição por finalidade

CITE 1



CITE 2



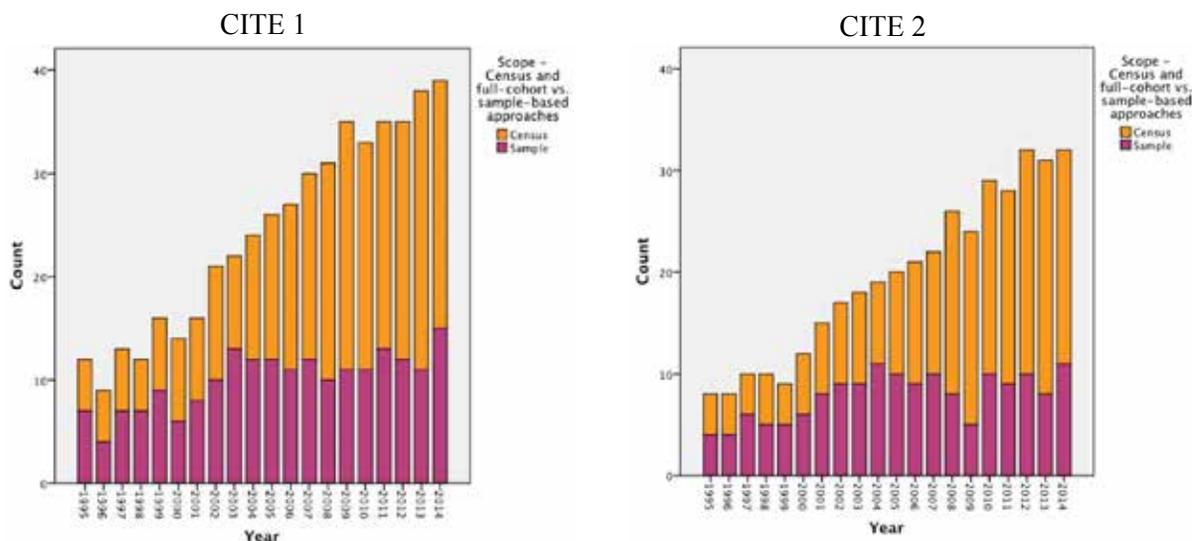
Fonte: Elaborado pelos autores deste artigo.

Nota: CITE 1 e CITE 2 correspondem, respectivamente, ao ensino primário (ou primeira fase do ensino básico) e ao ensino secundário inferior (ou segunda fase do ensino básico, que geralmente vai até aos 15/16 anos).

De fato, as modalidades particulares de avaliação que tiveram o maior crescimento são precisamente aquelas tipicamente associadas a metas de responsabilização, a saber: programas de avaliação que adotam a abordagem baseada no censo. Conforme mostrado na Figura 3, entre os programas com fins de monitoramento, avaliação e

diagnóstico, houve uma expansão particularmente notável de programas de testes baseados em censos ou de coorte total. Por outro lado, os programas de avaliação baseados em amostras tiveram um crescimento muito mais moderado. Mais uma vez, essa tendência se aplica tanto ao ensino primário quanto ao secundário inferior.

Figura 3 – Número de avaliações realizadas por ano (para fins de diagnóstico e avaliação) – distribuição por escopo.

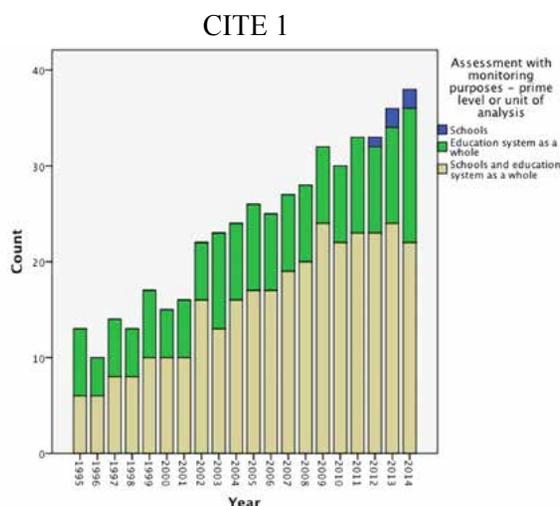


Fonte: Elaborado pelos autores deste artigo.

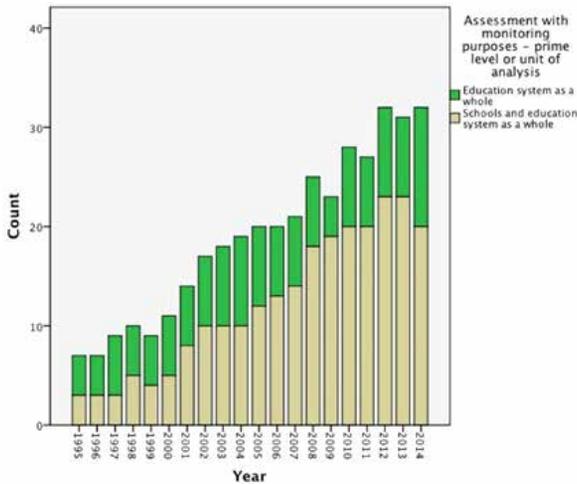
Nota: CITE 1 e CITE 2 correspondem, respectivamente, ao ensino primário (ou primeira fase do ensino básico) e ao ensino secundário inferior (ou segunda fase do ensino básico, que geralmente vai até aos 15/16 anos).

Os dados sobre a disseminação das NLSAs em relação à sua unidade principal de análise apontam para uma direção similar. Como mostra a Figura 4, embora o número absoluto de avaliações com foco no sistema educacional tenha aumentado apenas moderadamente, esse não é o caso de avaliações que enfocam tanto o sistema educacional quanto o nível escolar. A última variedade de avaliações, que visa agregar dados tanto em nível escolar como de sistema, teve um aumento mais pronunciado desde o final dos anos 1990.

Figura 4 – Número de avaliações realizadas por ano (para fins de avaliação) – distribuição por unidade principal de análise



CITE 2



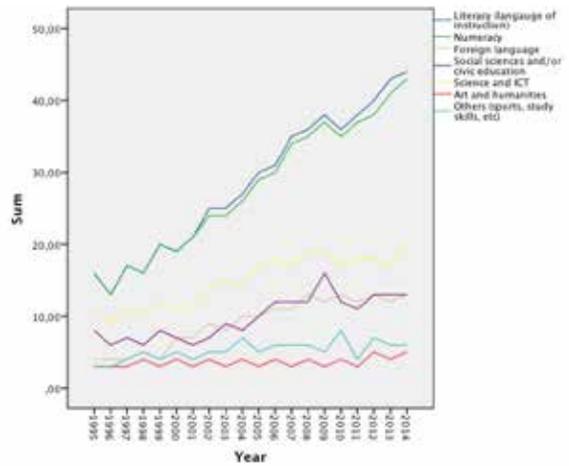
Fonte: Elaborado pelos autores deste artigo.

Nota: CITE 1 e CITE 2 correspondem, respectivamente, ao ensino primário (ou primeira fase do ensino básico) e ao ensino secundário inferior (ou segunda fase do ensino básico, que geralmente vai até aos 15/16 anos).

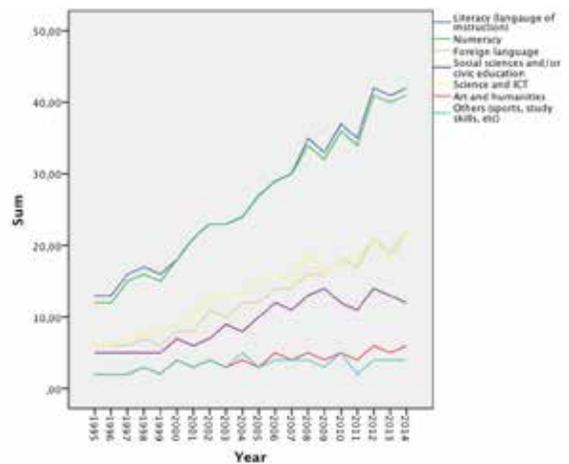
Finalmente, as áreas curriculares (ou disciplinas) avaliadas pelas NLSAs também passaram por uma transformação significativa. Como mostra a Figura 5, as avaliações introduzidas durante a última década tendem a focar um conjunto restrito de habilidades e assuntos básicos – especialmente numeracia e alfabetização. Em outras palavras, é menos provável que novas avaliações avaliem outros assuntos além da língua de instrução, da matemática e, em menor medida, da ciência e das TICs. Tal tendência é essencialmente a consequência de um aumento da participação das NLSAs baseadas em censos. Dada a necessidade de manter os recursos e o tempo investidos em testes em nível gerenciável, assim como de garantir que os dados coletados possam ser adequadamente comparados, as avaliações que adotam uma abordagem de censo têm maior probabilidade de se concentrar em um número limitado de áreas e, especialmente, em assuntos relacionados à numeracia e à alfabetização. Focar nesse grupo restrito de assuntos é também uma forma de alinhar as NLSAs com o que está sendo predominantemente medido em avaliações internacionais em larga escala, tais como o PISA (MEYER; BENAVIDES, 2013).

Figura 5 – Assuntos específicos – número total de avaliações por ano⁸

CITE 1



CITE 2



Fonte: Elaborado pelos autores deste artigo.

Nota 1: CITE 1 e CITE 2 correspondem, respectivamente, ao ensino primário (ou primeira fase do ensino básico) e ao ensino secundário inferior (ou segunda fase do ensino básico, que geralmente vai até aos 15/16 anos).

Em última análise, a crescente prevalência de modalidades específicas de NLSAs, como descrito até agora, é indicativa de uma mudança nos usos potenciais das avaliações: da certificação de estudantes ao monitoramento de escolas e professores. Avaliações em larga escala com propósitos de

8 Observe a sobreposição entre as linhas azul e verde (habilidades de alfabetização e numeracia, respectivamente).

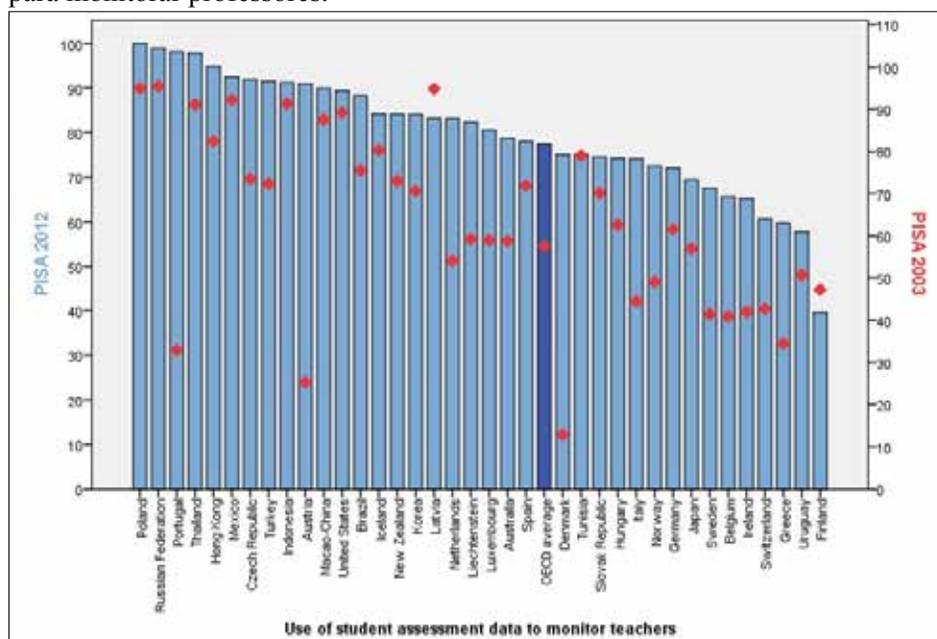
monitoramento podem servir a uma multiplicidade de propósitos, tais como informar a formulação e implementação de políticas, identificar ou alavancar prioridades políticas, ou responsabilizar escolas e professores (TOBIN et al., 2015). Nosso banco de dados sobre as NLSAs não inclui informações específicas sobre usos de responsabilização, mas mostra que as NLSAs atuais criam as condições necessárias para um eventual avanço da agenda de responsabilização. Na próxima seção, usamos os dados da OCDE/PISA para descobrir mais especificamente se os usos de responsabilização das NLSAs estão se intensificando nos países membros da OCDE e até que ponto isso está acontecendo.

As crescentes finalidades de responsabilização das NLSAs

Segundo Bovens (2007) e Hatch (2013), a avaliação externa de serviços públicos e práticas profissionais é uma condição necessária, mas não

suficiente, para aprimorar os sistemas de responsabilização. Para que a responsabilização ocorra, algumas “consequências” precisam derivar dos resultados da avaliação. No caso da educação, essas consequências incluem – mas não estão limitadas a – usar dados de teste para supervisionar de perto professores e/ou escolas, decidir sobre salários e promoção de professores, encorajar a competição escolar por meio da publicação de resultados de testes ou intervir na autonomia de escolas de baixo desempenho. O fato de que cada vez mais escolas e atores educacionais percebem a pressão performativa também pode ser interpretado como um aumento das relações de responsabilização dentro dos sistemas educacionais. Nesta seção examinamos os propósitos e práticas de responsabilização associados às NLSAs com base em dados provenientes DAS edições do PISA de 2003, 2006, 2012 e 2015 (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2004, 2007, 2013b, 2016).

Figura 6 – Mudança, entre 2003 e 2012, no uso de dados da avaliação de alunos para monitorar professores.



Fonte: Adaptado de Organisation for Economic Co-operation and Development (2004, 2013b). **Nota:** Colunas azuis representam dados do PISA 2012; pontos vermelhos representam dados do PISA 2003.

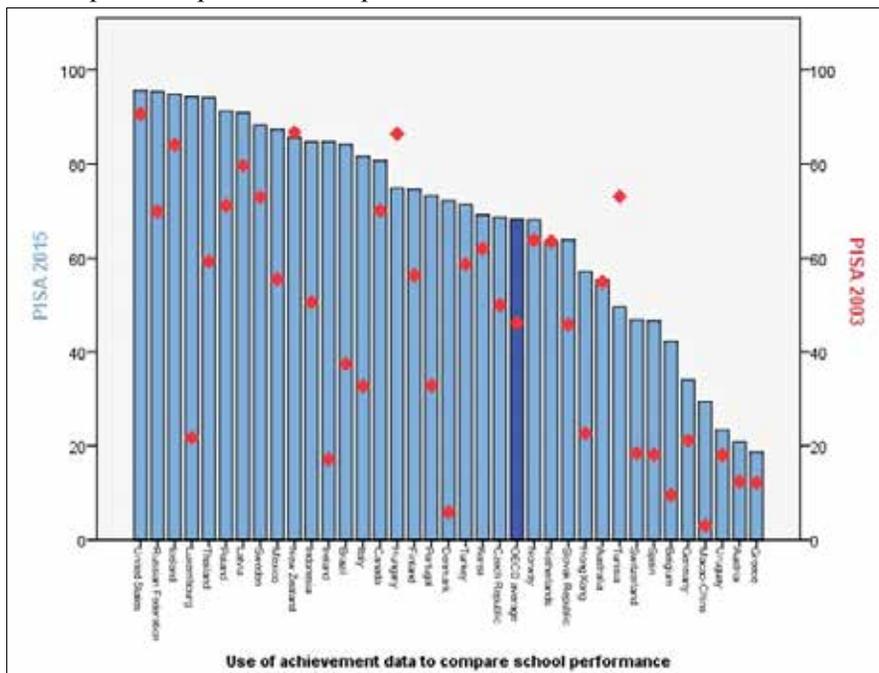
Nas edições do PISA de 2003 e 2012, os diretores das escolas foram questionados se os dados das NLSAs são usados para monitorar o trabalho dos

professores. Como mostra a Figura 6, em 2003, 17 países (de um total de 36 países que participaram das edições do PISA e tinham dados disponíveis

para essa questão) tinham menos de 60% dos seus alunos frequentando escolas que usavam os dados das NLSAs para monitorar as práticas dos professores, enquanto em 2012 havia apenas três países com menos de 60% (Finlândia, Uruguai e Grécia). De fato, em 2012, essas práticas poderiam ter aumentado mais de 10 pontos percentuais em relação ao ano de 2003 em 23 países da OCDE. Pelo contrário, a Finlândia e a Letônia foram os dois únicos países a mostrar uma redução substancial em relação à porcentagem de alunos que frequentavam escolas onde as avaliações eram usadas para monitorar o trabalho dos professores (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2013b). De forma geral, nossos dados mostram que as NLSAs estão se tornando cada vez mais centrais na avaliação das práticas dos professores. Smith e Kubacka (2017) observam uma tendência semelhante e acrescentam a esse argumento que as NLSAs têm minado a presença de outras formas de avaliação dos professores (como autoavaliação ou observações dos professores) e de outras fontes de dados no feedback recebido pelos professores.

A porcentagem de alunos que frequentam escolas que usam os dados das NLSAs para comparar o seu desempenho com o distrito, ou com as referências nacionais, também aumentou significativamente. Conforme mostrado na Figura 7, a média da OCDE nesse item – que de alguma forma indica a conscientização das escolas quanto a padrões e pressões performativas – aumentou 22 pontos percentuais entre o PISA 2003 e o PISA 2015. O número de países onde mais de 80% dos alunos frequentam escolas que usam dados da avaliação de alunos para esse fim passou de 04 (no ano de 2003) para 15 (no ano de 2015). Os países com maior porcentagem de escolas que usam dados das NLSAs para comparar o desempenho escolar ou distrital são: Estados Unidos, Rússia, Nova Zelândia, Turquia, Letônia, Suécia, Tailândia, Holanda, Reino Unido, Portugal, Polônia, Brasil, Canadá, Islândia, Indonésia, Irlanda, Luxemburgo, México e Itália. Em 2015, apenas a Grécia tinha menos de 20% dos alunos frequentando escolas que usam as NLSAs para comparar o desempenho escolar.

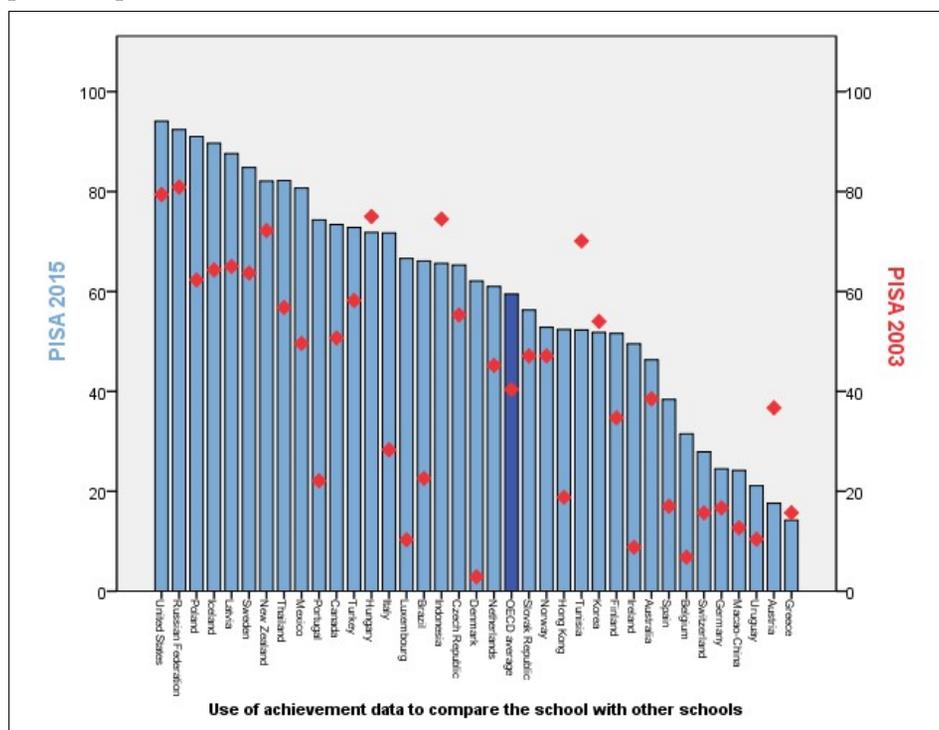
Figura 7 – Mudança, entre 2003 e 2015, no uso dos dados da avaliação de alunos para comparar o desempenho da escola com o distrital ou nacional.



O aumento das pressões competitivas sobre as escolas, derivadas da administração das NLSAs, também é evidente na Figura 8, que mostra como as avaliações nacionais são cada vez mais usadas para comparar o desempenho escolar com o de outras escolas. Em outras palavras, as NLSAs têm contribuído para aumentar a consciência competitiva e as pressões dentro dos mercados educacionais locais. A média da OCDE, que aumentou cerca de 20 pontos percentuais entre 2003 e 2015, também reflete essa

tendência global. Além disso, o número de países com mais de 80% de alunos frequentando escolas cujos diretores usam as NLSAs para comparar a sua escola com outras passou de 01 (em 2003) para 09 (em 2015). Por outro lado, de acordo com a edição do PISA de 2015 (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2016), existem apenas dois países com menos de 20% dos estudantes matriculados em escolas que aplicam esse tipo de prática, a saber: Áustria e Grécia.

Figura 8 – Mudança, entre 2003 e 2015, no uso de dados da avaliação de alunos para comparar uma escola com outras.



Fonte: Adaptado de Organisation for Economic Co-operation and Development (2004, 2016). **Nota:** Colunas azuis representam dados do PISA 2012; pontos vermelhos representam dados do PISA 2003.

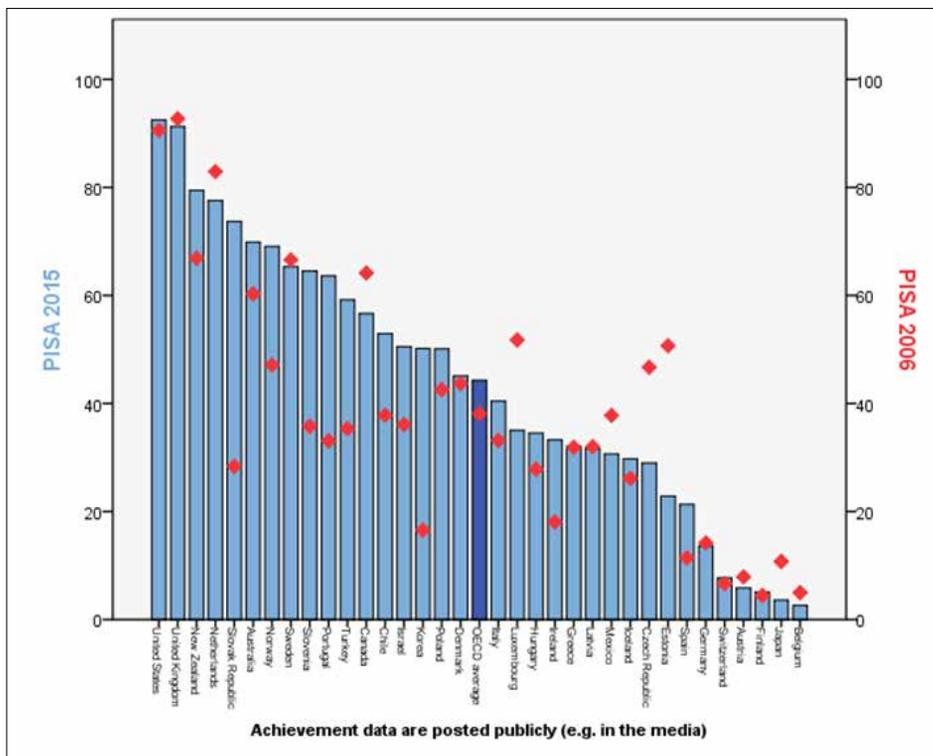
Entre 2006 e 2015, as pressões de responsabilização entre as escolas dos países da OCDE aumentaram ligeiramente, conforme medido pela publicação dos resultados escolares nas NLSAs. Em relação a esse indicador, a média da OCDE aumentou 6,2 pontos percentuais entre 2006 e 2015. A Figura 9 mostra que o número de países com mais de 60% dos alunos frequentando escolas nas quais os dados de desempenho são divulgados

publicamente passou de 07 (em 2006) para 10 (em 2015). A publicação dos dados de desempenho das escolas é uma variável controversa quando se projeta sistemas de responsabilização baseados em testes. Enquanto acadêmicos como Boarini e Lüdemann (2009) sugerem que a divulgação pública dos resultados escolares pode contribuir para a melhoria do desempenho acadêmico, outros estudiosos acham que a publicação de resultados

escolares – especialmente quando adotam a forma de rankings – contribui para estigmatizar escolas, ou minar a cooperação entre escolas (JONES; EGLEY, 2004). No entanto, o fato de a publicação dos resultados das escolas se traduzir, ou não, em pressão de responsabilização do mercado depende da configuração do sistema educacional adotado em cada país ou região. Assim, enquanto esse tipo de medida só pode desencadear pressões não competi-

tivas ou, no máximo, reputacionais para a melhoria da escola em países que não contam com escolha escolar (por exemplo, a Noruega), em países onde o sistema educacional combina liberdade de escolha escolar e esquemas de financiamento do lado da demanda (por exemplo, o Chile), a publicação dos resultados das escolas exerce efeitos mais diretos no aumento das pressões de responsabilização do mercado.

Figura 9 – Mudança, entre 2006 e 2015, no uso dos dados de aproveitamento da seguinte forma: Os dados de aproveitamento são disponibilizados publicamente (por exemplo, na mídia).

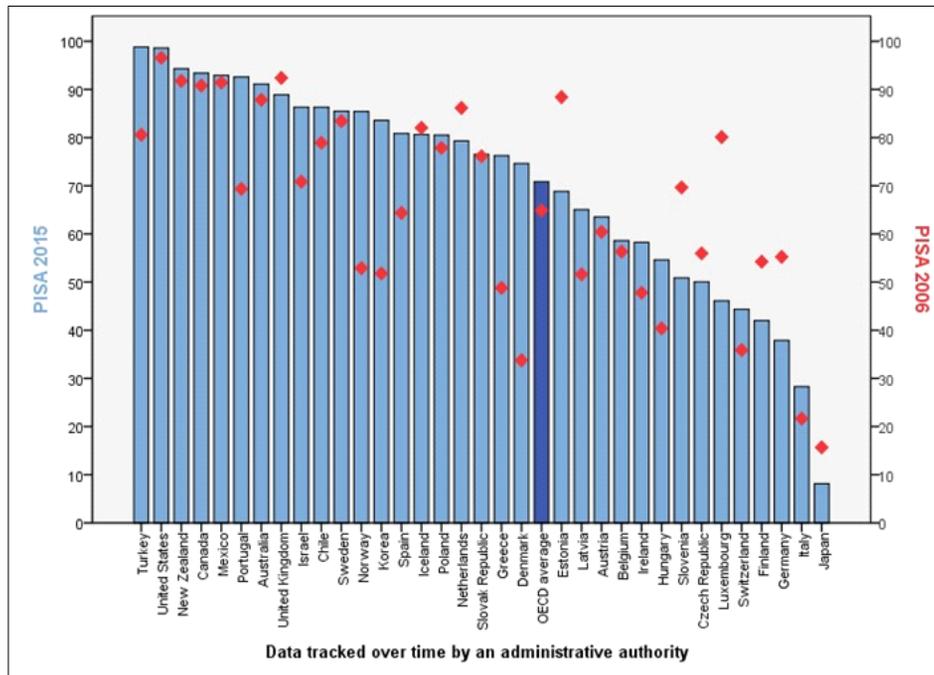


Fonte: Adaptado de Organisation for Economic Co-operation and Development (2007, 2016). **Nota:** Colunas azuis representam dados do PISA 2015; pontos vermelhos representam dados do PISA 2006.

Por fim, os resultados das NLSAs também podem ser usados para fortalecer a dinâmica da responsabilização administrativa. Na Figura 10, podemos observar que, entre 2006 e 2015, a média da OCDE em relação ao percentual de alunos matriculados em escolas que usam os dados de aproveitamento para fins de responsabilização administrativa aumentou 6 pontos percentuais. Em

2015, havia 14 países da OCDE com 80%, ou mais, de seus alunos frequentando escolas cujos dados de aproveitamento eram monitorados ao longo do tempo. Por outro lado, o Japão é o único país da OCDE onde menos de 20% dos alunos frequentam escolas onde os dados de aproveitamento são monitorados por uma autoridade administrativa para fins de responsabilização.

Figura 10 – Mudança, entre 2006 e 2015, no uso dos dados de aproveitamento da seguinte forma: Os dados de aproveitamento são monitorados ao longo do tempo por uma autoridade administrativa.



Fonte: Adaptado de Organisation for Economic Co-operation and Development (2007, 2016). **Nota:** Colunas azuis representam dados do PISA 2015; pontos vermelhos representam dados do PISA 2006.

Concluindo, os dados do PISA indicam que o desempenho das escolas nas NLSAs está se tornando cada vez mais central para gerir os sistemas escolares e as percepções e comportamentos dos atores escolares. No entanto, com base nesses dados, não podemos inferir diretamente que haja um aumento crescente nos riscos associados aos resultados das escolas nas NLSAs e, em particular, no aumento dos chamados grandes riscos. As evidências aqui apresentadas indicam que muitos países não estão apenas implementando as NLSAs para fins de diagnóstico educacional, eles também as usam com mais frequência para intensificar as pressões administrativas, performativas e/ou de mercado sobre as escolas.

Compreendendo a disseminação internacional das NLSAs e da TBA

As avaliações nacionais em larga escala têm se expandido internacionalmente como parte de um

movimento de reforma educacional global que promove currículos padronizados, descentralização e responsabilização em sistemas educacionais. No entanto, o GERM é mais do que um *explanans*, é algo que precisa ser explicado (ou um *explandum*). As teorias predominantes de difusão e globalização política (tais como racionalismo, neoinstitucionalismo e economia política internacional) podem contribuir para captar os mecanismos macroeconômicos subjacentes à disseminação internacional do GERM e, em particular, para o crescimento e a disseminação tanto das NLSAs como da TBA.

De acordo com uma abordagem racionalista de transferência de políticas, as políticas que gozam de ampla circulação internacional também funcionam melhor e são mais eficazes. Essa abordagem pressupõe que os tomadores de decisão são guiados principalmente por questões de eficiência funcional, de modo que são propensos a adotar políticas e programas cuja eficácia é respaldada pelas experi-

ências de outros países e por evidências empíricas. As NLSAs e as medidas de responsabilização que lhes são inerentes são percebidas por muitos como uma forma eficaz de promover e garantir a qualidade da educação. No entanto, tal como acontece com outros modelos de reforma educacional globalizados, as evidências sobre os efeitos positivos de políticas de responsabilização baseadas nas NLSAs são, no mínimo, inconclusivas. Até hoje, as pesquisas empíricas sobre responsabilização e educação obtiveram resultados muito diferentes – e até mesmo contraditórios – em relação aos efeitos da TBA no aumento dos níveis de aprendizado, na melhoria dos processos educacionais ou na redução das desigualdades educacionais. Além disso, um crescente número de estudos tem mostrado que certos sistemas de responsabilização baseados em testes podem gerar comportamentos indesejáveis ou oportunistas em nível escolar, os quais acabam prejudicando a qualidade do ensino e/ou a inclusão das escolas (ALLAN; ARTILES, 2016; AU, 2007). Portanto, pelo menos do ponto de vista teórico, o racionalismo seria insuficiente para compreendermos a globalização da responsabilização baseada em testes na educação.

O neoinstitucionalismo desafia os pressupostos básicos do racionalismo, argumentando que a legitimidade é um fator mais importante do que a eficiência funcional na adoção de políticas (DOBBINS; KNILL, 2009). As avaliações nacionais em larga escala e a TBA são vistas como instrumentos políticos legítimos, uma vez que estão intrinsecamente ligadas à modernização dos serviços públicos e à promoção da transparência e da meritocracia na educação pública.⁹ Além disso, o desenvolvimento e a presença crescente de avaliações internacionais em larga escala, tais como o PISA, o Estudo Internacional das Tendências em Matemática e Ciências (*Trends in International Mathematics and Science Study* – TIMSS) ou o Estudo Internacional de Alfabetização e Leitura (*Progress in International Reading Literacy Study* – PIRLS), também encorajaram muitos governos a adotar as NLSAs. Isso se deve ao fato de a realização das NLSAs ser

uma forma de os países se envolverem de maneira mais apropriada nos debates sobre políticas educacionais globais e sobre a “corrida educacional” gerada pelas avaliações internacionais em larga escala (MEYER; BENAVIDOT, 2013). De forma geral, segundo o neoinstitucionalismo, as avaliações nacionais seriam instrumentos políticos tão atraentes por conferirem legitimidade aos governos nos fóruns internacionais, bem como em face à sua própria população.

Por outro lado, as abordagens da economia política internacional dão mais ênfase a fatores de natureza material e econômica. De acordo com essa perspectiva, os processos de disseminação de políticas estão relacionados às transformações e pressões econômicas predominantes, bem como às ambições dos países de ocuparem melhor posição em uma economia global cada vez mais competitiva. Em tal ambiente econômico, a adoção das NLSAs e da TBA permitiria alcançar um triplo objetivo econômico: em primeiro lugar, melhorar os resultados de aprendizagem dos estudantes, que muitos veem como sinônimo de promover o crescimento econômico dos países e das economias baseadas no conhecimento, e atrair investimentos estrangeiros; em segundo lugar, promover a concorrência entre os estabelecimentos de ensino e os ganhos de eficiência em termos de custos nos serviços de educação pública; e, em terceiro lugar, favorecer a abertura do setor educacional aos interesses comerciais de uma indústria emergente de testes educacionais e de melhoria escolar (BALL, 2008; CARNOY, 2016; VERGER; PARCERISA, 2018).

Uma abordagem de sociologia política para instrumentos de política

As abordagens macro aplicadas à transferência de políticas descritas acima refletem sobre os mecanismos globais (isto é, evidência internacional, legitimação e competitividade econômica) por trás da disseminação e do crescimento internacional das NLSAs e da TBA. Nesta seção, refletimos sobre como uma abordagem de sociologia política aplicada a instrumentos de política pode complementar tais macroteorias ao trazer fatores de nível “meso” para a análise. A abordagem da sociologia política aplicada aos instrumentos de política visa

⁹ Do ponto de vista da meritocracia, as avaliações nacionais seriam fundamentais para identificar e recompensar talentos, bem como para motivar os alunos a trabalhar com mais diligência (HUTT; SCHNEIDER, 2018).

compreender, por um lado, “como os instrumentos são escolhidos, como se desenvolvem e como são operacionalizados” (KASSIM; LE GALÈS, 2010, p. 02), e, por outro, quais são as implicações políticas das escolhas dos instrumentos políticos (PETERS, 2002). Essa abordagem é contextualmente fundamentada, no sentido de que é compatível com premissas institucionais históricas sobre o papel das instituições na mediação de forças e agendas globais, mas também no sentido de dar, aos atores que operam em diferentes escalas, voz e agência na compreensão da adoção de políticas. Devido à sua inserção dentro de uma epistemologia construtivista, a abordagem da sociologia política aplicada aos instrumentos de política enfatiza que os processos de construção de significado interagem de forma importante com fatores políticos, institucionais e econômicos na produção de políticas e mudanças de políticas.

Nessa perspectiva, a natureza politicamente recompensadora, economicamente conveniente e semioticamente maleável das NLSAs e da TBA são alguns dos fatores que poderiam explicar a disseminação e o crescimento desses instrumentos de política:

a) *Politicamente recompensadora.* Adotar sistemas de garantia de qualidade e de responsabilização na educação permite que os políticos sinalizem para seus públicos que estão trabalhando seriamente para mudar a educação e que estão preocupados com a qualidade da educação, com os resultados da aprendizagem e com o futuro das crianças, ao mesmo tempo em que colocam a pressão da reforma nas escolas e nos professores em vez de no governo. Além disso, para os atores políticos, é mais fácil discutir e concordar com a adoção de instrumentos políticos do que com transformações políticas mais profundas ou mais fundamentais (KASSIM; LE GALÈS, 2010). De forma geral, a adoção de novos instrumentos pode ser uma forma de evitar abordar debates políticos controversos, ou de abordar debates complexos (como as causas e as soluções para as desigualdades educacionais) sob a perspectiva de escolhas técnicas e instrumentais. Por exemplo, é bastante comum

que as avaliações externas nas atuais narrativas políticas sejam retratadas como dispositivos-chave para abordar as “lacunas de aprendizado” e para assegurar que “todas as crianças” alcancem padrões mínimos de aprendizado.

b) *Economicamente conveniente.* Os instrumentos de política são mais fáceis de adotar se forem administráveis e economicamente viáveis ou, pelo menos, se o seu custo não for visível para o público (PETERS, 2002). As reformas de responsabilização são, de certa forma, “mais baratas e mais rápidas do que as reformas alternativas” (SMITH et al., 2004, p. 50) e, em particular, do que a adoção de reformas de equidade mais profundas e estruturais. Algo similar poderia ser dito sobre a adoção de outros componentes do GERM, tais como os padrões curriculares que, segundo Hargreaves e Shirley (2009 apud SAHLBERG, 2016, p. 131), se expandem como “incêndios florestais” porque são “fáceis de escrever e baratos de financiar”.

c) *Semioticamente maleável.* Os instrumentos de política se disseminam mais rapidamente se puderem ser associados a símbolos positivos (como “garantia de qualidade”) e a slogans politicamente poderosos (como “enfrentar a crise de aprendizado” ou “lacuna de aprendizado”). Esses instrumentos também se expandem mais rapidamente quando têm natureza polissêmica e/ou podem operar como significantes vazios, ou seja, o mesmo instrumento pode ter objetivos muito diferentes e até mesmo contraditórios (STEINER-KHAMSI, 2016). Esse é o caso de instrumentos como as NLSAs e a TBA, os quais permitem que diferentes grupos políticos avancem através de objetivos diferentes. Por exemplo, o fato de o mesmo instrumento de responsabilização poder contribuir para atingir uma diversidade de objetivos como qualidade, equidade, transparência, escolha de escolas etc., facilita que governos com orientações ideológicas diferentes convirjam em conceber a responsabilização baseada em

testes como um instrumento adequado para articular a mudança educacional.

A seleção de instrumentos de política é contingente a fatores políticos, discursivos e econômicos, como os que acabamos de mencionar. Não obstante, os instrumentos de política estão longe de serem tecnologias técnicas e neutras que os atores políticos possam manipular à vontade (LASCOURMES; LE GALÈS, 2007). Apesar de seu caráter socialmente construído, ou talvez por causa disso, os instrumentos de política, uma vez adotados, têm o potencial de afetar a política, a semiótica e a economia de políticas educacionais. Algumas premissas específicas podem ser testadas em relação ao papel constitutivo dos instrumentos de política: *em primeiro lugar*, “os instrumentos de política têm uma existência independente das decisões que os criaram” (KASSIM; LE GALÈS, 2010, p. 11). É difícil prever a forma que qualquer instrumento assumirá, uma vez que os “instrumentos de política criam as suas próprias estruturas de oportunidade de maneira que são tipicamente imprevisíveis quando são adotadas” (KASSIM; LE GALÈS, 2010, p. 12), e geram atividade política e consequências políticas que vão além da ontologia inicial do instrumento (PETERS, 2002).

Basicamente, nossos dados indicam que as NLSAs tendem a ser cada vez mais usadas para fins de responsabilização e monitoramento, mesmo em lugares onde elas não foram inicialmente adotadas com esses objetivos. Uma hipótese defendida aqui é que as NLSAs, no momento de sua adoção, não são vistas como instrumentos intrusivos pelos atores da escola devido a seus objetivos de diagnóstico educacional e/ou de classificação de estudantes; no entanto, uma vez que a complexa infraestrutura de avaliação em larga escala está em vigor, as políticas de responsabilização começam a ser adotadas e são cada vez mais associadas aos resultados da avaliação. Esse é, por exemplo, o caso do teste de fim de ensino primário na Holanda. Embora o teste tenha sido originalmente destinado a fornecer informações sobre o tipo de escola secundária mais apropriada para cada aluno, novos usos foram associados a ele nos últimos anos. Mais especificamente, o teste se tornou um instrumento para fins de avaliação escolar, uma vez que os resultados agregados são agora usados pelos inspetores como

indicadores da qualidade da escola (NUSCHE et al., 2014; ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2013a; SCHEERENS et al., 2012). Há também atores que, além dos governos nacionais, podem começar a fazer políticas de responsabilização com as NLSAs. Na Dinamarca e na Noruega, o governo parou de produzir classificações com pontuações escolares devido à contestação das organizações de professores e de outras partes educacionais interessadas. No entanto, a mídia e alguns governos locais, aproveitando as regras de transparência na administração pública, têm acesso aos dados das NLSAs para produzir e disseminar os seus próprios rankings escolares (HATCH, 2013; YDESEN; ANDREASEN, 2014).

Em segundo lugar, uma vez selecionados, os instrumentos de política privilegiam determinados atores e seus interesses ao determinar a alocação de recursos, o acesso ao processo de políticas e as representações de problemas (MENON; SEDELMEIER, 2010). Aqui, é importante observar que atores educacionais são capacitados pelos sistemas de TBA, e quais atores não têm poder, bem como se esses sistemas contribuem para o surgimento de novas subjetividades políticas, tais como o movimento de autoexclusão (*opt-out*), que está cada vez mais influente em várias localidades dos Estados Unidos (HUTT; SCHNEIDER, 2018; PIZMONY-LEVY; SARAISKY, 2016). Interesses econômicos também emergem em torno de novos instrumentos de política, tais como produtores terceirizados (por exemplo, testando empresas encarregadas das NLSAs) e partes apoiando a implementação de políticas (por exemplo, consultores oferecendo serviços de análise de dados, de melhoria instrucional ou de preparação de testes para escolas no intuito de melhorar as suas pontuações). Quando esses interesses econômicos são fortes, há mais razões para esperar que os instrumentos de política perdurem, independentemente do quão efetivos eles sejam.

Finalmente, os efeitos produzidos pelos instrumentos de política “dependem de como os objetivos e propósitos atribuídos a eles, e os significados e representações que eles carregam, são percebidos, compreendidos e respondidos por atores-chave” (SKEDSMO, 2011, p. 07). De forma geral, uma abordagem de sociologia política aplicada a ins-

trumentos de política visa ir além do estudo do impacto das políticas (em termos de, por exemplo, acesso educacional, realização de aprendizagem, e assim por diante). Em vez disso, ela visa analisar os tipos de relações sociais, comportamentos e respostas que os instrumentos desencadeiam entre as comunidades de prática (LASCOUNES; LE GALÈS, 2007). Instrumentos de política são formas de poder que visam disciplinar o comportamento dos atores, mas também sofrem resistência e/ou são criativamente transformadas pelos atores educacionais. Assim, para entender como as NLSAs evoluem, precisamos ir além do nível da formulação formal de políticas e observar como os praticantes também “fazem políticas” (BALL; MAGUIRE; BRAUN, 2012) ao adotar esses instrumentos e os mandatos educacionais implícitos a eles, bem como se o fazem de forma imprevista ou não intencional.

Conclusão

Na era do “governar por números” (GREK, 2009; OZGA, 2009), as NLSAs se tornaram um instrumento fundamental na promoção de uma abordagem de reforma educacional estruturada em torno de três princípios políticos principais, a saber: padrões, responsabilização e descentralização. As avaliações nacionais em larga escala são um componente central dessa abordagem de reforma e qualificadas por muitos como o GERM. As NLSAs geram os dados necessários para que o Estado responsabilize diferentes atores educacionais pelo aproveitamento dos alunos e por sua adesão aos padrões curriculares nacionais. Tanto o número de países que adotam as NLSAs quanto o número de NLSAs administradas no espaço político da OCDE cresceram substancialmente nas duas últimas décadas.

Além disso, os usos de monitoramento e responsabilização dessas avaliações também se intensificaram, mesmo em países da Europa continental que, até recentemente, pareciam ser imunes ao GERM. O nosso banco de dados mostra que a disseminação das NLSAs abre uma janela de oportunidade para a adoção de sistemas de TBA (e outros componentes centrais do GERM), apesar de não especificar os usos finais desses instrumentos de política pelos países. Na verdade, os dados de

desempenho das escolas gerados pelas NLSAs podem ser usados como uma ferramenta de avaliação formativa (ou seja, como uma ferramenta que apoia a melhoria instrucional) ou como um dispositivo para acionar a competição entre escolas, assim como para controlar e sancionar os atores escolares. Uma análise mais precisa dos usos internacionalmente específicos das NLSAs exigiria o acesso a novas fontes de dados secundários ou, devido à sua natureza de mudança constante, à condução de estudos de casos qualitativos e históricos que analisem o desenho político em evolução das configurações de NLSA/TBA e sua recontextualização em nível escolar.

As abordagens teóricas predominantes para a transferência de políticas (a saber: racionalismo, neoinstitucionalismo e economia política internacional) se referem aos macromecanismos e fatores globais por trás da disseminação internacional das NLSAs e da TBA. No entanto, uma abordagem de sociologia política aplicada a instrumentos de política é mais apropriada para entendermos quais configurações específicas de TBA estão sendo retidas em sistemas educacionais, e como elas são transformadas, ressignificadas e implementadas pelos atores educacionais que operam em múltiplas escalas. As avaliações nacionais em larga escala e a TBA se disseminam devido a mandatos e pressões mutáveis da educação global, mas também porque representam um modelo de reforma da educação politicamente gratificante, economicamente conveniente e politicamente maleável em nível doméstico. Essa abordagem de pesquisa também concebe as NLSAs e a TBA como dispositivos autônomos que – além de instrumentos de garantia de qualidade simples ou neutros – podem contribuir para mudar as configurações dos atores da educação, os comportamentos escolares e os objetivos das organizações educacionais de maneiras não previstas por seus promotores e criadores. A abordagem da sociologia política aplicada a instrumentos de política se encaixa dentro de uma ontologia da economia política cultural, no sentido de que está interessada em observar como a semiose e a construção de significado fazem parte da configuração político-econômica mais ampla das instituições educacionais. Ela também se encaixa em uma abordagem multiescalar da

política educacional, no sentido de que concebe que fatores e atores que operam em múltiplas escalas interagem criticamente na produção de políticas: desde o estabelecimento de agendas políticas globais até as práticas políticas cotidianas. As avaliações nacionais em larga escala, os mecanismos de garantia de qualidade e a TBA começaram a ser analisados a partir de uma pers-

pectiva semelhante (GABLE; LINGARD, 2015; KAUKO; RINNE; TAKALA, 2018; SKEDSMO, 2011), mas é necessário realizar pesquisas mais empíricas a fim de viabilizar um entendimento mais complexo de como, e por que, essas políticas evoluem e se expandem em todo o mundo, assim como se seus múltiplos efeitos nos sistemas educacionais endossam essa expansão.

REFERÊNCIAS

- ALLAN, J.; ARTILES, A. J. (Ed.). **World Yearbook of Education 2017: assessment inequalities**. New York: Routledge, 2016.
- AU, W. “High-Stakes testing and curricular control: a qualitative metasynthesis.” **Educational Researcher**, v. 36, n. 5, p. 258-267, 2007.
- BALL, S. J. **The education debate**. Bristol: The Policy Press, 2008.
- BALL, S. J.; MAGUIRE, M.; BRAUN, A. **How schools do policy**. Policy enactments in secondary schools. Abingdon: Routledge, 2012.
- BALL, S. J.; JUNEMANN, C.; SANTORI, D. **Edu.net**. Globalisation and education policy mobility. Abingdon: Routledge, 2017.
- BENAVOT, A.; TANNER, E. **The growth of national learning assessments in the world, 1995-2006**. Background paper prepared for the 2008 Global Monitoring Report. Paris: UNESCO-GMR, 2007. Disponível em: <<http://unesdoc.unesco.org/images/0015/001555/155507e.pdf>>. Acesso em:
- BOARINI, R.; LÜDEMANN, E. The role of teacher compensation and selected accountability policies on learning outcomes: an empirical analysis for OECD countries. **OECD Journal: Economic Studies**, v. 1, p. 1-20, 2009.
- BOVENS, M. Analysing and assessing accountability: a conceptual framework. **European Law Journal**, v. 13, n. 4, p. 447-468, 2007.
- CARNOY, M. Educational policies in the face of globalization: whither the nation state? In: MUNDY, K. Et al (Ed.). **Handbook of global policy and policy-making in education**. West Sussex: Wiley-Blackwell, 2016. p. 27-42,
- DALE, R. Specifying globalisation effects on national policy: A focus on the mechanisms. **Journal of Education Policy**, v. 14, n. 1, p. 1-17, 1999.
- DOBBINS, M.; KNILL, C. Higher education policies in central and eastern europe: convergence toward a common model? **Governance**, v. 22, n. 3, p. 397-430, 2009.
- EURYDICE; EDUCATION, AUDIOVISUAL AND CULTURE EXECUTIVE AGENCY – EACEA. **National testing of pupils in Europe: objectives, organisation and use of results**. Brussels: Eurydice/EACEA, 2009. Disponível em: <http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/109EN.pdf>. Acesso em:
- GABLE, A.; LINGARD, B. NAPLAN Data: a new policy assemblage and mode of governance in australian schooling. **Policy Studies**, v. 37, p. 6, p. 568-582, 2015.
- GREK, S. Governing by Numbers: The PISA “effect” in Europe. **Journal of education policy**, v. 24, n. 1, p. 23-37, 2009.
- GUNTER, H. M. Et al. NPM and the dynamics of education policy and practice in Europe. In: GUNTER, H. M. Et al (Ed.). **New public management and the reform of education: european lessons for policy and practice**. London: Routledge, 2016. p. 3-17.
- HAMILTON, L. S.; STECHER, B. M.; KLEIN, S. P. (Ed.). **Making sense of test-based accountability in education**. Santa Monica: RAND Corporation, 2002. Disponível em: <https://www.rand.org/content/dam/rand/pubs/monograph_reports/2002/MR1554.pdf>. Acesso em:

- HARGREAVES, A. Et al. **Learning to change**. Teaching beyond subjects and standards. San Francisco: Jossey-Bass, 2001.
- HATCH, T. Beneath the surface of accountability: answerability, reponsibility and capacity-building in recent education reforms in Norway. **Journal of Educational Change**, v. 14, n. 2, p. 113-138, 2013.
- HUTT, E.; SCHNEIDER, J. **A thin line between love and hate**: educational assessment in the United States. 2018. Acesso em: <https://www.holycross.edu/sites/default/files/educ/thin_line.pdf>. Acesso em:
- JONES, L. V. A history of the national assessment of educational progress and some questions about its future. **Educational Researcher**, v. 25, n. 7, p. 15-22, 1996.
- JONES, B. D.; EGLEY, R. J. Voices from the frontlines: teachers' perceptions of high-stakes testing. **Education Policy Analysis Archives**, v. 12, n. 39, 2004.
- KAMENS, D. H.; MCNEELY, C. L. Globalization and the growth of international educational testing and national assessment. **Comparative Education Review**, v. 54, n. 1, p. 5-25, 2010.
- KASSIM, H.; LE GALÈS, P. Exploring governance in a multi-level polity: a policy instruments approach. **West European Politics**, v. 33, n. 1, p. 1-21, 2010.
- KAUKO, J.; RINNE, R.; TAKALA, T. **Politics of quality in education**: a comparative study on Brazil, China, and Russia. New York: Routledge, 2018.
- KLEES, S. J. A quarter century of neoliberal thinking in education: misleading analyses and failed policies. **Globalisation, Societies and Education**, v. 6, n. 4, p. 311-348, 2008.
- LASCOUMES, P.; LE GALÈS, P. Introduction: understanding public policy through its instruments. From the nature of instruments to the sociology of public policy instrumentation. **Governance**, v. 20, n. 1, p. 1-21, 2007.
- MENON, A.; SEDELMEIER, U. Instruments and intentionality: civilian crisis management and enlargement conditionality in EU security policy. **West European Politics**, v. 33, n. 1, p. 75-92, 2010.
- MEYER, H-D.; BENAVIDOT, A. PISA and the globalization of education governance: some puzzles and problems. In: _____. (Ed.). **PISA, power, and policy**: the emergence of global educational governance. Oxford: Symposium Books, 2013. p. 9-26.
- NUSCHE, D. Et al. **OECD reviews of evaluation and assessment in education**: Netherlands 2014. Paris: OECD Publishing, 2014. Disponível em: <<http://www.oecd.org/education/school/OECD-Evaluation-Assessment-Review-Netherlands.pdf>>. Acesso em:
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT – OECD. **Learning for tomorrow's world**. First results from PISA 2003. Paris: OECD, 2004. Disponível em: <<http://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/34002216.pdf>>. Acesso em:
- _____. **PISA 2006**. Science competencies for tomorrow's world. Paris: OECD, 2007. Disponível em: <http://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/pisa2006results.htm#tables_figures_dbase>. Acesso em:
- _____. **Synergies for better learning**. An international perspective on evaluation and assessment. OECD Reviews of Evaluation and Assessment in Education. Paris: OECD, 2013a. Disponível em: <<https://www.oecd-ilibrary.org/docserver/9789264190658-en.pdf?expires=1523571951&id=id&accname=guest&checksum=DBDA81CBDD-068F947A077B99A8005765>>. Acesso em:
- _____. **Results: What makes a school successful?** Resources, policies and practices. Vol. 6. Paris: OECD, 2013b. Disponível em: <https://www.oecd-ilibrary.org/education/pisa-2012-results-what-makes-a-school-successful-volume-iv_9789264201156-en>. Acesso em:
- _____. **PISA Results**: policies for successful schools. Vol. 2. Paris: OECD, 2016. Disponível em: <https://www.oecd-ilibrary.org/education/pisa-2015-results-volume-ii_9789264267510-en>. Acesso em:
- OLMEDO, A.; WILKINS, A. Governing through parents: a genealogical enquiry of education policy and the construction of neoliberal subjectivities in England. **Discourse: Studies in the Cultural Politics of Education**, v. 38, n. 4, p. 573-589, 2017.
- OSBORNE, D.; GAEBLER, T. **Reinventing government**: how the entrepreneurial spirit is transforming the public

sector. New York: Addison-Wesley, 1992.

OZGA, J. Governing education through data in England: from regulation to self-evaluation. **Journal of Education Policy**, v. 24, n. 2, p. 149-163, 2009.

PETERS, B. G. The politics of tool choice. In: SALAMON, L. M. (Ed.). **The tools of government: a guide to the new governance**. Oxford: Oxford University Press, 2002. p. 552-564.

PIZMONY-LEVY, O.; SARAISKY, N. G. **Who opts out and why?** Results from a national survey on opting out of standardized tests. New York: Teachers College/Columbia University, 2016.

POPHAM, J. Why standardized tests don't measure educational quality. **Educational Leadership**, v. 56, n. 6, p. 8-15, 1991. Disponível em: <<http://www.ascd.org/publications/educational-leadership/mar99/vol56/num06/Why-Standardized-Tests-Don%27t-Measure-Educational-Quality.aspx>>. Acesso em:

ROBERTSON, S. L. Researching global education policy: angles in/on/out. In: VERGER, A.; NOVELLI, M.; ALTINYELKEN, H. (Ed.). **Global education policy and international development: new agendas, issues and practices**. London: Continuum, 2012. p. 33-53.

_____. What teachers need to know about the global education reform movement. In: LITTLE, G. (Ed.). **Global education "reform"**. Building resistance and solidarity. Croydon: Manifesto Press, 2015. p. 10-17.

SAHLBERG, P. Education reform for raising economic competitiveness. **Journal of Educational Change**, v. 7, n. 4, p. 259-287, 2005.

_____. The global educational reform movement and its impact on schooling. In: MUNDI, K. Et al. (Ed.). **The handbook of global education policy**. West Sussex: Wiley-Blackwell, 2016. p. 128-144.

SCHEERENS, J. Et al. **OECD review on evaluation and assessment frameworks for improving school outcomes**. Country background for the Netherlands. Paris: OECD, 2012. Disponível em: <http://www.oecd.org/education/school/NLD_CBR_Evaluation_and_Assessment.pdf>. Acesso em:

SKEDSMO, G. Formulation and realisation of evaluation policy: inconcistencies and problematic issues. **Educational Assessment, Evaluation and Accountability**, v. 23, n. 1, p. 5-20, 2011.

SMITH, W. C. An introduction to the global testing culture. In: _____ (Ed.). **The global testing culture: shaping education policy, perceptions, and practice**. London: Symposium Books, 2016. p. 7-23.

SMITH, W. C.; KUBACKA, K. The emphasis of student test scores in teacher appraisal systems. **Education Policy Analysis Archives**, v. 25, n. 86, p. 1-29, 2017.

SMITH, M. L. Et al. **Political spectacle and the fate of american schools**. New York: Routledge, 2004.

STEINER-KHAMSI, G. New directions in policy borrowing research. **Asia Pacific Education Review**, v. 17, n. 3, p. 381-390, 2016.

TOBIN, M. Et al. **Using large-scale assessments of students' learning to inform education policy. Insights from the Asia-Pacific region**. Melbourne/Bangkok: ACER/UNESCO, 2015. Disponível em: <https://research.acer.edu.au/cgi/viewcontent.cgi?article=1020&context=monitoring_learning>. Acesso em:

UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION – UNESCO. **Strong foundations**. Early childhood care and education. Education for all global monitoring report 2007. Paris: UNESCO Publishing, 2007. Disponível em: <<http://unesdoc.unesco.org/images/0014/001477/147794e.pdf>>. Acesso em:

_____. **Education for all 2000-2015: achievement and challenges**. Education for all global monitoring report 2007. Paris: UNESCO Publishing, 2015. Disponível em: <<http://unesdoc.unesco.org/images/0023/002322/232205e.pdf>>. Acesso em:

VERGER, A.; PARCERISA, L. Test-based accountability and the rise of regulatory governance in education: a review of global drivers. In: WILKINS, A.; OLMEDO, A. (Ed.). **Education governance and social theory: interdisciplinary approaches to research**. London: Bloomsbury, 2018.

VERGER, A.; NOVELLI, M.; ALTINYELKEN, H. K. Global education policy and international development: a revisited introduction. In: _____ (Ed.). **Global education policy and international development new agendas, issues and policies**. 2. ed. London: Bloomsbury, 2018.

YDESEN, C.; ANDREASEN, K. E. Accountability practices in the history of Danish primary public education from the 1660s to the present. **Education Policy Analysis Archives**, v. 22, n. 120, 2014.

WEILER, H. Comparative perspectives on educational decentralization: an exercise in contradiction? **Educational Evaluation and Policy Analysis**, v. 12, n. 4, p. 433-448, 1990.

Recebido em: 10/07/2018

Aprovado em: 10/09/2018