

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO

GABRIELA MIRANDA MORICONI

**Medindo a eficácia dos professores: o uso de modelos de valor agregado para
estimar o efeito do professor sobre o desempenho dos alunos**

SÃO PAULO

2012

GABRIELA MIRANDA MORICONI

**Medindo a eficácia dos professores: o uso de modelos de valor agregado para
estimar o efeito do professor sobre o desempenho dos alunos**

Tese apresentada à Escola de
Administração de Empresas de São
Paulo da Fundação Getulio Vargas,
como requisito para obtenção do título
de Doutora em Administração Pública
e Governo.

Área de Concentração:

Política e Economia do Setor Público

Orientador: Prof. Dr. Nelson Marconi

SÃO PAULO

2012

Moriconi, Gabriela Miranda.

Medindo a eficácia dos professores: o uso de modelos de valor agregado para estimar o efeito do professor sobre o desempenho dos alunos / Gabriela Miranda Moriconi. - 2012.

114 f.

Orientador: Nelson Marconi

Tese (CDAPG) - Escola de Administração de Empresas de São Paulo.

1. Ensino fundamental - São Paulo (SP). 2. Escolas públicas - São Paulo (SP). 3. Professores - Avaliação. 4. Valor adicionado. I. Marconi, Nelson. II. Tese (CDAPG) - Escola de Administração de Empresas de São Paulo. III. Título.

CDU 371.12

GABRIELA MIRANDA MORICONI

**Medindo a eficácia dos professores: o uso de modelos de valor agregado para
estimar o efeito do professor sobre o desempenho dos alunos**

Tese apresentada à Escola de
Administração de Empresas de São
Paulo da Fundação Getúlio Vargas,
como requisito para obtenção do título
de Doutora em Administração Pública
e Governo.

Área de Concentração:
Política e Economia do Setor Público

Data de aprovação:

___/___/___

Banca examinadora:

Prof. Dr. Nelson Marconi (Orientador)
EAESP-FGV

Prof. Dr. André Portela de Souza
EESP-FGV

Prof. Dr. José Francisco Soares
FAE-UFMG

Prof^a. Dr^a. Paula Baptista Louzano
FE-USP

Prof. Dr. Reynaldo Fernandes
FEA/RP-USP

Ao Renato, meu esposo e companheiro
em todos os momentos.

AGRADECIMENTOS

A colaboração direta e indireta de uma série de pessoas e instituições foi essencial para o longo e importante processo de desenvolvimento desta tese. Agradeço sinceramente a cada um deles.

À FGV-EAESP, aos professores do curso de Doutorado em Administração Pública e Governo e, em especial, aos professores da linha de pesquisa em Política e Economia do Setor Público, pelo estímulo e apoio ao meu desenvolvimento como pesquisadora.

Ao GVPesquisa e à CAPES, pela concessão das bolsas que possibilitaram a realização deste curso.

À Fátima Alves e à Alicia Bonamino pela colaboração para obtenção dos dados da Pesquisa GERES, e ao David Pitombeira pela colaboração para obtenção dos dados do SPAECE. Não foi possível utilizar ambos os bancos de dados para os fins desta pesquisa – no entanto, foi grande o empenho dos três para disponibilizar os dados.

À Secretaria Municipal de Educação de São Paulo, em especial ao Secretário Alexandre Schneider e à Alejandra Velasco, pela autorização do uso dos dados da Prova São Paulo. À Paula Louzano, pela intermediação nesse contato e pelas várias conversas sobre a educação brasileira que muito contribuíram para a elaboração desta tese, e à Ana Carolina Zoghbi, pela total disponibilidade em ajudar no tratamento dos dados.

Ao Nelson Marconi, meu orientador, pela disponibilidade para orientar a tese e ajudar nas decisões profissionais, mesmo à distância, com toda a calma e paciência necessárias.

Ao Reynaldo Fernandes, com quem muito aprendi sobre indicadores e modelos de valor agregado, pelas discussões metodológicas iniciadas no INEP e continuadas durante a elaboração desta tese. A ele e à Elaine Pazello, agradeço pela oportunidade de trabalho, aprendizado e amizade no INEP.

A todos os colegas de trabalho e amigos feitos no e por intermédio do INEP, pelas inúmeras discussões e trabalhos compartilhados e pelo aprendizado a partir dos diferentes pontos de vista, mas da mesma dedicação e bom humor para enfrentar os desafios da educação do país.

Aos amigos do CEPESP, com quem convivi especialmente no período do mestrado e no início do doutorado, por compartilhar as dúvidas, as dificuldades e os achados desse começo de vida na pesquisa acadêmica.

A todos os amigos que de alguma forma participaram desse período da minha vida e o tornaram muito mais leve e divertido.

A toda a minha família, em especial aos meus pais, Ivson e Marisa, e à minha irmã, Carol, pelo carinho, apoio e incentivo.

Ao Renato, meu esposo, pelos ouvidos, ombros e palavras certos, nos momentos certos. E foram muitos.

RESUMO

Dadas a importância dos professores para o processo educacional e a disponibilidade de dados para a estimação de medidas de eficácia dos professores, entendida como a capacidade em produzir resultados em termos da aprendizagem dos alunos, este trabalho buscou aplicar um modelo de valor agregado para estimar e analisar a eficácia de uma amostra dos professores que lecionaram para a 4ª série da rede municipal de ensino de São Paulo no ano de 2010. Obtivemos evidências de que a variação na eficácia dos professores explicaria cerca de 9% da variação nas notas dos alunos da amostra analisada, menos do que a variação nas variáveis de *background* dos alunos, em torno de 15%, mas mais do que a variação nas variáveis de escola, em torno de 5%. Na estimação dos efeitos professor individuais, os resultados indicam que um aumento de um desvio-padrão nos efeitos professor levaria a um aumento de, no mínimo, 0,062 desvio-padrão na proficiência dos alunos em Língua Portuguesa e 0,049 desvio-padrão em Matemática, que correspondem a, 2,79 pontos em Língua Portuguesa e 2,20 pontos em Matemática. Na análise dos fatores associados aos efeitos professor estimados, foram encontradas associações positivas entre as seguintes variáveis e os efeitos professor: o tempo dedicado ao trabalho pedagógico fora da escola, a frequência com que o professor passa lição de casa e o uso dos Cadernos de Apoio e Aprendizagem. As análises acerca da confiabilidade e da estabilidade das medidas de efeito professor estimadas indicam que os dados utilizados têm uma capacidade limitada para subsidiar recomendações em relação às políticas de pessoal, porém permitem identificar que cerca de 13% dos professores tiveram efeitos estimados distintos da média. Dada essa distinção, esses professores são o público ideal para pesquisas futuras sobre suas práticas docentes, em especial sobre os fatores para os quais foi observada uma associação positiva com os efeitos professor estimados.

Palavras-chave: eficácia, efeito professor, modelos de valor agregado

ABSTRACT

Given the importance of teachers to the educational process and availability of data for estimating measures of teacher effectiveness, understood as the ability to produce results in terms of student learning, this study sought to apply a value-added model to estimate and analyse the effectiveness of a sample of teachers who taught fourth grade in São Paulo's municipal schools. We obtained evidence that variation in teacher effectiveness explains about 9% of the variation in students scores from the analysed sample, less than the variation in student background variables, around 15%, but more than variation in school variables, around 5%. For estimating teacher individual effects, results indicate that an increase of one standard-deviation in teacher effects would lead to a minimum increase of 0,062 standard deviation on student scores on Portuguese and 0,049 standard deviation on Mathematics, which represent 2,79 points in Portuguese and 2,20 points in Mathematics. In the analysis of factors associated to the teacher effects were found positive associations between the following variables and teacher effects: time devoted to teaching job out of the school, how often the teacher asks homework, and use of Learning and Support Workbooks. The analysis of reliability and stability of estimated teacher effect measurements indicate that this data have a limited capacity to support recommendations regarding personnel policies, but allow to identify that about 13% of the teachers had different effects from the estimated average. Given this distinction, these teachers are the ideal target to future research on their teaching practices, especially on the factors for which we observed a positive association with the estimated teacher effects.

Keywords: effectiveness, teacher effects, value-added models

SUMÁRIO

| | | |
|--------|--|----|
| 1. | INTRODUÇÃO..... | 11 |
| 2. | O USO DE RESULTADOS DOS ALUNOS EM TESTES PADRONIZADOS: DAS PESQUISAS EDUCACIONAIS À TOMADA DE DECISÃO EM POLÍTICAS DE GESTÃO DE PESSOAS..... | 15 |
| 2.1. | A Função de Produção Educacional..... | 15 |
| 2.2. | Em Busca de Características dos Professores que Melhorem o Desempenho dos Alunos..... | 19 |
| 2.3. | Estimação dos Efeitos do Conjunto de Professores..... | 24 |
| 2.4. | Estimação do Efeito Professor Individual Para a Tomada de Decisões em Gestão de Pessoas..... | 26 |
| 3. | ESTIMANDO MODELOS DE VALOR AGREGADO: DESAFIOS METODOLÓGICOS..... | 30 |
| 3.1. | Causalidade..... | 30 |
| 3.2. | Confiabilidade..... | 36 |
| 3.3. | Estabilidade..... | 38 |
| 4. | DADOS..... | 41 |
| 4.1. | A Prova São Paulo..... | 41 |
| 4.2. | Definição da Amostra..... | 46 |
| 4.3. | Descrição do Perfil de Alunos e Professores..... | 51 |
| 5. | ESTRATÉGIAS EMPÍRICAS..... | 57 |
| 5.1. | Do Modelo Cumulativo de Aprendizagem a Modelos de Valor Agregado..... | 57 |
| 5.2. | Especificação do Modelo para Estimar os Efeitos Professor..... | 61 |
| 5.3. | Análise de Fatores Associados aos Efeitos Professor..... | 65 |
| 6. | RESULTADOS..... | 69 |
| 6.1. | Efeitos do Conjunto de Professores..... | 69 |
| 6.2. | Efeito Professor Individual..... | 72 |
| 6.2.1. | Análise do Viés Devido à Alocação Não Aleatória dos Alunos..... | 76 |
| 6.2.2. | Análise de Confiabilidade..... | 83 |

| | |
|--|-----|
| 6.2.3. Análise de Estabilidade | 85 |
| 6.2.4. Decomposição dos Efeitos Professor Estimados..... | 88 |
| 6.3. Análise de Fatores Associados aos Efeitos Professor Estimados | 92 |
| 7. CONCLUSÃO | 98 |
| REFERÊNCIAS | 102 |
| APÊNDICES | 108 |

1. INTRODUÇÃO

Os professores têm um papel central no processo educacional: eles interagem diretamente com os alunos todos os dias, conduzindo ou intermediando os processos de ensino e aprendizagem. Seu poder discricionário em relação à execução das políticas educacionais é imenso, pois decidem como realizarão o gerenciamento da classe, quais métodos utilizarão para apresentar idéias e desenvolver habilidades e como se comunicarão com os alunos, entre tantas outras decisões críticas para os resultados em relação à qualidade da educação.

Não é de se estranhar, portanto, a atenção por parte de gestores públicos e pesquisadores em educação dada aos temas relativos ao magistério. Em especial, há um amplo debate sobre a eficácia dos professores, ou seja, sua capacidade em produzir resultados em termos da aprendizagem dos alunos, e os fatores que seriam capazes de influenciar essa eficácia. Encontrando os fatores que são capazes de explicar a eficácia dos professores, a idéia seria propor políticas com base nesses fatores, buscando melhorar a eficácia dos professores e, conseqüentemente, os resultados dos alunos.

As atenções da literatura têm sido voltadas essencialmente para fatores relativos aos professores como experiência, nível de escolaridade, área de formação e notas em testes de ingresso na carreira. Esse interesse pode ser explicado pela disponibilidade aos pesquisadores de variáveis que medem esses fatores, mas especialmente pela importância desses critérios para a tomada de decisões relativas às políticas do magistério. Diplomas em cursos específicos na área de formação de professores e aprovação em provas sobre conhecimentos pedagógicos e sobre os conteúdos que os professores irão lecionar são pré-requisitos comuns exigidos por sistemas de ensino no momento da contratação. O tempo de experiência e os títulos adicionais, por sua vez, são critérios recorrentes para crescimento na carreira e na remuneração.

Apesar de serem encontrados diversos estudos que indicariam relações positivas entre fatores como experiência e a certificação de professores e o desempenho dos alunos, os impactos desses fatores seriam muito pequenos, não contribuindo para explicar de forma satisfatória o desempenho dos alunos.

Dado esse cenário, uma série de pesquisadores passou a investigar os efeitos individuais dos professores olhando para as diferenças no crescimento do desempenho dos alunos entre professores. Eles partiram do princípio que a falha em encontrar efeitos de um conjunto de características mensuráveis disponíveis dos professores não quer dizer que a eficácia dos professores é igual, mas sim que se está medindo as características erradas ou, ainda, que as medidas existentes sejam ruins.

Os resultados obtidos por diversos estudos com base em dados de estados americanos indicaram que a variação na eficácia dos professores era responsável por uma considerável parcela da variação nos resultados dos alunos nos testes, e apresentaram grandes diferenças de resultados entre alunos com professores entre os mais eficazes e os menos eficazes. Isso indica que há, sim, diferenças de eficácia entre os professores.

A junção de evidências de que professores diferem consideravelmente em termos de eficácia com as dúvidas acerca do que faz com que um professor seja mais ou menos efetivo, levou a propostas do uso direto dessas medidas de eficácia em políticas para o magistério. Se a preocupação é com o desempenho dos alunos, as políticas relativas aos professores devem ser atreladas à sua capacidade de agregar valor ao desempenho dos alunos e não a fatores como tempo de experiência e títulos, já que não há evidências fortes de que estes fatores estejam relacionados à eficácia dos professores, afirmam os defensores dessas propostas.

Nos Estados Unidos, buscou-se colocar essas propostas em prática, em grande medida, com o programa *Race to the Top*, criado em 2009. O programa concede recursos financeiros aos estados que adotem as políticas de reforma educacional indicadas pelo governo federal, dentre elas “o uso das avaliações com base nos resultados dos alunos para subsidiar uma série de decisões em relação aos profissionais, como compensação, promoção, retenção, e também em relação a outras políticas, como o credenciamento de instituições de formação de professores” (*United States of America*, 2009).

A ênfase do governo federal americano em políticas referenciadas pelos resultados dos alunos nos testes padronizados e em medidas de eficácia dos professores com base nesses resultados impulsionou o desenvolvimento da literatura sobre os desafios e limites metodológicos da estimação de modelos de valor agregado para gerar essas medidas.

No Brasil, não há informação de nenhum estado ou município que tenha implementado políticas de alto impacto tendo como critério medidas de eficácia dos professores com base no desempenho dos alunos. Aqui, são raras as experiências de avaliação em larga escala que testam os mesmos alunos em anos consecutivos, o que permitiria a construção de medidas de valor agregado para cada professor.

Portanto, enquanto a motivação dos pesquisadores nos Estados Unidos está relacionada, especialmente, às preocupações em torno dos impactos diretos das políticas com base nessas medidas sobre a vida dos indivíduos, no caso deste estudo a motivação se deve ao potencial dessas metodologias em fornecer evidências que permitam aprofundar os conhecimentos existentes acerca da eficácia dos professores no Brasil.

Além das análises sobre a importância dos efeitos professor e seus determinantes, entende-se que as reflexões acerca das medidas de eficácia criadas podem oferecer indicações sobre o potencial delas para indicar professores mais ou menos eficazes cujas práticas mereçam uma observação mais aprofundada, de modo a compreender os fatores que possam explicar essa eficácia que não estejam sendo capturados por variáveis presentes nos bancos de dados disponíveis.

O objetivo deste trabalho é, portanto, criar e analisar medidas de eficácia individual para os professores de uma amostra da rede municipal de ensino de São Paulo, a partir da estimação de um modelo de valor agregado com base nos resultados dos seus respectivos alunos na Prova São Paulo.

Pretende-se realizar uma série de análises a partir dos efeitos professor estimados. Em primeiro lugar, a idéia é identificar a proporção da variação nas notas dos alunos que se deve à variação dos efeitos professor, de modo a analisar a importância dos professores como um dos fatores intra-escolares que afetam o desempenho dos alunos em testes padronizados.

Em segundo lugar, a partir da distribuição dos efeitos professor estimados, pretende-se simular o impacto que mudanças na eficácia dos professores teriam sobre os resultados dos alunos nos testes, de modo a quantificar a importância de políticas que alterem a eficácia dos professores.

Adicionalmente, os efeitos professor estimados permitirão estimar o viés devido ao caráter não aleatório de alocação de alunos a turmas e analisar a confiabilidade e a estabilidade das estimativas, que se mostram como características importantes para que se analise a utilidade dessas medidas.

Por fim, pretende-se realizar uma análise de fatores associados às estimativas de efeito professor, buscando evidências sobre características dos professores que estejam relacionadas aos efeitos professor estimados no primeiro estágio. Assim, espera-se construir um retrato relevante sobre a eficácia dos professores da rede municipal de ensino de São Paulo, contribuindo para o desenvolvimento e a avaliação de políticas voltadas para o magistério.

Esta pesquisa terá mais seis seções além desta introdução. Na segunda seção, é apresentada uma revisão da literatura a respeito do uso dos resultados dos testes padronizados dos alunos. Na terceira seção, passamos a uma discussão acerca dos desafios metodológicos da estimação de modelos de valor agregado. Na quarta seção, são apresentados a Prova São Paulo e seus dados, os quais servem de base para a realização das análises deste trabalho. Na quinta e na sexta seções são apresentadas as estratégias empíricas empregadas e os resultados das análises realizadas, respectivamente. Por fim, a sétima seção contém as conclusões a respeito da eficácia dos professores da amostra da rede municipal de ensino de São Paulo analisada.

2. O USO DE RESULTADOS DOS ALUNOS EM TESTES PADRONIZADOS: DAS PESQUISAS EDUCACIONAIS À TOMADA DE DECISÃO EM POLÍTICAS DE GESTÃO DE PESSOAS

É de conhecimento geral no âmbito da pesquisa educacional que um dos primeiros estudos com base em dados educacionais é o estudo intitulado *Equality of Educational Opportunity*, mais conhecido como Relatório Coleman, de 1966. Além de se tratar de uma descrição detalhada das escolas americanas, contendo informações sobre o desempenho e as características de mais de meio milhão de alunos e suas escolas, a notoriedade de Coleman et al (1966) se deve às evidências apresentadas no Relatório, que indicaram que as diferenças nas escolas tinham pouca influência sobre o desempenho dos alunos: o *background* familiar e as características dos pares pareciam ter importância muito maior.

Os achados de Coleman et al (1966) geraram uma série de críticas e debates sobre a importância de fatores escolares e de políticas educacionais para o desempenho educacional dos alunos. E também impulsionaram a literatura com base na função de produção educacional sobre os determinantes intra-escolares e extra-escolares do desempenho dos alunos, discutida a seguir.

2.1. A Função de Produção Educacional

Amplamente utilizadas na literatura econômica, as funções de produção descrevem a forma como os insumos são conjugados em processos de produção para gerar produtos. Elas são derivadas de relações de engenharia que refletem processos tecnológicos exogenamente dados. A firma decide sobre um conjunto de insumos e a função de produção é o melhor processo para combinar esses insumos – aquele que produzirá a maior quantidade de produtos. Ou, alternativamente, a firma decide sobre uma quantidade de produtos a ser gerada, e a partir da função de produção, obtém a menor quantidade de insumos que deverão ser empregados para produzi-la.

Enquanto a teoria de produção padrão se concentra na variação de quantidades de um produto homogêneo, isso não é traduzido facilmente para diversas áreas fora da indústria, como é o caso da educação. De acordo com Hanushek (1979), por meio da educação quantidades fixas de indivíduos (insumos) seriam transformados em indivíduos com atributos com qualidades diferentes (produtos).

Para medir a qualidade de atributos dos indivíduos, os melhores instrumentos que se tem conhecimento são os testes padronizados de larga escala. Esses testes geram medidas de habilidades cognitivas dos indivíduos, como a leitura e resolução de problemas matemáticos, as quais permitem com que esses atributos sejam quantificados e os alunos recebam um valor que traduz a qualidade de seu atributo em uma escala numérica. Por transformar a qualidade dos atributos dos indivíduos em quantidades que podem ser medidas antes e depois de um processo educacional, esses resultados têm grande potencial para serem usados como insumos e produtos nas funções de produção educacionais – como de fato se dá na grande maioria dos estudos da área.

Apesar desse potencial, há significativos questionamentos em torno do uso dos resultados dos alunos em testes cognitivos como produtos a serem maximizados no processo educacional.

A primeira questão diz respeito à importância das habilidades cognitivas em relação às capacidades futuras dos alunos ao desempenhar um papel na sociedade depois da escola. Como descreve Hanushek (1979), naquele momento já havia uma série de estudos relacionando a quantidade de escolarização sobre resultados da vida profissional e social dos indivíduos: salários, emprego, participação em eleições, socialização, criminalidade, entre outros. Mas não havia evidências sobre se esses resultados variariam em relação às habilidades cognitivas. Trinta anos depois, Hanushek e Rivkin (2010) já afirmavam que essa questão já estaria bem respondida pela literatura, com autores como Murnane, Willett e Levy (1995) e Hanushek e Woessmann (2008), por exemplo, demonstrando que as notas em testes padronizados para medir habilidades cognitivas seriam altamente relacionadas a fatores como os rendimentos e os resultados econômicos agregados.

Especialmente em razão da dificuldade em obter bancos de dados históricos que permitam a relacionar os resultados educacionais de indivíduos com

seu sucesso profissional e pessoal muito tempo depois, são raras as análises confiáveis sobre essa relação. Mesmo assim, com ou sem evidências empíricas sobre os efeitos posteriores das habilidades cognitivas, fato é que o desenvolvimento de uma série de habilidades cognitivas consta dos currículos de um modo geral como um dos objetivos a serem alcançados com a escolarização. Pressupõe-se daí que, seja qual for a razão – seja ela instrumental ou intrínseca – os sistemas educacionais valorizam o desenvolvimento dessas habilidades cognitivas, o que já valeria a pena sua medição de modo a verificar se um dos objetivos do sistema está sendo alcançado.

Mas outra questão – essa mais controversa – diz respeito ao aspecto restritivo no uso de habilidades cognitivas mensuradas como produto educacional. Na grande maioria dos casos de sistemas de avaliação nacionais e locais, se testa a habilidade de leitura e de resolução de problemas matemáticos. As demais habilidades cognitivas, como aquelas relacionadas às áreas de ciências naturais, ciências sociais e artes, por exemplo, por vezes não são objeto dos testes padronizados – como no caso brasileiro. Mais raros ainda são testes baseados em outros tipos de fatores que possam ser igualmente valorizados, tais como a criatividade, a capacidade de colaboração, o pensamento crítico. Nesses casos, um grande complicador é a dificuldade de obter medidas que ao mesmo tempo sejam eficazes para capturar o construto e objetivas para comparar os resultados dos indivíduos. De qualquer forma, ao restringir os produtos do processo educacional aos resultados em testes de leitura e matemática, há que se explicitar os pressupostos que estão sendo feitos. Pressupõe-se que há uma forte correlação entre os resultados dessas habilidades cognitivas e das demais habilidades, de modo que a medição apenas da leitura e da matemática bastaria? Ou que basta garantir a aprendizagem de leitura e matemática e as demais habilidades teriam importância secundária, não valendo o custo de serem aferidas?

De qualquer forma, vale ressaltar a importância de que algum destes pressupostos seja definido ao analisar os resultados de aplicações da função de produção educacional ou que se explicita que nenhum dos pressupostos foi adotado, e que somente se está avaliando os determinantes da habilidade de leitura ou de resolução de problemas matemáticos por parte de um grupo de alunos, sem que se realize nenhum tipo de extrapolação aos demais produtos dos sistemas educacionais.

No caso do presente trabalho, busca-se estimar a eficácia dos professores apenas no ensino relativo à leitura e à resolução de problemas matemáticos, dados os resultados dos alunos nesses atributos. Isto se deve ao fato de serem os resultados disponíveis para a análise, mas também pelo reconhecimento dessas habilidades como primordiais para o desenvolvimento de uma série de outros atributos por parte dos alunos.

Em relação às tecnologias existentes que determinariam as relações entre insumos e produtos na função de produção educacional, Hanushek (1979) afirma que esse seria o papel das teorias da aprendizagem. No entanto, ele afirma que quase todas as análises educacionais começam com lamentos sobre não ter nenhuma teoria de aprendizagem que sirva para guiar a análise de insumo-produto. Nesse sentido, ele aponta que boa parte das críticas ao uso de funções de produção educacionais se deve ao fato de que a especificação dos demais insumos, além da medida inicial do atributo do aluno, tem recebido pouca atenção. Desde o próprio Relatório Coleman até a publicação desse artigo, Hanushek (1979) descreve a falta de clareza conceitual das análises e crítica a escolha dos insumos que, em sua opinião, pareceria, algumas vezes explicitamente, ser guiada mais pelos dados disponíveis do que pelo que seria desejável conceitualmente.

Como podemos observar no decorrer dos anos, pouco se alterou nessa tendência da literatura baseada na função de produção educacional: o mais comum é não encontrar declarações de modelos conceituais nos quais estariam baseadas as análises, sendo o objetivo destes buscar relacionar potenciais insumos cujos dados estejam disponíveis nos bancos de dados com o desempenho dos alunos em testes padronizados. Uma possível explicação para essa situação seria que a elaboração dos questionários aplicados em conjunto com esses testes muitas vezes não é precedida do devido planejamento das pesquisas que serão realizadas com base nesses dados. Além disso, ainda há um considerável distanciamento entre os especialistas em teorias pedagógicas e os especialistas nos modelos estatísticos utilizados para analisar esses dados. Há um grande trabalho a ser realizado em relação a esses dois pontos para melhorar a qualidade dos estudos educacionais e possibilitar uma maior aplicação dos seus resultados como subsídios à realização e à avaliação de políticas educacionais.

2.2. Em Busca de Características dos Professores que Melhorem o Desempenho dos Alunos

De acordo com Hanushek e Rivkin (2004), desde 1966 mais de 400 estudos com base na função de produção educacional teriam sido publicados em revistas e livros.

O principal objetivo dessa literatura é identificar os insumos capazes de contribuir para melhores resultados educacionais, para com base nas evidências poder recomendar intervenções nos sistemas educacionais.

Nessa literatura, é grande o número de estudos que buscam analisar os efeitos de insumos relativos aos professores. Essa atenção especial aos professores não é de se estranhar, já que são eles que interagem diretamente com os alunos, conduzindo ou intermediando os processos de ensino e aprendizagem. Podem ser identificados com o que Lipsky (1980) denominou “burocracia do nível da rua” (*“street level bureaucreacy”*): agentes executores de políticas públicas cujo espaço para agir de forma discricionária é muito grande. Como observa Hanushek (1986), nessa discricionariedade encontram-se decisões como o gerenciamento de uma classe, a definição de métodos para apresentar idéias abstratas e o uso de habilidades de comunicação.

As atenções da literatura têm sido voltadas essencialmente para fatores como experiência, nível de escolaridade, área de formação, notas em testes e certificação dos professores. Assim como ocorre com outros insumos incluídos nas funções de produção educacional, esse interesse pode ser explicado parcialmente pela disponibilidade dessas variáveis, em especial da educação e da experiência, em bancos de dados administrativos das redes de ensino ou nos questionários aplicados junto aos testes dos alunos. Mas esses fatores recebem tamanha atenção especialmente porque uma série de decisões relativas às políticas de gestão de pessoas referentes ao magistério são tomadas com base nesses critérios. É comum que os sistemas de ensino exijam pré-requisitos para a contratação de professores que incluem um diploma em um curso de graduação ou pós-graduação específico na área de formação de professores e a aprovação em provas sobre conhecimentos pedagógicos e sobre os conteúdos que os professores irão lecionar. Assim como também é comum que os diplomas adquiridos pelos professores, bem como o tempo

de experiência deles na docência, façam parte dos critérios de crescimento na carreira e na remuneração.

Como existe uma gama muito ampla de estudos com essa temática, um bom ponto de partida é a revisão da literatura empírica com base em dados nacionais ou subnacionais dos Estados Unidos de Hanushek e Rivkin (2004), já que agrega os resultados de estudos empíricos já revisados por Hanushek (1997, 2003). Essa revisão é interessante também porque conta com estudos com base em modelos que podem ser considerados mais fracos, baseados na especificação contemporânea, sem a medida do desempenho prévio do aluno, até estimativas com base em modelos de valor agregado, considerados pelos autores como “estimativas de alta qualidade”. Nesse último caso, além de trabalhar com modelos de valor agregado, cada artigo é baseado em dados de apenas um estado americano, o que garantiria que as estimativas não estariam sendo influenciadas por políticas educacionais que variam entre redes de ensino distintas.

Dentre os fatores relativos aos professores, o nível de escolaridade dos professores é o que apresenta as evidências mais fracas na revisão da literatura empírica: em apenas 9% dos estudos considerados mais fracos metodologicamente o coeficiente relativo a esse fator estimado foi positivo e estatisticamente significativo. No conjunto das “estimativas de alta qualidade”, os coeficientes da escolaridade dos professores não se mostram positivos e significativos em nenhum dos estudos, sendo que em 9% deles os coeficientes chegam a ser negativos e significativos. Em artigos analisados que foram publicados após a revisão de Hanushek e Rivkin (2004), novamente o fator mais recorrente com efeitos negativos ou não significativos é o nível de escolaridade dos professores (Rivkin, Hanushek e Kain, 2005; Clotfelter, Ladd e Vigdor, 2007; Croninger et al, 2007).

Vale ressaltar que, no caso americano, quando os estudos lidam com o nível de escolaridade, geralmente estão comparando graduados com portadores de diploma de mestrado e doutorado. Portanto, no caso da escolaridade, é possível dizer que a literatura empírica para o caso americano tem produzido evidências consistentes de que o fato de professores terem mestrado ou doutorado não os faz produzirem melhores resultados em termos do desempenho dos seus alunos.

A experiência, por sua vez, apresenta uma das evidências mais fortes dentre os fatores analisados na literatura considerada mais fraca metodologicamente por Hanushek e Rivkin (2004), com 29% de estudos com coeficientes positivos e

significativos – perdendo apenas para as evidências apresentadas pelas notas em testes. Essa tendência se mantém nas “estimativas de alta qualidade”: os coeficientes da experiência se mostram positivos e significativos em 41% dos estudos, enquanto em 56% as estimativas se mostraram não significativas. Em se tratando dos artigos mais recentes, a experiência é o fator mais recorrente com efeitos positivos e significativos (Rockoff, 2004; Rivkin, Hanushek e Kain, 2005; Clotfelter, Ladd e Vigdor, 2006, 2007; Croninger et al, 2007; Goldhaber e Hansen, 2010). Em algumas estimativas de Rockoff (2004) e Rivkin, Hanushek e Kain (2005), o aumento do tempo de experiência dos professores produziu impactos somente nos primeiros anos da docência, não demonstrando impactos após o terceiro ano de docência.

As evidências também apresentam uma tendência forte da relação positiva entre a experiência dos professores e o crescimento do desempenho dos alunos com o uso de modelos de valor agregado. No caso da experiência, o principal questionamento sobre essas evidências relaciona-se ao fato de que professores mais experientes comumente têm a preferência na escolha das escolas e tenderiam a escolher aquelas em que o desempenho dos alunos é melhor (Hanushek, 1986; Hanushek, Kain e Rivkin, 2004). Isso gera a possibilidade de que aconteça o contrário: alunos com melhor desempenho atraem professores mais experientes, ou que a causalidade seja em ambas as direções. Questionamentos como esses só tenderão a diminuir se mais artigos como Rockoff (2004) encontrarem efeitos positivos da experiência comparando professores dentro das escolas, ou com trabalhos que consigam observar os mesmos professores em diferentes escolas ao longo do tempo.

As notas dos professores em testes realizados antes do ingresso na docência também apresentaram evidências fortes dentre os fatores analisados por Hanushek e Rivkin (2004), com 37% de estudos metodologicamente mais fracos com coeficientes positivos e significativos, e com 22% de “estimativas de alta qualidade” positivas e significativas, restando ainda 67% de estimativas não significativas. Porém, mesmo em relação a esses fatores, os autores ressaltam que esses percentuais não indicam grande consistência dos resultados. E apresentam outros problemas que enfraquecem ainda mais essas evidências. No caso das notas em testes, como se trata de diversos testes com conteúdos e metodologias

diferentes, apesar de fazê-lo, os autores não consideram que seria realmente adequado agregar essas evidências.

Outro fator que merece destaque pela importância que lhe é dada nas políticas de contratação de professores nos Estados Unidos é a certificação. Enquanto alguns estudos não encontram diferenças entre professores que possuem certificados tradicionais para lecionar e os que possuem certificados alternativos ou não possuem nenhum dos dois, como Ballou e Podgursky (2000), Hanushek, Rivkin e Kain (2005) e Croninger et al (2007), outros encontram diferenças positivas e significativas, como Darling-Hammond et al (2005), Clotfelter, Ladd e Vigdor (2007); Kane, Rockoff e Staiger (2006) e Rockoff et al (2008) – embora alguns possam considerar essas diferenças pequenas.

A revisão de Darling-Hammond (1999), que além de trabalhos com base na estimação da função de produção inclui também estudos qualitativos, destaca adicionalmente que professores com certificado nacional e com formação específica na área que lecionam estariam mais relacionados a maiores desempenhos dos alunos que professores com mestrado, por exemplo.

Rockoff et al (2008) é o trabalho encontrado com o maior esforço para expandir os fatores analisados, buscando testar também medidas que eles consideram não tradicionais, tais como notas em um teste de inteligência, pontuação em um instrumento que mede traços de personalidade, pontuação em instrumentos que medem crenças e valores que seriam indicativos do sucesso de professores em sala de aula, entre outros. Eles trabalham apenas com professores iniciantes na carreira, encontrando impactos positivos não significativos ou significativos mas pequenos das medidas tradicionais e não tradicionais para prever a eficácia dos professores.

Embora em menor quantidade, uma série de estudos também foram desenvolvidos para o caso brasileiro buscando analisar a relação entre fatores relativos aos professores, dentre outros fatores intra-escolares, e o desempenho dos alunos em exames padronizados no contexto nacional ou subnacional. Dentre eles, podemos destacar Barros et al (2001), Albernaz, Ferreira e Franco (2002), Soares, J. F. (2003), Soares, T. M. (2005), entre outros.

Esses estudos, com exceção de Soares, T. M. (2003), enfatizam o interesse nas variáveis relativas à formação do professor. Os dois primeiros encontram evidências de que o nível de escolaridade dos professores do ensino

fundamental aumentaria o desempenho médio de seus alunos. Soares, J. F. (2005), por sua vez, identifica evidências positivas do professor ter licenciatura em matemática sobre o desempenho dos seus alunos nessa disciplina. No caso brasileiro, geralmente os estudos comparando níveis de escolaridade incluem também a comparação com professores sem nível superior, o que era uma realidade bem presente em boa parte do país, em especial entre os professores dos anos iniciais do ensino fundamental, e ainda é, mesmo que em menores proporções, em algumas regiões do país.

Nas aplicações da função de produção educacional para o Brasil, o maior desafio é a escassez de estudos longitudinais no país. Muito recentemente começa-se a ter dados que permitam a estimação de modelos de valor agregado, contornando as dificuldades em relação à exigência de dados históricos dos alunos. Tanto é que nenhum dos estudos citados para o caso brasileiro pôde contar com essa configuração de dados, tendo que trabalhar com variações da especificação contemporânea.

Apesar de serem encontrados diversos estudos que indicariam relações positivas entre fatores como experiência e a certificação de professores e o desempenho dos alunos, os avanços metodológicos possibilitados por bancos de dados mais completos têm permitido confirmar uma afirmação que Hanushek (1986) já vinha fazendo há um certo tempo: que os impactos desses fatores, apesar de positivos, seriam muito pequenos, não contribuindo para explicar de forma satisfatória o desempenho dos alunos. A inclusão desses fatores em modelos para analisar os determinantes do desempenho dos alunos tem sido capaz de adicionar, no máximo, cerca de 3 pontos percentuais ao R^2 das estimações, demonstrando a pouca capacidade de contribuir para explicar os resultados dos alunos nos testes padronizados (Goldhaber e Brewer, 1997; Rivkin, Hanushek e Kain, 2005; Clotfelter, Ladd e Vigdor, 2006; Harris e Sass, 2006).

Nesse sentido, ao longo do tempo cresceram os questionamentos sobre a capacidade de contribuição das pesquisas com base em características observáveis dos professores. E cresceram também as dúvidas em relação à diferença de eficácia dos professores, ou seja, de sua capacidade de produzir melhores resultados em termos do desempenho dos alunos nos testes. Se as características observáveis não são capazes de explicar o desempenho dos alunos, será que a relação dessas

variáveis com a eficácia dos professores é fraca ou não há diferença de eficácia entre os professores?

2.3. Estimação dos Efeitos do Conjunto de Professores

Enquanto parte da literatura encontrava evidências fracas de fatores considerados importantes em relação aos docentes, diversos especialistas se concentraram nas medidas da eficácia dos professores baseadas somente nos resultados dos alunos.

Essa estratégia alternativa, chamada por alguns deles de *total teacher effects* – algo como efeitos do total de professores ou efeitos do conjunto de professores – baseia-se na investigação dos efeitos dos professores olhando para as diferenças no crescimento do desempenho dos alunos entre professores. Um bom professor seria o que obtém maiores crescimentos em termos das notas dos seus alunos, enquanto um professor ruim seria aquele que produziria menores crescimentos.

De um modo geral, os pesquisadores que realizam esse tipo de análise partem do princípio que a falha em encontrar efeitos de um conjunto de características mensuráveis disponíveis dos professores não quer dizer que a eficácia dos professores é igual. É possível que se esteja medindo as características erradas ou, ainda, que as medidas existentes sejam muito ruins.

Nye, Konstantopoulos e Hedges (2004) destacam que a vantagem do uso de análises de efeitos do conjunto dos professores é que essa abordagem não requer que o pesquisador identifique nem meça adequadamente as características do professor que estariam relacionadas ao desempenho dos alunos. Hanushek e Rivkin (2004) têm a mesma opinião, acrescentando ainda como vantagem o fato de não exigir conhecimento sobre como essas características interagem para produzir aprendizagem, já que a maioria dos estudos sobre características específicas assumem que elas entram linearmente e aditivamente na determinação do desempenho. Além disso, vêem a estimação dos efeitos do conjunto de professores como um *benchmark* da importância de variações na qualidade dos professores

contra qualquer outro tipo de política pública que possa ser pensada, como por exemplo, as políticas de redução do tamanho da classe.

Apesar de terem se tornado mais conhecidos na última década, o uso de modelos de valor agregado para estimar efeitos do conjunto de professores já vinha ocorrendo em algumas análises como as de Hanushek (1971), Armour (1976), Murnane e Philips (1981), Hanushek (1992) e Goldhaber e Brewer (1997).

Esses artigos, revisados por Nye, Konstantopoulos e Hedges (2004), têm em comum a estratégia de regredir características do *background* dos alunos e seu desempenho prévio no desempenho atual e depois fazer o mesmo incluindo uma *dummy* para cada professor da amostra. Os coeficientes das *dummies* indicariam os efeitos de cada professor, enquanto que a diferença no R² das duas estimações com a inclusão das variáveis dos professores representa o quanto da variação no desempenho dos alunos se deve aos efeitos do conjunto dos professores. As análises desses estudos, baseadas em diferentes amostras de alunos dos Estados Unidos, apresentaram resultados que indicam 7 a 21% da variação das notas dos alunos sendo associadas à variação na eficácia dos professores.

Considerando que os professores são um fator intra-escolar entre vários outros que se pressupõe terem influência sobre a aprendizagem dos alunos, esse padrão de resultados contribuiria com os que defendem que professores diferem sim ao agregar valor aos alunos, embora haja dificuldades em definir, com maior segurança, os fatores que explicam essas diferenças de eficácia.

O desenvolvimento de estudos para estimar o valor agregado por professores individuais e as diferenças de efetividade dos professores foi impulsionado a partir do *No Child Left Behind*, legislação americana de 2002 que, dentre outras definições, passou a exigir que todas as escolas públicas que recebem recursos federais devem testar todos os seus alunos da 3ª à 8ª séries anualmente e uma vez durante o ensino médio em leitura e matemática (Ravitch, 2010). Assim, multiplicaram-se os estados com dados longitudinais acompanhando os mesmos alunos em anos subseqüentes, permitindo a realização de estudos para estimar e analisar os efeitos dos professores.

Os significativos avanços metodológicos nas estratégias empíricas dos artigos devido à disponibilidade de dados longitudinais, como a adoção de efeitos fixos de escolas e até alunos, em alguns casos, contribuíram para dirimir o resto de

dúvida que pudesse existir em relação às diferenças significativas na eficácia dos professores, pelo menos para o caso americano.

Diferentemente dos estudos revisados por Nye, Konstantopoulos e Hedges (2004), esse conjunto de estudos mais recentes, em sua maioria, calcula a variação dos efeitos professor a partir dos próprios coeficientes das *dummies* dos professores individuais, e não pela diferença entre os R^2 dos modelos com e sem elas. A estratégia mais comum é calcular o desvio-padrão dos coeficientes dos efeitos professor estimados, corrigi-lo por conta dos erros amostrais e relacioná-lo às notas dos alunos, que são geralmente padronizadas com média zero e desvio-padrão 1 (Rockoff, 2004; Aaronson, Barrow e Sander, 2007; Goldhaber e Hansen, 2010; Rothstein, 2009, 2010; Koedel e Betts, 2011).

Os 11 estudos revisados por Hanushek e Rivkin (2010) publicados entre 2004 e 2010 com base em dados de estados ou cidades americanas com essa abordagem indicam variações de 0.08 a 0.36 desvio padrão estimado da eficácia dos professores, em termos do desvio padrão dos resultados dos alunos. Ou seja, são resultados que indicam que mover um desvio-padrão para cima na distribuição da eficácia dos professores, pode levar a aumentos de 0.08 a 0.36 desvio-padrão na distribuição das notas dos alunos, dependendo da análise que se tome por referência.

2.4. Estimação do Efeito Professor Individual Para a Tomada de Decisões em Gestão de Pessoas

Até a década de 1980, o único foco do desenvolvimento dos modelos de estimação dos efeitos de professores eram as pesquisas em busca dos determinantes do desempenho educacional dos alunos. Hanushek (1992) já afirmava que seus resultados indicavam que a diferença estimada no crescimento anual do desempenho com um professor bom e com um professor ruim podia ser maior do que o equivalente a uma série inteira. No entanto, estudos como esses somente tinham como consequência a recomendação de políticas a partir da compilação de resultados de diversas análises, não tratando de propor o uso dos resultados para avaliar professores individualmente, sugerindo-os inclusive como

subsídios para a tomada de decisões de pessoal, o que se tornou mais recorrente na década seguinte.

O primeiro sistema que se propôs a avaliar professores com base em um modelo de valor agregado em relação aos resultados dos alunos em testes padronizados é o Tennessee Value-Added Assessment System (TVAAS). Adotado pelo estado americano do Tennessee em 1992, o TVAAS é um modelo para determinar a eficácia de sistemas escolares, escolas e professores. Usa uma metodologia estatística de modelos mistos para permitir uma análise multivariada e longitudinal dos dados de desempenho dos alunos. Como descrevem Sanders e Horn (1998), o TVAAS utiliza o histórico de notas dos alunos ao longo do tempo para modelar seus padrões de aprendizagem. Aproveitando esse aspecto longitudinal dos dados, é possível notar quando há desvios do ritmo normal do seu crescimento. Os autores explicam que, ao acompanhar o crescimento ao longo do tempo, a criança serve como seu próprio controle, e é isso que permite particionar os efeitos do sistema escolar, escola e professor livres de fatores exógenos que influenciam o desempenho acadêmico e são consistentemente presentes em cada aluno ao longo do tempo.

O grande desafio de metodologias que modelam o crescimento acadêmico de cada aluno é a exigência massiva de dados. No caso do TVAAS, os bancos incluem notas dos alunos em testes administrados anualmente a todos os alunos da 3ª à 8ª séries em 5 áreas: matemática, ciências, estudos sociais, leitura e artes (Sanders e Horn, 1998).

Segundo Wright, Sanders e Horn (1997), as pesquisas feitas a partir dos dados do TVAAS mostram que a eficácia do professor é o maior determinante do progresso acadêmico dos alunos, superando inclusive os efeitos do nível socioeconômico dos alunos. Em um dos primeiros estudos com base nos dados do TVAAS, Sanders e Rivers (1996) encontraram evidências de que a distância entre as notas dos alunos alocados a professores do quintil mais alto de eficácia por 3 anos consecutivos e as notas dos alunos alocados a professores do quintil mais baixo por 3 anos consecutivos foi de 50 pontos percentuais.

Afirma Ravitch (2010) que, ao concluir que o fator mais importante afetando a aprendizagem dos alunos é o professor e que há grande variação na eficácia dos professores, a implicação imediata e clara de estudos como os

baseados no TVAAS é que mais pode ser feito para melhorar a educação melhorando a eficácia dos professores do que atuando sobre qualquer outro fator.

Como descrevem Gordon, Kane e Staiger (2006), o Ato *No Child Left Behind*, de 2002, adotou políticas tradicionais para o magistério, tentando melhorar a qualidade da força de trabalho por meio do aumento dos requisitos para aqueles que pretendem ingressar. O Ato exigiu que todos os professores das disciplinas centrais fossem altamente qualificados – com no mínimo bacharelado, uma licença estadual completa e certificação (geralmente exigindo que os professores se graduem em algum programa de formação de professores), e demonstrando competência específica da área (cursando uma disciplina ou passando em um teste).

Mas, como essas exigências não apresentam evidências consistentes de sua contribuição para a melhoria da aprendizagem dos alunos na visão de pesquisadores como Hanushek e Rivkin (2004) e Gordon, Kane e Staiger (2006), estes passaram a recomendar que se libere a oferta potencial de professores, permitindo com que profissionais interessados na docência, mas que não tenham todas essas credenciais, sejam admitidos como professores e que tenham a sua qualidade avaliada por meio dos resultados dos alunos. Ambos os artigos partem do mesmo princípio: se a preocupação é com o desempenho dos alunos, as políticas relativas aos professores devem ser atreladas ao desempenho dos alunos, entre outras medidas do desempenho do professor em sala de aula.

Eles também concordam que as políticas devem ser desenhadas de modo que professores altamente efetivos sejam incentivados a lecionar em escolas com altas proporções de alunos com baixa renda. Gordon, Kane e Staiger (2006) vão além, propondo que o governo federal garanta verbas para ajudar estados que associem o desempenho dos alunos com a eficácia dos professores individuais ao longo do tempo.

Nos Estados Unidos, buscou-se colocar essas propostas em prática, em grande medida, com o programa *Race to the Top*, criado em 2009 na gestão do Presidente Barack Obama, contando com a participação do próprio Robert Gordon, um dos autores do artigo citado. O programa concede um total de cerca de 4 bilhões de dólares aos estados que adotem as políticas de reforma educacional indicadas pelo governo federal.

O tema que recebe a maior pontuação entre os critérios de seleção dos estados para o programa *Race to the Top* é o relativo às políticas para professores e

diretores de escola. Dentre as ações em relação aos professores e diretores que os estados devem adotar para receber os recursos, podem ser destacadas: (1) a previsão de rotas alternativas para a certificação; (2) o desenho e a implementação de avaliação do desempenho dos profissionais que considerem, como um fator significativo, o crescimento das notas dos seus alunos; (3) a garantia de que alunos mais pobres tenham acesso igual aos dos demais alunos a professores de alta eficácia; (4) uso das avaliações com base nos resultados dos alunos para subsidiar uma série de decisões em relação aos profissionais, como compensação, promoção, retenção, e também em relação a outras políticas, como o credenciamento de instituições de formação de professores (*United States of America, 2009*).

Para se tornar elegível aos fundos, no momento em que o estado submete sua “candidatura” ao programa, não pode haver nenhuma barreira legal, estatutária ou regulatória no nível estadual que impeça relacionar os dados de desempenho dos alunos aos professores e diretores com o propósito de avaliá-los (*United States of America, 2009*).

De acordo com Banchemo e Kesmodel (2011) em reportagem publicada no *Wall Street Journal*, a partir do estímulo do programa *Race to the Top*, pelo menos 26 estados americanos concordaram em julgar os professores parcialmente com base nos resultados dos alunos em testes padronizados, e já haveria ao menos oito entidades no país desenvolvendo modelos de valor agregado como o TVAAS para apoiar as ações dos estados.

O aumento do número de estados com políticas de alto impacto com base em estimativas dos efeitos professor potencializou o desenvolvimento de uma série de estudos a respeito dos desafios metodológicos da aplicação de modelos de valor agregado para avaliar indivíduos e instituições. São análises acerca de aspectos como as relações de causalidade, a precisão das medidas e sua estabilidade ao longo do tempo. Essas análises não só tem contribuído para as discussões em torno dessas políticas de pessoal, mas também para o aprimoramento dos estudos sobre essa temática, como é o caso deste trabalho.

3. ESTIMANDO MODELOS DE VALOR AGREGADO: DESAFIOS METODOLÓGICOS

Na literatura de avaliação de políticas públicas, onde está inserida a avaliação de políticas educacionais, o conceito de eficácia¹ tem sido entendido como a produção de efeitos sobre uma realidade que se pretende modificar, também explicitado como a capacidade de gerar ou agregar valor a essa realidade.

As análises de eficácia são utilizadas para comparar os efeitos de dois ou mais insumos sobre os resultados alcançados. O melhor insumo seria aquele que teria conseguido maior êxito em produzir melhores resultados em termos dos objetivos propostos, ou seja, teria sido mais eficaz no alcance dos objetivos.

Como ressaltam Levin e McEwan (2002), análises de eficácia são fundamentalmente um esforço comparativo. Ou seja, permitem a escolha entre duas ou mais alternativas, mas não dão respostas no sentido de indicar se uma alternativa valeria a pena de forma absoluta.

3.1. Causalidade

Como bem destacam Rubin, Stuart e Zanutto (2004), o objetivo da literatura de valor agregado parece ser estimar os efeitos dos professores ou escolas, ou seja, determinar quanto um professor ou escola em particular “agregou de valor” às notas dos alunos. Estaria implícito que os efeitos sendo estimados são causais: a escola ou o professor causaria os efeitos estimados sobre o crescimento do desempenho dos alunos que foram submetidos a esse fator. E essa é uma premissa fundamental se o objetivo das estimações é gerar uma medida de qualidade do professor ou da escola que permita avaliá-los.

Nas análises de eficácia de um modo geral, o desafio fundamental é identificar a relação causal entre uma intervenção e um resultado. Para avaliar se os resultados são melhores com a intervenção ou tratamento, é necessário ter alguma

¹ Do inglês *effectiveness*.

estimativa de quais seriam os resultados sem essa intervenção (Levin e McEwan, 2002). Rubin, Stuart e Zanutto (2004) indicam que os efeitos causais são inerentemente comparações de resultados potenciais, medidos no mesmo momento do tempo em um conjunto comum de unidades. De modo que, para estimar o efeito de estar na escola A versus na escola B para um aluno Q, a nota do aluno Q na 5ª série na escola A, é comparada com a nota na 5ª série se o aluno estivesse na escola B. Como se sabe, o problema fundamental de qualquer inferência causal é que apenas um desses resultados potenciais pode ser observado, sob tratamento ou sob controle, pois ou o aluno esteve na escola A ou na B naquele ponto do tempo. Para obter a inferência, se estima os resultados potenciais não observados, implícita ou explicitamente.

Rubin, Stuart e Zanutto (2004) entendem que, assim como nas demais análises que buscam fazer inferências causais, também nos modelos para estimar o efeito professor dever-se-ia, como primeira tarefa, identificar as unidades, os tratamentos e os resultados potenciais. Para esses autores, um desafio para o estabelecimento de uma avaliação de valor agregado é a dificuldade de definir até esses conceitos fundamentais, e muitas vezes essas definições ficam implícitas nos modelos e não ganham a devida atenção.

Nas análises da eficácia dos professores, como se pretende analisar a viabilidade de se quantificar os efeitos de cada professor em particular, o tratamento pode ser entendido como ser submetido a ter aulas com um professor específico. As unidades, objetos aos quais os tratamentos são aplicados, são os alunos. Os resultados potenciais são as notas no final do ano no nível da unidade, o aluno. Os resultados potenciais sob o tratamento devem ser comparados aos resultados potenciais sob controle, ou seja, sob uma alternativa plausível, como indicam McCaffrey et al (2003).

Hanushek e Rivkin (2010) indicam que todos os artigos revisados por eles que buscam estimar os efeitos professor utilizaram como controle o professor médio da mesma escola, com exceção de apenas um artigo. Dessa forma, estariam admitindo, implicitamente, como situação plausível, o aluno ser submetido a um professor com a eficácia média dentre todos os professores de sua escola, em comparação com o seu professor de fato. No entanto, essa não é uma escolha óbvia: o controle pode ser outro professor em particular, o professor médio de todo o sistema de ensino, ou até uma situação em que não haveria professor.

Já que, como já observado, o problema fundamental das inferências causais é que apenas um dos resultados potenciais pode ser observado, pois uma mesma unidade só pode ser observada como unidade de tratamento ou de controle ao mesmo tempo, a identificação da situação contrafactual faz parte da essência da análise de eficácia.

Os alunos do professor X em uma determinada disciplina em um determinado ano não podem ser observados no mesmo ano com o professor Y ou com um professor médio nessa disciplina. Da mesma forma como não é possível alocar alunos exatamente iguais ao professor de tratamento e aos professores de controle.

Assim sendo, experimentos aleatórios seriam a melhor forma para estimar efeitos causais como os efeitos professor. Em uma situação ideal, alunos seriam alocados aleatoriamente aos professores e às escolas. Assim, se o grupo alocado a cada professor específico que se pretende medir o efeito professor, o grupo de tratamento, e o grupo alocado ao controle, forem grandes o suficiente, a aleatorização garantirá que os grupos são essencialmente iguais antes do tratamento ser aplicado. Assim, o grupo de controle serviria como uma boa estimativa do contrafactual, e quaisquer diferenças nos resultados observáveis entre os grupos seriam estimativas não viesadas do efeito real do professor em questão (Levin e McEwan, 2002; Rubin, Stuart e Zanutto, 2004; Rothstein, 2009).

A grande maioria dos estudos de eficácia não trabalha com experimentos, e sim com dados observacionais. Nesses casos, os pesquisadores não têm controle sobre a alocação das unidades e, portanto, usam técnicas estatísticas para tentar controlar as características individuais que afetam os resultados, mas também tentar determinar a probabilidade de indivíduos participarem da alternativa (Levin e McEwan, 2002).

Como bem destaca Braun (2005), na vida real, a distribuição de alunos e professores está longe de ser aleatória: pais influenciam a alocação de seus filhos a escolas, classes e professores, bem como professores podem escolher a escola e as classes que vão lecionar. As decisões de alocação dentro das escolas são tomadas, muitas vezes, pelos diretores e coordenadores, e geralmente os pesquisadores não possuem informações completas sobre como essas decisões são tomadas, se de forma acidental ou proposital; e, em sendo proposital, se com base somente em variáveis observadas pelo pesquisador, como as notas prévias

dos alunos, ou se baseadas em um conjunto composto por variáveis observadas e não observadas pelos pesquisadores, como pressupõe Rothstein (2010).

Desse modo, há grande probabilidade de que os alunos alocados a um professor não sejam representativos da população geral de alunos em relação ao seu nível e taxa de crescimento do desempenho, apoio dos pais, motivação, hábitos de estudo, dinâmica interpessoal e outras características relevantes, como descreve Braun (2005).

Sabendo disso, Rubin, Stuart e Zanutto (2004) afirmam que, com dados observacionais, um dos objetivos chave é encontrar unidades de tratamento e de controle que sejam o mais similar possível em relação às variáveis do *background* dos alunos. Se os grupos forem muito diferentes em relação às variáveis de *background*, os resultados provavelmente estarão baseados em pressupostos não testáveis e em extrapolação.

Os autores explicam que a extrapolação implícita é comum e particularmente difícil de diagnosticar em modelos complexos como os dessa literatura, porque os diagnósticos dos modelos comuns não avaliam a sobreposição das distribuições de covariadas. Como as escolas diferem muito nos valores de variáveis importantes como o “percentual de pobres”, é provável que as estimativas ajustando por essas covariadas usem modelos que dependem muito da extrapolação. Se a escola A não tem alunos “semelhantes” aos de outras escolas, seria impossível estimar o efeito da escola A em comparação com outras escolas sem fazer pressupostos heróicos, dizem eles.

A maioria dos estudos com base em modelos de valor agregado encontrados mais recentemente aborda a questão da distribuição aleatória de alunos a professores e escolas. As estratégias para verificar a hipótese de alocação aleatória utilizadas são variadas. Alguns aplicam testes qui-quadrado para verificar se a alocação dos alunos nas turmas são estatisticamente independentes de uma série de características dos alunos, tais como gênero, cor/raça, escolaridade dos pais, participação no programa de alimentação subsidiada e nota do aluno no ano anterior (Clotfelter, Ladd e Vigdor, 2006; Goldhaber e Hansen, 2010). Outros adotam testes de falsificação ao estimar os efeitos de professores do 5º ano sobre o desempenho dos alunos no 4º ano (Rothstein, 2009, 2010; Koedel e Betts, 2011). A hipótese testada é de associação zero, já que um fator futuro não pode ter efeitos sobre um resultado passado. Esses autores encontram evidências de “efeitos”,

rejeitando essa hipótese de associação zero. Segundo eles, o coeficiente desse “efeito professor” somente poderia ser diferente de zero se os erros do 4º ano forem correlacionados com a alocação do professor no 5º ano, ou seja, se a alocação do professor depende de resultados passados dos alunos – e portanto, não seria aleatória.

Em raros casos os pesquisadores podem contar com dados que demonstram terem sido originados a partir da alocação aleatória de alunos a turmas e professores, como é o caso de Nye, Konstantopoulos e Hedges (2004), que analisam os dados do Projeto STAR. Esse projeto, por ter o objetivo de avaliar os efeitos de tamanho de classe sobre o desempenho dos alunos, foi desenhado para distribuir alunos de forma aleatória.

Como depois de aplicado o tratamento não é possível aleatorizar a distribuição deste entre as unidades observadas, nos demais casos onde se identifica que a alocação não foi aleatória os pesquisadores buscam apenas tentar estimar o viés causado por essa alocação “intencional” e buscar formas de reduzi-lo, de modo a melhorar as estimativas de efeito professor.

Rothstein (2009, 2010) busca quantificar o viés que surge em modelos que falham em controlar os determinantes das alocações das turmas. A tarefa é simples caso o pressuposto seja de que a alocação dos alunos é aleatória condicional em variáveis observáveis: bastaria incluir essas variáveis como controles no modelo. Em sua análise, o viés é reduzido – com uma variância de cerca de 15% do tamanho dos efeitos professor reais – em um modelo de valor agregado que controla pelo desempenho passado, e tem maiores reduções quando outras notas passadas são incluídas como controles. Há um custo para isso: quanto mais notas passadas forem exigidas, mais alunos terão que ser excluídos da análise por *missing*.

A análise fica mais complexa quando ele abandona o pressuposto considerado por ele mesmo como irrealista de que toda a informação considerada pelo diretor ao alocar professores a turmas está disponível nos bancos de dados. Ele desenvolve métodos para estimar o viés quando se assume que o diretor tem acesso a uma quantidade limitada de informações que o pesquisador não observa. Modelos mais completos que controlam por todo o histórico de notas em lugar de apenas um desempenho passado reduzem o viés, mas apenas se o diretor tiver informações bem limitadas sobre o potencial dos alunos. Com pressupostos menos

restritivos, os vieses continuariam quantitativamente importantes mesmo em modelos de valor agregado mais completos.

Segundo Rothstein (2010), mesmo com modelos que consideram tantas variáveis que os diretores podem fazer uso, ainda existem os pedidos de pais e professores e outras influências pouco prováveis de oferecer variáveis disponíveis para serem utilizadas em modelos de valor agregado.

Em seus testes, Aaronson, Barrow e Sander (2007) e Clotfelter, Ladd e Vigdor (2007) obtiveram evidências de que a parcela do viés devida à alocação “intencional” entre escolas e a parcela devida à alocação “intencional” dentro das escolas seriam de magnitudes semelhantes. Ambos utilizaram a estratégia, muito adotada na literatura, de incluir efeitos fixos das escolas, o que implica que os coeficientes dos efeitos professor são identificados com base na variação na qualificação dos professores entre turmas dentro de cada escola, eliminando o viés associado com a alocação entre escolas.

Incluir efeitos fixos da escola pode ter ainda outra vantagem. A alocação aleatória de alunos a professores, com grandes quantidades de alunos sendo observados com cada professor, garantiria a expectativa de que, em média, as unidades de tratamento e de controle, antes do tratamento, fossem iguais. No entanto, não necessariamente garantem que ambos os grupos recebam as mesmas influências dos demais fatores que atuam durante o período estudado. Ou seja, não garantem que os alunos sejam expostos às mesmas condições físicas das escolas, ao mesmo currículo, recebam o mesmo tipo de material, tenham a mesma quantidade de horas de aula, entre outros fatores externos ao professor. Como alerta Braun (2005), como essas características variam sistematicamente entre escolas em uma rede, elas podem prejudicar a avaliação de professores com base no valor agregado. Como as chances de variação desses fatores dentro de uma mesma escola são bem menores que dentro de uma rede de ensino como um todo, estimar os efeitos professor dentro de cada escola reduz o viés devido a essas influências não contabilizadas.

Outra forma adotada em alguns trabalhos para mitigar esse viés é substituir as variáveis de aluno que não variam com o tempo por efeitos fixos dos alunos – em outras palavras, deixam os interceptos da equação de desempenho variar de acordo com o aluno, como fizeram Clotfelter, Ladd e Vigdor (2007). Assumindo que os efeitos são lineares, a inclusão desses efeitos fixos dos alunos

eliminariam qualquer viés associado à alocação não aleatória de alunos e professores. Isso é possível porque a presença dos efeitos fixos dos alunos significa que a única variação usada para estimar os coeficientes de interesse é a variação “dentro” e não entre alunos individuais.

Uma estratégia alternativa adotada em alguns trabalhos é identificar uma amostra dentre a população de alunos avaliada que apresenta evidências de ter sido alocada aleatoriamente e restringir parte das análises a essa amostra (Clotfelter, Ladd e Vigdor, 2006; Goldhaber e Hansen, 2010). A comparação dos resultados obtidos para a população e para essa amostra daria um indicador da magnitude do viés devido à alocação não aleatória.

A questão da aleatorização da alocação dos alunos a turmas, professores e escolas e da garantia de condições semelhantes nas quais ocorra o trabalho docente é primordial para a estimação do efeito individual dos professores a fim de avaliar seu desempenho. Quaisquer fatores que influenciem positiva ou negativamente o desempenho dos alunos de professores de forma diferente e não estejam devidamente incluídos nos modelos de valor agregado podem prejudicar a comparação entre esses professores e, conseqüentemente, a tomada de qualquer tipo de decisão relativa à gestão de pessoas com base nessa comparação.

3.2. Confiabilidade

Para que um indicador seja útil para orientar qualquer recomendação ou tomada de decisão, ele deve ser fiel à realidade que pretende representar. No caso dos efeitos professor, deve indicar, com máxima precisão possível, o nível de eficácia do professor avaliado em relação à situação contrafactual, geralmente o professor médio da escola ou da rede.

Para isso, não basta que a alocação dos alunos aos professores tenha sido realizada de forma aleatória. Mesmo que os alunos tenham sido distribuídos nas classes por sorteio, uma série de fatores não relacionados à qualidade dos professores pode influenciar os resultados dos alunos nos testes padronizados, que são as medidas nas quais está baseado todo o cálculo dos efeitos professor.

McCaffrey et al (2003) explicam que a variância residual nas notas dos alunos devida a erros de medida e a fontes de variabilidade no desempenho além dos professores, escolas e outros fatores do modelo resulta em erros amostrais nos parâmetros estimados do modelo (incluindo os componentes da variância) e nos efeitos professor estimados. Os parâmetros do modelo dependem da variabilidade da amostra inteira de alunos. O erro amostral na estimativa do efeito de um professor individualmente depende do erro amostral dos parâmetros do modelo e da variabilidade residual das notas dos alunos na classe desse professor.

Um mal-estar ou problema familiar no dia da prova são exemplos de fatores que afetam o desempenho de um aluno em um teste específico, sendo responsáveis por esses erros amostrais. Como em qualquer medida, aumentando o tamanho da amostra, diminui-se a influência de cada um dos erros sobre os efeitos estimados.

Portanto, como descrevem McCaffrey et al (2003), a variabilidade de um efeito professor individual devido a erro amostral vai diminuir com o aumento do tamanho da classe – efeitos estimados para professores com classes maiores tendem a ter variabilidade menor. A qualidade das inferências sobre os professores vai depender da razão entre a variabilidade devida aos erros amostrais e a variabilidade dos efeitos professor reais. Os autores afirmam que, dado que a variabilidade dos professores provavelmente representa apenas uma porção modesta da variabilidade nas notas dos alunos, a variabilidade residual deve ser bem maior do que a variabilidade devida aos professores. Além disso, muitos professores ensinam apenas números moderadamente pequenos de alunos. Então, a variabilidade no erro amostral nos efeitos estimados deve ser grande em relação à real variância dos efeitos professor. Lockwood, Louis e McCaffrey (2002), em sua análise a respeito da incerteza na estimação de *rankings* de eficácia de professores, descobriram a partir de simulações que, a menos que a razão entre a variância do erro amostral e a variância dos efeitos dos professores estimados seja menor que 0,1, os *rankings* estimados somente serão suficientemente precisos para distinguir entre os professores de eficácia mais extrema.

McCaffrey et al (2003) entendem que obter *rankings* precisos, tendo com base essa razão de 0,1 entre a variância do erro amostral e a variância dos efeitos professor é um objetivo difícil de alcançar. Em McCaffrey et al (2004), por exemplo, a variabilidade devida a erro amostral obtida foi de 0,2 a 0,4 da variabilidade estimada

dos efeitos professor, dependendo da série e do modelo empregado. Embora esse percentual não permita inferências sobre *rankings*, seguindo a referência de Lockwood, Louis e McCaffrey (2002), suas análises geraram evidências de que entre um terço e um quarto dos professores podem ser considerados como distintos da média.

A idéia deles é que, embora muitas vezes os erros amostrais não permitam estimativas precisas o suficiente para alguns usos como *rankings*, são precisas o suficiente para outros tipos de análise, como essa apresentada. Na opinião desse conjunto de autores, é preciso ter ou uma extrema heterogeneidade entre professores ou variâncias minúsculas nas estimativas do efeito professor específico, produzindo coeficientes de alta estabilidade, para que se considere a ordenação confiável.

3.3. Estabilidade

De acordo com Goldhaber e Hansen (2010), muitas vezes as políticas de seleção se baseiam no desempenho passado dos trabalhadores como um preditor confiável do desempenho futuro deles. É basicamente nessa idéia que se fundamenta a proposta de Hanushek e Rivkin (2004) e Gordon, Kane e Staiger (2006) de manter na carreira apenas os professores que demonstrarem estar acima de um certo nível no *ranking* de eficácia de professores. Goldhaber e Hansen (2010) entendem que essa idéia está baseada em um pressuposto não testado da estabilidade do desempenho dos professores. Eles afirmam que, caso se confirme que o desempenho individual do professor é uma medida extremamente estável, então poderia ser utilizada como instrumento para dispensar professores pouco efetivos.

Uma grande dificuldade das análises de estabilidade é definir um parâmetro aceitável em relação à correlação dos efeitos estimados para um mesmo professor ao longo dos anos.

Lockwood, Louis e McCaffrey (2002) afirmam que os estimadores e as regras de decisão não são realmente confiáveis até que o coeficiente de estabilidade esteja perto de 1. Então, para que eles pudessem considerar confiável, cada

professor individual teria que ter medidas de efeito professor distintas que fossem quase iguais para que a variância dessas medidas de efeito professor específico fosse próxima de zero.

Essa meta estaria longe de ser alcançada, como demonstram os estudos que buscaram levantar evidências de estabilidade das medidas de professor.

Aaronson, Barrow e Sander (2007) fazem um teste qui-quadrado que rejeita que a correlação entre as medidas dos efeitos de um mesmo professor seria zero – esse seria um caso extremo em que os *rankings* de eficácia de professores seriam aleatórios. Obtêm evidências de que a chance de um professor que está no topo do *ranking* em um ano aparecer no topo no outro ano é grande, mas o mesmo não ocorre com os que estão nos níveis mais baixos no primeiro ano.

McCaffrey et al (2009) também analisam a estabilidade e obtêm correlações nos efeitos professor estimados de 0,22 a 0,67. Segundo eles, essas correlações implicam *rankings* de professores com estabilidade apenas moderada: cerca de um terço dos professores do quintil mais alto permanecem no quintil mais alto no ano seguinte enquanto cerca de um décimo caem para o quintil mais baixo. Goldhaber e Hansen (2010) obtiveram correlações de 0,32 a 0,59. Segundo eles, o nível de variação observada é consistente com medidas de resultado de trabalhadores de outras áreas. Em uma meta-análise com ocupações de alta complexidade como a docência, as correlações entre as estimativas da qualidade dos trabalhadores em anos diferentes foram de 0,33 a 0,40, semelhantes aos encontrados por eles para leitura e menores que os de matemática, de cerca de 0,50.

Diferentemente do que afirmavam Lockwood, Louis e McCaffrey (2002), Goldhaber e Hansen (2010) reconhecem que não se espera que essa correlação seja perfeita. Além dos erros amostrais e do componente permanente da qualidade do professor estimado pela correlação entre duas medidas de efeito professor em dois momentos distintos, eles indicam a existência de um componente transitório nos efeitos professor. Esse componente pode capturar desde a variação no desempenho do professor em diferentes momentos da vida profissional até outros fatores no nível da classe naquele ano específico, como uma boa “química” entre professores e alunos dessa classe.

McCaffrey et al (2009) também passaram a analisar os efeitos professor estimados em cada ano como a soma de três variáveis independentes: o efeito

professor persistente, o efeito professor não persistente e os erros amostrais, seguindo a abordagem de Kane e Staiger (2002) na análise dos efeitos escola. Diferentemente dos erros amostrais e dos efeitos persistentes, ambos os estudos indicam que não se consegue estimar diretamente a variabilidade dos efeitos não persistentes. Em lugar disso, estima-se a variância dos efeitos não persistentes de forma indireta, sendo a parcela restante da variância dos efeitos professor estimados, após subtrair-se a parcela devida aos efeitos persistentes e aos erros amostrais.

Goldhaber e Hansen (2010) usam dados longitudinais de 12 anos, chegando a acompanhar o mesmo professor por 10 anos. Eles encontram evidências de que erros amostrais nas notas dos alunos têm um papel importante na determinação da estabilidade das estimativas dos efeitos professor ao longo do tempo, correspondendo a aproximadamente um terço da variação nos efeitos professor. Da variância restante, entre um e dois terços se deve à variação da eficácia do próprio professor ao longo do tempo, os efeitos persistentes, e entre um e dois terços se deve à variação dos efeitos não persistentes.

4. DADOS

A presente pesquisa buscará analisar a eficácia de uma amostra de professores da rede municipal de ensino de São Paulo e será desenvolvida com base nos dados da Prova São Paulo, descritos nesta seção.

4.1. A Prova São Paulo

A Prova São Paulo é uma avaliação em larga escala que compõe o Sistema de Avaliação de Aproveitamento Escolar dos Alunos da Rede Municipal de Ensino de São Paulo, instituído pela Lei 14.063, de 14 de outubro de 2005, com os seguintes objetivos:

I - desenvolver um sistema de avaliação do desempenho escolar dos alunos do ensino fundamental e médio da Rede Municipal de Ensino de São Paulo, que subsidie a Secretaria Municipal de Educação nas tomadas de decisão quanto à Política Educacional do Município;

II - verificar o desempenho dos alunos nas séries do ensino fundamental e médio, nos diferentes componentes curriculares, de modo a fornecer ao sistema de ensino, às equipes técnico-pedagógicas das Coordenadorias de Educação e às Unidades Educacionais informações que subsidiem:

a) a política de formação continuada dos recursos humanos do magistério;

b) a reorientação da proposta pedagógica desses níveis de ensino, de modo a aprimorá-la;

c) a viabilização da articulação dos resultados da avaliação com o planejamento escolar, a formação dos professores e o estabelecimento de metas para o projeto pedagógico de cada escola;

d) a orientação para os trabalhos desenvolvidos nas Salas de Apoio Pedagógico - SAPs das unidades escolares com os alunos que necessitam de reforço na aprendizagem. (SÃO PAULO, 2005)

Os testes da Prova São Paulo são elaborados com base em Matrizes de Referência construídas especificamente para esta avaliação municipal.

De acordo com o documento da Secretaria Municipal de Educação que apresenta as Matrizes da avaliação, foram contratados assessores-especialistas em avaliação que organizaram a proposta inicial de matrizes de Língua Portuguesa e Matemática, com base em documentos que balizam as avaliações nacionais e internacionais, bem como documentos oficiais da Secretaria de Educação do Município de São Paulo que estabelecem as diretrizes do Ensino Fundamental (SÃO PAULO, 2007).

O documento descreve que a primeira versão das Matrizes foi discutida e analisada criticamente em reuniões técnicas com professores e representantes de todas as coordenadorias técnico-pedagógicas da rede municipal.

A consolidação desse trabalho ainda passou por um período em que todos os professores e especialistas da rede municipal puderam indicar, em uma ficha de validação que acompanhou as matrizes, sua concordância ou discordância sobre cada uma das habilidades.

Os professores de cada ano/ciclo discutiram com seus pares as matrizes propostas em suas respectivas escolas e, estando de acordo, sintetizaram em uma única planilha suas opiniões.

Os dados foram analisados em reuniões técnicas com especialistas e orientaram a nova redação das Matrizes de Referência. Nessa última versão das Matrizes, foi considerada a necessidade de adequação às Matrizes de Referência do Saeb/Prova Brasil, para quarta e oitava séries, com a manutenção de habilidades consensualmente aceitas e já avaliadas pelo sistema nacional.

Nas provas do quarto ano dos Ciclos I e II – 4ª e 8ª séries do Ensino Fundamental – são incluídos blocos de itens já calibrados do Saeb, para que os resultados possam ter comparabilidade nacional. Os resultados de todas as escolas são colocados na escala nacional do Saeb, à qual são acrescentados os demais anos avaliados.

A primeira edição da Prova São Paulo foi aplicada em 2007 e buscou avaliar de forma censitária os alunos do 2º e 4º anos dos Ciclos I e II, em Língua Portuguesa e Matemática.

A segunda edição apresentou dois avanços. O primeiro foi a avaliação de forma censitária dos alunos nas seguintes turmas: 2º ano do Ciclo I, 3º ano PIC² (Programa Intensivo de Ciclo) do Ciclo I, 4º ano do Ciclo I, 4º ano PIC do Ciclo I, 1º ano do Ciclo II (alunos que apresentaram, em 2007, proficiências abaixo de 150 em Língua Portuguesa no 4º ano do Ciclo I) e 2º ano do Ciclo II; e, por amostragem para os alunos das seguintes turmas: 3º ano do Ciclo I, 1º ano do Ciclo II (demais proficiências), 3º ano do Ciclo II e 4º ano do Ciclo II.

Além disso, criou-se um ponto de corte, denominado “Não Satisfatório”, para identificar os alunos considerados com baixas proficiências. De acordo com a Secretaria o critério para a definição do corte foi amparado em analogias com cortes de outros sistemas de avaliação.

Nas edições de 2009 e 2010 os critérios de escolha dos alunos avaliados foram mantidos, mas houve uma alteração na apresentação dos resultados com a definição de quatro níveis de proficiência dos alunos: “Abaixo do Básico”, “Básico”, “Adequado” e “Avançado”, como descrito nas tabelas 1 e 2:

Tabela 1 – Níveis de Proficiência de Língua Portuguesa da Prova São Paulo

| Ano Escolar | Abaixo do Básico | Básico | Adequado | Avançado |
|-------------------|------------------|-----------|-----------|-------------|
| 2º ano – Ciclo I | menos de 115 | 115 a 149 | 150 a 199 | 200 ou mais |
| 3º ano – Ciclo I | menos de 135 | 135 a 174 | 175 a 224 | 225 ou mais |
| 4º ano – Ciclo I | menos de 150 | 150 a 199 | 200 a 249 | 250 ou mais |
| 1º ano – Ciclo II | menos de 165 | 165 a 214 | 215 a 264 | 265 ou mais |
| 2º ano – Ciclo II | menos de 175 | 175 a 224 | 225 a 274 | 275 ou mais |
| 3º ano – Ciclo II | menos de 185 | 185 a 249 | 250 a 299 | 300 ou mais |
| 4º ano – Ciclo II | menos de 200 | 200 a 274 | 275 a 324 | 325 ou mais |

Fonte: Elaboração própria com base em informações disponíveis em <http://portalsme.prefeitura.sp.gov.br/Projetos/nucleo/Default.aspx>.

² O Programa Intensivo de Ciclo (PIC) é voltado para alunos da 3ª e 4ª séries que não foram alfabetizados na idade esperada.

Tabela 2 – Níveis de Proficiência de Matemática da Prova São Paulo

| Ano Escolar | Abaixo do Básico | Básico | Adequado | Avançado |
|--------------------|-------------------------|---------------|-----------------|-----------------|
| 2º ano – Ciclo I | menos de 125 | 125 a 174 | 175 a 224 | 225 ou mais |
| 3º ano – Ciclo I | menos de 150 | 150 a 199 | 200 a 249 | 250 ou mais |
| 4º ano – Ciclo I | menos de 175 | 175 a 224 | 225 a 274 | 275 ou mais |
| 1º ano – Ciclo II | menos de 190 | 190 a 239 | 240 a 289 | 290 ou mais |
| 2º ano – Ciclo II | menos de 200 | 200 a 249 | 250 a 299 | 300 ou mais |
| 3º ano – Ciclo II | menos de 210 | 210 a 274 | 275 a 324 | 325 ou mais |
| 4º ano – Ciclo II | menos de 225 | 225 a 299 | 300 a 349 | 350 ou mais |

Fonte: Elaboração própria com base em informações disponíveis em <http://portalsme.prefeitura.sp.gov.br/Projetos/nucleo/Default.aspx>.

De acordo com informações obtidas no Portal Eletrônico da Prefeitura de São Paulo, esta maneira de agrupamento dos pontos da escala de proficiência teve como objetivo favorecer as intervenções pedagógicas por permitir entender melhor os processos de aprendizagem dos alunos.

Para compor as amostras a serem testadas, os alunos das séries ímpares são selecionados todos os anos de forma aleatória em cada série dentro de cada escola. As amostras são desenhadas para representar o desempenho de cada escola em cada série por ano, com o objetivo de acompanhar a aprendizagem dos alunos nesse nível de agregação. Portanto, possui as limitações de não representar as turmas nem acompanhar os mesmos alunos ao longo do período analisado.

A tabela a seguir apresenta o desenho da aplicação da Prova São Paulo de acordo com as séries entre os anos de 2007 e 2010 – período para o qual os dados encontram-se disponíveis:

Tabela 3 – Descrição dos Grupos de Alunos Submetidos à Prova São Paulo

| Classificação utilizada pela Secretaria Municipal de SP | Correspondência com o Ensino Fundamental de 8 anos | Correspondência com o Ensino Fundamental de 9 anos | 2007 | 2008 | 2009 | 2010 |
|---|--|--|------------|--|------|------|
| | | 1º ano | | | | |
| 1º ano – Ciclo I | 1ª série | 2º ano | | | | |
| 2º ano – Ciclo I | 2ª série | 3º ano | Censitário | | | |
| 3º ano – Ciclo I | 3ª série | 4º ano | | Censitário para os alunos do PIC e amostral para os demais alunos | | |
| 4º ano – Ciclo I | 4ª série | 5º ano | Censitário | | | |
| 1º ano – Ciclo II | 5ª série | 6º ano | | Censitário para os alunos com proficiência abaixo de 150 em Língua Portuguesa no ano anterior e amostral para os demais alunos | | |
| 2º ano – Ciclo II | 6ª série | 7º ano | Censitário | | | |
| 3º ano – Ciclo II | 7ª série | 8º ano | | Amostral | | |
| 4º ano – Ciclo II | 8ª série | 9º ano | Censitário | | | |

Fonte: Elaboração própria com base em informações disponíveis em <http://portalsme.prefeitura.sp.gov.br/Projetos/nucleo/Default.aspx>.

Ainda são raras as experiências no Brasil de avaliações em larga escala que testam os mesmos alunos em diferentes pontos do tempo, preferencialmente com um pequeno intervalo de tempo entre os testes, como em anos seguidos, permitindo a construção de banco de dados longitudinais e a estimação de modelos de valor agregado.

A Prova São Paulo, apesar de não testar todos os alunos nas séries ímpares e não ser representativa das turmas nessas séries, permite a construção de uma considerável base de dados onde é possível acompanhar uma parcela de alunos de um ano para o outro, para gerar a medida de valor agregado, além de ser possível identificar os professores que lecionaram para a grande maioria dos alunos testados.

Outra vantagem da Prova São Paulo é o fato de, como descrito anteriormente, as matrizes terem sido construídas contando com a análise e contribuição dos profissionais do magistério. Supõe-se com isso que tanto o desenho do conteúdo das matrizes, quanto os níveis de desempenho definidos como esperados para cada série, estejam intimamente relacionados ao currículo estabelecido na rede municipal e que se espera que seja cumprido por parte dos professores.

4.2. Definição da Amostra

Tendo disponíveis os dados de resultados dos alunos na Prova São Paulo dos anos de 2007 a 2010, optou-se por usar os dados de 2010 de alunos e professores para as estimativas principais, servindo como *baseline* os dados de resultados de alunos de 2009. Para as análises de estabilidade, são utilizados adicionalmente os dados de alunos e professores de 2009, com os dados de resultados de 2008 servindo como *baseline*.

Dois fatores levaram à opção de realizar as principais estimativas com os dados do ano de 2010. O primeiro deles está relacionado ao fato de serem os dados mais recentes relativos à Prova SP disponíveis para pesquisas, permitindo a obtenção de um retrato da eficácia dos professores municipais que esteja mais próximo da realidade atual. Adicionalmente, pesou o fato de, para 2010, estarem disponíveis respostas de questionários aplicados a alunos, pais e professores, enquanto para 2008 não há questionário de pais de alunos e para 2009 não há questionário de professores. Apesar de ser possível encontrar respostas de um desses atores no questionário de outros anos, é sabido que procedimentos como esse resultam em perdas significativas de dados.

Diferentemente de todos os trabalhos encontrados na literatura empírica sobre o tema, todos com base em dados de estados ou cidades americanas, os testes da Prova SP foram aplicados a amostras de alunos de cada série ímpar, e não para a totalidade dos alunos. Portanto, uma preocupação particular foi buscar as séries com a maior proporção de alunos testados em relação à população em questão.

Tabela 4 – Proporção de Alunos de 2010 com Notas na Prova SP por Série Testada

| Séries | Alunos Matriculados ³ | Língua Portuguesa | | Matemática | |
|---------------------------------|----------------------------------|--------------------------|---------------------------------|--------------------------|---------------------------------|
| | | Alunos com Notas em 2010 | Alunos com Notas em 2009 e 2010 | Alunos com Notas em 2010 | Alunos com Notas em 2009 e 2010 |
| 3º ano - Ciclo I (3ª série) | 57509 | 19079 33% | 16359 28% | 19272 34% | 16600 29% |
| 4º ano - Ciclo I (4ª série) | 71071 | 65999 93% | 25330 38% | 66207 93% | 25539 39% |
| 1º ano - Ciclo II (5ª série) | 67673 | 29501 44% | 26522 39% | 29460 44% | 26356 39% |
| 2º ano - Ciclo II (6ª série) | 64103 | 58383 91% | 26898 42% | 58187 91% | 26877 42% |
| 3º ano - Ciclo II (7ª série) | 59759 | 16002 27% | 14195 24% | 15488 26% | 13706 23% |
| 4º ano - Ciclo II (8ª série) | 57934 | 50114 87% | 14833 30% | 48725 84% | 14271 25% |

Fonte: Elaboração própria com base em dados da Secretaria Municipal de São Paulo e da Prova SP dos anos de 2009 e 2010.

Como apresentado na tabela 4, a 4ª, a 5ª e a 6ª séries são aquelas que melhor atendem esse critério, pois cerca de 40% dos alunos matriculados têm notas na Prova São Paulo em 2010 e 2009, o que é um pré-requisito para tornarem-se alvo das análises de valor agregado.

Como os demais professores do Ciclo II, os professores da 5ª e da 6ª séries apresentam uma vantagem em relação aos professores do Ciclo I pois, por lecionar apenas uma disciplina, trabalham com um número maior de turmas e alunos por ano, podendo gerar melhores medidas de valor agregado. Enquanto os professores do Ciclo I são encontrados no banco de dados com, no máximo, duas turmas da mesma série por ano, os professores do Ciclo II chegam a ser encontrados com até dez turmas de uma mesma série por ano, como pode ser observado na tabela 5, que compara a quantidade de turmas identificadas a professores de 5ª série, por exemplo⁴, com aquelas identificadas a professores de 4ª série.

³ Número de alunos matriculados na data-base 28/11/2010, aptos a realizar a Prova São Paulo, segundo informações da Secretaria Municipal de Educação de São Paulo.

⁴ Os dados relativos à 6ª série seguem um padrão similar aos da 5ª série, de modo que optou-se por apresentar a comparação apenas entre os dados da 4ª e da 5ª séries.

Tabela 5 – Distribuição dos Professores de Acordo com a Quantidade de Turmas na Amostra em 2010

| Nº de turmas por professor | Língua Portuguesa | | | | Matemática | | | |
|----------------------------|-------------------|---------|-------------|---------|-------------|---------|-------------|---------|
| | 4ª série | | 5ª série | | 4ª série | | 5ª série | |
| | Nº de prof. | % Acum. | Nº de prof. | % Acum. | Nº de prof. | % Acum. | Nº de prof. | % Acum. |
| 1 turma | 2207 | 100% | 212 | 100% | 2207 | 100% | 202 | 100% |
| 2 turmas | 57 | 2,5% | 198 | 73,2% | 57 | 2,5% | 180 | 73,6% |
| 3 turmas | 0 | 0,0% | 170 | 48,2% | 0 | 0,0% | 166 | 50,1% |
| 4 turmas | 0 | 0,0% | 121 | 26,7% | 0 | 0,0% | 123 | 28,4% |
| 5 turmas | 0 | 0,0% | 85 | 11,4% | 0 | 0,0% | 91 | 12,3% |
| Mais de 5 turmas | 0 | 0,0% | 5 | 0,6% | 0 | 0,0% | 3 | 0,4% |
| Total de Professores | 2264 | | 791 | | 2264 | | 765 | |

Fonte: Elaboração própria com base em dados da Prova SP do ano de 2010.

Há duas grandes razões para priorizar amostras de professores com maiores números de turmas. Em primeiro lugar por permitir que se tente distinguir o efeito do professor do efeito da turma. Em segundo lugar porque se espera que um maior número de turmas leve a maiores números de observações de notas de alunos, possibilitando a estimativa de melhores medidas de efeito professor.

Tabela 6 – Distribuição dos Professores de Acordo com a Quantidade de Alunos na Amostra em 2010

| Nº de Alunos por Professor | Língua Portuguesa | | | | Matemática | | | |
|----------------------------|-------------------|---------|-------------|---------|-------------|---------|-------------|---------|
| | 4ª série | | 5ª série | | 4ª série | | 5ª série | |
| | Nº de prof. | % Acum. | Nº de prof. | % Acum. | Nº de prof. | % Acum. | Nº de prof. | % Acum. |
| Até 5 alunos | 176 | 100% | 3 | 100% | 176 | 100% | 6 | 100% |
| 6 a 10 alunos | 883 | 92,0% | 63 | 99,6% | 874 | 92,0% | 52 | 99,2% |
| 11 a 20 alunos | 1059 | 52,0% | 167 | 91,7% | 1060 | 52,4% | 165 | 92,4% |
| 21 a 50 alunos | 89 | 4,0% | 419 | 70,5% | 97 | 4,4% | 404 | 70,8% |
| 51 a 100 alunos | 0 | 0,0% | 136 | 17,6% | 0 | 0,0% | 138 | 18,0% |
| Mais de 100 alunos | 0 | 0,0% | 3 | 0,4% | 0 | 0,0% | 0 | 0,0% |
| Total de Professores | 2207 | | 791 | | 2207 | | 765 | |

Fonte: Elaboração própria com base em dados da Prova SP do ano de 2010.

De fato, a tabela 6 indica que o número de professores na 5ª série corresponde a aproximadamente um terço do número de professores na 4ª série, cada um concentrando quantidades geralmente bem maiores que os professores da 4ª série. Na 4ª série, cerca de metade dos professores são encontrados com menos

de 10 alunos na amostra, enquanto na 5ª série apenas cerca de 10% dos professores estão nessa situação.

No entanto, a grande desvantagem de se trabalhar com o Ciclo II está em justamente ter menos professores por série em cada escola, já que um mesmo professor trabalha com diversas turmas. Nas escolas que tem um único professor de uma disciplina para todas as turmas de uma mesma série, não é possível distinguir o efeito do professor do efeito da escola – o que prejudica sensivelmente as análises de efeito professor.

Como pode ser observado na tabela 7, esse é o caso de mais da metade das escolas da amostra de análise, em relação à 5ª série. Portanto, para mais da metade das escolas não seria possível separar os efeitos do professor dos efeitos das escolas sobre o desempenho dos alunos na 5ª série.

Tabela 7 – Distribuição das Escolas de Acordo com a Quantidade de Professores de 4ª e 5ª séries na Amostra da Análise em 2010

| Nº de Professores por Escola | Língua Portuguesa | | | | Matemática | | | |
|------------------------------|-------------------|---------|-------------|---------|-------------|---------|-------------|---------|
| | 4ª série | | 5ª série | | 4ª série | | 5ª série | |
| | Nº de prof. | % Acum. | Nº de prof. | % Acum. | Nº de prof. | % Acum. | Nº de prof. | % Acum. |
| 1 professor | 10 | 100% | 296 | 100% | 10 | 100% | 307 | 100% |
| 2 professores | 45 | 98,1% | 192 | 43,7% | 45 | 98,1% | 174 | 41,2% |
| 3 professores | 119 | 89,7% | 32 | 7,2% | 119 | 89,7% | 36 | 7,9% |
| 4 professores | 180 | 67,5% | 6 | 1,1% | 180 | 67,5% | 5 | 1,0% |
| 5 professores | 95 | 33,8% | 0 | 0,0% | 95 | 33,8% | 0 | 0,0% |
| Mais de 5 professores | 86 | 16,1% | 0 | 0,0% | 86 | 16,1% | 0 | 0,0% |
| Total de Escolas | 535 | | 526 | | 535 | | 522 | |

Fonte: Elaboração própria com base em dados da Prova SP do ano de 2010.

Portanto, a escolha da série a ser analisada está relacionada a uma decisão em que se priorize a quantidade de alunos por professor, de modo a buscar maior precisão das estimativas do efeito de cada professor individual, ou se priorize a estimação do efeito escola para separá-lo do efeito professor, de maneira a produzir estimativas que sejam mais fidedignas ao desempenho dos professores.

Não há um consenso sobre o número mínimo de alunos necessário para o cálculo do efeito de um professor individual. Todos os estudos empíricos encontrados trabalham com dados de provas que foram aplicadas de forma censitária aos alunos de determinada rede ou distrito. Portanto, a preocupação de

muitos deles está em restringir a quantidade mínima e máxima de alunos por turma, de modo a representar as turmas típicas. A maioria estabelece um número mínimo de 10 ou 12 alunos por turma para que a turma permaneça na amostra (Goldhaber e Hansen, 2010; Rothstein, 2009, 2010; Harris e Sass, 2006; Rivkin, Hanushek e Kain, 2005). Alguns deles afirmam que o fazem para excluir “turmas atípicas” da amostra. Portanto, parece ser uma questão relacionada à representatividade, e não à precisão das medidas. Por se tratar de provas censitárias, são poucos os que estabelecem números mínimos de alunos por professor, como Croninger et al (2007), que exigem pelo menos 2 alunos por professor, e Koedel e Betts (2011), que restringem a amostra aos professores com observações de pelo menos 20 alunos.

Os artigos empíricos revisados, de um modo geral, não explicam as razões para terem escolhido estas ou aquelas séries como base para sua análise – o que leva a imaginar que a escolha se deu por questões estatísticas, para melhorar as estimações dos efeitos professor.

Porém, devido à questão da falta de aleatorização na distribuição de alunos a escolas e a turmas, a grande maioria dos estudos calcula o efeito professor somente dentro das escolas – ou seja, separando-o do efeito da escola (Rockoff, 2004; Nye, Konstantopoulos e Hedges, 2004; Aaronson, Barrow e Sander, 2007; Rothstein, 2009, 2010; entre outros). Eles entendem que a parcela da alocação intencional dos alunos às escolas é tão significativa quanto a parcela da alocação intencional dos alunos às turmas, dentro das escolas.

Portanto, para as análises deste estudo, serão estimados os efeitos dos professores sobre o desempenho dos alunos da 4ª série – 4º ano do Ciclo I – em Língua Portuguesa e de Matemática no ano de 2010, tomando como base de partida todos os alunos com notas no ano de 2010 e no ano de 2009, para os quais sejam identificados à turma e o professor em 2010, permitindo a aplicação dos modelos de valor agregado para produzir as estimativas de efeito professor.

Para permitir a distinção entre os efeitos escola e os efeitos professor, constarão todas as escolas com pelo menos 2 professores de 4ª série. Buscando garantir uma quantidade mínima de alunos para estimar os efeitos professor individuais, serão mantidos todos os professores da 4ª série com pelo menos 6 alunos constantes do banco de dados da Prova SP.

4.3. Descrição do Perfil de Alunos e Professores

Como já indicado, a Prova SP foi desenhada para ser aplicada de forma censitária na 4ª série em 2010. Da mesma forma que geralmente acontece nas demais aplicações de testes padronizados, nem todos os alunos compareceram no dia em que a prova foi aplicada e, dentre os que compareceram, nem todos de fato completaram os testes e preencheram os questionários.

Na tabela 4, observamos que cerca de 93% dos alunos da 4ª série que estariam aptos a participar da Prova SP em 2010 de fato constam nos bancos de dados com notas nos testes. É possível desenhar o perfil de todos esses alunos que constam nos bancos da Prova SP, pelo menos em relação a características como gênero e idade, as quais constam nos bancos de dados da Secretaria Municipal. Para as demais características, conta-se com as informações resultantes das respostas aos questionários aplicados em conjunto com testes da Prova SP.

É importante descrever o perfil desse grupo dos 93% dos alunos da 4ª série em 2010, o qual chamaremos de Amostra Geral, porque permite a comparação dessa amostra próxima do que seria a população de alunos da 4ª série com a descrição do perfil dos alunos que permaneceram na amostra que será utilizada para as análises, a Amostra Principal. Vale lembrar que compõem a Amostra Principal os alunos da 4ª série de 2010 com notas em 2010 e 2009, de professores com pelo menos 6 alunos no banco de dados, de escolas com pelo menos dois professores na 4ª série.

Tabela 8 – Estatística Descritiva dos Grupos de Alunos na 4ª Série em 2010
(continua)

| | Amostra Geral | | Amostra Principal | |
|---------------------------------|---------------|----------------|-------------------|----------------|
| | % de alunos | % de respostas | % de alunos | % de respostas |
| Gênero Feminino | 47,20% | 100,00% | 44,30% | 100,00% |
| <u>Idade</u> | | | | |
| Menos de 10 anos | 0,50% | 100,00% | 0,40% | 100,00% |
| 10 anos | 67,90% | 100,00% | 56,70% | 100,00% |
| 11 anos | 21,70% | 100,00% | 28,90% | 100,00% |
| 12 anos | 6,40% | 100,00% | 10,10% | 100,00% |
| Mais de 12 anos | 2,30% | 100,00% | 2,90% | 100,00% |
| Participa do PIC | 6,60% | 100,00% | 13,87% | 100,00% |
| Fez Pré-Escola | 56,30% | 69,10% | 55,70% | 71,70% |
| <u>Escolaridade do Pai</u> | | | | |
| Não estudou | 4,90% | 71,20% | 5,40% | 71,90% |
| Anos Iniciais do EF Incompletos | 17,50% | 71,20% | 18,80% | 71,90% |
| Anos Iniciais do EF Completos | 28,40% | 71,20% | 29,40% | 71,90% |
| Ensino Fundamental Completo | 20,70% | 71,20% | 20,50% | 71,90% |
| Ensino Médio Completo | 25,50% | 71,20% | 23,40% | 71,90% |
| Ensino Superior Completo | 2,90% | 71,20% | 2,60% | 71,90% |
| <u>Escolaridade da Mãe</u> | | | | |
| Não estudou | 3,20% | 75,50% | 3,70% | 76,60% |
| Anos Iniciais do EF Incompletos | 12,90% | 75,50% | 14,70% | 76,60% |
| Anos Iniciais do EF Completos | 28,90% | 75,50% | 30,20% | 76,60% |
| Ensino Fundamental Completo | 22,20% | 75,50% | 21,50% | 76,60% |
| Ensino Médio Completo | 30,30% | 75,50% | 27,70% | 76,60% |
| Ensino Superior Completo | 2,60% | 75,50% | 2,30% | 76,60% |
| <u>Renda Familiar</u> | | | | |
| Até R\$ 850,00 | 42,80% | 56,30% | 44,70% | 56,80% |
| R\$ 851,00 a R\$ 1275,00 | 31,10% | 56,30% | 30,50% | 56,80% |
| R\$ 1276,00 a R\$ 2125,00 | 18,40% | 56,30% | 17,60% | 56,80% |
| R\$ 2126,00 a R\$ 4250,00 | 6,50% | 56,30% | 5,90% | 56,80% |
| Mais de R\$ 4250,00 | 1,30% | 56,30% | 1,40% | 56,80% |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2009 e 2010.

Tabela 8 – Estatística Descritiva dos Grupos de Alunos na 4ª Série em 2010
(conclusão)

| | Amostra Geral | | Amostra Principal | |
|--|---------------|----------------|-------------------|----------------|
| | % de alunos | % de respostas | % de alunos | % de respostas |
| <u>Níveis de Proficiência em Língua Portuguesa em 2010</u> | | | | |
| Abaixo do Básico | 33,40% | 92,40% | 39,57% | 94,20% |
| Básico | 42,00% | 92,40% | 39,93% | 94,20% |
| Adequado | 20,90% | 92,40% | 17,37% | 94,20% |
| Avançado | 3,70% | 92,40% | 3,13% | 94,20% |
| <u>Níveis de Proficiência em Matemática em 2010</u> | | | | |
| Abaixo do Básico | 46,20% | 92,70% | 52,31% | 94,90% |
| Básico | 37,70% | 92,70% | 34,22% | 94,90% |
| Adequado | 14,30% | 92,70% | 11,81% | 94,90% |
| Avançado | 1,80% | 92,70% | 1,65% | 94,90% |
| <u>Níveis de Proficiência em Língua Portuguesa em 2009</u> | | | | |
| Abaixo do Básico | 37,60% | 38,00% | 36,48% | 94,20% |
| Básico | 31,60% | 38,00% | 31,81% | 94,20% |
| Adequado | 22,70% | 38,00% | 23,27% | 94,20% |
| Avançado | 8,20% | 38,00% | 8,43% | 94,20% |
| <u>Níveis de Proficiência em Matemática em 2009</u> | | | | |
| Abaixo do Básico | 48,30% | 38,20% | 47,29% | 94,90% |
| Básico | 34,90% | 38,20% | 35,47% | 94,90% |
| Adequado | 12,50% | 38,20% | 12,86% | 94,90% |
| Avançado | 4,30% | 38,20% | 4,37% | 94,90% |
| Nº de Alunos | | 70367 | | 25777 |
| Nº de Escolas | | 535 | | 519 |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2009 e 2010.

A primeira coluna da tabela 8 caracteriza os alunos da 4ª série da rede municipal de São Paulo, permitindo obter diversas informações sobre o seu perfil. Dentre elas, observa-se que há uma parcela um pouco menor de meninas que de meninos (47%), pouco mais da metade fez pré-escola (56,3%), e que a grande maioria está na idade adequada⁵ para a 4ª série (91,3% com idade até 11 anos), tem pais cuja escolaridade é menor ou igual ao ensino fundamental completo (71,6% dos pais e 67,1% das mães) e possui renda familiar de até R\$ 1275,00 (73,9%). Em 2010, a maioria dos alunos encontrava-se no nível “Básico” em Língua Portuguesa

⁵ Como “idade adequada” para a 4ª série considerou-se as idades de 10 e 11 anos, seguindo os critérios utilizados pelo INEP. Para o cálculo das taxas de distorção idade-série, o INEP considera “atrasados” os alunos com 2 anos a mais da idade adequada para a matrícula – nesse caso, 10 anos.

(42%) e no nível “Abaixo do Básico” em Matemática (46,2%). Do percentual de alunos da 4ª série em 2010 para os quais foram encontradas notas na edição de 2009 da Prova SP, a maioria deles encontrava-se “Abaixo do Básico” em Língua Portuguesa e Matemática (37,6% e 48,3%, respectivamente).

A composição da Amostra Principal, por sua vez, depende especialmente da composição da amostra submetida à Prova SP em 2009, já que nem todos os alunos testados na 4ª série em 2010 foram testados na 3ª série em 2009. Como o desenho da aplicação da Prova SP para a 3ª série inclui os alunos participantes do PIC de forma censitária e os demais de forma amostral, é de se esperar que os alunos do PIC estejam sobre-representados na Amostra Principal, em comparação com a Amostra Geral. E é de fato o que ocorre, pois a proporção de alunos do PIC é de 6,6% em relação à Amostra Geral e de 13,87% em relação à Amostra Principal.

Essa diferença se reflete na distribuição dos alunos de acordo com os níveis de proficiência, com maiores proporções de alunos no nível “Abaixo do Básico” na 4ª série na Amostra Principal que na Amostra Geral. Dado que o PIC é um programa para alunos que não foram alfabetizados na idade correta, maiores proporções de alunos do programa na amostra fazem com que a distribuição de notas seja deslocada para baixo na Amostra Principal.

No entanto, além da diferença na distribuição das proficiências dos alunos em Língua Portuguesa e em Matemática, não é possível dizer que há diferenças significativas entre o perfil dos alunos da Amostra Geral e o perfil da Amostra Principal. O que se observa são pequenas variações – em sua maioria de 1 a 3 pontos percentuais – entre as categorias das características, mantendo a forma da sua distribuição. Essas são evidências de que não se tratam de grupos de alunos com perfis distintos.

A redução do conjunto de alunos a um grupo menor para as estimativas criou também um subgrupo de professores – aqueles que lecionaram para os alunos da Amostra Principal – além do grupo com todos os professores que lecionaram na 4ª série em 2010. Os perfis dos dois grupos de professores encontram-se na tabela 9.

Tabela 9 – Estatística Descritiva dos Grupos de Professores da 4ª Série em 2010

| | Todos os Professores | | Professores da Amostra Principal | |
|--|----------------------|----------------|----------------------------------|----------------|
| | % de professores | % de respostas | % de professores | % de respostas |
| Gênero Feminino | 95,70% | 100,00% | 95,70% | 100,00% |
| Branco ou Amarelo | 69,90% | 87,10% | 70,10% | 87,30% |
| <u>Idade</u> | | | | |
| Até 25 anos | 2,30% | 100,00% | 2,30% | 100,00% |
| 26 a 35 anos | 17,40% | 100,00% | 17,50% | 100,00% |
| 36 a 45 anos | 39,20% | 100,00% | 38,90% | 100,00% |
| 46 a 55 anos | 28,60% | 100,00% | 28,50% | 100,00% |
| Mais de 55 anos | 12,60% | 100,00% | 12,90% | 100,00% |
| <u>Nível de Escolaridade</u> | | | | |
| Superior | 75,80% | 87,70% | 75,60% | 87,80% |
| Especialização | 23,40% | 87,70% | 23,70% | 87,80% |
| Mestrado | 0,80% | 87,70% | 0,80% | 87,80% |
| Doutorado | 0,10% | 87,70% | 0,00% | 87,80% |
| <u>Tempo de Experiência como Professor</u> | | | | |
| Até 5 anos | 6,80% | 88,50% | 6,70% | 88,70% |
| 6 a 10 anos | 12,00% | 88,50% | 11,90% | 88,70% |
| 11 a 15 anos | 17,20% | 88,50% | 17,20% | 88,70% |
| 16 a 20 anos | 21,80% | 88,50% | 22,00% | 88,70% |
| Mais de 20 anos | 42,20% | 88,50% | 42,10% | 88,70% |
| <u>Renda Familiar</u> | | | | |
| R\$ 1000,00 a R\$ 2000,00 | 2,10% | 85,50% | 2,10% | 85,60% |
| R\$ 2001,00 a R\$ 3000,00 | 15,60% | 85,50% | 15,20% | 85,60% |
| R\$ 3001,00 a R\$ 4000,00 | 28,10% | 85,50% | 28,60% | 85,60% |
| R\$ 4001,00 a R\$ 5000,00 | 25,30% | 85,50% | 24,90% | 85,60% |
| Mais de R\$ 5000,00 | 29,00% | 85,50% | 29,20% | 85,60% |
| <u>Escolaridade do Pai</u> | | | | |
| Não estudou | 5,80% | 86,90% | 5,90% | 87,00% |
| Anos Iniciais do EF Incompletos | 24,70% | 86,90% | 24,40% | 87,00% |
| Anos Iniciais do EF Completos | 37,10% | 86,90% | 37,10% | 87,00% |
| Ensino Fundamental Completo | 12,50% | 86,90% | 12,40% | 87,00% |
| Ensino Médio Completo | 13,30% | 86,90% | 13,30% | 87,00% |
| Ensino Superior Completo | 6,60% | 86,90% | 6,90% | 87,00% |
| Nº de Professores | | 2209 | | 2055 |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2009 e 2010.

O perfil do grupo completo de professores que lecionaram na 4ª série em 2010 é apresentado na primeira coluna. Dentre esses professores, quase todos são mulheres (95,7%), a maioria se considera branca ou amarela (69,9%), tem nível superior sem pós graduação (75,8%), tem mais de 15 anos de experiência como

professor (64%), tem renda familiar maior que R\$ 3000,00 reais (84,4%) e tem pai que completou no máximo a 4ª série do ensino fundamental completo (69,5%).

Apesar do grupo de alunos da Amostra Principal ter menos da metade do número de alunos da Amostra Geral, o mesmo não ocorre com os professores, dado que a amostra de alunos selecionados para realizar os testes nas séries ímpares está espalhada pelas turmas. Assim, os professores que lecionaram para os alunos da Amostra de Análise representam a grande maioria dos professores da 4ª série em 2010, cerca de 93% deles.

Ao comparar o perfil do grupo com todos os professores com o perfil dos professores dos alunos da Amostra Principal, pelo menos em relação às características que constam do banco de dados da Secretaria Municipal e às questões respondidas no questionário da Prova SP, os dados indicam não haver diferenças entre os perfis desses dois grupos de professores.

5. ESTRATÉGIAS EMPÍRICAS

O objetivo deste trabalho é estimar e analisar a eficácia dos professores da rede municipal de ensino de São Paulo, com base nos resultados dos seus respectivos alunos na Prova São Paulo.

A construção dos modelos aplicados para estimar os efeitos professor e para analisar os fatores associados a eles, bem como as estratégias utilizadas para verificar sua capacidade de gerar informações que sejam úteis para analisar a eficácia dos professores, são descritas nesta seção.

5.1. Do Modelo Cumulativo de Aprendizagem a Modelos de Valor Agregado

Vários modelos têm sido utilizados para representar e estimar a função de produção educacional. Como demonstrado por diversos autores como Boardman e Murnane (1979), Hanushek (1979), Todd e Wolpin (2003), Harris e Sass (2006) e Ishii e Rivkin (2009), esses modelos podem ser derivados de um modelo cumulativo de aprendizagem. Para chegar aos modelos de valor agregado, optou-se por seguir esses autores e partir dos modelos cumulativos de aprendizagem, os quais auxiliam a compreensão das relações entre os insumos e os resultados educacionais dos alunos.

O modelo cumulativo de aprendizagem pode ser encontrado em uma forma bem geral a seguir:

$$A_{it} = A_t[X_i(t), \mu_{i0}, \varepsilon_{it}] \quad (1)$$

Onde:

A_{it} é o resultado do aluno i no ano t ,

$X_i(t)$ é um vetor de características variáveis com o tempo que contém toda a história dos insumos escolares, familiares e pessoais do aluno i ;

μ_{i0} é o termo referente à dotação inicial do aluno i , ou seja, por características do aluno que são fixas no tempo, contendo tanto as observadas, como o gênero e a raça/cor, como as não observadas, como sua habilidade nata;

ε_{it} é o termo do erro aleatório.

O modelo é chamado de cumulativo pois pressupõe que a relação de produção educacional é cumulativa, tendo os insumos passados algum efeito duradouro sobre o desempenho dos alunos.

Hanushek (1979) afirma que, sem pressupostos extremamente fortes sobre a dinâmica da educação, a quantidade de dados exigidos para a estimação dessa equação é enorme, o que inviabiliza sua estimação diretamente. Boardman e Murnane (1979) fazem afirmação semelhante e apresentam mais dois argumentos sobre a impossibilidade de estimar essa equação: a multicolinearidade, mesmo com dados disponíveis, e o fato de que a habilidade nata dos indivíduos não é observável.

Assim sendo, uma série de pressupostos deve ser empregada para permitir a estimação de modelos de valor agregado, de modo a reduzir a exigência de dados.

Boardman e Murnane (1979) entendem que, em modelos de determinantes do desempenho, o tempo indica uma fase em particular da vida da criança que tem um sentido próprio, de forma que se espera que determinadas influências sejam mais importantes para as crianças em uma idade que em outra.

Eles assumem o pressuposto que o efeito do ambiente em qualquer período subsequente a t , $t + n$, será $\theta^n \alpha_t$. Esse pressuposto permite que o impacto inicial do insumo seja diferente em cada período. Porém, assume que a taxa de variação é a mesma em cada período. Em outras palavras, o impacto do insumo em um aluno com 7 anos sobre o desempenho nos anos subsequentes deve diminuir/aumentar – na mesma taxa que o impacto do insumo em um aluno de 9 anos diminui/aumenta.

Ishii e Rivkin (2009) afirmam que é bem provável que θ vá de 0 a 1: indo de nenhum efeito das experiências passadas até a persistência dos efeitos completos no futuro. No entanto, não foi encontrado nenhum estudo que apresente alguma restrição para o valor de θ . Embora seja pouco provável, é possível que

determinado insumo aplicado em um ponto do tempo tenha efeitos maiores a cada ano que passa sobre o desempenho dos alunos ou até mesmo que afete negativamente o desempenho deles, prejudicando seu aprendizado. Portanto, não será assumida nenhuma restrição com relação a θ , o qual será verificado empiricamente.

Isto posto, a equação pode ser escrita como segue:

$$A_{it} = \alpha_t X_{it} + \theta \alpha_{t-1} X_{it-1} + \dots + \theta^{t-1} \alpha_1 X_{i1} + \psi_t \mu_{i0} + \varepsilon_{it} \quad (2)$$

Assim sendo, a equação que descreve o resultado no período imediatamente anterior será:

$$A_{it-1} = \alpha_{t-1} X_{it-1} + \theta \alpha_{t-2} X_{it-2} + \dots + \theta^{t-2} \alpha_1 X_{i1} + \psi_{t-1} \mu_{i0} + \varepsilon_{it-1} \quad (3)$$

A idéia dos modelos de valor agregado é justamente utilizar as notas dos alunos em dois anos consecutivos para analisar os fatores que explicam o valor que foi agregado entre os anos. Portanto, subtraindo o resultado do aluno em dado período pelo resultado do aluno no período imediatamente anterior, multiplicado pela “taxa de variação do impacto” dos insumos de um ano para o outro, θ , teremos:

$$\begin{aligned} A_{it} - \theta A_{it-1} &= \alpha_t X_{it} + \theta \alpha_{t-1} X_{it-1} + \dots + \theta^{t-1} \alpha_1 X_{i1} \\ &+ \psi_t \mu_{i0} + \varepsilon_{it} - \theta \alpha_{t-1} X_{it-1} - \dots - \theta^{t-1} \alpha_1 X_{i1} - \theta \psi_{t-1} \mu_{i0} - \theta \varepsilon_{it-1} \end{aligned} \quad (4)$$

Simplificando e reorganizando, tem-se:

$$A_{it} - \theta A_{it-1} = \alpha_t X_{it} + (\psi_t - \theta \psi_{t-1}) \mu_{i0} + (\varepsilon_{it} - \theta \varepsilon_{it-1}) \quad (5)$$

Harris e Sass (2006) assumem que o impacto da dotação inicial sobre o desempenho, ψ_t , muda a uma taxa constante, então $(\psi_t - \theta \psi_{t-1})$ pode ser expresso como uma constante, ϖ . Com isso, pode-se reescrever a equação:

$$A_{it} - \theta A_{it-1} = \alpha_t X_{it} + \gamma_i + \eta_i \quad (6)$$

Onde $\gamma_i = \varpi \mu_{i0}$ é o efeito individual do aluno e $\eta_i = \varepsilon_{it} - \theta \varepsilon_{it-1}$ é o erro aleatório.

A partir da equação (6) são obtidos os dois modelos de valor agregado amplamente utilizados pela literatura, denominados de modelo de ajuste de covariadas e modelo de ganho de notas. Esses modelos são utilizados tanto em *cross-sections* com apenas 2 resultados de desempenho para cada aluno, quanto

em painéis a partir de dados longitudinais com uma quantidade maior de resultados para cada aluno.

Para obter o modelo de ajuste de covariadas, basta passar θA_{it-1} para o lado direito da equação, ficando:

$$A_{it} = \theta A_{it-1} + \alpha_t X_{it} + \gamma_i + \eta_i \quad (7)$$

Dado o pressuposto de taxa geométrica de variação do impacto entre os períodos, o que viabiliza a estimação das equações é que a variável que corresponde ao desempenho no período anterior seria suficiente para substituir todos os insumos passados. Assim, o desempenho atual seria uma função dos insumos atuais do aluno, além do desempenho prévio e um efeito individual específico, mais os erros aleatórios.

Já o modelo de ganho de notas pode ser compreendido como um caso específico de (7) em que se assume que $\theta = 1$, ficando:

$$A_{it} - A_{it-1} = \alpha_t X_{it} + \gamma_i + \eta_i \quad (8)$$

Nesse caso, em lugar de assumir uma taxa constante de variação do impacto dos insumos sobre o desempenho do aluno, assume-se que não há queda ou aumento no efeito dos insumos passados sobre o desempenho corrente. Como notado por Boardman e Murnane (1979), Todd e Wolpin (2003) e Harris e Sass (2006), isso implica que o efeito de cada insumo deve ser independente de qual momento é aplicado. Em outras palavras, cada insumo tem um impacto imediato sobre o desempenho que não diminui com o tempo. Por exemplo, o fato de ter estudado na pré-escola deve ter o mesmo impacto sobre o desempenho de um aluno quando tem 8 anos e quando tem 9, 10, 11 anos.

Na discussão sobre o uso dos modelos de ajuste de covariadas e de ganho de notas o argumento mais citado contra o segundo é que ele é mais restritivo que o primeiro justamente por assumir que o coeficiente do desempenho anterior sobre o desempenho atual é 1, não havendo variação nos efeitos dos insumos passados sobre o desempenho dos alunos (Boardman e Murnane, 1979; Ishii e Rivkin, 2009). Portanto, como ressaltam esses autores, quanto mais distante de 1 for o coeficiente, maior será essa variação e mais viesada será a estimativa.

McCaffrey et al (2003) chamam a atenção para o apelo intuitivo do modelo de ajuste de covariadas, pois este pode ser interpretado como se todos os

alunos estivessem no mesmo nível inicial de aprendizagem. Essa interpretação é importante não só pela questão intuitiva, mas especialmente porque pode haver diferenças entre os ganhos de nota de acordo com o nível de desempenho prévio dos alunos, diferenças essas que não se devem aos insumos aplicados no período. Por exemplo, é possível que seja mais fácil produzir maiores ganhos de notas com alunos que tenham notas prévias mais baixas, e que em níveis mais elevados de desempenho seja mais difícil produzir melhorias. O modelo de ganho de notas não permitirá uma representação adequada nesse caso, pois somente considerará as diferenças de notas, sem considerar as diferenças nos níveis prévios de desempenho.

Outra vantagem do modelo de ajuste de covariadas é que ele viabiliza a estimação da equação mesmo que as duas medidas de desempenho estejam em diferentes escalas, o que não é válido para o modelo de ganho de notas, onde uma nota tem que ser subtraída da outra (Hanushek, 1979; McCaffrey et al, 2003).

5.2. Especificação do Modelo para Estimar os Efeitos Professor

Dadas as vantagens indicadas em relação ao modelo de ajuste de covariadas, partiremos dele para detalhar a especificação do modelo empírico que será empregado para análise dos dados da Prova São Paulo, a começar pela modelagem do próprio efeito professor.

O interesse deste trabalho está no valor que cada professor individual agregou, e não apenas os efeitos totais do grupo de professores sobre os alunos em questão. Assim sendo, como os estudos encontrados com o mesmo propósito, trabalharemos com uma variável *dummy* para cada professor, indicando se o aluno foi ou não submetido àquele professor específico (Goldhaber e Hansen, 2010; Koedel e Betts, 2011; Rothstein, 2009; Ishii e Rivkin, 2009; Aaronson, Barrow e Sander, 2007).

Em se tratando das variáveis de controle, uma das maiores discussões diz respeito aos efeitos da escola. A grande maioria das análises de efeito professor encontradas na literatura opta por trabalhar com *dummies* de escola – com exceção de casos como Kane e Staiger (2008), que comparam dados experimentais com não

experimentais dos mesmos professores. A inclusão do efeito fixo da escola é defendida em razão da hipótese de que a alocação dos alunos nas escolas seja feita de modo mais dependente de características dos alunos e de suas famílias – e, portanto, menos aleatória – do que a alocação dos alunos aos professores dentro das escolas. Isso se daria tanto pelas condições socioeconômicas das famílias, que determinam onde elas moram e isso determina a alocação à escola, quanto em seu interesse e esforço para influenciar a definição da escola onde seus filhos vão estudar.

Quando os professores da amostra lecionam em apenas uma escola, a inclusão de efeito fixo da escola faz com que os efeitos professor variem somente dentro de cada escola. Embora o uso de efeito fixo da escola ajude a controlar a influência de todas as variáveis relativas a cada escola específica, pode esconder as diferenças entre os níveis de professores de cada escola: é possível que algumas escolas tenham professores muito melhores, em média, que outras escolas, o que não fica evidente quando se compara os professores dentro de cada escola. E mais: caso só haja um professor daquela série naquele ano em uma escola, não será possível a inclusão de efeitos fixos da escola.

Apesar dessas desvantagens, seguiremos a literatura ao inserir uma variável *dummy* para cada escola, indicando se o aluno estudou ou não naquela escola específica naquele ano. Com isso, assumiremos que essa variável capta todos os insumos escolares que não estão relacionados aos professores e classes, tais como condições físicas da escola, qualidade do diretor, projeto pedagógico da escola, entre outros. E também capta todos os insumos relacionados aos processos que ocorrem na classe, mas que se espera que sejam iguais para a escola como um todo, tais como os materiais didáticos, por exemplo.

Dentro da classe, além dos efeitos do professor e os efeitos da escola que não variam de acordo com as classes, os fatores referentes à turma à qual o aluno pertence também influenciam o seu resultado ao final do período. Apesar de ser um fator cujo impacto sobre a aprendizagem é muito debatido, o tamanho da turma foi encontrado nas especificações apenas dos estudos de Clotfelter, Ladd e Vigdor (2007) e Harris e Sass (2006). Tanto eles, como Rockoff (2004), Ishii e Rivkin (2009) e Aaronson, Barrow e Sander (2007) incluem em seus modelos características observáveis dos pares do aluno na turma daquele ano. Essas características dependem das variáveis que estiverem disponíveis para os

pesquisadores nas bases de dados. Geralmente não são muitas, compreendendo cor/raça, gênero e se recebem refeição de graça na escola, entre outras.

Da mesma forma, a limitação de dados pode inviabilizar a inclusão de variáveis referentes à composição das turmas. Como ressaltam Hanushek e Rivkin (2010), com professores sendo observados apenas em uma classe em cada ano, a inclusão de efeitos fixos de professor por ano necessariamente impede qualquer outra variável explicativa que não varie dentro da classe. É o que ocorre em alguns estudos, como o de Goldhaber e Hansen (2010) e Clotfelter, Ladd e Vigdor (2006), por exemplo, que trabalham com um ano de estimativas e não incluem variáveis relacionadas às turmas. A desvantagem, nesses casos, é que a influência dos fatores de pares que não puderam ser incluídos fica misturada com a qualidade do professor estimada pelo efeito fixo do professor.

A amostra utilizada neste trabalho é um exemplo da situação citada por Hanushek e Rivkin (2010), na qual quase todos os professores são observados com apenas uma classe no ano. Assim sendo, neste modelo não serão incluídas variáveis explicativas da classe. Não conseguir observar o mesmo professor com turmas diferentes faz com que calculemos algo que pode ser denominado “efeito classe”, que inclui o efeito do professor, e também inclui os efeitos da turma e da interação entre o professor e a turma. Assim sendo, um dos desafios deste trabalho é buscar distinguir o efeito do professor desses demais fatores.

Para captar os efeitos de cada aluno, as análises geralmente incluem variáveis fixas do aluno, como por exemplo gênero e cor/raça, mas também variáveis que se alteram com o tempo, como fatores socioeconômicos e repetência (Clotfelter, Ladd e Vigdor, 2006; Rockoff, 2004). Trabalharemos com uma especificação semelhante a essa mais comumente encontrada na literatura, com a inclusão de características fixas e variáveis dos alunos que possam influenciar o seu resultado no teste.

Existe uma série de modelos que trabalham com o efeito fixo dos alunos. São modelos que utilizam resultados prévios do aluno para modelar o crescimento do seu desempenho ao longo do tempo, em lugar de adotar pressupostos sobre a natureza desse crescimento. McCaffrey et al (2003) chamam esses modelos de multivariados, pois modelam simultaneamente a distribuição conjunta de todas as medidas, não trabalhando com equações separadas para cada par de medidas, como os demais modelos.

Os autores explicam que nesses modelos os efeitos professor são desvios da tendência de crescimento do aluno, que podem ser considerados persistentes ou não – se acumulam ou não na equação das séries posteriores.

Eles indicam que o Tennessee Value Added Assessment System (TVAAS) é um exemplo bem difundido de aplicação de um modelo multivariado, o modelo em camadas (*layered model*). Na aplicação do TVAAS para a avaliação de professores do estado americano do Tennessee, os efeitos professor são relativos ao professor médio do distrito, não incluindo efeitos da escola. Os professores afetam o desempenho dos alunos quando eles estão em suas classes, e o efeito persiste sem diminuição em todos os anos futuros de testes. Como os demais modelos multivariados, o TVAAS exige pelo menos 3 anos consecutivos de dados dos mesmos alunos para poder modelar o crescimento de sua aprendizagem. O Tennessee utiliza uma especificação que modela conjuntamente as notas de 5 áreas do conhecimento para as quais são aplicados os testes em alunos de várias coortes.

Essa é a principal desvantagem dos modelos multivariados e razão pela qual não serão aplicados neste trabalho: a exigência de dados que acompanhem um número significativo de alunos por diversos anos, de modo que se tenha uma série de medidas de resultado para o mesmo aluno, com variações nos demais fatores que afetam seu desempenho. Isso porque, por outro lado, como ressaltam McCaffrey et al (2003), os modelos multivariados apresentam uma série de vantagens em relação aos modelos que trabalham com apenas duas medidas de desempenho dos alunos: não são sensíveis ao ano escolhido como inicial; permitem acomodar dados parciais de alunos, sem ter que descartá-los; permitem explorar uma variedade de pressupostos sobre a persistência do efeito professor e da covariância residual dos resultados dos alunos e permitem obter estimativas mais eficientes e robustas.

Como os dados que possuímos não permitem acompanhar uma quantidade expressiva de alunos por três anos consecutivos, não será possível fazer análises que se baseiem na variação dos resultados de um mesmo aluno ao longo do tempo, o que caracteriza modelos multivariados.

Assim como grande parte dos trabalhos pesquisados, buscaremos estimar os efeitos professor utilizando o modelo de ajuste de covariadas, com a seguinte especificação, detalhada a partir da equação (8):

$$A_{ijst} = \theta A_{it-1} + \beta \gamma_i + \alpha L_{it} + \tau T_{ijt} + \delta S_{ist} + \eta_i \quad (9)$$

A equação (9) descreve o desempenho do aluno i ensinado pelo professor j na escola s no ano t , em função das seguintes variáveis:

A_{it-1} , que representa o resultado do aluno i no ano $t-1$;

γ_i , que representa um vetor de características fixas do aluno i que inclui o gênero e se o aluno frequentou pré-escola;

L_{it} , que representa um vetor de características variáveis do aluno i no ano t que inclui um indicador se ele possui a idade adequada para a série, se participa do PIC, variáveis de escolaridade dos pais e variáveis de renda familiar;

T_{ijt} , que representa um vetor de *dummies* que identificam o professor j que lecionou para o aluno i no ano t ;

S_{ist} , que representa um vetor de *dummies* que identificam a escola s na qual o aluno i estudou no ano t ;

η_i , que representa o erro aleatório.

O efeito estimado para cada professor j no ano t será capturado pelo coeficiente da *dummy* referente a esse professor no modelo.

As características fixas e variáveis referentes aos alunos incluídas no modelo puderam ser construídas a partir dos dados disponíveis nos bancos da Secretaria Municipal ou das respostas dos professores ao questionário aplicado na ocasião da aplicação da Prova São Paulo. Uma descrição dessas variáveis pode ser encontrada no Apêndice A.

5.3. Análise de Fatores Associados aos Efeitos Professor

Um dos principais motivos para se estimar os efeitos dos professores em modelos de valor agregado é a possibilidade de buscar analisar se as variáveis que caracterizam os professores e estão contidas nos bancos de dados disponíveis são capazes de explicar alguma parcela desses efeitos. Entender quais são os fatores

associados aos efeitos professor é de grande valia para as políticas públicas voltadas para o magistério, e por essa razão tem sido dada tamanha atenção a esse tipo de análise pela literatura sobre o tema.

A intenção é regredir os efeitos professor estimados para cada professor no primeiro estágio $\hat{\tau}_j$ em uma série de características aqui representadas por Z :

$$\hat{\tau}_j = \phi Z_j + u_j \quad (10)$$

Porém, como explicam Goldhaber e Hansen (2010), estimações de mínimos quadrados ordinários dessa relação geram erros-padrão muito pequenos pois ignoram o erro na variável dependente, o efeito professor, originado na estimação anterior.

Assim sendo, utilizaremos o mesmo procedimento que Aaronson, Barrow e Sander (2007), Koedel e Betts (2007) e Goldhaber e Hansen (2010), os quais se basearam em Borjas (1987) e Borjas e Sueyoshi (1994) ao utilizar uma abordagem de mínimos quadrados generalizados, levando em conta a incerteza da variável dependente ao ponderar as observações de acordo com a proporção da confiabilidade de cada efeito professor individual estimado.

Com base nessa estratégia, a matriz de variância-covariância adequada para ponderar a regressão de mínimos quadrados generalizados, segundo esses autores, é dada por:

$$\Omega = \hat{\sigma}_u^2 I_j + \hat{V} \quad (11)$$

Onde a primeira porção corresponde à matriz de covariância derivada da estimativa de mínimos quadrados ordinários da equação (10) e a segunda porção corresponde à matriz cujos elementos são da diagonal da matriz de variância-covariância dos erros da equação (9) relativa às estimativas dos coeficientes dos efeitos professor. \hat{V} é obtido a partir da matriz de variância dos erros dos coeficientes dos efeitos professor, enquanto $\hat{\sigma}_u^2$ é obtido seguindo os procedimentos descritos por Borjas e Sueyoshi (1994).

Segundo Borjas e Sueyoshi (1994), o método dos mínimos quadrados ordinários aplicado à regressão de segundo estágio gera estimativas da variância do erro que contêm componentes de u/σ_η e do erro de estimação associado ao $\hat{\tau}$.

Se \tilde{u}_j é o resíduo estimado da regressão de mínimos quadrados ordinários de segundo estágio, então a estimativa da variância residual é:

$\tilde{\sigma}^2 = \sum_{j=1}^J \tilde{u}_j^2 / (J - M)$, sendo M o número de variáveis explicativas dos efeitos professor testadas. $\hat{\sigma}_u^2$ pode ser estimado por $\hat{\sigma}_u^2 = \tilde{\sigma}^2 - \sum_{j=1}^J \hat{\sigma}_j^2 / (J - M)$,

onde $\hat{\sigma}_j$ é o erro padrão da variável *dummy* j do procedimento de primeiro estágio.

Tendo calculado os pesos para ponderar a regressão de mínimos quadrados generalizados, foi possível estimar a relação entre os efeitos professor estimados para as duas amostras referentes às duas disciplinas, e uma série de variáveis que puderam ser construídas a partir dos dados disponíveis nos bancos da Secretaria Municipal ou das respostas dos professores ao questionário aplicado na ocasião da aplicação da Prova São Paulo.

Em se tratando das variáveis explicativas, observa-se que nos artigos revisados que realizam esse tipo de análise, todos com base em dados de estados americanos, é comum a inclusão de variáveis referentes ao nível de escolaridade, área de formação, tempo de experiência, possuir algum tipo de certificação, as notas em testes de certificação e algum indicador de qualidade relativo à instituição de ensino superior na qual se graduou. Para este trabalho, somente estavam disponíveis as respostas dos professores às questões sobre seu nível de escolaridade e seu tempo de experiência como professor, as quais foram incluídas no modelo. Assim como nos modelos desses artigos, também foram incluídas variáveis referentes ao gênero e à cor/raça dos professores como controles.

O questionário respondido pelos professores na ocasião da Prova São Paulo de 2010 pode ser considerado extenso: possui quase 200 itens, abordando temas que vão desde características socioeconômicas dos professores até sua opinião sobre o ambiente escolar, passando por tipos de atividades desenvolvidas e recursos utilizados.

Na escolha das questões para compor o modelo de fatores associados aos efeitos professor, buscou-se priorizar questões relacionadas diretamente ao professor e seu trabalho, e excluindo-se as relacionadas a terceiros, que também podem influenciar indiretamente o trabalho docente, mas não são o foco desta análise. Adicionalmente, foram priorizadas as questões sobre atitudes e práticas que

tenham um caráter que pôde ser considerado minimamente objetivo, excluindo-se os itens sobre opiniões e percepções dos professores.

Em relação às práticas dos professores, optou-se por trabalhar apenas com atividades e opções pedagógicas mais simples e gerais, como é o caso do uso de computadores e da frequência com que pede lição de casa. Entende-se que a utilização dos itens referentes às estratégias específicas de Língua Portuguesa e Matemática exige conhecimentos específicos sobre o ensino dessas disciplinas para a construção de variáveis que realmente tenham sentido pedagógico.

Não foram incluídos na análise fatores para os quais praticamente todas as respostas foram iguais, como por exemplo sobre livros didáticos, que praticamente todos os professores responderam utilizar. Não havendo variação no item, não haveria como analisar a relação entre esse fator e os efeitos professor.

A descrição das variáveis que compõem Z e foram utilizadas para estimar a equação (10) está contida no Apêndice B.

6. RESULTADOS

Os resultados das estratégias empíricas empregadas com base nos resultados dos alunos da 4ª série da rede municipal de ensino de São Paulo na Prova SP são apresentados nesta seção, permitindo as devidas análises acerca da eficácia dos professores dessa amostra.

6.1. Efeitos do Conjunto de Professores

Para identificar a proporção da variação nas notas dos alunos que se deve à variação dos efeitos professor, de modo a analisar a importância dos professores relativamente a outros fatores que também influenciam o desempenho dos alunos, foi estimada a equação (9) e mais quatro modelos que nada mais são que variações da equação (9) nos quais uma ou mais variáveis são excluídas do modelo.

O primeiro modelo é baseado em uma especificação com apenas as características de alunos e famílias; o segundo adiciona a nota prévia dos alunos; o terceiro adiciona as *dummies* de professores; o quarto adiciona as *dummies* de escolas – trata-se do modelo referente à equação (9) completa; e o quinto exclui as *dummies* de professores mas mantém todas as demais.

A tabela 10 apresenta os valores de R^2 para cinco modelos que buscam explicar o desempenho dos alunos em Língua Portuguesa e Matemática⁶.

⁶ Os resultados completos das regressões referentes às especificações 1, 2, 3 e 5 podem ser encontrados no Apêndice C. Os resultados completos referentes à especificação 4, o modelo escolhido para análise, serão apresentados na tabela 11.

Tabela 10 – Capacidade Explicativa dos Modelos de Acordo com as Especificações

| | Especificações do Modelo | | | | |
|--------------------------------------|--------------------------|--------|--------|--------|--------|
| | 1 | 2 | 3 | 4 | 5 |
| Características de Alunos e Famílias | sim | sim | sim | sim | sim |
| Nota no Ano Anterior | não | sim | sim | sim | sim |
| Dummies de Professores | não | não | sim | sim | não |
| Dummies de Escolas | não | não | não | sim | sim |
| <u>Língua Portuguesa</u> | | | | | |
| R ² | 0,1609 | 0,5210 | 0,6629 | 0,6653 | 0,5744 |
| Nº de Alunos | 9425 | 9425 | 9425 | 9425 | 9425 |
| Nº de Professores | 2055 | 2055 | 2055 | 2055 | 2055 |
| <u>Matemática</u> | | | | | |
| R ² | 0,1389 | 0,5302 | 0,6750 | 0,6770 | 0,5874 |
| Nº de Alunos | 9567 | 9567 | 9567 | 9567 | 9567 |
| Nº de Professores | 2055 | 2055 | 2055 | 2055 | 2055 |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2009 e 2010.

Os R² obtidos a partir dos modelos aplicados aos resultados dos alunos em Língua Portuguesa e Matemática apresentam padrões semelhantes, como indica a tabela 10. A maior diferença entre os resultados dos alunos nas duas disciplinas foi encontrada na especificação 1, onde as características de alunos e famílias têm maior poder explicativo das notas de Língua Portuguesa que de Matemática. A explicação recorrente na literatura de função de produção educacional para resultados que indicam que o ambiente em que o aluno se encontra teria maior influência em sua habilidade de leitura seria porque essa habilidade seria mais exercida no ambiente externo que sua habilidade de resolver problemas matemáticos – algo que dependeria mais dos fatores escolares para se desenvolver.

A considerável semelhança entre os resultados permite com que a análise dos R² obtidos para as diferentes especificações seja feita de forma conjunta para ambas as disciplinas.

O R² da especificação 1 indica o limitado poder explicativo de especificações que não incluem a nota prévia dos alunos como controle. Embora ainda pudessem ser agregadas ao modelo variáveis que caracterizam professores e escolas, como é prática comum nos estudos com base em especificações contemporâneas – aquelas que não incluem a nota prévia do aluno –, o poder explicativo agregado por essas variáveis está bem longe do agregado pela inclusão da nota prévia dos alunos ao modelo.

O R^2 da especificação 2 demonstra que a nota prévia é o fator cuja inclusão mais agrega poder explicativo à análise. Esse resultado faz todo o sentido ao interpretarmos a variável relativa à nota prévia como aquela que captura o histórico de fatores que influenciaram o desempenho do aluno até o ano anterior, com base no modelo cumulativo de aprendizagem.

Já a comparação dos R^2 da especificação 2 para a especificação 3 indicaria o quanto da variação do resultado dos alunos nos testes as diferenças de eficácia dos professores são capazes de explicar, como está descrito na literatura revisada por Nye, Konstantopoulos e Hedges (2004). Nesses artigos revisados, os modelos utilizados como equivalentes a esta especificação 2 incluem variáveis de alunos e de famílias, e alguns ainda incluem variáveis que caracterizam as escolas.

As diferenças entre os R^2 das duas especificações obtidas foram de cerca de 0,14 em ambas as disciplinas, valor compreendido dentro da amplitude de resultados encontrados nos estudos revisados por Nye, Konstantopoulos e Hedges (2004), que apresentaram diferenças de 0,07 a 0,21.

Na especificação 4, são incluídas as *dummies* de escola para permitir a separação dos efeitos dos professores dos efeitos das escolas nas quais eles lecionam, chegando ao modelo principal deste estudo, o qual será utilizado para as análises posteriores. Como se pode observar, a inclusão das variáveis das escolas quase não adiciona poder explicativo ao modelo. Já que é improvável que as escolas tenham um papel tão reduzido na explicação dos resultados dos alunos nos testes, uma possibilidade é que os professores que constavam da especificação 3 já estavam capturando parte dos efeitos das escolas, já que praticamente todos os professores estão cada um em uma única escola. A especificação 5 inclui as *dummies* de escolas mas exclui as *dummies* de professores justamente para checar essa possibilidade.

De fato, ao comparar a especificação 2 com a especificação 5, percebe-se que a inclusão somente das *dummies* de escolas também agrega um significativo poder explicativo, embora menor que os professores. A diferença entre os R^2 das especificações 5 e 2 indicam que a variação da eficácia das escolas representaria cerca de 0,05 da variação nas notas dos alunos. Portanto, parece que realmente quando professores e escolas entram no modelo de forma separada, refletem o seu poder explicativo e ainda um pouco do poder explicativo do outro fator.

Essa evidência leva a crer que os resultados dos trabalhos revisados por Nye, Konstantopoulos e Hedges (2004) superestimaram a importância da variação da eficácia dos professores para explicar a variação dos resultados dos alunos, em razão de não incluírem as *dummies* de escola em seus modelos.

Portanto, a estratégia descrita por Nye, Konstantopoulos e Hedges (2004) e utilizada por esses artigos será aplicada novamente, mas agora para calcular a diferença entre os R^2 nas especificações 5 e 4, ou seja, da especificação que contém as *dummies* de escola – além de características de alunos e famílias e da nota prévia – para a especificação que inclui as *dummies* de professor.

Enquanto as estimativas das diferenças entre os R^2 entre as especificações 3 e 2, sem a inclusão das variáveis referentes às escolas, era de 0,14 para ambas disciplinas, as estimativas das diferenças de R^2 entre as especificações 4 e 5 são de aproximadamente 0,09, indicando que a variação na eficácia dos professores explicaria cerca de 9% da variação nas notas dos alunos da amostra analisada. Esse percentual é menor que o encontrado na inclusão das variáveis relativas ao *background* dos alunos, em torno de 15%, mas maior que o obtido na inclusão das variáveis de escola, em torno de 5%. Portanto, considerando essas estimativas, é possível afirmar que a variação dos efeitos dos professores dentro de cada escola teria maior capacidade explicativa da variação nas notas dos alunos da 4ª série da rede municipal de São Paulo que a variação dos efeitos das escolas, por exemplo.

6.2. Efeito Professor Individual

A análise dos efeitos professor individuais será realizada com base na estimação da equação (9), que chamamos de especificação 4 na comparação anterior. Os resultados da estimação desse modelo escolhido para explicar o desempenho dos alunos da 4ª série na Prova SP em 2010 encontram-se na tabela 11 a seguir.

Tabela 11 – Resultados da Estimação da Equação (9) com a Amostra Principal

| | Língua Portuguesa | Matemática |
|---|---------------------|---------------------|
| Nota Prévia | 0,645*** [0,009] | 0,640*** [0,008] |
| Gênero Feminino | 0,100*** [0,015] | -0,016 [0,015] |
| Idade Adequada | 0,082*** [0,027] | 0,115*** [0,026] |
| Participa do PIC | -0,044 [0,288] | 0,228 [0,256] |
| Fez Pré-Escola | -0,006 [0,015] | -0,010 [0,015] |
| <u>Escolaridade da Mãe</u> | | |
| Anos Iniciais do Ensino Fundamental Completos | 0,016 [0,026] | 0,052** [0,025] |
| Ensino Fundamental Completo | 0,044 [0,028] | 0,039 [0,027] |
| Ensino Médio Completo | 0,071*** [0,028] | 0,117*** [0,027] |
| Ensino Superior Completo | 0,158*** [0,053] | 0,132*** [0,051] |
| <u>Escolaridade do Pai</u> | | |
| Anos Iniciais do Ensino Fundamental Completos | 0,038* [0,023] | 0,037* [0,022] |
| Ensino Fundamental Completo | 0,002 [0,025] | 0,074*** [0,024] |
| Ensino Médio Completo | 0,049* [0,026] | 0,084*** [0,025] |
| Ensino Superior Completo | 0,126** [0,050] | 0,168*** [0,048] |
| <u>Renda Familiar</u> | | |
| R\$ 851,00 a R\$ 1275,00 | 0,050*** [0,018] | 0,038** [0,018] |
| R\$ 1276,00 a R\$ 2125,00 | 0,085*** [0,022] | 0,067*** [0,022] |
| R\$ 2126,00 a R\$ 4250,00 | 0,146*** [0,034] | 0,119*** [0,033] |
| Mais de R\$ 4250,00 | 0,067 [0,067] | 0,058 [0,064] |
| <u>Desvio-Padrão</u> | | |
| Dos Efeitos Professor | 0,622 | 0,621 |
| Dos Efeitos Escola | 0,500 | 0,483 |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2009 e 2010.
Nota: O R² e o número de observações dessas regressões encontram-se na tabela 13.

Como se esperava, o coeficiente da nota prévia dos alunos é positivo e estatisticamente significativo. No entanto, é bem menor que 1, o que reforça a escolha do modelo de ajuste de covariadas e não de ganho de notas. Isto porque, como já descrito, o segundo modelo, ao adotar como variável dependente o ganho de notas de um ano para o outro, implicitamente assume que o coeficiente da nota prévia do aluno é 1.

Como já observado em outras análises desse tipo, meninas apresentam melhores resultados em Língua Portuguesa. Em Matemática, porém, não é possível afirmar que haja diferenças entre meninos e meninas pois os coeficientes foram muito pequenos e não significativos estatisticamente.

O fato de o aluno estar na idade correta para série, ou seja, não estar atrasado na escola, também apresentou impacto positivo e significativo nos resultados de ambas as disciplinas.

Os coeficientes das variáveis referentes ao aluno participar do PIC e ter feito pré-escola obtidos não são estatisticamente significativos. Especialmente em relação à pré-escola, esperava-se que efeitos positivos e significativos, como os estudos sobre o tema vêm demonstrando. No entanto, no caso deste estudo específico, a falha em obter impactos positivos desse fator provavelmente decorre da perda de respostas dos alunos, pois cerca de 30% dos alunos não respondeu a esse item do questionário e 10% responderam que não sabem ou não se lembram, restando apenas 60% de alunos com respostas válidas.

Os coeficientes de todas as variáveis relativas à escolaridade do pai e da mãe do aluno são positivos, tanto em Língua Portuguesa, quanto em Matemática, e são significativos estatisticamente para a maioria dessas variáveis.

Os coeficientes das variáveis de renda familiar, por sua vez, foram todos positivos, sendo significativos para todas as faixas de renda em relação aos alunos com renda abaixo de R\$ 850,00. A única faixa cujos coeficientes não foram estatisticamente significativos foi a de maior renda: acima de R\$ 4250,00.

A inclusão de variáveis *dummy* permite que se busque capturar o efeito de cada professor individualmente, estimados pelo coeficientes $\hat{\tau}_j$, e também analisar o quanto esses efeitos variam entre os professores pesquisados, permitindo estimar os impactos de mudanças na eficácia dos professores sobre o desempenho dos alunos nos testes.

Como as proficiências dos alunos nos testes foram transformadas em variáveis padronizadas com média zero e desvio-padrão um antes da estimação das regressões, o cálculo do desvio-padrão do conjunto de efeitos estimados para cada professor individual permite com que se estime o impacto de melhorias nos efeitos-professor sobre o desempenho dos alunos.

A partir dos resultados dos alunos da 4ª série na Prova São Paulo, foram obtidos desvios-padrão dos efeitos professor de aproximadamente 0,62 em ambas as disciplinas.

Para entender o quanto isso representa em termos das notas dos alunos da 4ª série da rede municipal de São Paulo em 2010, a tabela 12 apresenta as médias e os desvios-padrão dos alunos dos grupos analisados:

Tabela 12 – Médias e Desvios-Padrão das Notas dos Alunos da 4ª série em 2010

| | Amostra Geral | Amostra de Análise |
|--------------------------|---------------|--------------------|
| <u>Língua Portuguesa</u> | | |
| Média | 168,31 | 162,21 |
| Desvio-Padrão | 44,61 | 44,61 |
| <u>Matemática</u> | | |
| Média | 179,45 | 173,78 |
| Desvio-Padrão | 44,54 | 44,67 |

Fonte: Elaboração própria com base em dados da Prova SP do ano de 2010.

A tabela 12 indica que um desvio-padrão na proficiência dos alunos equivale a aproximadamente 45 pontos na escala da Prova São Paulo. Portanto, os resultados apresentados na tabela 11 equivaleriam a afirmar que um aumento de um desvio-padrão na eficácia dos professores, estimada a partir das estratégias descritas anteriormente, levaria a um aumento de 28 pontos nas notas dos alunos, tanto em Língua Portuguesa quanto em Matemática – impactos muito grandes para um único fator escolar, mesmo em se tratando dos professores.

Os estudos empíricos com análises semelhantes revisados indicam, porém, que o simples cálculo dos desvios-padrão dos efeitos professor estimados gera medidas que superestimam os efeitos que, de fato, são causados pelos professores. As análises desenvolvidas nas próximas três subseções deste trabalho terão por objetivo identificar e quantificar fatores que possam estar viesando essas estimativas.

6.2.1. Análise do Viés Devido à Alocação Não Aleatória dos Alunos

Como discutido anteriormente neste trabalho, análises de eficácia dos professores buscam estimar o tamanho do efeito que cada professor específico causou sobre o desempenho dos alunos para os quais ele lecionou.

Como não se pode alocar alunos exatamente iguais a todos os professores, de modo a garantir condições prévias iguais para sua atuação, a recomendação é que se trabalhe com situações em que alunos sejam alocados de forma aleatória a escolas e a professores dentro das escolas (Levin e McEwan, 2002; Rubin, Stuart e Zanutto, 2004; Rothstein, 2009). No entanto, sabemos que na vida real, a distribuição de alunos a escolas e turmas raramente é aleatória, sendo influenciada por decisões de pais, diretores e dos próprios alunos.

Isto posto, buscou-se mitigar o viés devido à alocação não aleatória de alunos a escolas com a inclusão dos efeitos fixos da escola, como recomenda a literatura sobre o tema (Aaronson, Barrow e Sander, 2007; Koedel e Betts, 2007; Ishii e Rivkin, 2009; Goldhaber e Hansen, 2010).

Devido ao caráter dos dados disponíveis, não será possível adotar estratégias como a inclusão de efeitos fixos dos alunos para mitigar o viés devido à alocação não aleatória de alunos a turmas. É possível, porém, produzir uma estimativa do tamanho desse viés.

Para investigar o quanto a alocação de alunos a turmas dentro das escolas ocorreu de forma aleatória na 4ª série da rede municipal de São Paulo, foram seguidos os procedimentos utilizados por Clotfelter, Ladd e Vigdor (2006).

Foram realizados dois testes qui-quadrado em cada uma das 519 escolas da amostra para buscar evidências de que a alocação dos alunos nas turmas tenha sido feita de forma estatisticamente independente das notas prévias dos alunos, um teste com base na nota prévia em Língua Portuguesa e outro com base na nota prévia em Matemática. Assim como Clotfelter, Ladd e Vigdor (2006), optou-se por trabalhar com apenas duas categorias em relação às características testadas. Enquanto os autores dividiram os alunos entre os que obtiveram notas abaixo da média do ano anterior e acima da média do ano anterior, neste estudo optou-se por dividir os alunos entre os que obtiveram notas classificadas no nível “Abaixo do

Básico” e os que obtiveram notas classificadas nos níveis “Básico”, “Adequado” ou “Avançado”. Isto porque, a partir do desenho das amostras da Prova SP e de informações obtidas no Portal da Prefeitura, percebe-se o destaque aos alunos que se encontram “Abaixo do Básico”, indicando um certo “ponto de corte” que alerta sobre o alcance de uma proficiência considerada “mínima”, “básica”.

A hipótese nula em cada teste é que os alunos foram distribuídos de forma aleatória entre as turmas dentro da escola em relação a estar “Abaixo do Básico” ou pelo menos no “Básico”, ou seja, que a distribuição dos alunos nas turmas não dependeu da nota prévia. Assim como Clotfelter, Ladd e Vigdor (2006), foi escolhido o nível de significância de 10% para os testes – nível considerado relativamente conservador pelos autores.

Além das notas prévias, os autores testaram se a alocação dos alunos nas turmas foi realizada de forma independente de outras características dos alunos como a cor, a participação em programas de alimentação subsidiada e a escolaridade dos pais. Neste estudo, optou-se por não realizar testes com variáveis como a escolaridade dos pais e a renda familiar pois, apesar de constarem do questionário aplicado aos pais na Prova SP, só foram respondidas por cerca de 75% e 55% dos pais, respectivamente – o que inviabiliza a realização dos testes com base nessas variáveis para uma considerável quantidade de alunos e, conseqüentemente, de escolas.

Tabela 13 – Sumário dos Resultados das Escolas nos Testes Qui-Quadrado de Alocação Aleatória de Alunos a Turmas

| Variáveis Testadas | Nº de Escolas com Êxito nos Testes | % |
|---|---|----------|
| Nível da Nota Anterior em Língua Portuguesa | 283 | 55% |
| Nível da Nota Anterior em Matemática | 309 | 60% |
| Ambos as Notas | 244 | 47% |
| Nº Total de Escolas | 519 | 100% |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2009 e 2010.

Como apresentado na tabela 13, uma considerável parcela das escolas não obteve êxito nos testes aplicados. Mais da metade delas apresentou evidências de ter distribuído os alunos nas turmas levando em consideração a nota prévia dos alunos em Língua Portuguesa ou em Matemática.

Essas evidências não permitem afirmar que tenha havido uma distribuição de todos os alunos da 4ª série nas turmas dessas escolas de forma a privilegiar determinados professores com alunos com notas prévias maiores que outros professores. Só é possível afirmar que os alunos que constam na amostra a ser utilizada neste trabalho indicam que os professores das escolas que falharam nos testes receberam números não balanceados de alunos de acordo com suas notas prévias, o que pode contribuir para que as estimativas dos efeitos professor individuais dessas escolas provavelmente sejam viesadas.

Portanto, assim como fizeram Clotfelter, Ladd e Vigdor (2006), estimamos novamente a equação (9), só que agora a amostra utilizada contém somente alunos das 244 escolas que tiveram êxito nos dois testes de alocação aleatória de alunos a turmas, denominada Amostra Balanceada. Espera-se que a comparação entre a variação nos efeitos professor da Amostra Principal com a variação nos efeitos da Amostra Balanceada permita com que se tenha uma noção do tamanho desse viés.

Antes de apresentar os resultados dessa estimação, porém, vale apresentar o perfil dos alunos da Amostra Balanceada, a fim de verificar se este difere de forma significativa do perfil dos alunos da Amostra Principal, o que pode afetar as estimativas.

Tabela 14 – Estatística Descritiva da Amostra Balanceada da 4ª Série em 2010
(continua)

| | % de alunos | % de respostas |
|---------------------------------|-------------|----------------|
| Gênero Feminino | 44,90% | 100,00% |
| <u>Idade</u> | | |
| Menos de 10 anos | 0,50% | 100,00% |
| 10 anos | 59,40% | 100,00% |
| 11 anos | 27,30% | 100,00% |
| 12 anos | 9,20% | 100,00% |
| Mais de 12 anos | 2,70% | 100,00% |
| Participa do PIC | 3,29% | 100,00% |
| Fez Pré-Escola | 55,90% | 72,60% |
| <u>Escolaridade do Pai</u> | | |
| Não estudou | 4,90% | 72,90% |
| Anos Iniciais do EF Incompletos | 18,10% | 72,90% |
| Anos Iniciais do EF Completos | 29,10% | 72,90% |
| Ensino Fundamental Completo | 20,40% | 72,90% |
| Ensino Médio Completo | 24,70% | 72,90% |
| Ensino Superior Completo | 2,80% | 72,90% |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2009 e 2010.

Tabela 14 – Estatística Descritiva da Amostra Balanceada da 4ª Série em 2010
(conclusão)

| | % de alunos | % de respostas |
|--|-------------|----------------|
| <u>Escolaridade da Mãe</u> | | |
| Não estudou | 3,50% | 77,70% |
| Anos Iniciais do EF Incompletos | 14,40% | 77,70% |
| Anos Iniciais do EF Completos | 29,20% | 77,70% |
| Ensino Fundamental Completo | 21,30% | 77,70% |
| Ensino Médio Completo | 29,10% | 77,70% |
| Ensino Superior Completo | 2,60% | 77,70% |
| <u>Renda Familiar</u> | | |
| Até R\$ 850,00 | 41,70% | 58,20% |
| R\$ 851,00 a R\$ 1275,00 | 31,30% | 58,20% |
| R\$ 1276,00 a R\$ 2125,00 | 18,60% | 58,20% |
| R\$ 2126,00 a R\$ 4250,00 | 7,10% | 58,20% |
| Mais de R\$ 4250,00 | 1,30% | 58,20% |
| <u>Níveis de Proficiência em Língua Portuguesa em 2010</u> | | |
| Abaixo do Básico | 38,40% | 94,70% |
| Básico | 39,10% | 94,70% |
| Adequado | 18,80% | 94,70% |
| Avançado | 3,70% | 94,70% |
| <u>Níveis de Proficiência em Matemática em 2010</u> | | |
| Abaixo do Básico | 50,80% | 95,30% |
| Básico | 34,60% | 95,30% |
| Adequado | 12,80% | 95,30% |
| Avançado | 1,80% | 95,30% |
| <u>Níveis de Proficiência em Língua Portuguesa em 2009</u> | | |
| Abaixo do Básico | 33,60% | 94,70% |
| Básico | 32,20% | 94,70% |
| Adequado | 25,10% | 94,70% |
| Avançado | 9,20% | 94,70% |
| <u>Níveis de Proficiência em Matemática em 2009</u> | | |
| Abaixo do Básico | 44,20% | 95,30% |
| Básico | 36,80% | 95,30% |
| Adequado | 13,80% | 95,30% |
| Avançado | 5,20% | 95,30% |
| Nº de Alunos | | 10938 |
| Nº de Escolas | | 244 |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2009 e 2010.

Apesar de os alunos do PIC estarem sobre-representados na Amostra Principal, por conta de terem sido testados de forma censitária na 3ª série, como já indicado, o mesmo não acontece na Amostra Balanceada – pelo contrário, estão até sub-representados nessa Amostra.

Como os alunos que não foram alfabetizados na idade correta ficam em turmas separadas do programa PIC, essas turmas têm grandes concentrações de alunos com nível de proficiência “Abaixo do Básico”. Portanto, essas turmas diferem consideravelmente das turmas medianas das escolas em termos da distribuição da proficiência prévia dos alunos. Em razão dessa diferença na distribuição dos alunos nas turmas de acordo com as notas prévias, a maioria das escolas com turmas do programa PIC acabam sendo excluídas da Amostra Balanceada, por não passarem nos testes realizados.

De 219 escolas – 42,2% das 519 escolas da Amostra Geral – restaram na Amostra Balanceada apenas 24 escolas com turmas do PIC e turmas regulares, 9,8% das 244 escolas da Amostra Balanceada. Nesse pequeno grupo de escolas, há turmas do PIC e turmas regulares, no entanto essas turmas não apresentam diferenças na distribuição de alunos em termos de sua proficiência prévia, permanecendo na Amostra Balanceada.

A menor proporção de alunos do PIC, porém, não altera o perfil de alunos da Amostra Balanceada em relação à Amostra Principal. Da mesma forma que na comparação entre a Amostra Geral e a Amostra Principal, novamente se observa apenas pequenas variações entre as categorias das características, mantendo a forma da sua distribuição.

Após verificar que não existem diferenças relevantes entre as características observadas dos alunos da Amostra Balanceada e da Amostra Principal, pode-se estimar a equação (9) com base na Amostra Balanceada. Os resultados dessa estimação são apresentados na tabela 15.

Tabela 15 – Resultados da Estimação da Equação (9) com a Amostra Balanceada

| | Língua Portuguesa | Matemática |
|---|--------------------------|---------------------|
| Nota Prévia | 0,683*** [0,013] | 0,653*** [0,012] |
| Gênero Feminino | 0,104*** [0,023] | -0,025 [0,022] |
| Idade Adequada | 0,022 [0,042] | 0,126*** [0,040] |
| Participa do PIC | 0,292 [0,441] | 0,684 [0,433] |
| Fez Pré-Escola | 0,017 [0,023] | -0,013 [0,023] |
| <u>Escolaridade da Mãe</u> | | |
| Anos Iniciais do Ensino Fundamental Completos | 0,052 [0,039] | 0,054 [0,038] |
| Ensino Fundamental Completo | 0,058 [0,042] | 0,026 [0,041] |
| Ensino Médio Completo | 0,064 [0,042] | 0,135*** [0,041] |
| Ensino Superior Completo | 0,170** [0,078] | 0,107 [0,076] |
| <u>Escolaridade do Pai</u> | | |
| Anos Iniciais do Ensino Fundamental Completos | 0,030 [0,035] | 0,045 [0,034] |
| Ensino Fundamental Completo | -0,020 [0,039] | 0,076** [0,037] |
| Ensino Médio Completo | 0,034 [0,039] | 0,120*** [0,038] |
| Ensino Superior Completo | 0,141* [0,075] | 0,218*** [0,072] |
| <u>Renda Familiar</u> | | |
| R\$ 851,00 a R\$ 1275,00 | 0,075*** [0,028] | 0,060** [0,027] |
| R\$ 1276,00 a R\$ 2125,00 | 0,117*** [0,034] | 0,052 [0,033] |
| R\$ 2126,00 a R\$ 4250,00 | 0,179*** [0,048] | 0,132*** [0,047] |
| Mais de R\$ 4250,00 | 0,150 [0,098] | 0,105 [0,095] |
| <u>Desvio-Padrão</u> | | |
| Dos Efeitos Professor | 0,549 | 0,542 |
| Dos Efeitos Escola | 0,490 | 0,405 |
| R ² | 0,690 | 0,687 |
| Nº de Alunos | 4141 | 4210 |
| Nº de Professores | 899 | 899 |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2009 e 2010.

A partir da tabela 15, é possível afirmar que os resultados dos coeficientes da estimação da equação (9) para a Amostra Balanceada são semelhantes àqueles obtidos para a Amostra Principal, com exceção de poucas variáveis, cujos coeficientes se apresentaram estatisticamente significativos na estimativa com a amostra original e não significativos com a amostra mais restrita.

Enquanto o desvio-padrão dos efeitos professor estimados para a Amostra Principal foi de cerca de 0,62 para ambas as disciplinas, os desvios-padrão dos efeitos professor estimados para a Amostra Balanceada foram de cerca de 0,55 em Língua Portuguesa e 0,54 em Matemática.

Há duas possíveis explicações para essas diferenças. Pode ser que os 899 professores da Amostra Balanceada sejam mais parecidos em termos de sua eficácia real que os 2055 professores da Amostra Principal, de modo que a distribuição de seus efeitos reais de fato seja mais concentrada em torno da média que a distribuição dos professores da Amostra Principal. Ou ainda pode ser que essa distribuição da eficácia real seja igual para os dois grupos, mas no caso da Amostra Principal outros fatores que influenciam positivamente ou negativamente o resultado dos alunos que não o professor – os quais podem ser representados pela nota prévia, que condicionou a distribuição dos alunos nas turmas das escolas que não foram excluídas para gerar a Amostra Balanceada – podem estar inflando seus efeitos, de modo que o desvio-padrão seja maior.

Caso a primeira hipótese seja verdadeira, as estimativas referentes à Amostra Principal refletem de forma mais fidedigna a variação dos efeitos de todos os professores da 4ª série da rede municipal de São Paulo. Caso a segunda hipótese seja verdadeira, as estimativas mais confiáveis seriam as referentes à Amostra Balanceada. Dadas as várias evidências apresentadas pela literatura de vieses oriundos da alocação não aleatória de alunos a turmas e professores, acredita-se ser mais provável que a segunda hipótese seja verdadeira.

No entanto, como não é possível afirmar qual hipótese reflete a situação real dos professores da rede municipal de ensino de São Paulo, as demais análises serão realizadas considerando as duas amostras de alunos e professores.

6.2.2. Análise de Confiabilidade

Como já discutido, sabe-se que a variância residual nas notas dos alunos devida a erros de medida e a fontes de variabilidade no desempenho além dos professores, escolas e outros fatores do modelo resulta em erros amostrais nos efeitos professor estimados. Assim sendo, parte do desvio-padrão dos efeitos professor estimados deve-se a erros amostrais. Tanto é que todos os artigos revisados que estimam os efeitos professor corrigem a variância dos efeitos professor estimados pelo tamanho do erro amostral (Rockoff, 2004; Nye, Konstantopoulos e Hedges, 2004; Aaronson, Barrow e Sander, 2007; Rothstein, 2009; Goldhaber e Hansen, 2010; Koedel e Betts, 2011; entre outros).

Aaronson, Barrow e Sander (2007) explicam que o termo do erro é particularmente problemático quando as estimativas dos efeitos dos professores são baseadas em populações pequenas. Quando isso ocorre, as variações amostrais podem superestimar ou subestimar os efeitos dos professores, pois poucas observações com resultados próximos dos extremos da distribuição podem influenciar fortemente os efeitos professor estimados. Consequentemente, o desvio-padrão da distribuição dos efeitos professor estimados seria quase sempre inflado.

Para ajustar o desvio-padrão estimado, adotou-se o mesmo procedimento que Aaronson, Barrow e Sander (2007), Goldhaber e Hansen (2010) e Koedel e Betts (2011), de acordo com o qual a variância dos efeitos professor estimados foi ajustada pelos erros amostrais, assumindo que o efeito professor estimado na equação (9) é a soma do efeito professor real mais um termo de erro, o qual não é correlacionado com o efeito professor real:

$$\hat{\tau}_j = \tau_j + \lambda_j \quad (12)$$

Da mesma forma, pode-se decompor a variância dos efeitos professor estimados em uma parcela referente à variância real dos efeitos professor e outra referente à variância dos erros amostrais:

$$Var(\hat{\tau}) = Var(\tau) + Var(\lambda) \quad (13)$$

Assim como esses autores, neste trabalho foi utilizada a média dos quadrados dos erros-padrão dos efeitos professor estimados como uma estimativa

da variância do erro amostral, subtraindo esse fator da variância observada dos efeitos professor para obter a variância ajustada e, assim, o desvio-padrão ajustado.

Com essa correção, encontramos um desvio-padrão ajustado dos efeitos professor estimados de cerca de 0,53 em ambas as disciplinas para a Amostra Principal e de cerca de 0,45 em Língua Portuguesa e 0,46 em Matemática para a Amostra Balanceada.

Para se ter uma idéia de como essa variação nos efeitos professor ajustados dos professores da 4ª série da rede municipal de ensino de São Paulo ainda pode ser considerada grande, os dez estudos com análises similares revisados por Hanushek e Rivkin (2010) e também citados neste trabalho, obtiveram desvios-padrão ajustados dos efeitos professor que foram de 0,08 a 0,26 em Língua Portuguesa e 0,11 a 0,36 em Matemática.

Para analisar o peso dos erros amostrais nas estimativas, McCaffrey et al (2009) trabalham com um coeficiente de confiabilidade que corresponde à razão entre a variância dos efeitos professor reais, estimada a partir do modelo descrito, e a variância dos efeitos professor estimados.

Neste estudo, o coeficiente de confiabilidade dos efeitos professor estimados com a Amostra Principal foram de cerca de 0,70 em Língua Portuguesa e 0,72 em Matemática, e os estimados com a Amostra Balanceada foram de 0,65 em Língua Portuguesa e 0,67 em Matemática.

Seguindo as recomendações de Lockwood, Louis e McCaffrey (2002) a partir de suas simulações a respeito da incerteza na estimação de *rankings* de eficácia de professores, a razão entre a variância do erro amostral e a variância dos efeitos professor estimados deveria ser menor que 0,1 para que o *ranking* estimado seja preciso para distinguir a eficácia dos professores.

Assim como em McCaffrey et al (2004), na qual a variabilidade devido a erro amostral obtida foi de 0,2 a 0,4 da variabilidade estimada dos efeitos professor, neste trabalho também os percentuais não permitem a estimação de *rankings* precisos, segundo esse critério. Foram obtidos percentuais de 0,28 a 0,35 – muito além do 0,1 indicado por Lockwood, Louis e McCaffrey (2002).

Do mesmo modo que McCaffrey et al (2004), foram feitos testes Wald para verificar se os coeficientes dos efeitos obtidos para cada professor podem ser considerados diferentes de zero, o que significa que esses professores podem ser

considerados distintos da média dos professores analisados, tanto para baixo quanto para cima.

No entanto, enquanto as análises desses autores permitiram obter evidências que indicam que entre um terço e um quarto dos professores podem ser considerados distintos da média, os resultados dos testes do presente trabalho indicam que apenas 12% dos professores tiveram efeitos estimados distintos da média em Língua Portuguesa, em ambas as amostras, e apenas 15% e 13% dos professores das Amostras Principal e Balanceada, respectivamente, tiveram efeitos estimados distintos da média em Matemática.

Deste modo, os dados nos quais este trabalho se baseia somente permitem distinguir como mais ou menos eficazes uma parcela pequena dos professores da rede municipal de ensino de São Paulo, demonstrando ter uma capacidade limitada para subsidiar recomendações em relação a políticas de pessoal.

6.2.3. Análise de Estabilidade

A forma mais simples de analisar a estabilidade dos efeitos individuais dos professores é calculando a correlação entre os efeitos estimados para os professores em um ano e os efeitos estimados para esses professores em um ano subsequente. Trabalhos como o de Goldhaber e Hansen (2010), que contaram com doze anos de dados, puderam realizar análises mais aprofundadas sobre essa questão. No caso deste trabalho, dada a restrição dos dados, somente foi possível estimar a medida mais simples de coeficiente de estabilidade, que é dada justamente pela correlação dos efeitos professor de dois anos consecutivos (McCaffrey et al, 2009).

Em todos os anos de sua realização, a Prova São Paulo foi aplicada para a 4ª série, estando disponíveis os resultados dos anos de 2007 a 2010. Já para a 3ª série, que serve de *baseline* para as análises deste trabalho, a Prova São Paulo foi aplicada a partir de 2008, estando disponíveis os resultados dos anos de 2008 a 2010. Assim sendo, além das estimativas dos efeitos dos professores da 4ª série referentes ao ano de 2010, tendo as notas dos alunos na 3ª série em 2009 como

baseline, somente é possível estimar os efeitos dos professores da 4ª série referentes ao ano de 2009, tendo as notas dos alunos na 3ª série em 2008 como *baseline*.

Em razão do atrito de professores na rede, de variação nas séries em que os professores lecionam, e ainda das restrições impostas às amostras de modo a realizar as estimativas, somente 714 dos 2055 professores da Amostra Principal de 2010 são encontrados lecionando para a Amostra Principal da 4ª série em 2009. Portanto, somente é possível calcular a correlação entre os efeitos estimados para esses professores, pois somente eles possuem estimativas nos dois anos.

Para analisar se esses professores constituem uma amostra semelhante ao conjunto completo dos professores analisados em 2010, a tabela 16 apresenta o perfil desse grupo comum aos dois anos, com base nas respostas dadas ao questionário de 2010.

Tabela 16 – Estatística Descritiva dos Professores da 4ª Série Comuns a 2009 e 2010

| | % de professores | % de respostas |
|--|-------------------------|-----------------------|
| Gênero Feminino | 95,24% | 100,00% |
| Branco ou Amarelo | 72,10% | 86,83% |
| <u>Idade</u> | | |
| Até 25 anos | 0,98% | 100,00% |
| 26 a 35 anos | 9,66% | 100,00% |
| 36 a 45 anos | 40,20% | 100,00% |
| 46 a 55 anos | 33,76% | 100,00% |
| Mais de 55 anos | 15,41% | 100,00% |
| <u>Nível de Escolaridade</u> | | |
| Superior | 76,59% | 87,95% |
| Especialização | 21,66% | 87,95% |
| Mestrado | 1,75% | 87,95% |
| Doutorado | 0,00% | 87,95% |
| <u>Tempo de Experiência como Professor</u> | | |
| Até 5 anos | 2,53% | 88,52% |
| 6 a 10 anos | 6,01% | 88,52% |
| 11 a 15 anos | 18,20% | 88,52% |
| 16 a 20 anos | 24,84% | 88,52% |
| Mais de 20 anos | 48,42% | 88,52% |
| <u>Renda Familiar</u> | | |
| R\$ 1000,00 a R\$ 2000,00 | 1,16% | 84,73% |
| R\$ 2001,00 a R\$ 3000,00 | 11,90% | 84,73% |
| R\$ 3001,00 a R\$ 4000,00 | 24,96% | 84,73% |
| R\$ 4001,00 a R\$ 5000,00 | 26,78% | 84,73% |
| Mais de R\$ 5000,00 | 35,21% | 84,73% |
| <u>Escolaridade do Pai</u> | | |
| Não estudou | 5,84% | 86,27% |
| Anos Iniciais do EF Incompletos | 24,19% | 86,27% |
| Anos Iniciais do EF Completos | 37,66% | 86,27% |
| Ensino Fundamental Completo | 13,47% | 86,27% |
| Ensino Médio Completo | 12,18% | 86,27% |
| Ensino Superior Completo | 6,66% | 86,27% |
| Nº de Professores | | 714 |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2008 a 2010.

Como se pode observar, o percentual de professores do grupo comum aos dois anos com menos de 35 anos é cerca de metade do percentual de professores do conjunto total de 2010. Enquanto isso, o percentual de professores com mais de 46 anos sobe de 31% para 39%. Além de serem mais velhos, em média, os professores do grupo comum têm mais tempo de experiência e maior renda familiar, como era de se esperar em razão da diferença de idade. Em relação

às demais características, não foram encontradas diferenças entre os dois grupos de professores.

Já que se trata de grupos com algumas diferenças de perfil, vale analisar os desvios-padrão dos efeitos professor estimados para esse grupo comum de professores em 2010, de forma a verificar se a variação destes efeitos difere muito da variação dos efeitos para o grupo completo de professores em 2010, já apresentada e discutida anteriormente.

O desvio-padrão dos efeitos professor calculado para os professores do grupo comum, em 2010, foi de 0,64, em Língua Portuguesa e de 0,61 em Matemática, valores bem próximos ao desvio-padrão de 0,62, obtido para o conjunto completo dos professores em ambas as disciplinas. Os desvios-padrão ajustados em razão dos erros amostrais, por sua vez, foram de 0,54 em Língua Portuguesa e 0,51 em Matemática, também próximos ao desvio-padrão de 0,53, obtido para o conjunto completo dos professores em 2010. Portanto, há evidências de que a variação dos efeitos professor desse grupo é capaz de representar a variação dos efeitos professor do conjunto completo de professores de 2010.

A correlação entre os efeitos professor estimados para os professores comuns aos dois anos foi calculada em 0,097 em Língua Portuguesa e 0,081 em Matemática. Essas correlações estimadas indicam uma estabilidade bem menor dos efeitos professor estimados para o presente trabalho que para os estudos revisados, tais como os de McCaffrey et al (2009), que obtiveram correlações nos efeitos professor estimados de 0,22 a 0,67, e Goldhaber e Hansen (2010), que obtiveram correlações de 0,32 a 0,59.

6.2.4. Decomposição dos Efeitos Professor Estimados

Para buscar compreender melhor a variação nos efeitos estimados para os professores da 4ª série da rede municipal ensino de São Paulo em 2010, as estimativas realizadas até o presente momento serão utilizadas para decompor a variação dos efeitos professor, como propõem McCaffrey et al (2009).

Os efeitos professor estimados em determinado ano podem ser decompostos da seguinte forma:

$$\hat{\tau}_j = \tau_j pers + \tau_j npers + \lambda_j \quad (14)$$

A diferença entre a equação (14) e a equação (12), que também decompõe o efeito professor estimado, é que na segunda o efeito real dos professores, ou seja, a parcela do efeito estimado que não se deve aos erros amostrais, é composta por um componente persistente e um componente não persistente. Enquanto o componente persistente se refere à parcela comum a todos os anos de estimações, o componente não persistente se refere à parcela específica dos efeitos do ano em questão.

Como já comentado, esse componente transitório pode capturar variações no desempenho do professor em diferentes momentos da vida profissional. Mas também pode capturar outros fatores no nível da classe naquele ano específico. No caso deste trabalho, onde cada professor é observado com apenas uma classe por ano, não sendo possível adotar estratégias para tentar separar o efeito professor do efeito turma, é bem provável que fatores relacionados à turma tenham grande responsabilidade por esse componente transitório dos efeitos professor. Por essa razão, deixaremos de denominar a parcela dos efeitos professor estimados que não se deve aos erros amostrais de “efeitos professor reais”, pois provavelmente inclui influências de fatores externos ao professor, como discutiremos mais adiante.

A variação dos efeitos professor estimados, portanto, também pode ser decomposta:

$$Var(\hat{\tau}) = Var(\tau pers) + Var(\tau npers) + Var(\lambda) \quad (15)$$

Como estamos trabalhando com os dados de 2010 como base principal das análises, e somente temos a variância dos efeitos persistentes para o grupo restrito de professores que é comum aos dois anos, faremos uma decomposição da variância dos efeitos professor estimados em 2010 para esse grupo restrito.

Da equação (15), já temos a variância dos efeitos professor estimados para esse grupo de professores – tanto é que já foram apresentados os desvios-padrão desses efeitos. A variância dos efeitos persistentes, por sua vez, pode ser estimada utilizando a correlação entre os efeitos professor de 2009 e 2010. Como a correlação entre os efeitos professor de dois anos consecutivos pode ser considerada uma medida do coeficiente de estabilidade dos professores, e esse coeficiente pode ser entendido como a razão entre a variância dos efeitos

persistentes sobre a variância dos efeitos professor estimados, como descrevem McCaffrey et al (2009), é possível obter uma estimativa da variância dos efeitos persistentes dos professores.

O mesmo procedimento adotado para calcular a variância dos erros amostrais para os efeitos professor do conjunto total de professores pode ser adotado para calcular a variância dos erros amostrais desse grupo. Com todos os demais fatores conhecidos, pode-se obter a variância dos efeitos não persistentes a partir da equação (15).

A decomposição da variância dos efeitos professor estimados em 2010 para o grupo de professores comuns a 2009 e 2010 é apresentada na tabela 17.

Tabela 17 – Decomposição da Variância dos Efeitos Professor Estimados

| | Língua Portuguesa | | Matemática | |
|----------------------|-------------------|-------|------------|-------|
| | Variância | % | Variância | % |
| $Var(\hat{\tau})$ | 0,403 | 100% | 0,368 | 100% |
| <u>Decomposição:</u> | | | | |
| $Var(\tau_{pers})$ | 0,039 | 9,7% | 0,030 | 8,1% |
| $Var(\pi_{pers})$ | 0,250 | 61,9% | 0,230 | 62,7% |
| $Var(\lambda)$ | 0,114 | 28,4% | 0,108 | 29,2% |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2008 a 2010.

Na análise da decomposição observa-se que apenas 9,7% da variação dos efeitos professor estimados em Língua Portuguesa e 8,1% em Matemática se deve a efeitos persistentes, ou seja, só é possível afirmar, com maior confiança, que cerca de 9% da variância dos efeitos professor estimados correspondem, de fato, àquele professor específico. Esse valor é importante não só porque apresenta uma estimativa da parcela dos efeitos professor que não se alteram com o passar dos anos, mas também porque pode ser visto como um limite inferior para a variação dos efeitos estimados que, de fato, são de responsabilidade dos professores e podem ser considerados os efeitos professor reais.

Por outro lado, identificando e excluindo a parcela da variância relativa aos erros amostrais, é possível estimar um limite superior para a variância dos efeitos professor que realmente são causados pelos professores. Dado que 28,4% da variação dos efeitos professor estimados em Língua Portuguesa e 29,2% em Matemática se devem aos erros amostrais, os efeitos professor não persistentes

correspondem a 61,9% da variação dos efeitos professor em Língua Portuguesa e 62,7% em Matemática. Como não se consegue distinguir o quanto dessa variação se deve aos professores isoladamente, às turmas isoladamente e à interação entre os professores e as turmas, no limite é possível que toda essa parte dos efeitos não persistentes se deva aos professores. Esse limite superior seria, então, de 71,6% em Língua Portuguesa e 70,8% em Matemática.

Para compreender o quanto essa variação representa em termos das notas dos alunos na Prova São Paulo, podemos trabalhar novamente com os desvios-padrão dos efeitos, já que as notas dos alunos estão padronizadas.

Tomando o desvio-padrão dos efeitos professor estimados, percentuais de 9,7% e 8,1% de variação em Língua Portuguesa e Matemática, respectivamente, representam 0,062 e 0,049 desvios-padrão, e percentuais de 71,6% e 70,8% de variação em Língua Portuguesa e Matemática, respectivamente, representam 0,454 e 0,429 desvios-padrão.

Isso indica que um aumento de um desvio-padrão nos efeitos professor levaria a um aumento entre 0,062 e 0,454 desvio-padrão na proficiência dos alunos em Língua Portuguesa e entre 0,049 e 0,429 desvio-padrão em Matemática, que correspondem a, aproximadamente, entre 2,79 e 31,94 pontos em Língua Portuguesa e entre 2,20 e 31,53 pontos em Matemática.

Dado que em uma distribuição normal 68% dos valores encontram-se a uma distância da média que é inferior a um desvio-padrão, olhando para a distribuição dos professores por ordem de eficácia estimada pode-se dizer que trocar um professor do 16º percentil de eficácia por um professor do 84º percentil, por exemplo, elevaria as notas dos alunos em, no mínimo, 2,79 pontos em Língua Portuguesa e 2,20 pontos em Matemática.

Como é pouco provável que toda a variação dos efeitos não persistentes seja referente apenas aos professores, provavelmente se está superestimando o limite superior da variação dos efeitos professor estimados que são de fato causados pelo professor. Em uma mera simulação, podemos supor, por exemplo, uma situação em que essa parcela dos efeitos não persistentes se deva de igual modo a professores, turmas e à interação entre professores e turmas, ficando cada um desses fatores com um terço da variação devida a componentes transitórios. Neste caso, que parece mais próximo do que seria real, pois não desconsidera a interação entre professores e turmas e as turmas isoladamente, a parcela da

variância dos efeitos professor estimados devida aos professores seria de 30,3% em Língua Portuguesa e 29% em Matemática. Nessa situação hipotética, um aumento de um desvio de um desvio-padrão nos efeitos professor levaria a um aumento de 0,349 desvio-padrão na proficiência dos alunos em Língua Portuguesa e 0,327 desvio-padrão em Matemática, que correspondem a, aproximadamente, 15,59 pontos em Língua Portuguesa e 14,55 pontos em Matemática.

6.3. Análise de Fatores Associados aos Efeitos Professor Estimados

Passamos, portanto, à análise dos fatores associados aos efeitos dos professores da 4ª série da rede municipal de ensino de São Paulo estimados pela equação (9) para as Amostras Principal e Balanceada, a fim de buscar levantar evidências de que as variáveis que caracterizam os professores e estão contidas nos bancos de dados da Secretaria são capazes de explicar alguma parcela desses efeitos. Essa análise é feita a partir da estimação da equação (10).

Antes de analisar os resultados obtidos por meio dessa regressão, é importante fazer algumas ressalvas. Em primeiro lugar, destaca-se a redução da quantidade de professores do primeiro para o segundo estágio quase pela metade, em razão da falta de respostas a itens do questionário por parte de alguns professores, o que faz com que as observações sejam descartadas pelo *software* estatístico *Stata*, na estimação da regressão. Além disso, vale ressaltar que se trata de uma *cross-section* na qual os efeitos professor em um determinado ano são regredidos em suas características referentes a esse mesmo ano. Portanto, não é possível tratar as evidências atribuindo um sentido de causalidade entre estes e os efeitos professor, pois não há como saber como eram os efeitos dos professores antes do “tratamento”, ou seja, antes deles passarem a ter a determinada característica ou realizar determinada prática pedagógica.

Isto posto, entende-se que os resultados obtidos, apresentados na tabela 18, devem ser tratados com a devida cautela, como evidências de um estudo exploratório sobre a associação entre esses fatores e os efeitos professor estimados.

Tabela 18 – Análise de Fatores Associados às Estimativas de Efeito Professor
(continua)

| | Língua Portuguesa | | Matemática | |
|---|----------------------|---------------------|--------------------|----------------------|
| | Amostra Principal | Amostra Balanceada | Amostra Principal | Amostra Balanceada |
| Feminino | 0,196** [0,088] | 0,098 [0,116] | -0,064 [0,083] | 0,162 [0,116] |
| Branco ou Amarelo | -0,075* [0,039] | -0,035 [0,056] | 0,026 [0,039] | -0,121** [0,053] |
| Especialização | -0,103*** [0,040] | -0,088* [0,053] | -0,072* [0,041] | -0,264*** [0,052] |
| Mestrado | -0,254 [0,186] | -0,366 [0,315] | -0,030 [0,178] | -0,072 [0,216] |
| Tem 1 a 5 anos de experiência | 0,219 [0,220] | -0,062 [0,393] | 0,067 [0,224] | 0,026 [0,398] |
| Tem 6 a 10 anos de experiência | 0,201 [0,216] | 0,050 [0,392] | 0,154 [0,221] | 0,110 [0,397] |
| Tem 11 a 15 anos de experiência | 0,153 [0,216] | 0,038 [0,391] | 0,097 [0,220] | 0,048 [0,396] |
| Tem 16 a 20 anos de experiência | 0,190 [0,217] | 0,091 [0,392] | 0,154 [0,221] | 0,108 [0,395] |
| Tem mais de 20 de experiência | 0,146 [0,216] | 0,145 [0,391] | 0,118 [0,220] | 0,123 [0,394] |
| Está 1 a 5 anos nessa escola | -0,019 [0,051] | -0,037 [0,067] | 0,040 [0,051] | 0,129* [0,067] |
| Está 6 a 10 anos nessa escola | 0,039 [0,061] | -0,051 [0,079] | 0,053 [0,060] | 0,071 [0,079] |
| Está 11 a 15 anos nessa escola | -0,000 [0,066] | 0,021 [0,086] | 0,002 [0,065] | 0,067 [0,082] |
| Está 16 a 20 anos nessa escola | 0,074 [0,088] | -0,090 [0,120] | -0,005 [0,086] | 0,194 [0,121] |
| Está mais de 20 anos nessa escola | -0,026 [0,106] | 0,070 [0,163] | -0,026 [0,100] | 0,165 [0,157] |
| Trabalha em mais de 1 escola | -0,031 [0,037] | 0,029 [0,049] | -0,008 [0,037] | 0,021 [0,048] |
| Tem outra atividade profissional além da docência na educação básica | 0,045 [0,110] | 0,043 [0,139] | -0,109 [0,105] | 0,223* [0,134] |
| Dedica 4 a 8 horas semanais fora da escola ao trabalho pedagógico | 0,055 [0,047] | 0,139** [0,061] | 0,025 [0,046] | 0,030 [0,062] |
| Dedica mais de 8 horas semanais fora da escola ao trabalho pedagógico | 0,071 [0,058] | 0,226*** [0,075] | -0,025 [0,057] | 0,084 [0,075] |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2009 e 2010.
Níveis de Significância: *** indica 1%; ** indica 5%; * indica 10%.

**Tabela 18 – Análise de Fatores Associados às Estimativas de Efeito Professor
(conclusão)**

| | Língua Portuguesa | | Matemática | |
|--|---------------------|--------------------|--------------------|--------------------|
| | Amostra Principal | Amostra Balanceada | Amostra Principal | Amostra Balanceada |
| Usa lousa para cópia de textos | -0,023 [0,040] | 0,028 [0,052] | 0,017 [0,040] | -0,026 [0,051] |
| Usa computador e internet nas aulas | -0,013 [0,038] | -0,045 [0,051] | 0,020 [0,038] | 0,019 [0,050] |
| Costuma convidar os pais para conversar sobre os filhos | 0,101 [0,144] | -0,010 [0,189] | 0,145 [0,144] | -0,118 [0,182] |
| Passa lição de casa sempre ou quase sempre | 0,039 [0,046] | -0,036 [0,066] | 0,106** [0,045] | -0,008 [0,064] |
| Usa Cadernos de Apoio e Aprendizagem de 1 a 2 vezes por semana | 0,219* [0,125] | 0,100 [0,261] | -0,027 [0,122] | -0,218 [0,275] |
| Usa Cadernos de Apoio e Aprendizagem de 3 a 4 vezes por semana | 0,176 [0,126] | 0,208 [0,262] | -0,016 [0,122] | -0,230 [0,275] |
| Usa Cadernos de Apoio e Aprendizagem todos os dias | 0,284** [0,140] | 0,316 [0,276] | 0,016 [0,137] | -0,133 [0,288] |
| Cumpriu de 50 a 70% do currículo | 0,119 [0,247] | -0,198 [0,654] | 0,245 [0,215] | -0,684 [0,652] |
| Cumpriu mais de 70% do currículo | 0,120 [0,246] | -0,201 [0,653] | 0,255 [0,212] | -0,494 [0,652] |
| Gasta de 20 a 30 minutos com a organização da turma | 0,060 [0,081] | 0,165 [0,114] | -0,109 [0,081] | 0,153 [0,104] |
| Gasta de 10 a 20 minutos com a organização da turma | 0,049 [0,070] | 0,116 [0,101] | -0,086 [0,072] | 0,048 [0,090] |
| Gasta menos de 10 minutos com a organização da turma | 0,109 [0,075] | 0,159 [0,107] | -0,040 [0,077] | 0,042 [0,096] |
| Constante | -0,784** [0,375] | -0,285 [0,839] | -0,436 [0,368] | 0,591 [0,835] |
| R ² | 0,031 | 0,081 | 0,024 | 0,113 |
| Nº de professores | 1161 | 482 | 1155 | 503 |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2009 e 2010.
Níveis de Significância: *** indica 1%; ** indica 5%; * indica 10%.

Em primeiro lugar, deve-se destacar que o R^2 das estimações chegou ao valor máximo de 0,11, indicando que a vasta maioria da variação na eficácia dos professores não é explicada pelas características dos professores – evidências semelhantes às encontradas por Aaronson, Barrow e Sander (2007) e Koedel e Betts (2007).

No entanto, como esses autores explicam, o R^2 é uma subestimação do poder explicativo da relação, dado que uma porção significativa, talvez um terço, da variação em $\hat{\tau}_j$ se deva a erros amostrais. Assim como fizeram Aaronson, Barrow e Sander (2007), se multiplicarmos a soma total dos quadrados por uma taxa de 50% para contabilizar os erros amostrais – taxa considerada conservadora pelos autores – o R^2 dobraria. Ainda assim, chegaria a no máximo, 0,22, um valor que pode ser considerado baixo para o poder explicativo de uma regressão.

Em se tratando das variáveis tradicionalmente analisadas nos artigos sobre essa temática, visto que são comumente utilizadas como critérios para crescimento na carreira do magistério, não foi encontrada associação entre o tempo de experiência e a eficácia do professor. Por outro lado, destaca-se a relação negativa e significativa para professores com especialização – em relação àqueles que possuem somente o nível superior completo – observada em ambas as amostras de ambas as disciplinas.

Esperava-se que essa relação – entre a especialização e os efeitos professor – fosse positiva ou que não houvesse associação, mas não que ela fosse negativa, significativa e recorrente em todas as análises. Esse resultado indica que professores com especialização têm, em média, efeitos professor menores. Como os professores podem realizar uma série de cursos de especialização de várias temáticas e abordagens, é difícil buscar explicações para esse resultado que sejam válidas para todos os cursos. Essa associação se apresentou com evidências fortes se comparada às outras características dos professores analisados, de modo que valeria a pena buscar levantar informações que possam auxiliar na explicação desses resultados, tais como fatores que sejam comuns aos cursos de especialização realizados, tais como algum tipo de orientação pedagógica, ou, ainda, características dos professores que estejam relacionados à realização desses cursos. Por exemplo, pode-se começar analisando os critérios de seleção para os cursos de especialização que são ofertados ou pagos pela Secretaria, para verificar

se tendem a selecionar um perfil distinto de profissionais menos efetivos para os cursos que, por sua vez, não teriam êxito em melhorar a eficácia dos professores selecionados.

As variáveis referentes ao tempo que o professor leciona naquela mesma escola apresentaram-se positivas e significativas estatisticamente somente em relação aos efeitos professor em Matemática, na Amostra Balanceada. Pode-se afirmar que há considerável consenso em torno da preferência pela permanência dos professores nas mesmas escolas durante maiores períodos de tempo, em comparação com a alta rotatividade destes entre as escolas. Entende-se que permanecer por curtos períodos de tempo não permite com que o professor conheça e crie relações adequadas com a equipe escolar e com os alunos, e se acostume com a dinâmica daquela escola específica, prejudicando a qualidade de seu trabalho. As maiores controvérsias, nesse caso, estão em como incentivar ou restringir os professores a permanecer nas mesmas escolas.

O coeficiente da variável relativa a ter outra atividade profissional, fora da educação básica, apresentou-se positivo e significativo apenas em Matemática, na Amostra Balanceada. Esperava-se que não houvesse evidências da associação entre essa variável e os efeitos professor ou que esta fosse negativa, indicando que professores com outra atividade profissional fora da educação básica produziram efeitos professor menores. Uma possível interpretação para o resultado obtido é que a outra atividade desses professores seja, por exemplo, relativa ao ensino superior, permitindo com que estes desenvolvam estudos que levem sua eficácia a ser maior que a dos demais professores.

As variáveis referentes ao tempo dedicado ao trabalho pedagógico fora da escola apresentaram-se positivas e significativas em relação aos efeitos professor em Língua Portuguesa, na Amostra Balanceada. Uma associação positiva entre esse fator e os efeitos professor era de fato esperada, podendo ter vários significados, como por exemplo que quanto maior o tempo dedicado ao trabalho pedagógico fora da escola, maior seriam os efeitos professor obtidos. Ou, por exemplo, pode ser que essas variáveis sejam indicadores do esforço dos professores, ou seja, que os professores mais esforçados gastam mais tempo no trabalho pedagógico fora da escola e em uma série de outras atividades que façam com que o seu desempenho seja melhor. De qualquer forma, esses resultados indicam que esse é um fator que merece ser alvo de estudos mais detalhados em

relação ao perfil desses professores e quais atividades eles realizam no tempo adicional de trabalho pedagógico fora da escola.

A variável referente ao professor passar lição de casa sempre ou quase sempre se mostrou positiva e significativa apenas em Matemática para a Amostra Principal. Esperava-se o mesmo resultado para Língua Portuguesa, e para ambas as amostras, mas há que se reconhecer o caráter restrito desse item do questionário em torno da frequência da lição de casa, e não da adequação ou da qualidade das tarefas propostas, o que limita um pouco a realização de análises sobre esse fator.

Outras variáveis que se destacaram foram aquelas que dizem respeito ao uso dos Cadernos de Apoio e Aprendizagem, neste caso em relação aos efeitos professor em Língua Portuguesa. Na Amostra Principal, o uso dos Cadernos todos os dias apresentou resultados positivos e significativos a 5%, enquanto que o uso de 1 a 2 dias por semana apresentou resultados positivos e significativos a 10%. Em Matemática, não foram encontrados resultados significativos estatisticamente para nenhuma das Amostras, assim como para a Amostra Balanceada em Língua Portuguesa. Assim sendo, entende-se que nesse caso também será válido realizar estudos detalhados sobre o uso dos Cadernos de Apoio e Aprendizagem em Língua Portuguesa para compreender o que leva os professores que os usam com maior frequência terem maiores efeitos professor. Novamente, existe uma série de possíveis explicações, como por exemplo a qualidade desse recurso didático em comparação com outros recursos utilizados pelos professores, a participação do professor em capacitações específicas para o uso desse material que aprimoram sua capacidade de ensino de Língua Portuguesa, ou ainda o fato de os Cadernos de Apoio e Aprendizagem terem um caráter de “material estruturado” facilite a organização do trabalho do professor, permitindo com que ele aproveite melhor o tempo para o ensino de Língua Portuguesa.

Apesar de reconhecer a falta de consistência dos resultados obtidos entre as amostras analisadas, entende-se que as evidências obtidas são relevantes para justificar a recomendação para que seja dada especial atenção aos fatores que se mostraram significativos e que sejam realizados estudos que permitam aprofundar os conhecimentos acerca da relação entre esses fatores e a eficácia dos professores.

7. CONCLUSÃO

Dada a importância dos professores para o processo educacional e a disponibilidade de dados que permitissem a estimação de medidas de eficácia dos professores com base nos resultados dos alunos, este trabalho buscou aplicar modelos de valor agregado para estimar e analisar a eficácia de uma amostra dos professores que lecionaram para a 4ª série da rede municipal de ensino de São Paulo no ano de 2010.

A partir das análises, foi possível obter uma série de evidências acerca dos efeitos desses professores sobre os resultados de seus respectivos alunos na Prova São Paulo.

Em primeiro lugar, ao estimar os efeitos totais dos professores sobre a variação dos resultados dos alunos, obteve-se uma diferença de 0,09 entre os R^2 das estimações dos modelos com e sem as variáveis que identificam os professores, tanto em Língua Portuguesa quanto em Matemática. Isso indica que a variação na eficácia dos professores explicaria cerca de 9% da variação nas notas dos alunos da amostra analisada, menos do que a variação nas variáveis de *background* dos alunos, em torno de 15%, mas mais do que a variação nas variáveis de escola, em torno de 5%.

Ao estimar os efeitos professor individuais dessa amostra, foram obtidos desvios-padrão dos efeitos professor estimados de aproximadamente 0,62 em ambas as disciplinas. Isso equivaleria a dizer que um aumento de um desvio-padrão na eficácia dos professores levaria a um aumento de 28 pontos nas notas dos alunos, tanto em Língua Portuguesa quanto em Matemática.

Sabendo que essas evidências representam impactos muito grandes para um único fator escolar, mesmo no caso dos professores, foi realizado um esforço para decompor os efeitos professor estimados, de modo a buscar separar a parcela deles que, de fato, se deve a cada professor individual, e não a outros fatores que possam influenciar as estimativas.

Cerca de 30% da variância dos efeitos professor estimados devem-se aos erros amostrais, fazendo com que o coeficiente de confiabilidade dos efeitos professor estimados estivesse entre 0,65 e 0,72, de acordo com a amostra e a disciplina.

Dados os consideráveis erros amostrais, apenas 12% dos professores tiveram efeitos estimados distintos da média em Língua Portuguesa, e em torno de 13% tiveram efeitos estimados distintos da média em Matemática.

Do restante da variação dos efeitos professor estimados, fazem parte a variação do componente persistente dos efeitos professor e a parcela da variação do componente não persistente dos efeitos professor.

A variação do componente persistente dos efeitos professor, aquele que corresponde à parcela da eficácia do professor que seria permanente ao longo dos anos, foi calculada em cerca de 9% da variância dos efeitos professor estimados. Como só estavam disponíveis dois anos de estimativas de efeito professor, essa medida foi dada pela correlação entre os efeitos professor estimados para 2009 e 2010, e permite indicar também que o coeficiente de estabilidade encontrado foi de 0,09.

Como esse componente é único que pode ser atribuído totalmente aos professores individuais, foi considerado o limite inferior da variação dos efeitos professor reais para o ano de 2010. Já os efeitos professor não persistentes podem conter, além da eficácia do professor, a influência de outros fatores relativos à turma ou à interação entre o professor e a turma, podendo ser considerados como um efeito classe.

Dados os limites inferior e superior calculados para os efeitos professor reais, os resultados indicam que um aumento de um desvio-padrão nos efeitos professor levaria a um aumento mínimo de 0,062 desvio-padrão na proficiência dos alunos em Língua Portuguesa e de 0,049 desvio-padrão em Matemática, que correspondem a 2,79 pontos em Língua Portuguesa e 2,20 pontos em Matemática. Esse aumento seria de, no máximo 0,454 desvio-padrão ou 31,94 pontos em Língua Portuguesa e 0,429 desvio-padrão ou 31,52 pontos em Matemática – valores que correspondem ao efeito classe, por englobarem os efeitos dos professores com os efeitos da turma e da interação entre eles.

Essas evidências se aproximam das obtidas nos 11 estudos revisados por Hanushek e Rivkin (2010) publicados entre 2004 e 2010 referentes a estados ou cidades americanas, os quais indicaram que aumento de um desvio-padrão nos efeitos professor levaria a aumentos entre 0,08 a 0,36 desvio-padrão nos resultados dos alunos.

Buscando explicar o que estaria relacionado à eficácia dos professores da amostra pesquisada, foi realizada uma análise dos fatores associados aos efeitos estimados para os professores da 4ª série da rede municipal de ensino de São Paulo.

Em se tratando das variáveis comumente utilizadas como critérios para crescimento na carreira do magistério, não foi encontrada associação entre o tempo de experiência e a eficácia do professor. Por outro lado, foi encontrada uma relação negativa e significativa entre professores com especialização – em relação àqueles que possuem somente o nível superior completo.

Já em relação às demais variáveis testadas, algumas chamaram a atenção em razão das evidências de associações positivas entre elas e os efeitos professor em pelo menos uma das amostras analisadas. São elas: o tempo dedicado ao trabalho pedagógico fora da escola, a frequência com que o professor passa lição de casa e o uso dos Cadernos de Apoio e Aprendizagem.

Com base nos critérios sugeridos pelas pesquisas revisadas que se debruçaram sobre a qualidade das estimativas de eficácia dos professores, como os trabalhos de Lockwood, Louis e McCaffrey (2002), McCaffrey et al (2009) e Goldhaber e Hansen (2010), é possível afirmar que os dados utilizados têm uma capacidade limitada para subsidiar recomendações em relação às políticas de pessoal. Segundo o primeiro estudo citado, a razão entre a variância dos erros amostrais e a variância dos efeitos estimados dos professores deveria ser menor que 0,1 para que o *ranking* estimado fosse preciso para distinguir a eficácia dos professores, o que não ocorre neste trabalho, no qual foram obtidos percentuais de 0,28 a 0,35.

Da mesma forma, o coeficiente de estabilidade, estimado em 0,097 em Língua Portuguesa e 0,081 em Matemática, indica uma estabilidade bem menor dos efeitos professor estimados para o presente trabalho que para os estudos revisados, tais como os de McCaffrey et al (2009), que obtiveram correlações nos efeitos professor estimados de 0,22 a 0,67, e Goldhaber e Hansen (2010), que obtiveram correlações de 0,32 a 0,59. A pouca estabilidade indica que um professor considerado altamente eficaz pela medida obtida em 2009 tem grandes chances de não ser considerado altamente eficaz em 2010, o que limita o uso dessas medidas para prever os efeitos professor futuros.

Apesar das limitações, as medidas de eficácia obtidas neste trabalho oferecem evidências que abrem consideráveis possibilidades para pesquisas futuras em relação à eficácia dos professores da rede municipal de ensino de São Paulo.

Foi possível verificar que 12% dos professores tiveram efeitos estimados distintos da média em Língua Portuguesa e cerca de 14% dos professores tiveram efeitos estimados distintos da média em Matemática. Isso evidencia que cada um dos professores desse pequeno percentual pode ser considerado mais ou menos eficaz que os demais professores analisados.

Dada essa distinção, esses professores são o público ideal para que se faça uma análise mais aprofundada acerca de suas práticas docentes na interação com os alunos, como os tipos de atividades propostas para o tempo dos alunos em classe, como o uso de materiais como os Cadernos de Apoio e Aprendizagem, e as atividades propostas para o tempo dos alunos fora da classe; assim como suas práticas docentes solitárias, tais como o planejamento de suas atividades com a turma, sua formação continuada – incluindo os cursos de especialização – e seu trabalho pedagógico fora da sala de aula.

Entende-se que a análise das medidas de eficácia dos professores com base nos resultados dos alunos aliada à observação das práticas docentes permitirá uma melhor compreensão dos fatores envolvidos na eficácia dos professores, subsidiando ações que busquem sua melhoria.

REFERÊNCIAS

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teacher and Student Achievement in Chicago Public High Schools. *Journal of Labor Economics*, 25(1), 95-135.
- Albernaz, A., Ferreira, F., & Franco, C. (2002). Qualidade e equidade no ensino fundamental brasileiro. *Pesquisa e Planejamento Econômico*, 32(3), 453-476.
- Armour, D. T. (1976). *Analysis of the school preferred reading program in selected Los Angeles minority schools*. (R-2007-LAUDS). Santa Monica, CA: Rand Corporation.
- Ballou, D., & Podgursky, M. (2000). Reforming Teacher Preparation and Licensing: What is the Evidence? *Teachers College Record*, 102(1), 5-27.
- Banchero, S., & Kesmodel, D. (2011). Teachers Are Put to the Test: More States Tie Tenure, Bonuses to New Formulas for Measuring Test Scores. *Wall Street Journal*, September 13, 2011.
- Barros, R. P., Mendonça, R., Santos, D. D., & Quintaes, G. (2001). Determinantes do desempenho educacional no Brasil. *Pesquisa e Planejamento Econômico*, 31(1), 1-42.
- Boardman, A. E., & Murnane, R. J. (1979). Using Panel Data to Improve Estimates of the Determinants of Educational Achievement. *Sociology of Education*, 52(2), 113-121.
- Borjas, G. J. (1987) Self-selection and the earnings of immigrants. *American Economic Review*, 77, 531-553.
- Borjas, G. J., & Sueyoshi, G. T. (1994). A two-stage estimator for probit models with structural group effects. *Journal of Econometrics*, 64, 165-182.
- Braun, H. I. (2005) Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models. *A Policy Information Center's Report*. Princeton, NJ: Educational Testing Service (ETS).

- Clotfelter, C., Ladd, H., & Vigdor, J. (2006) Teacher-student matching and the assessment of teacher effectiveness. *The Journal of Human Resources*, 41(4), 778-820.
- Clotfelter, C., Ladd, H., & Vigdor, J. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6), 673-682.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of Educational Opportunity*. Washington: U.S. Government Printing Office.
- Croninger, R., Rice, J. K., Rathbun, A., & Nishio, M. (2007). Teacher qualifications and early learning: effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review*, 26(3), 312-324.
- Darling-Hammond, L. (1999). *Teacher Quality and Student Achievement: a Review of State Policy Evidence*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.
- Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness. *Education Policy Analysis Archives*, 13(42), 1-51.
- Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources*, 32(3), 505-523.
- Goldhaber, D. D., & Hansen, M. (2010). Is It Just a Bad Class? Assessing the Stability of Measured Teacher Performance. *CEDR Working Paper 2010-3*. Seattle, WA: University of Washington.
- Gordon, R., Kane, T. J., & Staiger, D. O. (2006). Identifying Effective Teachers Using Performance on the Job. In Furman, J., & Bordoff, J. E., *Path to Prosperity: Hamilton project ideas on income security, education, and taxes*. Washington: Brookings Institution Press.
- Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: estimation using micro data. *American Economic Review*, 61(2), 280-288.

- Hanushek, E. (1979). Conceptual and Empirical Issues in the Estimation of Educational Production Functions. *Journal of Human Resources*, 14(3), 351–388.
- Hanushek, E. (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, 24(3), 1141-1177.
- Hanushek, E. A. (1992). The tradeoff between child quantity and quality: Some empirical evidence. *Journal of Political Economy*, 100(1), 84-117.
- Hanushek, E. A. (1997). Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis*, 19(2), 141-164.
- Hanushek, E. A. (2003). The Failure of Input-Based Schooling Policies. *The Economic Journal*, 113(485), F64-F98.
- Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004) Why Public Schools Lose Teachers. *Education Next*, 4(1), 77-82.
- Hanushek, E. A., & Rivkin, S. G. (2004). How to Improve the Supply of High-Quality Teachers. *Brookings Papers on Education Policy*, 7, 7-25.
- Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). The market for teacher quality". *NBER Working Paper* 11154.
- Hanushek, E. A., & Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, 46(3), 607–68.
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about Using Value-Added Measures of Teacher Quality. *American Economic Review*, 100(2), 267–271.
- Harris, D., & Sass, T. R. (2006). Value-added models and the measurement of teacher quality. *Unpublished manuscript*.
- Ishii, J., & Rivkin, S. G. (2009). Impediments to the Estimation of Teacher Value Added. *Education Finance and Policy*, 4(4), 520-536.

- Kane, T. J., Rockoff, J. E. & Staiger, D. O. (2006). What does certification tell us about teacher effectiveness? Evidence from New York City. *NBER Working Paper* 12155.
- Kane, T. J. & Staiger, D. O. (2002). The promises and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, 16(4), 91-114.
- Kane, T. J., & Staiger, D. O. (2008). Estimating teacher effects on student achievement: an experimental evaluation. *NBER Working Paper* 14607.
- Koedel, C., & Betts, J. (2007). Re-examining the role of teacher quality in the educational production function. *Working Paper*. San Diego, CA: University of Missouri.
- Koedel, C., & Betts, J. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and Policy*, 6(1), 18-42.
- Levin, H. M., & McEwan, P. J. (2002). Cost-Effectiveness and Educational Policy. In Levin, H. M., & McEwan, P. J., *Cost-Effectiveness and Educational Policy: 2002 Yearbook of the American Education Finance Association*. Larchmont, NY: Eye on Education.
- Lipsky, M. (1980). *Street-level Bureaucracy: dilemmas of the individual in public services*. New York: Russell Sage Foundation.
- Lockwood, J. R., Louis, T. A., & McCaffrey, D. F. (2002). Uncertainty in rank estimation: implications for value-added modeling accountability systems. *Journal of Educational and Behavioral Statistics*, 27(3), 255-270.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND Corporation.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., Louis, T. A., & Hamilton, L. S. (2004). Models for Value-Added Modeling of Teacher Effects. *Journal of Educational and Behavioral Statistics*, 29(1), 67-101.

- McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effects estimates. *Education Finance and Policy*, 4(4), 572-606.
- Murnane, R. J., & Phillips, B. R. (1981). What do effective teachers of inner-city children have in common? *Social Science Research*, 10(1), 83-100.
- Murnane, R. J., Willett, J. B., & Levy, F. (1995). The Growing Importance of Cognitive Skills in Wage Determination. *Review of Economics and Statistics*, 77(2), 251-66.
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teachers effects? *Educational Evaluation and Policy Analysis*, 26(3), 237-257.
- Ravitch, D. (2010). *The death and life of the great American school system: how testing and choice are undermining education*. New York: Basic Books.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: evidence from panel data. *The American Economic Review*, 94(2), 247-252.
- Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2008). Can you recognize an effective teacher when you recruit one? *NBER Working Paper 14485*.
- Rothstein, J. (2009). Student sorting bias in value added estimation: selection on observables and unobservables. *Education Finance and Policy*, 4(4), 537-571.
- Rothstein, J. (2010). Teacher quality in educational production: tracking, decay and student achievement. *The Quarterly Journal of Economics*, 125(1), 175-214.
- Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004). A Potential Outcomes View of Value-Added Assessment in Education. *Journal of Educational and Behavioral Statistics*, 29(1), 103-116.
- Sanders, W. L., & Horn, S. P. (1998). Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research. *Journal of Personnel Evaluation in Education*, 12(3), 247-256.

Sanders, W. L., & Rivers, J. C. (1996). Cumulative and Residual Effects of Teachers on Future Student Academic Achievement. *Research Progress Report*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

São Paulo (Município). Secretaria Municipal de Educação (2005). *Lei nº 14.603, de 14 de outubro de 2005: Institui o Sistema de Avaliação de Aproveitamento Escolar dos Alunos da Rede Municipal de Ensino de São Paulo, sob responsabilidade da Secretaria Municipal de Educação*. Diário Oficial da Cidade de São Paulo, São Paulo, 15 out. 2005. Folha 1.

São Paulo (Município). Secretaria Municipal de Educação (2007). *Matrizes de referência para a avaliação do rendimento escolar / Secretaria Municipal de Educação*. São Paulo: SME.

Soares, J. F. (2005). Qualidade e equidade na educação básica brasileira: fatos e possibilidades. In Brock, C., & Schwartzman, S., *Os Desafios da Educação no Brasil*. Rio de Janeiro: Nova Fronteira, 91-117.

Soares, T. M. (2003). Influência do professor e do ambiente em sala de aula sobre a proficiência alcançada pelos alunos avaliados pelo Simave 2002. *Estudos em Avaliação Educacional*, 28, 103-123.

Todd, P. E. & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485), F3-F33.

United States of America. Department of Education (2009). *Race to the Top Program: Executive Summary*. Disponível em: <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>. Acessado em 07 de dezembro de 2011.

Wright, S. P.; Horn, S. P.; Sanders, W. L. (1997) "Teacher and Classroom Context Effects on Student Achievement: Implications for Teacher Evaluation". *Journal of Personnel Evaluation in Education*, 11, 57-67.

APÊNDICES

Apêndice A – Descrição das Variáveis dos Alunos Contidas na Tabela 11

(continua)

| Questão | Construção da Variável | Nome da Variável | Observação |
|---------------------------------------|---|-------------------------------|--|
| Nota Prévia | Nota em 2009 em Língua Portuguesa ou em Matemática, padronizada para média 0 e desvio-padrão 1 | Nota Prévia | Não há. |
| Sexo | Sexo = F recebe 1; Sexo = M recebe 1 | Feminino | Não há. |
| Idade | Até 11 anos recebe 1; Mais de 11 anos recebe 0 | Idade Adequada | Seguiu-se os critérios utilizados pelo INEP para o cálculo das taxas de distorção idade-série: são “atrasados” os alunos com 2 anos a mais da idade adequada para a matrícula – nesse caso, 10 anos. |
| PIC | Está em turma do PIC recebe 1; não está recebe 0 | Participa do PIC | Consta do banco de dados da Secretaria Municipal. |
| Qual é o grau de escolaridade da mãe? | Ensino fundamental (1ª a 4ª série) completo e Ensino fundamental (5ª a 8ª série) incompleto recebem 1; Demais recebem 0 | Anos Iniciais do EF Completos | Não estudou e ensino fundamental (1ª a 4ª série) incompleto são o <i>default</i> nessa categoria. |
| | Ensino fundamental (5ª a 8ª série) completo e Ensino médio (antigo 2º grau) incompleto recebem 1; Demais recebem 0 | Ensino Fundamental Completo | |
| | Ensino médio (antigo 2º grau) completo e ensino superior (faculdade) incompleto recebem 1; Demais recebem 0 | Ensino Médio Completo | |
| | Ensino superior (faculdade) completo e pós-graduação recebem 1; Demais recebem 0 | Ensino Superior Completo | |

Fonte: Elaboração própria com base nos dados e nos questionários aplicados a pais e alunos na Prova SP do ano de 2010.

Apêndice A – Descrição das Variáveis dos Alunos Contidas na Tabela 11

| | | | (conclusão) |
|---|---|-------------------------------|---|
| Questão | Construção da Variável | Nome da Variável | Observação |
| Qual é o grau de escolaridade do pai? | Ensino fundamental (1ª a 4ª série) completo e Ensino fundamental (5ª a 8ª série) incompleto recebem 1; Demais recebem 0 | Anos Iniciais do EF Completos | |
| | Ensino fundamental (5ª a 8ª série) completo e Ensino médio (antigo 2º grau) incompleto recebem 1; Demais recebem 0 | Ensino Fundamental Completo | Não estudou e ensino fundamental (1ª a 4ª série) incompleto são o <i>default</i> nessa categoria. |
| | Ensino médio (antigo 2º grau) completo e ensino superior (faculdade) incompleto recebem 1; Demais recebem 0 | Ensino Médio Completo | |
| | Ensino superior (faculdade) completo e pós-graduação recebem 1; Demais recebem 0 | Ensino Superior Completo | |
| Qual é a renda familiar, ou seja, a soma dos salários dos que moram com você? | De R\$ 851,00 a R\$ 1275,00 recebe 1; Demais recebem 0 | R\$ 851,00 a R\$ 1275,00 | Até R\$ 850,00 é o <i>default</i> nessa categoria. |
| | De R\$ 1276,00 a R\$ 2125,00 recebe 1; Demais recebem 0 | R\$ 1276,00 a R\$ 2125,00 | |
| | De R\$ 2126,00 a R\$ 4250,00 recebe 1; Demais recebem 0 | R\$ 2126,00 a R\$ 4250,00 | |
| | Mais de R\$ 4250,00 recebe 1; Demais recebem 0 | Mais de R\$ 4250,00 | |

Fonte: Elaboração própria com base nos dados e nos questionários aplicados a pais e alunos na Prova SP do ano de 2010.

Apêndice B – Descrição das Variáveis dos Professores Contidas na Tabela 19

(continua)

| Questão | Construção da Variável | Nome da Variável | Observação |
|--|---|-----------------------------------|---|
| Sexo | Sexo = F recebe 1; Sexo = M recebe 1 | Feminino | Consta do banco de dados da Secretaria Municipal. |
| Qual é a sua cor ou raça? | Branco ou Amarelo recebe 1; Pardo, Preto ou Indígena recebe 0 | Branco ou Amarelo | |
| Qual o seu nível de escolaridade mais alto? | Especialização recebe 1; Superior e Mestrado recebe 0 | Especialização | Nas amostras não restou nenhum professor com doutorado. |
| | Mestrado recebe 1; Superior e Especialização recebe 0 | Mestrado | |
| Há quantos anos você é professor? | 1 a 5 anos recebe 1; Demais recebem 0 | Tem 1 a 5 anos de experiência | Menos de 1 ano de experiência é o <i>default</i> nessa categoria. |
| | 6 a 10 anos recebe 1; Demais recebem 0 | Tem 6 a 10 anos de experiência | |
| | 11 a 15 anos recebe 1; Demais recebem 0 | Tem 11 a 15 anos de experiência | |
| | 16 a 20 anos recebe 1; Demais recebem 0 | Tem 16 a 20 anos de experiência | |
| | Mais de 20 anos recebe 1; Demais recebem 0 | Tem mais de 20 de experiência | |
| Há quanto tempo você é professor nesta escola? | 1 a 5 anos recebe 1; Demais recebem 0 | Está 1 a 5 anos nessa escola | Menos de 1 ano nessa escola é o <i>default</i> nessa categoria. |
| | 6 a 10 anos recebe 1; Demais recebem 0 | Está 6 a 10 anos nessa escola | |
| | 11 a 15 anos recebe 1; Demais recebem 0 | Está 11 a 15 anos nessa escola | |
| | 16 a 20 anos recebe 1; Demais recebem 0 | Está 16 a 20 anos nessa escola | |
| | Mais de 20 anos recebe 1; Demais recebem 0 | Está mais de 20 anos nessa escola | |
| Em quantas escolas você trabalha como professor? | 2, 3, 4 ou mais escolas recebe 1; 1 escola recebe 0 | Trabalha em mais de 1 escola | |

Fonte: Elaboração própria com base nos dados e nos questionários aplicados aos professores na Prova SP do ano de 2010.

Apêndice B – Descrição das Variáveis dos Professores Contidas na Tabela 19

(continuação)

| Questão | Construção da Variável | Nome da Variável | Observação |
|--|--|--|--|
| Você exerce outra atividade profissional? | "Sim, outras" recebe 1; "Sim, na rede municipal de ensino de SP", "Sim, na rede estadual de ensino de SP", "Sim, em outra rede municipal" e "Sim, docência em escola privada" recebem 0 | Tem outra atividade profissional fora da educação básica | Todas as atividades que recebem 0 foram consideradas como educação básica. |
| Quantas horas por semana, fora da escola, você dedica às atividades relacionadas ao trabalho pedagógico (planejamento das aulas, correção de provas e trabalhos etc.)? | 4 a 8 horas semanais recebe 1; Demais recebem 0 Mais de 8 horas semanais recebe 1; Demais recebem 0 | Dedica 4 a 8 horas semanais fora da escola ao trabalho pedagógico Dedica mais de 8 horas semanais fora da escola ao trabalho pedagógico | Menos de 4 horas dedicadas é o <i>default</i> nessa categoria. |
| Com que frequência você passa lição de casa? | Sempre e quase sempre recebem 1; Nunca, quase nunca e às vezes recebem 0 | Passa lição de casa sempre ou quase sempre | |
| Você costuma convidar os pais ou os responsáveis para tratar de assuntos relacionados aos alunos? | Sim recebe 1; Não recebe 0 | Costuma convidar os pais para conversar sobre os filhos | |
| Quanto do conteúdo do programa curricular previsto para este ano letivo você conseguiu desenvolver até este momento? | Entre 50% e 70% recebe 1; Demais recebem 0 Entre 71% e 90% e Mais de 90% recebem 1; Demais recebe 0 | Cumpriu de 50 a 70% do currículo Cumpriu mais de 70% do currículo | Menos de 50% é o <i>default</i> nessa categoria. |
| Quanto tempo da sua aula você gasta para tratar de atividades como organização da turma, chamada, avisos, problemas disciplinares? | Entre 20 e 30 minutos recebe 1; Demais recebem 0 Entre 10 e 20 minutos recebe 1; Demais recebem 0 Menos de 10 minutos e Não gasto tempo com estas atividades recebem 1; Demais recebem 0 | Gasta de 20 a 30 minutos com a organização da turma Gasta de 10 a 20 minutos com a organização da turma Gasta menos de 10 minutos com a organização da turma | Mais de 30 minutos é o <i>default</i> nessa categoria. |

Fonte: Elaboração própria com base nos dados e nos questionários aplicados aos professores na Prova SP do ano de 2010.

Apêndice B – Descrição das Variáveis dos Professores Contidas na Tabela 19

(conclusão)

| Questão | Construção da Variável | Nome da Variável | Observação |
|--|---|--|--|
| Indique quais dos equipamentos e recursos pedagógicos você utiliza em suas aulas. | Sim para computador e acesso à internet recebe 1; Sim só para computador, só para internet ou para nenhum dos dois recebe 0 | Usa computador e internet nas aulas | |
| | Sim para lousa para cópia de textos recebe 1; não recebe 0 | Usa lousa para cópia de textos | |
| (a) Indique quais dos equipamentos e recursos pedagógicos abaixo você utiliza em suas aulas. (b) Caso você tenha respondido sim na questão (a), responda quantas vezes na semana você utiliza os Cadernos de Apoio e Aprendizagem de Língua Portuguesa / Matemática | Sim para Cadernos de Apoio e Aprendizagem de Língua Portuguesa / Matemática na questão (a) e 1 a 2 vezes por semana na questão (b) recebe 1; Demais combinações recebem 0 | Usa Cadernos de Apoio e Aprendizagem de 1 a 2 vezes por semana | Não para Cadernos de Apoio e Aprendizagem de Língua Portuguesa / Matemática na questão (a) é o <i>default</i> nessas categorias. Há três variáveis nesse formato para Língua Portuguesa e três variáveis para Matemática, sendo utilizadas de modo separado para explicar os efeitos professor com base nos resultados de língua portuguesa e matemática, respectivamente. |
| | Sim para Cadernos de Apoio e Aprendizagem de Língua Portuguesa / Matemática na questão (a) e 3 a 4 vezes por semana na questão (b) recebe 1; Demais combinações recebem 0 | Usa Cadernos de Apoio e Aprendizagem de 3 a 4 vezes por semana | |
| | Sim para Cadernos de Apoio e Aprendizagem de Língua Portuguesa / Matemática na questão (a) e todos os dias na questão (b) recebe 1; Demais combinações recebem 0 | Usa Cadernos de Apoio e Aprendizagem todos os dias | |

Fonte: Elaboração própria com base nos dados e nos questionários aplicados aos professores na Prova SP do ano de 2010.

Apêndice C – Resultados das Estimções das Especificações Indicadas na Tabela 10 para Língua Portuguesa

(continua)

| | Especificações do Modelo | | | |
|---|--------------------------|----------|----------|----------|
| | (1) | (2) | (3) | (5) |
| Nota Prévia | - | 0,664*** | 0,644*** | 0,655*** |
| | - | [0,008] | [0,009] | [0,008] |
| Gênero Feminino | 0,231*** | 0,073*** | 0,101*** | 0,093*** |
| | [0,019] | [0,014] | [0,015] | [0,015] |
| Idade Adequada | 0,060** | 0,000 | 0,084*** | 0,088*** |
| | [0,025] | [0,019] | [0,027] | [0,025] |
| Participa do PIC | -0,034* | -0,021 | -0,005 | -0,001 |
| | [0,019] | [0,014] | [0,015] | [0,015] |
| Fez Pré-Escola | -0,483*** | 0,106*** | 0,271 | 0,111*** |
| | [0,029] | [0,023] | [0,226] | [0,027] |
| <u>Escolaridade da Mãe</u> | | | | |
| Anos Iniciais do Ensino Fundamental Completos | -0,116*** | -0,038* | 0,015 | 0,020 |
| | [0,030] | [0,023] | [0,226] | [0,024] |
| Ensino Fundamental Completo | -0,007 | 0,001 | 0,041 | 0,049* |
| | [0,033] | [0,025] | [0,028] | [0,026] |
| Ensino Médio Completo | 0,112*** | 0,039 | 0,071** | 0,076*** |
| | [0,033] | [0,025] | [0,028] | [0,026] |
| Ensino Superior Completo | 0,298*** | 0,125** | 0,151*** | 0,172*** |
| | [0,066] | [0,050] | [0,053] | [0,051] |
| <u>Escolaridade do Pai</u> | | | | |
| Anos Iniciais do Ensino Fundamental Completos | -0,036 | 0,000 | 0,035 | 0,022 |
| | [0,028] | [0,021] | [0,023] | [0,022] |
| Ensino Fundamental Completo | -0,027 | -0,008 | 0,001 | 0,012 |
| | [0,031] | [0,024] | [0,025] | [0,024] |
| Ensino Médio Completo | 0,115*** | 0,048** | 0,049* | 0,053** |
| | [0,032] | [0,024] | [0,026] | [0,024] |
| Ensino Superior Completo | 0,311*** | 0,137*** | 0,129*** | 0,117** |
| | [0,063] | [0,048] | [0,050] | [0,048] |
| <u>Renda Familiar</u> | | | | |
| R\$ 851,00 a R\$ 1275,00 | 0,131*** | 0,061*** | 0,051*** | 0,060*** |
| | [0,023] | [0,017] | [0,018] | [0,017] |
| R\$ 1276,00 a R\$ 2125,00 | 0,295*** | 0,096*** | 0,084*** | 0,089*** |
| | [0,028] | [0,021] | [0,022] | [0,021] |
| R\$ 2126,00 a R\$ 4250,00 | 0,392*** | 0,162*** | 0,142*** | 0,139*** |
| | [0,043] | [0,032] | [0,034] | [0,033] |
| Mais de R\$ 4250,00 | 0,103 | 0,079 | 0,061 | 0,078 |
| | [0,085] | [0,064] | [0,067] | [0,064] |
| Dummies de Professores | Não | Não | Sim | Não |
| Dummies de Escolas | Não | Não | Não | Sim |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2009 e 2010.

Apêndice C – Resultados das Estimações das Especificações Indicadas na Tabela 10 para Matemática

(conclusão)

| | Especificações do Modelo | | | |
|---|--------------------------|-----------|-----------|-----------|
| | (1) | (2) | (3) | (5) |
| Nota Prévia | - | 0,601*** | 0,577*** | 0,594*** |
| | - | [0,009] | [0,010] | [0,009] |
| Gênero Feminino | -0,018 | -0,161*** | -0,135*** | -0,146*** |
| | [0,019] | [0,016] | [0,017] | [0,016] |
| Idade Adequada | 0,075*** | 0,035* | 0,095*** | 0,092*** |
| | [0,025] | [0,020] | [0,030] | [0,027] |
| Participa do PIC | -0,028 | -0,007 | -0,010 | 0,005 |
| | [0,018] | [0,015] | [0,017] | [0,016] |
| Fez Pré-Escola | -0,447*** | 0,091*** | 0,258 | 0,082*** |
| | [0,029] | [0,025] | [0,242] | [0,029] |
| <u>Escolaridade da Mãe</u> | | | | |
| Anos Iniciais do Ensino Fundamental Completos | -0,054* | 0,014 | 0,048* | 0,042 |
| | [0,030] | [0,024] | [0,028] | [0,026] |
| Ensino Fundamental Completo | 0,035 | 0,031 | 0,045 | 0,058** |
| | [0,033] | [0,027] | [0,030] | [0,028] |
| Ensino Médio Completo | 0,175*** | 0,100*** | 0,124*** | 0,114*** |
| | [0,033] | [0,027] | [0,030] | [0,028] |
| Ensino Superior Completo | 0,302*** | 0,135** | 0,179*** | 0,158*** |
| | [0,066] | [0,054] | [0,057] | [0,055] |
| <u>Escolaridade do Pai</u> | | | | |
| Anos Iniciais do Ensino Fundamental Completos | 0,016 | 0,044* | 0,048* | 0,053** |
| | [0,028] | [0,023] | [0,025] | [0,024] |
| Ensino Fundamental Completo | 0,036 | 0,044* | 0,072*** | 0,060** |
| | [0,031] | [0,026] | [0,028] | [0,026] |
| Ensino Médio Completo | 0,149*** | 0,087*** | 0,075*** | 0,081*** |
| | [0,032] | [0,026] | [0,028] | [0,026] |
| Ensino Superior Completo | 0,317*** | 0,163*** | 0,126** | 0,128** |
| | [0,063] | [0,052] | [0,054] | [0,052] |
| <u>Renda Familiar</u> | | | | |
| R\$ 851,00 a R\$ 1275,00 | 0,133*** | 0,070*** | 0,047** | 0,060*** |
| | [0,022] | [0,019] | [0,020] | [0,019] |
| R\$ 1276,00 a R\$ 2125,00 | 0,300*** | 0,124*** | 0,095*** | 0,101*** |
| | [0,028] | [0,023] | [0,024] | [0,023] |
| R\$ 2126,00 a R\$ 4250,00 | 0,413*** | 0,196*** | 0,164*** | 0,160*** |
| | [0,042] | [0,035] | [0,037] | [0,035] |
| Mais de R\$ 4250,00 | 0,149* | 0,111 | 0,083 | 0,113 |
| | [0,084] | [0,069] | [0,072] | [0,069] |
| Dummies de Professores | Não | Não | Sim | Não |
| Dummies de Escolas | Não | Não | Não | Sim |

Fonte: Elaboração própria com base em dados da Prova SP dos anos de 2009 e 2010.