



Ensaio: Avaliação e Políticas Públicas em
Educação
ISSN: 0104-4036
ensaio@cesgranrio.org.br
Fundação Cesgranrio
Brasil

Ferrão, Maria Eugénia; Couto, Alcino
Indicador de valor acrescentado e tópicos sobre consistência e estabilidade: uma
aplicação ao Brasil
Ensaio: Avaliação e Políticas Públicas em Educação, vol. 21, núm. 78, enero-marzo,
2013, pp. 131-164
Fundação Cesgranrio
Rio de Janeiro, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=399538144008>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

Indicador de valor acrescentado e tópicos sobre consistência e estabilidade: uma aplicação ao Brasil

Maria Eugénia Ferrão*

Alcino Couto**

Resumo

Este artigo examina a abordagem de valor acrescentado no contexto da literatura de eficácia escolar. Debatidos aspectos relevantes do estado-da-arte sobre o uso do indicador de valor acrescentado, o enfoque recai sobre os problemas de escolha do modelo estatístico e das propriedades de consistência e estabilidade do indicador, considerando o contexto educacional brasileiro. São apresentadas as vantagens e limitações que, por sua vez, elucidam quanto ao uso que lhe possa ser dado. A componente empírica é desenvolvida através da aplicação de modelo multinível de componentes de variância aos dados do GERES 2005 referentes ao município de Campinas, dado este apresentar, simultaneamente, maior percentagem de escolas por estrato e menor taxa de atrito. A evidência empírica revela que o indicador produzido pelo modelo de resultados contextualizados é muito diferente do de valor acrescentado, sustentando a necessidade de adoção de estudos longitudinais. Os níveis de consistência e de estabilidade encontrados face aos dois modelos de valor acrescentado testados sugerem a sua utilidade para atempadamente diagnosticar as escolas que sistematicamente apresentam desempenho próprio de escolas eficazes ou de escolas não eficazes.

Palavras-chave: Desempenho escolar. Avaliação. Valor acrescentado. Modelo de regressão multinível. GERES 2005

Value-added indicator and topics on consistency and stability: an application to Brazil

Abstract

The paper presents a value-added approach in the context of the literature of school effectiveness. Relevant aspects such as the use of the value-added indicator, the choice of the statistical model, and the scientific properties of consistency and stability in Brazilian education are discussed. The empirical evidence presented is based on the

* Doutora em Teoria de Controle e Estatística, PUC-Rio. Universidade da Beira Interior e CEMAPRE/Instituto Superior de Economia e Gestão *E-mail:* meferrao@ubi.pt

** Professor Auxiliar de Economia da Universidade da Beira Interior *E-mail:* acouto@ubi.pt

variance component models applied to GERES 2005 data collected in the municipality of Campinas. The results show that the indicator obtained by the contextualized results model is rather different from the one obtained by a value-added model, supporting the need of conducting longitudinal data studies. The findings about consistency and stability suggest the value-added usefulness to timely diagnose schools that systematically perform as effective schools or as ineffective schools.

Keywords: School performance. Evaluation. Value added. Multilevel model. GERES 2005.

Indicador de Valor adicionado y Tópicos Sobre Consistencia y Estabilidad: una aplicación a Brasil

Resumen

Este artículo analiza el abordaje de valor adicionado en el contexto de la literatura de eficacia escolar. Discutidos aspectos importantes de los últimos avances actuales sobre el uso del indicador de valor adicionado, el enfoque recae sobre los problemas surgidos al elegir el modelo estadístico y las propiedades de consistencia y estabilidad del indicador, dentro del contexto educacional brasileño. Se presentan las ventajas y limitaciones que, a su vez, esclarecen el uso que se le puede dar. El componente empírico se desarrolla a través de la aplicación de un modelo multinivel de componentes de varianza a los datos del GERES 2005 (Estudio de la Generación Escolar 2005) referentes al municipio de Campinas (Brasil), ya que presenta, simultáneamente, mayor porcentaje de escuelas por estrato y menor tasa de desavenencia. La evidencia empírica revela que el indicador que el modelo de resultados contextualizados produce es muy diferente del de valor adicionado, lo que fundamenta la necesidad de adopción de estudios longitudinales. Los niveles de consistencia y de estabilidad encontrados frente a los modelos de valor adicionado verificados sugieren su utilidad para diagnosticar a tiempo las escuelas que sistemáticamente presentan desempeño propio de escuelas eficaces o de no eficaces.

Palabras clave: Desempeño escolar. Evaluación. Valor adicionado. Modelo de regresión multinivel. GERES 2005.

Introdução

A qualidade da educação e a equidade na distribuição das oportunidades de aprendizagem têm suscitado um debate controverso e complexo sobre a sua natureza, as suas causas, o desempenho relativo das escolas e de outras unidades educativas. Nele incluem-se controvérsias sobre as implicações quanto às políticas educativas, às reformas, à conceptualização dos modelos e instrumentos de avaliação, sua validade e fiabilidade e seus efeitos nas decisões dos agentes e na eficácia educacional.

Há muito tempo que a relevância da avaliação educacional é matéria de concordância nesse debate, bem como a pertinência de um sistema de avaliação educacional baseado em indicadores de desempenho. (WHOLEY, 1979) Os indicadores

de desempenho dos serviços públicos são usados tipicamente para aferir o impacto das políticas ou programas nesses serviços ou para comparar o desempenho institucional ou profissional. Os indicadores de desempenho também podem ser usados no âmbito de políticas de responsabilização – prestação de contas e, nesse sentido, o Governo assume o duplo papel de monitorizar o desempenho das instituições públicas e ser ele próprio monitorizado pelos indicadores de desempenho.

Principalmente devido a esse duplo papel, a concepção, planeamento e apuramento dos indicadores de desempenho têm de obedecer aos princípios e requisitos metodológicos inerentes à produção de estatísticas oficiais, tais como são os princípios de independência profissional, de objectividade e imparcialidade, de relevância, de precisão e fiabilidade, de coerência e comparabilidade. Para efeito da comparabilidade do desempenho das escolas, o indicador baseado no conceito de valor acrescentado (VA) é amplamente consensual (BRAUN; CHUDOWSKY; KOENIG, 2010; GOLDSTEIN; SPIEGELHALTER, 1996; OECD, 2008). Todavia, aspetos referentes à sua operacionalização ainda emergem como obstáculo à adopção e uso para determinadas finalidades. Alguns desses aspetos dizem respeito à especificação do modelo estatístico e ao estudo das propriedades de consistência e estabilidade do indicador.

Este artigo tem como principal objetivo trazer à discussão alguns desses aspetos, nomeadamente o da escolha do modelo estatístico e das propriedades de consistência e estabilidade do indicador, considerando o contexto educacional brasileiro; são apresentadas as vantagens e limitações que, por sua vez, elucidam quanto ao uso que a ele possa ser dado. A componente empírica é desenvolvida através da aplicação de modelo multinível aos dados do GERES 2005. Concretamente, o modelo de componentes de variância foi aplicado aos dados referentes ao município de Campinas, por ser o que simultaneamente apresenta maior percentagem de escolas por estrato e menor taxa de atrito.

Na seção seguinte é apresentada uma revisão sintética da literatura em eficácia educacional e áreas associadas. A seção três é constituída pela descrição metodológica do estudo empírico; na seção quatro, apresentam-se os resultados; e na quinta, a sua discussão e, por fim, algumas considerações relativamente às opções metodológicas tomadas na conduta do estudo.

Modelo de Valor Acrescentado: conceito, estabilidade e consistência

Braun e Wainer (2007) definem modelo de valor acrescentado (MVA) como sendo a família de modelos estatísticos que são usados para produzir inferência sobre a eficácia das unidades educativas, usualmente escolas e/ou professores¹. O relatório da OCDE apresenta uma definição mais detalhada do ponto de vista da implementação

¹ “[...] a family of statistical models that are employed to make inferences about the effectiveness of educational units, usually schools and /or teachers”.

dos modelos: "A class of statistical models that estimate the contributions of schools to student progress in stated or prescribed education objectives (e.g. cognitive achievement) measured at least two points in time" (OECD, 2008, p. 17). Decorre destas definições que o VA representa a contribuição da escola para o progresso dos alunos, no que se refere aos objetivos educativos estabelecidos e que essa contribuição está livre da influência de outros factores extra-escolares que contribuem para o progresso educativo dos alunos. Sammons, Thomas e Mortimore (1997, p.24) definem-no como sendo uma indicação de quanto a escola contribui para o progresso de todos os seus alunos num conjunto de áreas (académicas ou não), durante determinado período de tempo, comparativamente com a contribuição das outras escolas. Conceptualmente encontra-se concordância na literatura relativamente à definição de VA (ARIAS ; SOTO, 2009; BALLOU; SANDERS; WRIGHT, 2004; BRAUN; CHUDOWNSKY; KOENING, 2010; BRIGGS; WEEKS, 2011; FERRÃO, 2009a; 2012a; FERRÃO; GOLDSTEIN, 2009; FIELDING; YANG; GOLDSTEIN, 2003; GOLDSTEIN, 1997; GOLDSTEIN; SAMMONS, 1997; GOLDSTEIN; SPIEGELHALTER, 1996; LADD; WALSH, 2002; LOCKWOOD et al., 2007; MARTINEAU, 2006; MORTIMORE; SAMMONS; THOMAS, 1994; NEWTON et al., 2010; RAUDENBUSH; WILLMS, 1995; REARDON; RAUDENBUSH, 2009; RUBIN; STUART; ZANUTTO, 2004; SAMMONS; THOMAS; MORTIMORE, 1997; TEKWE et al., 2004; THOMAS; PENG; GRAY, 2007; WEST, 2010). A sua quantificação ou medição é o computo da contribuição das escolas para o progresso dos alunos (SAUNDERS, 1999). Todavia, encontram-se diversas abordagens no tratamento estatístico para operacionalizar o conceito, ao nível do desempenho quer institucional, quer profissional. Apesar disso, há consenso amplo relativamente a um conjunto de requisitos metodológicos para a obtenção das estimativas de VA na educação, a saber:

1. aferir o nível de conhecimento do aluno à entrada na escola ou no início do período de tempo em avaliação (desempenho prévio);
2. adoptar plano de recolha de dados longitudinal que envolva no mínimo três coortes de alunos;
3. considerar a estrutura hierárquica da população em estudo, adequadamente representada nos dados colectados;
4. considerar os factores extra-escolares com influência nos resultados de aprendizagem dos alunos, nomeadamente o estatuto socioeconómico e cultural dos alunos (AITKIN; LONGFORD, 1986; BIRD et al., 2005; BRAUN et al., 2010; GOLDSTEIN; SPIEGELHALTER, 1996; GRAY et al., 1995; RUBIN et al., 2004).

As estimativas de VA das escolas dependem do tipo de modelo estatístico usado, do conjunto de variáveis contextuais incluídas no modelo, da dimensão das escolas, do período temporal compreendido (isto é, do número de anos lectivos entre a aferição do desempenho prévio e do resultado escolar do aluno), do conjunto de áreas (académicas ou não) consideradas, e da cobertura da amostra. De acordo com a revisão efectuada

por (FERRÃO, 2012a), a consistência diz respeito ao grau de concordância entre as estimativas de VA baseadas em diferentes conjuntos de variáveis e especificações do modelo. Usualmente a consistência é aferida ou pela correlação entre posições de ordem ou entre as estimativas de VA. A literatura apresenta evidência de consistência elevada. Por exemplo, Thomas e Mortimore (1996) compararam cinco modelos multinível, cada um deles com especificação e complexidade diferentes, para averiguar qual o mais adequado. Um dos modelos inclui as seguintes variáveis de perfil do aluno: desempenho prévio verbal e quantitativo e também capacidade cognitiva não-verbal, sexo, étnia, mobilidade, situação face a apoio social (free school meals). A consistência reportada é forte. Também Fielding, Yang e Goldstein (2003) compararam estimativas de VA baseando-se em especificações diferentes do modelo multinível. Um deles é um modelo cuja variável resposta tem distribuição condicionalmente normal, e o outro é um modelo cuja variável resposta é ordinal. A correlação entre as estimativas de VA obtidas através desses modelos é maior do que 0.96. Tekwe et al. (2004) compararam os resultados de quatro modelos estatísticos aplicados aos dados de alunos nos 3º, 4º, 5º anos de 22 escolas elementares da Flórida. Os modelos variam em complexidade e nenhum satisfaz simultaneamente os requisitos acima enunciados. O valor da correlação entre as estimativas obtidas dos diferentes modelos é maior do que 0.90, com exceção das que envolvem o modelo multinível mais complexo cujos valores são maiores do que 0.70. Com base em dados recolhidos numa amostra aleatória de alunos dos 1º, 3º, 5º, 7º e 8º anos de escolaridade da região da Cova da Beira, Portugal, Ferrão e Goldstein (2009) reportam correlação com valores superiores a 0.88, quando as estimativas de VA comparadas são as obtidas através de um modelo multinível que inclui como variáveis de perfil do aluno, o seu desempenho prévio e o seu nível socioeconómico (modelo designado pelos autores como MVA tradicional), e através de um modelo que inclui adicionalmente as seguintes variáveis: sexo, frequência de pré-escola ou jardim escola, aluno com necessidades educativas especiais, se ficou retido pelo menos um ano em algum ano de escolaridade.

Apesar da evidência tender a suportar a consistência do indicador de VA, deve ter-se em conta que se trata duma propriedade estudada até à data a partir de dados delimitados no tempo. Isto é, a comparação entre o VA das escolas tem por referência determinado coorte de alunos em determinado momento da sua trajectória escolar. A sua utilidade é, por isso, limitada do ponto de vista da transferência do seu significado à prática e da sua tradução em medidas que favoreçam a melhoria da educação. A sobreposição dos respectivos intervalos de confiança sugere que apenas as escolas nos extremos da distribuição se diferenciam entre si (BIRD et al., 2005; GOLDSTEIN ; HEALY, 1995; GOLDSTEIN; SPIEGELHALTER, 1996). Leckie e Goldstein (2009) mostram que, devido ao reduzido número de alunos por escola-coorte, apenas 63.2% das escolas apresentam VA estatisticamente diferente da média. Assim, as estimativas baseadas em apenas um ano são de uso limitado se o interesse é a comparação da performance institucional.

Tal facto sugere a necessidade de identificação das escolas que mantêm a sua posição de ordem ao longo do tempo, bem como as que permanentemente apresentam evidência de melhoria, de tal modo que seja possível estudá-las em detalhe. É a este aspeto que a propriedade estabilidade concerne. A estabilidade diz respeito ao grau de concordância entre as estimativas de VA ao longo do tempo (FERRÃO, 2012a). A complexidade dos dados necessários à quantificação da estabilidade reduz consideravelmente o número de países e de contextos educativos para os quais é possível estudar esta propriedade. Segundo a revisão efectuada sobre o tema num dos primeiros artigos que o abordaram (GRAY et al., 1995), nenhum dos doze estudos realizados em termos internacionais até 1995 satisfazia os requisitos metodológicos acima enunciados. Mesmo o estudo reportado em Gray et al. (1995) não incluiu o nível socioeconómico dos alunos como variável de controle. Talvez por isso, a correlação entre as estimativas de VA referentes a um grupo de escolas secundárias no ano de 1990 e 1991 e entre 1991 e 1992 é, respectivamente, 0.94 e 0.96. No estudo mais recente de 63 escolas secundárias de Lancashire, Thomas, Peng e Gray (2007) apresentam valores de correlação que variam entre 0.79 e 0.89 com base em modelos VA que incluem o nível socioeconómico entre as variáveis de perfil do aluno.

Até ao momento, os estudos indicam valores moderados da estabilidade, em particular no ensino primário. Considerando o ensino secundário, a estabilidade quantificada através da correlação entre as estimativas de VA para o intervalo de tempo de quatro anos é em torno de 0.56 (GRAY; GOLDSTEIN; JESSON, 1996; GRAY; GOLDSTEIN ; THOMAS, 2001; YANG; WOODHOUSE, 2001). Leckie e Goldstein (2009) reportam valores da correlação que decrescem com o aumento do número de anos de permeio (0.89, 0.87, 0.76, 0.70, 0.64). De referir que os valores da correlação reportados para o período 1990-1994 por Gray, Goldstein e Jesson (1996) também são decrescentes.

Um estudo da educação primária cipriota realizado por Kyriakides e Creemers (2008) mostra que 20 em 28 (71.4%) escolas permanecem no mesmo nível de VA (estatisticamente menor/igual/maior do que zero) ao longo de quatro anos lectivos. O estudo da estabilidade do VA em turmas do ensino primário do sistema educativo português entre 2005/6 e 2007/8 mostra que 65% permanecem no mesmo nível (quartil) de VA por, pelo menos, dois anos lectivos consecutivos e 11% permanecem no mesmo nível de VA pelos três anos lectivos. Estes resultados foram obtidos com base na amostra representativa do coorte de alunos que começaram o 1º ano de escolaridade em 2005/2006 na região NUT III Cova da Beira (FERRÃO, 2012a).

Valor Acrecentado e Eficácia Escolar: fundamentos, modelagem e uso

A seção apresenta a revisão sobre valor acrescentado da escola a partir da literatura em eficácia escolar, aspectos técnicos da sua modelagem, bem como uma breve discussão sobre o uso do indicador de valor acrescentado.

Valor Acrescentado e Eficácia Escolar

A definição VA anteriormente apresentada remete para a definição de eficácia educacional (*educational effectiveness*) que agrega a investigação sobre a eficácia escolar (*school effectiveness*) e eficácia docente (*teacher effectiveness*), na orientação dada por Scheerens & Bosker (1997) para abranger a eficácia do sistema educativo em geral. Eficácia escolar designa a área de investigação científica em Educação que é dedicada à estimação do efeito-escola (nas dimensões valor agregado, eficácia diferencial e equidade social), à identificação dos fatores que contribuem para que uma escola seja eficaz (FERRÃO et al., 2005; MORTIMORE; SAMMONS; THOMAS, 1994; TOWNSEND, 2007) e abrange ainda o campo de trabalho cujo enfoque é a procura de métodos adequados e fiáveis para medir a qualidade da escola (MORTIMORE, 1998, p. 258). Define-se usualmente *escola eficaz* como aquela onde o progresso do aluno vai além do que seria esperado, levando em consideração as suas características à entrada na escola (MORTIMORE, 1998, p. 237). Nesse sentido, a investigação em eficácia educacional tem procurado compreender, analisar e clarificar o papel da escola como pilar da aprendizagem dos alunos e do desempenho do sistema educativo, através do efeito-escola e do efeito-professor. O *efeito-escola* é um conceito ao qual estão associadas diversas definições ao longo dos últimos 40 anos de investigação (TEDDLIE; REYNOLDS, 2000, p. 65-68). O objetivo consiste não apenas em identificar e dissociar as variáveis controláveis pela escola das não controláveis e medir o contributo das primeiras para os resultados de aprendizagem dos alunos, mas também investigar políticas e práticas conducentes à sua transformação e melhoria continuada do seu desempenho. No âmbito deste artigo debruçamo-nos sobre o VA. Todavia, deixamos nota da importância do estudo do efeito-escola segundo a dimensão da eficácia diferencial e equidade social. Alguns dos estudos que mostram evidência da ocorrência de eficácia diferencial (*differential school effects*), isto é o efeito-escola pode ser influenciado pelo perfil do aluno à entrada na escola (*students' intake*), revelam ainda que a composição da população discente (*school composition effects*) também influencia os resultados atingidos. Estes colocam em evidência o facto do desempenho individual dos alunos de uma escola se encontrar associado ao perfil dominante dos seus alunos (*aggregated students' intake*). Sammons (1999, p. 102) argumenta "disadvantaged students in the most effective schools can end up with higher achievements than their advantaged peers in the less effective schools". Caso a "desvantagem" dos alunos seja de natureza socioeconómica, o efeito pode ser estudado na perspectiva da equidade social da escola.

A história da investigação em eficácia educacional está descrita em muitos artigos e livros (BROOKE; SOARES, 2008; MORTIMORE, 1998; MORTIMORE et al., 1988; MURILLO, 2003; RUTTER et al., 1979; SAMMONS; THOMAS; MORTIMORE, 1997; SCHEERENS; BOSKER, 1997; TEDDLIE; REYNOLDS, 2000; TEDDLIE; STRINGFIELD, 1993; TOWNSEND,

2007) e não é possível nem necessário revê-la no âmbito deste artigo. Pela influência que teve no Brasil, vale uma nota de destaque à revisão dos factores de eficácia elaborada por Sammons, Hillman e Mortimore (1995). A revisão foi o enquadramento teórico e conceptual para os primeiros trabalhos de eficácia escolar realizados no Brasil. Outra menção de destaque refere-se ao desenvolvimento do software estatístico para os modelos multinível (GOLDSTEIN, 1997, 2011), MLn esteve na origem do actual MIwiN, que viabilizou o aumento substancial no número de estudos, em contextos diversos e países diferentes. De modo geral, tais estudos têm vindo a corroborar a existência de diferenças significativas, tanto na acepção estatística como na educativa, no que se refere à contribuição das escolas para o desempenho dos alunos.

A investigação sobre eficácia educacional partilha de perspectivas multidimensional e multinível quanto às determinantes do desempenho escolar dos alunos. As diferenças observadas entre escolas nos resultados de aprendizagem são moldadas por um espectro abrangente e interactivo de variáveis tais como o ambiente da escola, liderança (MOLLER, 2009; ROBINSON; LLOYD; ROWE, 2008), qualidade do currículum, cultura de avaliação e monitorização (HOFMAN; DIJKSTRA; HOFMAN, 2009), cultura e experiência de cooperação entre escolas (WEST, 2010), cooperação entre os actores da escola e a cooperação com os encarregados de educação e a comunidade (HULPIA; VALCKE, 2004). Alguns estudos defendem a análise centrada no professor como elemento determinante do desempenho da escola (HILL; ROWE, 1994), donde decorre o argumento segundo o qual "teachers teach and schools operate" (OECD, 2008, p. 11) ou "... high quality teachers are the most important asset of schools" (HANUSHEK, 2010, p. 466). Esses estudos sustentam ainda que, para as escolas com desempenho relativo inferior melhorarem o seu desempenho, é fundamental contarem com professores experientes e qualificados. Contudo, a realidade mostra a incapacidade dessas escolas em atrair e reter professores mais qualificados (CLOTFELTER et al., 2004).

A Modelagem Estatística de VA

A literatura regista uma grande diversidade de análise e modelagem estatísticas para a comparação do desempenho institucional. Essa diversidade faz aumentar a importância de clarificar o que é o MVA, cuja definição foi apresentada na secção anterior. A sua operacionalização estatística tem levado alguns investigadores e analistas a designar de VA indicadores de resultados contextualizados (OECD, 2008, p. 11, 17). Os indicadores de resultados contextualizados e os de resultados escolares circunscritos aos resultados dos testes e dos exames (*raw scores*) são indicadores de *status*. O indicador VA diferencia-se, assim, conceptualmente dos indicadores *status* já que incluem a medida resultante da aferição do desempenho prévio do aluno.

Do ponto de vista da eficácia da escola e da sua melhoria, podemos salientar quatro aspectos importantes associados aos indicadores VA. Primeiro, permitem a identificação das escolas que mais contribuem para a aprendizagem dos alunos. Segundo, possibilitam ainda que escolas inseridas em contextos sociais, económicos

e culturais adversos possam ser reconhecidas pelo seu desempenho. Terceiro, tornam possível a identificação de situações críticas e de sucesso que possam apoiar a adopção de políticas de melhoria da eficácia da escola e de redução das desigualdades na educação. Finalmente, permite que a avaliação educacional possa estabelecer uma relação mais rigorosa e justa entre desempenho e incentivos do que a viabilizada apenas pelos indicadores de *status*.

Os atributos dos indicadores VA acima enunciados sugerem maior robustez na avaliação educacional das unidades fundamentais do sistema – escolas e professores – do que a oferecida pelos indicadores de *status*. Estes refletem o desempenho num determinado momento do tempo e a sua eventual utilização como critério de um sistema de incentivos tende a premiar escolas com reduzido VA e a desvalorizar a prestação das escolas inseridas em contextos adversos (BRAUN *et al.*, 2010; OECD, 2008; GOLDSTEIN, 1997; SAMMONS; THOMAS; MORTIMORE; BOSKER, 1997; SCHEERENS; BOSKER, 1997; RAUDENBUSH; WILLMS, 1995; MORTIMORE *et al.*, 1988). Algumas das escolas, onde os alunos atingem níveis de desempenho elevado, pouco contribuem para esse desempenho: tendencialmente são escolas frequentadas por alunos com elevado nível de desempenho à entrada (escolas selectivas) e provenientes de grupos sociais favorecidos. Em contrapartida, as escolas que recebem alunos com baixo desempenho e onde, no final do período sob avaliação, o desempenho se mantém aquém da meta educativa estabelecida, podem, apesar disso, ter contribuído fortemente para alavancar a aprendizagem desses alunos e, portanto, apresentam elevado VA. A sua contribuição não está reflectida no indicador de *status* mas sim no indicador de VA (FERRÃO, 2012b).

Uso do indicador VA

Quanto ao uso, o debate prende-se com a finalidade e a delimitação do universo de aplicação dos indicadores no âmbito de um sistema de avaliação educacional holístico. Invoca-se o artigo clássico (CRONBACH, 1963) sobre a natureza da avaliação, o seu uso e finalidade, em que o autor diz que uma finalidade importante da avaliação é "course improvement", ou seja, o objetivo da avaliação não deve ser apenas o juízo de valor final sobre a eficácia do currículo mas também fornecer informação para as modificações necessárias visando à melhoria dos cursos em apreço. É a partir desta ideia que, posteriormente, Scriven (1966) distingue entre avaliação formativa e sumativa. Ora, sobre o uso do indicador de VA no âmbito de um processo de avaliação, tal analogia conduz à identificação de duas perspectivas com finalidades instrumentais diferenciadas.

Por um lado, a corrente da *melhoria da escola*, utiliza os indicadores de VA como ferramenta para identificar quer boas práticas quer casos de insucesso, e retirar implicações de política e de ação orientadas para a melhoria da escola. O enfoque é colocado na criação de um ambiente incubador de dinâmicas de aprendizagem virtuosas, através de mudanças nas práticas organizacional, de ensino e pedagógicas, bem como o desenvolvimento de uma cultura e de políticas de auto-avaliação.

Por outro lado, e em confronto, o seu uso, como indicador de desempenho preferencial de um modelo de sistema educativo, assente numa lógica de *quasi*-mercado, tanto ao nível da avaliação das escolas como dos seus profissionais com consequências de elevado impacto (high-stakes). Modelo em que os princípios da responsabilização e da prestação de contas (*school accountability*) orientam a afectação de incentivos/penalização, da liberdade de escolha da escola (*school choice*), com recurso a mecanismos de incentivos do tipo cheques-ensino e a *charter schools*, parece reforçarem-se mutuamente como pilares das políticas de reforma orientada para a eficácia educacional (ANDRADE, 2009; MCCAFFREY et al., 2004). Contudo, observa-se que o primado da sua utilização instrumental na perspectiva de *quasi*-mercado carece de alguma precaução e de investigação adicional. O problema reside na sua associação às políticas de responsabilização e da prestação de contas, cujos resultados sobre a eficácia escolar são amplamente controversos face à ocorrência de efeitos não desejáveis comprometedores de ganhos de eficácia educacional (HANUSHEK; RAYMOND, 2005; MCCAFFREY et al., 2004). Estudos reportam a ocorrência de comportamentos oportunistas, em particular quando as duas políticas convergem no sentido do reforço da competição decorrente da divulgação dos resultados escolares e da criação de *ranking* de escolas. Por exemplo, estudos sobre No Child Left Behind Act of 2001 revelam uma elevada concentração de recursos e de esforço de aprendizagem em torno de alunos no limiar crítico de proficiência, com prejuízo para os alunos com elevada ou reduzida performance (BALLOU; SPRINGER, 2008; NICHOLS; BERLINER, 2007). Nichols e Berliner (2007) fazem referência a práticas pedagógicas centradas na preparação para exames, na redução de conteúdos programáticos e no incentivo ao abandono de estudantes com reduzidos resultados de aprendizagem. Por sua vez, na Inglaterra, a escolha da escola pelos pais conduziu a maior segregação social (ALLEN; VIGNOLES, 2007; GOLDSTEIN; NODEN, 2003, 2004; GORARD, 2004; NODEN; GOLDSTEIN, 2007). Tal parece suscitar a procura das melhores escolas por parte dos melhores alunos e a segregação dos alunos com mais baixo aproveitamento, reforçando a permanência num ciclo vicioso das escolas com maior proporção de alunos com défice educativo.

Apesar de se reconhecer que o desenvolvimento de MVA para avaliações com consequências de elevado impacto (high-stakes), no âmbito de políticas de responsabilização e prestação de contas, carece de investigação adicional de tal magnitude que impõe uma agenda de investigação científica (BRAUN et al., 2010, p. 25, 38), partilhamos do argumento de Robert Gordon, segundo o qual "although many important technical issues still need to be resolved, it is not realistic to think that policy makers will wait 20 years until all of the difficulties are worked out before making use of such methods" (BRAUN et al., 2010, p. 55).

Um estudo empírico sobre o Brasil

A seção apresenta o estudo longitudinal GERES-2005, cujos dados são usados para estimar o valor acrescentado escolar. Adicionalmente, apresenta-se uma breve caracterização descritiva dos dados modelados.

População e Amostra

Os dados recolhidos no âmbito do projecto GERES 2005, uma pesquisa longitudinal de eficácia escolar (BROOKE; BONAMINO, 2011) são usados neste trabalho. O projecto GERES 2005 teve o apoio institucional de PUC-Rio (LAED - Laboratório da Avaliação da Educação), UEMS, UFBA (Linha de Pesquisa de Avaliação da Educação do Programa de Pós-Graduação em Educação), UFJF (CAEd - Centro de Políticas Públicas e Avaliação da Educação), UFMG (GAME - Grupo de Avaliação e Medidas Educacionais), Unicamp (LOED - Laboratório de Observação e Estudos Descritivos). Contou com o apoio financeiro de Fundação Ford, INEP, FAPERJ e FAPEMIG.

A população alvo é constituída pelos alunos das turmas regulares, no segundo ano do período diurno das escolas urbanas públicas com, pelo menos, 20 alunos e das escolas particulares com, pelo menos, 10 alunos. Todas elas com, no máximo, três turmas e localizadas nos municípios de Belo Horizonte, Campinas, Campo Grande, Rio de Janeiro, Salvador; alunos em Escolas Federais ou Colégios de Aplicação de Universidades Públicas também foram considerados pelo GERES 2005. A população estudada resulta da população alvo com as seguintes exclusões: escolas com apenas turmas multisseriadas no segundo ano; escolas localizadas na zona rural e também as escolas com alunos de segundo ano somente no período noturno (BROOKE; BONAMINO, 2011). "A primeira onda de medida foi tomada em março de 2005, com a observação de alunos da 1^a série (2º ano) do Ensino Fundamental (ou seu equivalente, quando a organização do ensino era em ciclos), e a segunda onda de medida ocorreu em outubro/novembro de 2005. O painel foi observado também em novembro de 2006, 2007 e 2008, viabilizando o acompanhamento da amostra de alunos ao longo de quatro anos letivos". (BROOKE; BONAMINO, 2011).

Doravante passaremos a designar os momentos de recolha de dados por onda 1 (março/2005), onda 2 (novembro/2005), onda 3 (novembro/2006), onda 4 (novembro/2007), onda 5 (novembro/2008). Todos os alunos frequentavam o 2ºano nas ondas 1 e 2, e a maioria dos alunos frequentavam o ano teórico, dada a idade, nas restantes ondas. Isto é, a onda 3 corresponde ao 3ºano teórico, a onda 4, ao 4ºano teórico e a onda 5 ao 5º ano teórico.

O cadastro para amostragem foi o Censo Escolar de 2003, e a amostra selecionada refere-se aos alunos do 2º ano.

Tabela 1 - Critérios de Estratificação e Número de Escolas por Estrato

Estrato	Cidade	Dependência Administrativa	Tamanho do Estrato	Escolas Selecionadas	
				Nº	%
1	Rio de Janeiro	Municipal	765	30	3.9
2		Privada	805	30	3.7
3	Belo Horizonte	Estadual	155	20	12.9
4		Municipal	135	20	14.8
5		Privada	144	20	13.9
6	Campinas	Estadual	95	20	21.1
7		Municipal	39	20	51.3
8		Privada	47	20	42.6
9	Campo Grande	Estadual	70	20	28.6
10		Municipal	76	19	25.0
11		Privada	80	20	25.0
12	Salvador	Estadual*	67	21	31.3
13		Municipal**	332	20	6.0
14		Particular	278	20	7.2
15	Escolas Federais ou Colégios de Aplicação de Universidades Públicas		9	9	100.0
Total			3.097	310	

Fonte: Brooke e Bonamino (2011).

* A amostra de escolas estaduais foi dividida entre escolas sem previsão de municipalização e escolas a serem municipalizadas em 2005/2006.

** A amostra de escolas municipais foi dividida entre escolas originalmente municipais e escolas municipalizadas entre 2002/2004

Tabela 2 - Número de alunos com proficiência estimada ao longo do tempo por município²

Município	Amostral	Número de Alunos									
		Onda: 1		Onda:1,2		Onda:1,2,3		Onda:1,2,3,4		Onda:1,2,3,4,5	
		Mat	Por	Mat	Por	Mat	Por	Mat	Por	Mat	Por
Belo Horizonte	4387	4084	4076	3675	3678	2965	2972	2480	2507	2061	2075
Campinas	4881	4382	4382	3844	3899	2965	3021	2533	2588	1987	2036
Campo Grande	3605	3205	3196	2736	2729	1747	1746	1211	1217	924	925
Rio de Janeiro	3559	3954	3942	3468	3454	2739	2732	2269	2265	1815	1831
Salvador	4268	3339	3330	2656	2667	1916	1935	1385	1404	0	0
Total	20700	18964	18926	16379	16427	12332	12406	9878	9981	6787	6867

Fonte: Brooke e Bonamino (2011).

A tabela 2 inclui o número de alunos da onda 1 que se mantiveram na amostra ao longo do tempo. A ilustração gráfica é apresentada na Figura 1, através da qual se pode observar o decaimento do número de elementos da amostra, fenômeno próprio dos estudos longitudinais, tecnicamente designado por atrito ou mortalidade experimental quando se trata de um plano experimental. Apesar de serem tomadas medidas adequadas, quer na fase de planeamento dos estudos longitudinais, quer na fase de monitoramento do trabalho de campo, o atrito pode inviabilizar a análise de dados para determinados propósitos, no caso de ultrapassar níveis que comprometam a representatividade da amostra.

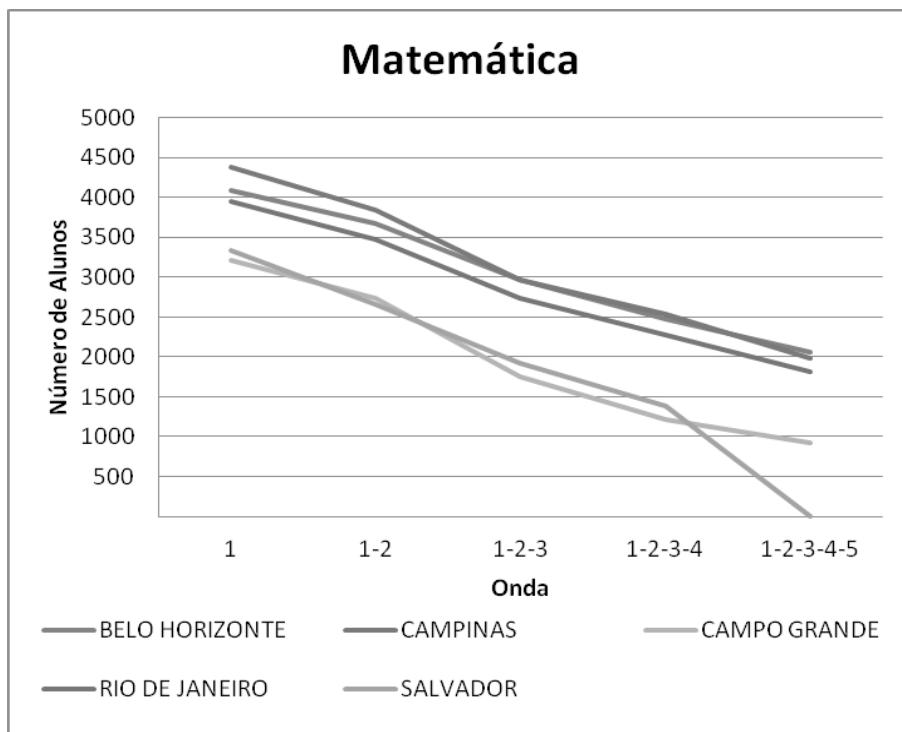
No estudo em análise, a amostra inicial foi reduzida³ em torno de 50% nos municípios de Belo Horizonte, Campinas e Rio de Janeiro, 74%, em Campo Grande. Em Salvador não se verificou a recolha de dados na última onda⁴. O atrito deve-se fundamentalmente à mobilidade de alunos entre escolas, apesar do esforço operacional que representou 'que alunos que não foram aprovados continuaram sendo observados desde que permanecessem na mesma escola ou se transferissem para outra escola da amostra' (BROOKE; BONAMINO, 2011; Tabela 5.36).

² O número de alunos amostral não inclui os alunos das escolas federais e colégios de aplicação (fonte: Brooke e Bonamino, 2011). Número de alunos nas ondas apurado de acordo com os registros da base de dados.

³ Considerando a aplicação dos testes de Matemática.

⁴ 'Os alunos de Salvador iniciaram a pesquisa na 2^a série; assim, a última observação ocorreu em 2007, quando os alunos estavam concluindo a 4^a série' (Brooke & Bonamino, 2011; p.51).

Figura 1 – Decaimento do número de alunos



Fonte: Brooke e Bonamino (2011).

Há evidência de que a perda de alunos ao longo do tempo não ocorreu aleatoriamente. Consideremos, por exemplo, a amostra de alunos⁵ na primeira onda (A1) e o subconjunto desses alunos que permaneceu em todas as ondas (A2). A tabela 3 contém algumas estatísticas referentes a A1 e A2 que permitem diagnosticar a existência de viés de seleção. Verifica-se que a média do nível socioeconómico de A1 é estatisticamente menor do que a média de A2, e que a percentagem de alunos submetidos ao teste versão 'fácil'⁶ é estatisticamente maior em A1 do que em A2. Em síntese, o atrito ocorreu com maior incidência entre o grupo de alunos de menor nível socioeconómico e de menor desempenho esperado. Parece-nos plausível que também se verifiquem diferenças por idade, devido à retenção, e por raça-cor. Todavia, não foi possível testar tais diferenças no âmbito do exercício ora apresentado, porque a recolha dos dados necessários a tal comparação não se verificou antes da onda 4. A caracterização detalhada do viés de seleção e do estudo das suas implicações na análise de dados estão fora do âmbito deste trabalho.

⁵ Para a comparação excluímos os alunos do município de Salvador.

⁶ De acordo com o nível de desenvolvimento dos alunos, foram aplicadas duas versões com nível de dificuldade diferente.

Tabela 3 - Diagnóstico do viés de selecção

Estatística		A1	A2
Nível Socioeconómico	Média amostral	-0.02	0.14
	Desvio padrão	0.556	0.630
	n	27340	6779
	IC95%	[-0.03;-0.01]	[0.13;0.15]
Aplicação de teste 'fácil' (Matemática)	Percentagem	68%	64.1%

Fonte: Autores (2012).

Nas diversas ondas entraram novos alunos para o estudo, de tal modo que o número de alunos envolvidos por onda situou-se entre 4084 e 4457 em Belo Horizonte, entre 4382 e 5020 em Campinas, entre 3205 e 3368 em Campo Grande, entre 3878 e 5261 no Rio de Janeiro e entre 3325 e 3946 em Salvador.

Dados e variáveis

Para o propósito deste artigo foram usados os dados referentes ao município de Campinas. Os critérios para esta opção foram dois. O primeiro por ser o município que apresenta maior percentagem de escolas por estrato; o segundo, por ser o município que apresenta menor taxa de atrito. A selecção das variáveis incluídas na modelagem está justificada na próxima secção. As variáveis são as seguintes:

- Resultados escolares, score em Matemática e Leitura no final do período de escolarização em análise;
- Desempenho prévio, score em Matemática e Leitura no início do período de escolarização em análise;
- Nível socioeconómico do aluno, índice GERES;
- Raça-cor do aluno⁷, auto-declaração (categoria de referência: branco);
- Sexo do aluno⁸ (categoria de referência: feminino);
- Frequência da pré-escola ou jardim de infância⁹ (sim vs. não);
- Composição do nível socioeconómico da escola, obtido a partir da média do nível socioeconómico dos respectivos alunos;
- Rede da escola, Municipal/Estadual/Particular (categoria de referência: estadual).

⁷ Variável disponível a partir da onda 4.

⁸ Variável disponível a partir da onda 4.

⁹ Variável disponível na onda 5.

As estatísticas¹⁰ que se apresentam em seguida foram obtidas da subamostra de 2036 alunos que participaram¹¹ em todas as ondas, dos quais 50.6% são do sexo feminino, 90.4% declaram ter frequentado pré-escolar, maternal ou jardim de infância antes de entrar na primeira série. A distribuição por raça-cor é a seguinte: 39.4% declara-se branco; 41.4%, pardo; 10.9%, preto; 4.2%, amarelo; 4%, indígena¹².

A metodologia usada e opções feitas para a produção das escalas de proficiência e dos scores em Português e em Matemática está descrita em Brooke e Bonamino (2011, cap. 6). Estatísticas descritivas dos scores são apresentadas nas tabelas 4 e 5. De modo geral, essas estatísticas sugerem diferenças estatisticamente significativas na comparação da média do desempenho entre redes. Particularmente em matemática, de notar que a variabilidade dos scores tem tendência a aumentar ao longo das ondas quando analisada através do desvio padrão. Por exemplo, o desvio padrão na onda 5 face à onda 1 mais do que duplica; mesmo considerando a influência da média dos respectivos anos, através do coeficiente de variação, subsistem algumas questões. Trata-se do aumento real da média e da heterogeneidade dos alunos relativamente às aprendizagens realizadas? Trata-se do efeito das opções metodológicas na produção das escalas e dos scores? Ou de ambas as razões? Está fora do domínio deste trabalho dar resposta a essas perguntas, apesar de elas poderem afetar alguns dos resultados adiante reportados. Doravante assume-se que os scores têm a fiabilidade e precisão desejáveis a estudos desta natureza.

Tabela 4 - Estatísticas Descritivas, Português

Rede da Turma na Onda 1		Onda 1	Onda 2	Onda 3	Onda 4	Onda 5
ESTADUAL	Média	92.09	116.61	134.63	152.18	165.04
	N	787	787	787	787	787
	Desvio Padrão	21.69	20.45	25.49	26.77	25.46
	Coeficiente de Variação	0.24	0.18	0.19	0.18	0.15
	Erro Padrão	0.77	0.73	0.91	0.95	0.91

¹⁰ Casos válidos na onda 5.

¹¹ Com scores em Matemática.

¹² Na onda 4 a distribuição é a seguinte: 48% declara-se branco, 28.6% pardo, 15.1% negro, 3.6% amarelo, 4.7% indígena.

MUNICIPAL	Média	89.21	114.47	128.43	146.92	159.52
	N	822	822	822	822	822
	Desvio Padrão	17.94	17.10	22.56	25.75	24.35
	Coeficiente de Variação	0.20	0.15	0.18	0.18	0.15
	Erro Padrão	0.63	0.60	0.79	0.90	0.85
PRIVADA	Média	126.74	145.74	163.67	178.10	188.38
	N	427	427	427	427	427
	Desvio Padrão	17.93	16.28	20.36	19.25	15.42
	Coeficiente de Variação	0.14	0.11	0.12	0.11	0.08
	Erro Padrão	0.87	0.79	0.99	0.93	0.75
Total	Média	98.19	121.86	138.22	155.49	167.70
	N	2036	2036	2036	2036	2036
	Desvio Padrão	24.43	22.08	26.88	27.63	25.67
	Coeficiente de Variação	0.25	0.18	0.19	0.18	0.15
	Erro Padrão	0.54	0.49	0.60	0.61	0.57

Fonte: Autores (2012).

Tabela 5 - Estatísticas Descritivas, Matemática

Rede da Turma na Onda 1		Onda 1	Onda 2	Onda 3	Onda 4	Onda 5
ESTADUAL	Média	97.77	118.80	139.90	187.95	237.57
	n	787	787	787	787	787
	Desvio Padrão	27.16	48.45	59.02	64.84	68.00
	Coeficiente de Variação	0.28	0.41	0.42	0.35	0.29
	Erro Padrão	0.97	1.73	2.10	2.31	2.42
MUNICIPAL	Média	95.39	117.70	132.72	172.44	218.39
	n	822	822	822	822	822
	Desvio Padrão	24.35	41.62	51.14	58.69	66.43
	Coeficiente de Variação	0.26	0.35	0.39	0.34	0.30
	Erro Padrão	0.85	1.45	1.78	2.05	2.32

PRIVADA	Média	132.40	163.77	206.28	253.65	300.61
	n	427	427	427	427	427
	Desvio Padrão	25.36	29.89	43.25	47.57	57.25
	Coeficiente de Variação	0.19	0.18	0.21	0.19	0.19
	Erro Padrão	1.23	1.45	2.09	2.30	2.77
Total	Média	104.07	127.79	150.92	195.46	243.05
	n	2036	2036	2036	2036	2036
	Desvio Padrão	29.55	46.22	60.13	66.59	72.14
	Coeficiente de Variação	0.28	0.36	0.40	0.34	0.30
	Erro Padrão	0.65	1.02	1.33	1.48	1.60

Fonte: Autores (2012).

Modelo de Valor Acresentado

Decorre do exposto, nas secções 2 e 3, que a expressão "o modelo de valor acresentado" denota uma escolha entre diversos modelos estatísticos possíveis. Se o objetivo é usar o indicador VA como critério em um sistema de avaliação institucional, a escolha deve ser realizada em concordância com o critério de comparação estabelecido pelo *designer* do sistema de avaliação, de modo adequado ao contexto educativo, social, económico e cultural em que o modelo vai ser aplicado, articuladamente com as opções políticas e/ou programáticas (FERRÃO, 2012b). Os MVA e a abordagem seguida para o estudo da consistência e estabilidade dos indicadores VA, aplicados aos dados GERES de Campinas são os descritos em Ferrão e Goldstein (2009), Ferrão e Couto (n.d.). Nesses trabalhos, os autores debruçam-se sobre a temática no contexto português e obtêm estimativas de VA a partir dos dados do estudo longitudinal realizado no âmbito do projecto de investigação 3EM (FERRÃO, 2009b; FERRÃO, et al., 2009; FERRÃO et al., 2005; LOUREIRO et al., 2011; VICENTE, 2007).

A especificação do modelo, no que se refere à sua forma funcional, à escolha de variáveis e das unidades da estrutura hierárquica, bem como à dimensão temporal, atende a condições pré-estabelecidas e que se resumem à obtenção de estimativas do VA da escola:

1. anuais;
2. considerando diferentes medidas de resultados escolares (Português e Matemática);
3. considerando a heterogeneidade da população discente e da composição da escola;

4. passível de ser generalizado e aplicado, minimizando o efeito do viés de seleção devido ao atrito.

Trata-se de um modelo multinível de componentes de variância com estrutura hierárquica de dois níveis – alunos (i) agrupados em escolas (j) foi aplicado, separadamente, aos dados GERES de Matemática e Português do 2º ao 5º ano, isto é, envolvendo todas as ondas, $o=1,..,5$, com diferentes conjuntos de variáveis independentes. O modelo [1] é exemplo de um modelo de resultados contextualizados e inclui a variável independente, nível socioeconómico do aluno. O modelo [2] é um MVA que inclui as duas variáveis essenciais sobre o perfil do aluno à entrada, ou seja, *score* de conhecimento prévio e nível socioeconómico. O modelo [3] também é um MVA que inclui, adicionalmente, variáveis individuais, tais como autodeclaração de raça-cor, sexo, frequência de pré-escolar, e variáveis referentes à escola, tais como composição da escola e rede. A variável dependente representa o resultado das aprendizagens do aluno em Português e Matemática, *score* do conhecimento no final do ano letivo. A título de exemplo, para o 2º ano, o modelo [2] especifica-se através da seguinte equação:

$$y_{2(ij)} = \beta_0 + \beta_1 y_{1(ij)} + \beta_2 nse_a_{(ij)} + u_{0j} + e_{ij}$$

onde e_{ij} representa a componente aleatória de nível 1, e u_{0j} a componente aleatória do nível 2. Assumem-se os pressupostos de que as componentes aleatórias seguem distribuição normal com média zero e variância constante, e a covariância entre elas é nula,

$$\begin{aligned} e_{ij} &\sim N(0, \sigma_e^2) \\ u_{0j} &\sim N(0, \sigma_{u0}^2) \\ COV(e_{ij}, u_{0j}) &= 0. \end{aligned}$$

Assim, os parâmetros fixos a serem estimados são $\beta = (\beta_0, \beta_1, \beta_2)$ e $\sigma_e^2, \sigma_{u0}^2$, os parâmetros aleatórios. Segundo essa equação, o desempenho do aluno i na escola j no final do 2º ano é condicional ao conhecimento prévio do aluno no início do 2º ano, e do seu nível socioeconómico, ou seja, as variáveis que definem o perfil do aluno no início do ano letivo. A estimativa do VA das escolas é dada por (GOLDSTEIN; SPIEGELHALTER, 1996; GOLDSTEIN, 1997).

Resultados

A seção contém os principais resultados obtidos da análise e modelagem estatística de dados.

Estimativas dos parâmetros fixos e aleatórios

Os resultados obtidos da aplicação dos modelos constam das tabelas A1 e A2 no anexo A, respectivamente para Português e Matemática. Os modelos

foram ajustados, considerando o conjunto de variáveis mencionadas na secção anterior e foram excluídas do modelo aquelas cujos coeficientes se revelaram estatisticamente iguais a zero (nível de significância de 5%). Para a leitura simplificada, nas Tabelas 6 e 7 estão assinalados os parâmetros fixos estatisticamente diferentes de zero com o respectivo sinal indicando a associação positiva ou negativa da variável com o desempenho do aluno no final do período de escolarização em causa.

Tabela 6 - Resultados sintéticos para parâmetros fixos: Português

Português	2ºano			3ºano			4ºano			5ºano		
Modelo =>	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]
Nse_aluno	+	+	+	+	+	+	+	+	+	+	+	+
Score prévio		+	+		+	+		+	+		+	+
Masculino						-			-			-
Pardo												
Preto			-			-			-			-
Amarelo												
Indígena						-			+			-
Jardim-escola						+			+			
Nse_escola			+									+
Rede Municipal						-						
Rede Particular												

Fonte: Autores (2012).

Tabela 7 - Resultados sintéticos para parâmetros fixos: Matemática

Matemática	2ºano			3ºano			4ºano			5ºano		
Modelo =>	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]
Nse_aluno	+	+		+	+	+	+	+	+	+	+	+
Score prévio		+	+		+	+		+	+		+	+
Masculino												
Pardo						-						
Preto						-			-			
Amarelo												
Indígena												
Jardim-escola						+						

Nse_escola			+			+					+
Rede Municipal								-			
Rede Particular											

Fonte: Autores (2012).

Os parâmetros fixos associados a nível socioeconómico do aluno e desempenho prévio são estatisticamente maiores do que zero em todos os modelos, à excepção do modelo [3] para Matemática, 2º ano. Contudo, nesse caso, a composição social e económica da escola tem relação linear positiva com o desempenho individual dos seus alunos. Relativamente às demais variáveis de perfil do aluno, aparece um padrão, de tendência negativa, associado aos alunos do sexo masculino e aos alunos autodeclarados raça-cor pretos. O padrão é menos persistente em Matemática do que em Português. Relativamente a Português, as estimativas sugerem que os meninos atingem scores menores do que os alcançados pelas meninas, do 3º ano em diante. Algo semelhante acontece com os alunos autodeclarados de raça-cor pretos a partir da 2º ano.

Considerando os resultados obtidos para o MVA [3], as estimativas dos parâmetros fixos associados às variáveis de escola – rede e composição da escola, mostram um padrão sobre o qual vale a pena reflectir. Pode verificar-se que um dos dois parâmetros é estatisticamente diferente de zero (com excepção do modelo referente a Português no 4º ano). Isto é, ou o parâmetro associado a rede ou o parâmetro associado à composição da escola apresentam relação estatisticamente significativa com os resultados escolares, que é negativa no que se refere a rede Municipal e positiva no que se refere à composição da escola. O fenómeno subjacente a estes resultados pode ser o mesmo. Ou seja, as escolas da rede Municipal servirem população discente, cuja composição e efeito composicional nos resultados se diferencia da composição e do efeito composicional das escolas das outras redes. Esse aspeto e suas implicações precisam ser investigados em futuros trabalhos.

O desempenho escolar dos alunos no final do ano letivo varia muito de escola para escola. O teste de hipóteses para a significância do MVA (teste da *deviance*) sugere que a variância entre escolas é estatisticamente diferente de zero em todos em todos os modelos ajustados, à excepção do modelo de Português no 4º ano. A Tabela 8 apresenta os resultados do coeficiente de partição da variância calculado para todos os modelos e permite verificar, também, que as variáveis de perfil do aluno, em particular a variável de desempenho prévio, contribuem fortemente para explicar essa variabilidade. Observa-se que, quando calculado com base no modelo de resultados contextualizados, o coeficiente varia entre 0.14 e 0.42; quando calculado com base nos MVA o valor do coeficiente varia entre zero e 0.12. Nota-se que, em geral, o valor do coeficiente de partição da variância é maior nos anos iniciais e atenua-se ao longo do tempo. De notar,

no modelo [2], a "quebra" que se verifica no valor do coeficiente do 2º para o 3º ano de escolaridade. A interpretação e o significado educativo desse resultado deve ter em conta que a subamostra dos alunos que permanecem no estudo é, tendencialmente, constituída por alunos de sucesso educativo; portanto, a "realidade" que conduziu a esses resultados é a "realidade" desse subgrupo de alunos. Adicionalmente, também importa considerar que a recolha de dados na 1ª e na 2ª onda se verificou no mesmo ano letivo, respectivamente o score de conhecimento prévio e o score de conhecimento no final do 2º ano, enquanto tal não ocorreu nas demais ondas.

Tabela 8 - Coeficiente de participação da variância

Modelo →	Resultados Contextualizados [1]				Valor Acrescentado [2]				Valor Acrescentado [3]			
	2º	3º	4º	5º	2º	3º	4º	5º	2º	3º	4º	5º
Português	0.42	0.24	0.14	0.15	0.12	0.05	0.00	0.03	0.04	0.02	0.00	0.02
Matemática	0.32	0.21	0.17	0.18	0.11	0.05	0.04	0.06	0.04	0.04	0.03	0.05

Fonte: Autores (2012).

Consistência e Estabilidade

O conceito de consistência apresentada na seção 2 é usualmente operacionalizado através de estatísticas de associação entre as estimativas de VA obtidas a partir de diferentes especificações do modelo VA. Para o propósito deste artigo compararam-se as estimativas obtidas do modelo de resultados contextualizados [1] e dos MVA [2], [3] através da correlação entre as estimativas respectivamente obtidas. Adicionalmente, repete-se o exercício usando a posição de ordem ocupada pelas escolas na lista ordenada baseada no indicador VA.

A correlação entre as estimativas de VA dos modelos [1] e [2] ou [1] e [3] é de fraca a moderada em todos os anos de escolaridade, tanto em Português como Matemática, à exceção do 2º ano, em que a correlação entre [1] e [2] é 0.88 em Português e 0.90 em Matemática. Isto significa que, à exceção do 2º ano, verifica-se uma diferença considerável nos indicadores de *performance* das escolas caso seja usado o modelo de resultados contextualizados ou o MVA como critério de avaliação.

A correlação entre as estimativas de VA dos modelos [2] e [3] é forte ($r > 0.88$) em todos os anos de escolaridade, tanto em Português como em Matemática, à exceção do 2º ano em que a correlação é aproximadamente 0.70, ou seja, é possível encontrar escolas cujo indicador de *performance* varie dependendo do modelo VA usado mas, na maioria dos casos, os resultados do modelo [2] ou do modelo [3] conduzem às mesmas conclusões. A escolha de um ou outro modelo tem implicações diferenciadas apenas para

algumas escolas. É necessário trabalho adicional que permita discutir essas implicações e sobre a identificação/caracterização das escolas mais afetadas pela escolha do modelo.

Caso o modelo [2] seja adotado como o modelo VA adequado à realidade educacional e social brasileira, a correlação entre as estimativas de VA apuradas para cada escola em Português e Matemática é 0.67 no 2º ano; 0.66, no 3º ano; 0.43, no 4º ano e 0.70 no 5º ano. Quer isto dizer que, na generalidade, há uma tendência de concordância positiva apesar de se encontrarem escolas com elevado VA numa disciplina e que podem ter menor VA na outra.

O estudo da estabilidade dos indicadores VA pode ser influenciado por opções metodológicas da ligação vertical das escalas e pela heterocedasticidade das escalas ao longo do tempo, tal como aludimos anteriormente. Todavia, é possível identificar padrões no desempenho das escolas ao longo do tempo. Por exemplo, usando a mediana da distribuição VA como referência para a comparação, pode verificar-se esse padrão, através da frequência das escolas que apresentam sistematicamente VA maior ou menor VA do que a mediana (Tabelas 9 e 10). Considerando-se a disciplina de Matemática, entre as 56 escolas que compõem a amostra, 18 têm VA acima da mediana nos 2º e 3º anos, 12 escolas, nos 2º, 3º e 4º anos e 8 escolas apresentam esse desempenho em todos os anos. Adicionalmente, 16 escolas têm VA menor do que a mediana nos 2º e 3º anos, 8 escolas, nos 2º, 3º e 4º anos e 6 escolas apresentam esse desempenho em todos os anos de escolaridade. No que concerne à disciplina de Português, 16 têm VA acima da mediana nos 2º e 3º anos; 10 escolas, nos 2º, 3º e 4º anos e 5 escolas apresentam esse desempenho em todos os anos de escolaridade. Com desempenho menor do que a mediana, encontram-se 14 escolas nos 2º e 3º anos, 8 escolas nos 2º, 3º e 4º anos e 4 escolas apresentam esse desempenho em todos os anos. Os resultados apresentados nas tabelas também mostram que, de modo geral, são as escolas da rede privada que maioritariamente se mantêm com valores positivos de VA, e as escolas da rede Municipal que maioritariamente se mantêm com valores negativos de VA. Este resultado está ligado ao efeito "rede" e "composição da escola" mencionado anteriormente.

Tabela 9 - Frequência das escolas com VA maior do que a mediana, por Rede

	Português			Matemática		
	Ano			Ano		
Rede	2º e 3º	2º, 3º e 4º	2º, 3º, 4º e 5º	2º e 3º	2º, 3º e 4º	2º, 3º, 4º e 5º
Estadual	4	2	0	5	2	1
Municipal	1	0	0	4	2	1
Privada	11	8	5	9	8	6
Total	16	10	5	18	12	8

Fonte: Autores (2012).

Tabela 10 - Frequência das escolas com VA menor do que a mediana, por Rede

	Português			Matemática		
	Ano			Ano		
Rede	2 ^a e 3 ^a	2 ^a , 3 ^a e 4 ^a	2 ^a , 3 ^a , 4 ^a e 5 ^a	2 ^a e 3 ^a	2 ^a , 3 ^a e 4 ^a	2 ^a , 3 ^a , 4 ^a e 5 ^a
Estadual	5	4	0	4	1	1
Municipal	8	5	4	10	7	5
Privada	1	0	0	2	0	0
Total	14	9	4	16	8	6

Fonte: Autores (2012).

O teste de hipótese do qui-quadrado aplicado para verificar a independência da classificação das escolas ao longo do tempo em Português e em Matemática, bem como a associação entre essas classificações quantificada através de coeficientes não paramétricos, tais como o coeficiente de contingência e o coeficiente de Cramer, corroboram que a classificação das escolas não é estatisticamente independente e que o grau da associação entre essas classificações é estatisticamente maior do que zero (nível de significância de 5%), variando entre 0.38 e 0.66, e considerando todas as combinações acima mencionadas.

Discussão

A evidência de consistência e estabilidade encontrada neste artigo face aos dois MVA testados sugere a sua utilidade para atempadamente diagnosticar as escolas que sistematicamente são classificadas como escolas eficazes ou como escolas não eficazes, de acordo com as definições apresentadas. Esse diagnóstico abre a possibilidade de investigar aprofundadamente as suas características e, consequentemente, promover mudanças que visem à melhoria da educação. No ensino primário, os resultados tendem a corroborar elevada consistência e moderada estabilidade nos indicadores VA, tal como reportado em estudos realizados em outros países (FERRÃO, 2012a; FERRÃO; GOLDSTEIN, 2009; KYRIAKIDES; CREEMERS, 2008). Por isso, dever-se-á tentar identificar, as escolas que mantêm a sua posição de ordem ao longo do tempo, bem como as que permanentemente apresentam evidência de melhoria, de tal modo que seja possível estudá-las em detalhe e fomentar a mudança, por exemplo, através da partilha de experiências e colaboração entre escolas.

No nosso entender, o trabalho colaborativo entre escolas e entre profissionais, essencial à melhoria da educação, é obstaculizado por políticas ou medidas que visem a estimular a competitividade através de mecanismos de incentivo tais como prémios de desempenho. Exemplos de outros países sugerem que as consequências podem ser de deterioração e segregação, pelo que se torna relevante a existência de um sistema de monitorização de resultados e de desempenho.

Mostra-se que indicadores produzidos pelo modelo de resultados contextualizados são muito diferentes dos de VA, evidência que remete para a adopção de estudos longitudinais.

No que se refere à escolha do modelo a adoptar, os resultados sugerem que com base na consistência e estabilidade o modelo mais parcimonioso poderia ser o escolhido. Contudo, deve levar-se em consideração aspectos tais como: a importância da informação dada pelo modelo referente ao desempenho relativo de grupos minoritários, tais como o desempenho do grupo constituído por alunos autodeclarados de raça-cor preto; o número de anos do estudo, o número de coortes e de escolas é limitado para generalizar qualquer conclusão sobre a escolha do modelo.

Considerações finais

Os resultados apresentados estão ainda limitados pela ocorrência de erros não amostrais que ultrapassam o âmbito deste artigo mas que é fundamental investigar em trabalhos futuros. Um é o do erro da medida, e o outro é o do erro de representatividade ou viés de seleção devido ao atrito. A evidência descriptiva apresentada, heterocedasticidade da proficiência nas diversas ondas, sugere que o erro da medida pode atingir nível acima do recomendável; por outro lado, verificou-se que o atrito ocorreu com maior incidência entre o grupo de alunos de menor nível socioeconómico e de menor desempenho esperado, não tendo sido possível testá-lo para outras variáveis sociodemográficas tais como idade e raça-cor. Uma implicação importante da evidência de viés de seleção é que o pressuposto, segundo o qual os dados são "missing at random", que é usualmente invocado para a estimação dos pesos de não resposta e imputação em estudos longitudinais é talvez demasiado forte. É necessário trabalho complementar, inclusivamente com a utilização da amostra GERES 2005 e não apenas com os casos completos, para testar a sensibilidade dos resultados obtidos.

Os autores agradecem as sugestões e comentários valiosos de Adilson Dalben, Glauco Aguiar, Nigel Brooke, Ruben Klein e de dois pareceristas anónimos. Todas as insuficiências apresentadas pelo texto são da plena responsabilidade dos autores. Os autores querem expressar ainda o seu reconhecimento a Alicia Bonamino e a Nigel Brooke pelo empenho e perseverança na coordenação geral do projeto GERES 2005, cuja base de dados criada constitui inquestionavelmente uma ferramenta valiosa para o estudo das políticas educativas, destacando a relevância da ação determinada de Nigel Brooke desde a fase embrionária do projeto que remonta a 2002.

Referências

AITKIN, M.; LONGFORD, N. *Statistical modeling issues in school effectiveness studies*. Journal of the Royal Statistical Society: Series A, London, v. 149, p. 1-43, 1986.

ALLEN, R.; VIGNOLES, A. *What should an index of school segregation measure?* Oxford Review of Education, Dorchester on Thames, v. 33, n. 5, p. 643–668, 2007. Disponível em: <<http://dx.doi.org/10.1080/03054980701366306>>. Acesso em: mar. 2012.

ANDRADE, E. *Alternativa de política educacional para o Brasil: school accountability*. Revista de Economia Política, São Paulo, v. 29, n. 4, p. 454–472, 2009.

ARIAS, R. M; SOTO, J. G. *Concepto y evolución de los modelos de valor añadido en educación*. Revista de Educación, Madrid, v. 348, p. 15–45, enero/abr, 2009.

BALLOU; D., SANDERS, W.; WRIGHT, P. *Controlling for Student Background in Value-Added Assessment of Teachers*. Journal of Educational and Behavioral Statistics, Washington, D.C v. 29, n. 1, p. 37–65, 2004.
DOI:10.3102/10769986029001037

BALLOU, D.; SPRINGER, M. *Achievement Trade-Offs and No Child Left Behind*. Nashville, Tennessee, Vanderbilt University, 2008. Disponível em : <<http://www.vanderbilt.edu/schoolchoice/documents/achievement>>. Acesso em mar. 2012.

BIRD, S. M. et al. *Performance indicators: good, bad, and ugly*. Journal of the Royal Statistical Society: Series A, London, v. 168, n. 1, p. 1–27, 2005.
DOI:10.1111/j.1467-985X.2004.00333.x.

BRAUN, H.; CHUDOWSKY; N.; KOENIG, J. *Getting value out of value-added. Social Sciences*. Washington, DC: National Academies Press, 2010. Disponível em: <<http://216.78.200.159/Documents/RandD/Other/Getting Value out of Value-Added.pdf>>. Acesso em: mar. 2012.

BRAUN, H.; WAINER, H. *Value-Added Modeling*. In: Rao, C. R.; S. Sinharay (Eds.), *Handbook of Statistics: Psychometrics*. Amsterdam: Elsevier, 2007, v. 26, p. 867–892 DOI: [http://dx.doi.org/10.1016/S0169-7161\(06\)26027-9](http://dx.doi.org/10.1016/S0169-7161(06)26027-9)

BRIGGS, D. C.; WEEKS, J. P. *The Persistence of School-Level Value-Added*. Journal of Educational and Behavioral Statistics. Washington, D.C, v. 36, n. 5, p. 616–637, 2011. DOI:10.3102/1076998610396887

BROOKE, N.; BONAMINO, A. (Eds.). *GERES 2005: razões e resultados de uma pesquisa longitudinal sobre eficácia escolar*. Rio de Janeiro: Walprint, 2011. Disponível em: <http://dc314.4shared.com/doc/DbyLSnul/preview.html>. Acesso em: mar. 2012.

BROOKE, Nigel; SOARES, J. F. *Pesquisa em eficácia escolar: origem e trajetórias*. Belo Horizonte: UFMG, 2008

CLOTFELTER, C. T. et al. *Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers?* Journal of Policy Analysis and Management, New York, v. 23, n. 2, p. 251–271, 2004. DOI:10.1002/pam.20003

CRONBACH, L. *Course Improvements through evaluation*. Teachers College Record, New York, v. 64, n. 8, p. 672, 1963.

FERRÃO, M. E. *Avaliação educacional e modelos de valor acrescentado: tópicos de reflexão*. Educação & Sociedade, Campinas (SP), v. 33, n. 119, p. 455-469, 2012b.

_____. *Modelo de valor acrescentado: algumas evidências a partir do projecto 3EM*, com discussão. Covilhã: Universidade da Beira Interior, 2009b.

_____. *On the stability of value added indicators*. Quality & Quantity, The Netherlands, v.46, n. 2, p. 627–637, 2012. DOI:10.1007/s11135-010-9417-6

_____. *Sensibilidad de las especificaciones del modelo de valor añadido: midiendo el estatus socioeconómico*. Revista de Educación, Madrid, n. 348, p.137–152, 2009a.

FERRÃO, M. E.; COUTO, A. *The use of a school value added model for educational improvement: a case study from the Portuguese primary education system*. School Effectiveness and School Improvement, accepted. 2013.

FERRÃO, M. E.; GOLDSTEIN, H. *Adjusting for measurement error in the value added model: evidence from Portugal, The Netherlands*. Quality & Quantity, v. 43, n. 6, p. 951–963, 2009. DOI:10.1007/s11135-008-9171-1

FERRÃO, M. E. et al. *Aferição das aprendizagens em matemática no Ensino Básico: a proposta 3EMat* [Learning assessment in maths: the proposal 3EMat]. Covilhã: Universidade da Beira Interior, 2009.

_____. *À Procura da escola eficaz: referencial teórico do projecto de investigação eficácia escolar no ensino da matemática*. Covilhã: Universidade da Beira Interior, 2005.

FIELDING, A., YANG, M.; GOLDSTEIN, H. *Multilevel ordinal models for examination grades*. Statistical Modelling, London, v. 3, n. 2, p.127-153, 2003. DOI:10.1191/1471082X03st052oa

- GOLDSTEIN, H. *Methods in school effectiveness research*. School Effectiveness and School Improvement, Bristol, v. 8, n. 4, p. 369-395, 1997. DOI:10.1080/0924345970080401
- _____. *Multilevel statistical models*, 4th ed. Chichester: Wiley, 2011.
- GOLDSTEIN, H.; HEALY, M. *The graphical presentation of collection of means*. *Journal of the Royal Statistical Society. Series A*, v. 158, n. 1, p. 175-177, 1995.
- GOLDSTEIN, H.; NODEN, P. *Modeling social segregation*. Oxford Review of Education. Dorchester and Thames, v. 29, n. 2, p. 225-237, 2003. DOI:10.1080/0305498032000080693
- _____. : _____. *A response to Gorard on social segregation*. Oxford Review of Education, Dorchester and Thames, v. 30, n. 3, p. 441- 442, 2004. DOI:10.1080/0305498042000260539
- GOLDSTEIN, H.; SAMMONS, P. *The influence of secondary and junior schools on sixteen year examination performance: a cross-classified multilevel analysis*. School Effectiveness and School Improvement, Bristol, v. 8, n.2, p. 219-230, 1997.
- GOLDSTEIN, H.; SPIEGELHALTER, D. *League tables and their limitations: statistical issues in comparisons of institutional performance*. Journal of the Royal Statistical Society: Series A, London, n. 159, v. 3, p. 385-443. 1996. Disponível em: <<http://www.jstor.org/pss/2983325>>. Acesso em: mar. 2012.
- GORARD, S. *Comments on "Modeling social segregation" by Goldstein and Noden*. Oxford Review of Education, Dorchester and Thames, v. 30, n. 3, p. 435-440, 2004. DOI:10.1080/0305498042000260520.
- GRAY, J.; GOLDSTEIN, H.; JESSON, D. *Changes and improvements in schools' effectiveness: trends over five years*. Research Papers in Education, Exeter, n. 11, p. 35-51, 1996.
- GRAY, J.; GOLDSTEIN; H.; THOMAS, S. *Predicting the future: the role of past performance in determining trends in institutional effectiveness at A level*. British Educational Research Journal, London, n. 27, p. 391-405, 2001.
- GRAY, J., et al. *A multi-level analysis of school improvement: changes in schools' performance over time*. School Effectiveness and School Improvement, Bristol, v. 6, n. 2, p. 97-114, 1995.

HANUSHEK, E.; RAYMOND, M. *Does school accountability lead to improved student performance?* Journal of Policy Analysis and Management, New York, v. 24, n. 2, p. 297-327, 2005.

HANUSHEK, E. A. *The economic value of higher teacher quality.* Economics of Education Review, Elsevier, Amsterdam, v. 30, n. 3, p. 466-479, june 2011.

HILL, P.; ROWE, K. *Multilevel modeling of school effectiveness research.* In: International Congress for School Effectiveness and Improvement, 7., Melbourne, 1994.

HOFMAN, R.; DIJKSTRA, N.; HOFMAN, W. *School self-evaluation and student achievement.* School Effectiveness and School Improvement, Bristol, v.20, n.1, p. 47-68, 2009.

HULPIA, H.; VALCKE, M. *The use of performance indicators in a school improvement policy: the theoretical and empirical context.* Evaluation & Research in Education, [s.l.], v. 18, p. 102-119, 2004.

KYRIAKIDES, L.; CREAMERS, B. P. M. *A longitudinal study on the stability over time of school and teacher effects on student outcomes.* Oxford Review of Education, Dorchester on Thames, v. 34, p. 521-545, 2008.

LADD, H. F.; WALSH, R. P. *Implementing value-added measures of school effectiveness: getting the incentives right.* Economics of Education Review, Elsevier, Amsterdam, v. 21, n. 1, p. 1-17, 2002. DOI:10.1016/S0272-7757(00)00039-X.

LECKIE, G.; GOLDSTEIN, H. *The limitations of using school league tables to inform school choice.* Journal of the Royal Statistical Society: Series A, London, v. 172, n. 4, p. 835-851, 2009. DOI:10.1111/j.1467-985X.2009.00597.x.

LOCKWOOD, J. R. et al. *The sensitivity of value-added teacher effect estimates to different mathematics achievement measures.* Journal of Educational Measurement, Washington, D.C., v. 44, n.1, p. 47-67, 2007. DOI:10.1111/j.1745-3984.2007.00026.x.

LOUREIRO, M. J. et al. *Avaliação do autoconceito Infantil.* Revista Psychologica, Coimbra, v. 52, n.1, p. 469-478, 2001.

MARTINEAU, J. A. *Distorting value added: The Use of Longitudinal, vertically scaled student achievement data for growth-based, value-added accountability.* Journal of Educational and Behavioral Statistics, Washington (D.C.), v. 31, n.1, p. 35-62, 2006. DOI:10.3102/10769986031001035.

MCCAFFREY et al. *Evaluating value-added models for teacher accountability*. Santa Monica, Califórnia, RAND Education, 2004. Disponível em: <http://www.rand.org/pubs/monographs/2004/RAND_MG158.pdf>. Acesso em: mar. 2012.

MORTIMORE, P. et al. *The road to Improvement: reflections on school effectiveness*. Lisse: Swets & Zeitlinger Publishers, 1998.

_____. *School matters*. Somerset: Open Books, 1998.

MORTIMORE, P., SAMMONS, P.; THOMAS, S. *School effectiveness and value added measures*. Assessment in education. principles, policy & practice, [S.I.], v. 1, n. 3, p. 315-332, 1994. DOI:10.1080/0969594940010307.

MURILLO, F. J. *Una panorámica de la investigación iberoamericana sobre eficacia escolar*. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, Bilboa, Espanha, v. 1, n. 1, 2003. Disponível em: <<http://www.ice.deusto.es/rinace/reice/vol1n1/Murillo.pdf>>. Acesso em: mar. 2012.

MOLLER, J. *School leadership in an age of accountability: tensions between managerial and professional accountability*. Journal of Educational Change, New York, v. 10, n. 2, p. 37-46. 2009.

NEWTON, X. A. et al. *Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts*. Education Policy Analysis Archives, Arizona, United States, v. 18, n. 23, p. 23. 2010. Disponível em: <<http://epaa.asu.edu/ojs/article/view/810>>. Acesso em: mar. 2012.

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

NODEN, P.; GOLDSTEIN, H. *A brief response to Gorard and Fitz*. British Educational Research Journal, London, v. 33, n. 2, p. 273-274. 2007. DOI:10.1080/01411920701292906.

OECD. *Measuring Improvements in learning outcomes*. Paris: OECD Publishing, 2008. DOI:10.1787/9789264050259-en.

RAUDENBUSH, S. W.; WILLMS, J. D. *The estimation of school effects*. Journal of Educational and Behavioral Statistics, Washington, D.C, v. 20, n. 4, p. 307-335, 1995. DOI:10.2307/1165304.

- REARDON, S. F.; RAUDENBUSH, S. W. *Assumptions of value-added models for estimating school effects*. Education Finance and Policy, Massachusetts v. 4, n. 4, p. 492-519. 2009. Disponível em: <<http://www.mitpressjournals.org/doi/abs/10.1162/edfp.2009.4.4.492>>. Acesso em: mar. 2012.
- ROBINSON, V.; LLOYD, C.; ROWE, K. *The impact of leadership on student outcomes: an analysis of the differential effects of leadership types*. Educational Administration Quarterly, [S.I.], v. 44, n. 5, p. 635-674, 2008.
- RUBIN, D. B.; STUART; E. A.; ZANUTTO, E. L. *A potential outcomes view of value-added assessment in education*. Journal of Educational and Behavioral Statistics, Washington, D.C., v. 29, n.1, p. 103-116, 2004. DOI:10.3102/10769986029001103.
- RUTTER, M., et al. *Fifteen Thousand Hours*. London: Open Books, 1979. p. 279.
- SAMMONS, P.; HILLMAN; J.; MORTIMORE, P. *Key Characteristics of Effective Schools: a review of school effectiveness research*. London, Institute of Education, 1995.
- SAMMONS, P; *School Effectiveness: Coning of age in the 21 st Century*.Lisse, Swets e Zeithing Publishers, 1999.
- SAMMONS, P.; THOMAS, S.; MORTIMORE, P. *Forging links: effective schools and effective departments*. London: Paul Chapman, 1997.
- SAUNDERS, L. *A brief history of educational "value added"*: How did we get to where we are? School Effectiveness and School Improvement, Bristol, v.10, n.2, p. 233-256, 1999. DOI:10.1076/seis.10.2.233.3507.
- SCHEERENS, J.; BOSKER, R. *The Foundations of Educational Effectiveness*. Oxford: Pergamon, p. 347 , 1997.
- SCRIVEN, M. *The methodology of evaluation*. lafayette. Ind.: Purdue University, 1196, p. 61.
- TEDDLIE, C.; REYNOLDS, D. *The international handbook of school effectiveness research*. London; Routledge; New York: Falmer, 2000, p. 411..
- TEDDLIE, C.; STRINGFIELD, S. *Schools make a difference: lessons learned from a 10-year study of school effects* . New York: Teacher; London: College Press, 1993, p. 272.

TEKWE, C. D. et al. *An empirical comparison of statistical models for value-added assessment of schools performance*. Journal of Educational and Behavioral Statistics, Washington, D.C., v. 29, n.1, p. 11-36, 2004.

THOMAS, S.; MORTIMORE, P. *Comparison of value-added models for secondary school effectiveness*. Research Papers in Education, London, v. 11, n. 1, p. 5-33, 1996. DOI:10.1080/0267152960110103.

THOMAS, S.; PENG, W.; GRAY, J. *Modeling patterns of improvement over time: value added trends in English secondary school performance across ten cohorts*. Oxford Review of Education, Dorchester on Thames, v. 33, n. 3, p. 261-295, 2007. DOI:10.1080/03054980701366116.

TOWNSEND, T. *International handbook of school effectiveness and improvement*. New York: Springer, 2007.

VICENTE, P. *Plano amostral do projecto 3EM – eficácia escolar no ensino da matemática* [Sampling Design of 3EM Survey - School Effectiveness in Maths]. In: FERRÃO, M. E., NUNES, C.; BRAUMANN, C. (Ed.). *Estatística: Ciência Interdisciplinar*. Lisboa: Sociedade Portuguesa de Estatística, 2007. p. 849-858.

WEST, M. *School-to-school cooperation as a strategy for improving student outcomes in challenging contexts*. School Effectiveness and School Improvement, Briston, v. 21, n.1, p. 93-112, 2010. DOI:10.1080/09243450903569767.

WHOLEY, J. S. *Evaluation - promise and performance*. Washington, D.C.: The Urban Institute. Disponível em: <<http://www.ncjrs.gov/App/publications/abstract.aspx?ID=61382>>. Acesso em: mar. 2012.

YANG, M.; WOODHOUSE, G. *Progress from GCSE to A and AS level: institutional and gender differences, and trends over time*. British Educational Research Journal, London, v. 27, n. 3, p. 245-267, 2001.

Recebido em: 19/12/2012

Aceito para publicação em: 18/04/2013

Anexo

Português	2 ^a ano			3 ^a ano			4 ^a ano			5 ^a ano		
	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]
Constante	124,52 (1,84)	71,80 (1,87)	73,03 (1,76)	139,82 (1,72)	30,562 (2,61)	35,79 (3,02)	156,01 (1,41)	42,83 (2,03)	43,52 (2,35)	168,01 (1,34)	49,48 (1,91)	54,25 (2,04)
Nse_aluno	5,54 (0,81)	3,22 (0,67)	1,55 (0,03)	9,76 (1,06)	4,30 (0,73)	3,22 (0,75)	12,27 (1,10)	3,24 (0,61)	2,95 (0,62)	10,88 (1,03)	2,42 (0,56)	1,56 (0,69)
Score prévio	0,52 (0,02)	0,49 (0,02)		0,88 (0,02)	0,85 (0,02)		0,81 (0,02)	0,80 (0,02)		0,76 (0,01)	0,74 (0,01)	
Masculino							-3,15 (0,73)		-2,03 (0,68)		-2,05 (0,57)	
Pardo												
Preto			-2,56 (0,99)			-5,77 (1,18)		-3,50 (1,11)			-2,13 (0,93)	
Amarelo												
Indígena						-4,69 (1,86)		4,01 (1,74)			-2,95 (1,46)	
Jardim-escola						3,02 (1,28)		3,26 (1,19)				
Nse_escola			10,08 (1,29)							2,51 (1,04)		
Rede Municipal						-3,72 (1,12)						
Rede Particular												
σ^2_{u0}	174,90 (35,09)	25,04 (6,01)	8,12 (2,67)	143,18 (30,41)	8,55 (3,14)	6,04 (2,62)	85,97 (19,97)	0,72 (1,19)	0,13 (1,03)	77,34 (17,93)	4,22 (1,72)	3,42 (1,54)
σ^2_e	243,81 (7,75)	181,66 (5,77)	180,18 (14,17)	445,58 (8,25)	159,83 (8,16)	252,80 (16,56)	520,62 (7,26)	229,01 (7,17)	452,56 (14,39)	162,57 (5,16)	155,88 (5,01)	

Matemática	2 ^a ano			3 ^a ano			4 ^a ano			5 ^a ano		
	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]
Constante	135.90 (2.39)	66.58 (2.45)	65.57 (2.23)	155.64 (3.38)	0.29 (4.30)	1.64 (5.08)	197.62 (3.52)	61.23 (2.85)	67.19 (3.25)	245.94 (3.90)	78.80 (3.24)	79.11 (3.12)
Nse_aluno	8.57 (1.27)	4.24 (1.02)		23.04 (2.23)	14.48 (1.62)	10.45 (1.94)	30.22 (2.56)	10.16 (1.58)	9.06 (1.60)	27.16 (2.75)	3.60 (1.64)	
Score prévio		0.64 (0.02)	0.63 (0.02)		1.14 (0.03)	1.10 (0.03)		0.88 (0.02)	0.86 (0.02)		0.85 (0.02)	0.84 (0.02)
Masculino												
Pardo						-3.60 (1.75)						
Preto							-16.59 (2.77)			-6.11 (2.57)		
Amarelo												
Indígena												
Jardim-escola							7.75 (1.83)					
Nse_escola				15.36 (1.62)			8.01 (3.26)					
Rede Municipal										-7.26 (2.51)		
Rede Particular												
σ^2_{uo}	287.58 (59.14)	53.28 (13.02)	18.52 (6.12)	537.65 (116.23)	64.77 (20.09)	49.38 (16.85)	559.64 (125.36)	45.75 (15.90)	33.32 (13.36)	686.80 (153.12)	77.65 (22.43)	59.60 (18.70)
σ^2_{e}	601.87 (19.36)	424.73 (13.67)	424.18 (13.63)	1980.76 (63.24)	1220.37 (39.37)	1196.90 (39.10)	2667.63 (85.92)	1197.72 (38.23)	1199.80 (38.59)	3088.40 (98.42)	1204.33 (38.42)	1204.23 (38.39)