

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE CIÊNCIAS ECONÔMICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA**

**JULIANA CAMARGO**

**O EFEITO DO TAMANHO DA TURMA SOBRE O DESEMPENHO ESCOLAR: UMA  
AVALIAÇÃO DO IMPACTO DA “ENTURMAÇÃO” NO ENSINO FUNDAMENTAL DO  
RIO GRANDE DE SUL**

**Porto Alegre**

**2012**

**JULIANA CAMARGO**

**O EFEITO DO TAMANHO DA TURMA SOBRE O DESEMPENHO ESCOLAR: UMA  
AVALIAÇÃO DO IMPACTO DA “ENTURMAÇÃO” NO ENSINO FUNDAMENTAL DO  
RIO GRANDE DE SUL**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do título de Mestre em Economia, com ênfase em Economia Aplicada.

Orientador: Prof. Dr. Sabino da Silva Pôrto Júnior

**Porto Alegre**

**2012**

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)

Responsável: Biblioteca Gládis Wiebelling do Amaral, Faculdade de Ciências Econômicas da UFRGS

C173e Camargo, Juliana

O efeito do tamanho da turma sobre o desempenho escolar : uma avaliação do impacto da "enturmação" no ensino fundamental do Rio Grande do Sul / Marcelo Chaine. – Porto Alegre, 2012.

67 f. : il.

Orientador: Sabino da Silva Porto Júnior.

Ênfase em Economia Aplicada.

Dissertação (Mestrado em Economia) – Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Programa de Pós-Graduação em Economia, Porto Alegre, 2012.

1. Economia da educação. 2. Qualidade da educação. 3. Modelo econométrico. I. Porto Júnior, Sabino da Silva. II. Universidade Federal do Rio Grande do Sul. Faculdade de Ciências Econômicas. Programa de Pós-Graduação em Economia. III. Título.

CDU 37.015.6

**JULIANA CAMARGO**

**O EFEITO DO TAMANHO DA TURMA SOBRE O DESEMPENHO ESCOLAR: UMA  
AVALIAÇÃO DO IMPACTO DA “ENTURMAÇÃO” NO ENSINO FUNDAMENTAL DO  
RIO GRANDE DE SUL**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do título de Mestre em Economia, com ênfase em Economia Aplicada

Aprovada em: Porto Alegre, 06 de novembro de 2012.

BANCA EXAMINADORA:

---

Prof. Dr. Sabino da Silva Porto Jr. – Orientador  
UFRGS

---

Profª. Dra. Izete Pengo Bagolin  
PUCRS

---

Prof. Dr. Sérgio Marley Modesto Monteiro  
UFRGS

---

Prof. Dr. Stefano Florissi  
UFRGS

*Aos meus pais, José e Silvia Camargo.*

## AGRADECIMENTOS

Esta dissertação é a realização de um sonho. Na verdade, é a extrapolação de um sonho. Durante muitos anos da minha vida sonhei com o dia em que eu ingressaria no Ensino Superior. Muitas vezes tive dúvidas se realmente conseguiria alcançar este objetivo, dadas minhas circunstâncias, mas com muito esforço e dedicação consegui ser aprovada no vestibular da UFRGS para o curso de Ciências Econômicas e, finalmente, realizar o meu sonho. Mal sabia eu que o pior estava por vir. A permanência no curso foi um desafio ainda maior, mas não maior do que meu sonho e minha vontade de concluir minha graduação. Ao longo dessa etapa, obviamente, passei por muitas transformações, inclusive dos meus sonhos, e também por um amadurecimento tanto pessoal quanto profissional. Como resultado desse processo surgiu minha necessidade de aprofundar ainda mais meus conhecimentos e meu novo sonho: ingressar no Mestrado em Economia. Porém, assim como quando entrei na Graduação, mal sabia eu que o pior estava por vir. A adaptação ao novo nível de exigência não foi fácil, mas pude compartilhar essa jornada com muitas pessoas especiais, que me incentivaram e apoiaram, e, portanto, gostaria de agradecer-las.

Agradeço aos meus pais, José Camargo e Silvia Camargo, pelo amor incondicional e por sempre me apoiarem nas minhas decisões. Por todo exemplo pessoal e por sempre me inspirarem a ser uma pessoa melhor. Por sempre torcerem por mim, compreenderem minha ausência e aguentarem a saudade.

Agradeço à minha irmã, Joseane Camargo, pelo amor, pela amizade, pelo companheirismo, por ter me inspirado a me mudar para Porto Alegre e estudar na UFRGS, por relevar meus momentos de mau-humor e desespero, por sempre ter algum assunto diferente pra gente discutir, por sempre tornar minha vida mais alegre, leve e divertida. Agradeço também ao meu irmão, Rogner Camargo, pelo amor, pela amizade, por sempre estar presente mesmo com a distância, pelo exemplo de força, superação e alegria que ele é para mim. Por sempre estar torcendo por mim e me incentivando.

Agradeço ao meu namorado, Marcelo Chaine, por dividir comigo as dificuldades e alegrias do mestrado. Por todo apoio, carinho, compreensão, amor, companheirismo. Por muitas vezes acreditar em mim mais do que eu mesma. Por todo incentivo nas diferentes etapas que vivenciamos durante o curso. Obrigada por ser tão especial.

Agradeço ao meu orientador, Prof. Dr. Sabino da Silva Porto Jr., pela confiança e pela generosidade com que investiu em minha formação. Por todas as oportunidades, desde a Iniciação Científica na graduação, até a carta de recomendação para o Doutorado. Por me ensinar a amar a pesquisa e a docência e a fazê-las com dedicação. Foi um longo percurso e um aprendizado inestimável. Gostaria de agradecer principalmente pela amizade, pelos conselhos, pelas conversas e por todos os “toldos amarelos” que descobrimos durante esses anos. Também agradeço à Aladya Porto pela amizade.

Agradeço aos colegas de mestrado, pela companhia nas tardes de estudo na salinha do PPGE, pela união, pelo companheirismo, pelas amizades, pelos incontáveis cafés, pelos almoços muito divertidos e todos os momentos, tanto os alegres quanto os tensos, que dividimos durante o mestrado.

Agradeço aos meus amigos, dos diferentes círculos que frequento, por terem se somado à minha vida, por muitas vezes compreenderem minha ausência e mesmo com tantas recusas não terem desistido de mim. Obrigada pelo carinho, pelo apoio e por toda torcida. Em especial gostaria de agradecer à Marianne Stampe, pelo apoio durante o mestrado nos momentos mais difíceis pelos quais eu passei. Sem a tua amizade e a tua ajuda eu jamais teria conseguido. Obrigada por tudo.

Finalmente, agradeço ao Programa de Pós-Graduação em Economia da UFRGS, pela formação de qualidade. Obrigada aos professores, pelas ótimas aulas, por dividirem seu conhecimento conosco e pela dedicação. Aos funcionários da secretaria do PPGE, pela eficiência e solicitude de sempre. E agradeço à UFRGS como um todo, por ter me aberto as portas, feito parte da minha vida, por ter sido minha casa (literalmente) durante anos e por ter tornado meu sonho possível.

## RESUMO

A compreensão dos aspectos determinantes da qualidade educacional é uma tarefa difícil. Existe grande controvérsia na literatura acerca de quais variáveis de fato impactam na qualidade do ensino, principalmente no que se refere aos insumos escolares. O tamanho da turma tem estado no centro do debate, pois normalmente é considerado uma variável mais fácil de manipular. Porém, existe grande dificuldade em se mensurar o efeito do número de alunos por classe na proficiência dos mesmos. O objetivo deste estudo é investigar o impacto do tamanho da classe no desempenho dos alunos de escolas públicas estaduais no estado do Rio Grande do Sul. Para tanto, utiliza-se os métodos de Variáveis Instrumentais e Regressão com Descontinuidade para explorar uma variação exógena na regra de organização das turmas no ano de 2007. Os resultados obtidos mostram que não há evidência de impacto estatisticamente significativo da política de “enturmação” na proficiência dos alunos.

**Palavras-chave:** Educação. Qualidade da Educação. Tamanho da Turma.

## **ABSTRACT**

Understanding the key determinants of educational quality is a difficult task. There is no consensus in the literature about which variables actually impact the quality of education, especially when the investigation involves school inputs. The class size has been the key variable in the debate, because it is usually considered easier to manipulate. However, there is great difficulty in measuring the effect of the number of students in a class on their achievement. The objective of this study is to investigate the impact of class size on students' performance in public schools in the state of Rio Grande do Sul. Therefore, it was used the Instrumental Variables and Regression Discontinuity Design methods to explore an exogenous variation in the rule of organization of classes in the year of 2007. The results show that there is no evidence of statistically significant impact of the policy called "enturmação" on students' achievement.

**Keywords:** Education. Quality Education. Class Size.

## LISTA DE GRÁFICOS

Gráfico 1: Impacto de uma melhora de 0,5 d.p. no desempenho em testes padronizados sobre o crescimento do PIB per capita.....	14
Gráfico 2: Distribuição de alunos por turma para a quarta série – RS .....	47
Gráfico 3: Distribuição de alunos por turma para a oitava série – RS.....	47

## LISTA DE FIGURAS

Figura 1: Exemplos de gráficos com Sharp Regression Discontinuity Design .....	37
Figura 2: Probabilidade de Vencer a Eleição dada a Eleição Passada .....	38

## **LISTA DE TABELAS**

Tabela 1: Estatísticas descritivas .....	48
Tabela 2: Resultados das Estimações para quarta série .....	52
Tabela 3: Resultados das Estimações para oitava série .....	53

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>10</b>
<b>2</b>	<b>DETERMINANTES DA QUALIDADE DA EDUCAÇÃO .....</b>	<b>13</b>
2.1	Tamanho da Turma .....	16
2.2	A Política de Enturmação.....	20
2.3	Estratégia Empírica.....	22
<b>3</b>	<b>METODOLOGIA.....</b>	<b>26</b>
3.1	Variáveis Instrumentais .....	27
3.1.1	Abordagem de Mínimos Quadrados em Dois Estágios .....	28
3.1.2	Local Average Treatment Effects (LATE) .....	29
3.2	Desenho de Regressão com Descontinuidade (RDD).....	33
3.2.1	Sharp RD .....	35
3.2.2	Fuzzy RD .....	39
3.3	Base de Dados.....	45
3.3.1	Estatísticas Descritivas .....	46
<b>4</b>	<b>RESULTADOS .....</b>	<b>50</b>
<b>5</b>	<b>CONCLUSÃO .....</b>	<b>55</b>
	<b>REFERÊNCIAS .....</b>	<b>57</b>

## 1 INTRODUÇÃO

A importância dada ao valor econômico da Educação sofreu, segundo Bowmann (1966), uma verdadeira revolução a partir de 1960, quando se passou a considerar investimentos em educação como sendo investimentos em Capital Humano. Essa nova forma de avaliação influenciou governos, planejadores, agências internacionais e educadores do mundo inteiro.

Além disso, existem externalidades decorrentes dos investimentos em educação que trazem aumento do bem-estar social que, apesar das dificuldades de serem mensuradas, vem sendo abordadas pela literatura. De Walque (2004), sugere que, segundo a teoria do Capital Humano, educação é um investimento no futuro, dessa forma, os indivíduos aperfeiçoam suas análises na hora de tomarem decisões.

De acordo com Becker (1993), quando se emprega o termo “capital”, este se associa, normalmente, a uma idéia monetária. Entretanto, existem outras formas de capital, uma delas é o Capital Humano, que é, para muitos autores, o produto do investimento no próprio indivíduo.

Economistas sabem há muito tempo que as pessoas são parte importante da riqueza das nações, porém, devido aos valores e crenças vigentes, teóricos como John Stuart Mill e Alfred Marshall não enfatizaram o fato de que as pessoas investem nelas mesmas, pois não seria aceitável tratar o ser humano como bem de capital. No entanto, Adam Smith já afirmava que todas as habilidades adquiridas e úteis deveriam ser consideradas partes do capital.

Gastos diretos em educação, saúde, treinamento e migração constituem formas de capital humano, uma vez que não é possível separar a pessoa do seu conhecimento, das suas habilidades, da sua saúde ou dos valores que ela possui.

Capital Humano é tanto um elemento qualitativo quanto quantitativo. Qualitativo no sentido de que o indivíduo adquire habilidades, conhecimento, que contribuem para o seu bem-estar e quantitativo porque, ao adquirir essas habilidades, o indivíduo melhora a sua produtividade, sendo esta refletida na renda real do trabalhador.

Dessa forma, investimentos em Capital Humano se justificam, pois impactam de forma positiva em várias dimensões econômicas e sociais: aumento de salários, diminuição da criminalidade, melhoria nas condições de saúde e diminuição da probabilidade de ficar desempregado. Além disso, para o país como um todo, uma população mais educada traz um

crescimento econômico maior, aumenta a produtividade das empresas, e potencializa os efeitos da globalização (Menezes Filho, 2007).

A partir da década de 90 houve um maior esforço por parte dos formuladores de políticas públicas a fim de ampliar o acesso ao ensino fundamental. Entretanto, com as políticas de universalização do ensino básico, a grande preocupação atual reside na qualidade da educação que é ofertada. Apesar de existir certo consenso entre os economistas em relação à importância dos investimentos em educação, os determinantes da qualidade da educação geram certa controvérsia. Enquanto alguns trabalhos defendem a tese de que não existe relação entre o aumento de gastos em insumos escolares e o desempenho escolar (HANUSHEK, 1986 e 1998), outros argumentam que essa relação existe e é grande o suficiente para ser levada em consideração (HEDGES *et al.*, 1994; KRUEGER, 2003).

Uma das variáveis sobre a qual os pesquisadores têm depositado atenção é o tamanho da turma. Talvez por ela ser vista como um insumo mais fácil de se modificar. Porém, há divergência nos resultados dos estudos investigando os impactos do tamanho da turma na proficiência dos alunos. Uma das maiores dificuldades em se elaborar pesquisas sobre o tamanho da turma é a existência de endogeneidade, ou seja, esse insumo escolar pode estar correlacionado com fatores não observáveis que interferem no desempenho dos alunos. Por esse motivo é necessário elaborar estratégias de identificação do efeito causal do tamanho da classe sobre a proficiência. É necessário ter em mente que o processo de decisão de aumento de gastos em políticas públicas deve considerar a escassez dos recursos disponíveis, ou seja, a questão não é somente identificar se o aumento de gastos em insumos escolares impacta no desempenho dos alunos, mas também se os benefícios justificam os custos; ou ainda, se é possível realocar os insumos existentes, considerando que não seja possível aumentar gastos, de forma a conseguir melhor utilizar os recursos financeiros disponíveis e melhorar (ou simplesmente manter) a qualidade da educação.

A fim de contribuir com a literatura dos determinantes da qualidade da educação, este trabalho investiga o impacto do tamanho da classe no desempenho dos alunos para escolas públicas estaduais no estado do Rio Grande do Sul. Além da presente introdução, o segundo capítulo traz a apresentação da discussão sobre determinantes da educação e sobre o tamanho da turma. Além disso, é detalhada nesse capítulo a política “enturmação” que será analisada, em

seguida é apresentada a estratégia empírica. O terceiro capítulo apresenta de forma mais aprofundada a metodologia econométrica utilizada e discute as principais dificuldades em se estimar o impacto de programas ou políticas públicas tratadas na literatura e as estratégias utilizadas para que seja possível realizar tais avaliações. É apresentado o problema do viés de seleção e os métodos utilizados nesse trabalho para mensurar o impacto da política de “enturmação”: Local Average Treatment Effects (LATE) e Desenho de Regressão com Descontinuidade. O quarto capítulo expõe algumas estatísticas descritivas e os resultados encontrados neste estudo. Por fim, são realizadas as considerações e discussões finais na conclusão, indicando possíveis problemas que podem comprometer os resultados encontrados e pesquisas futuras.

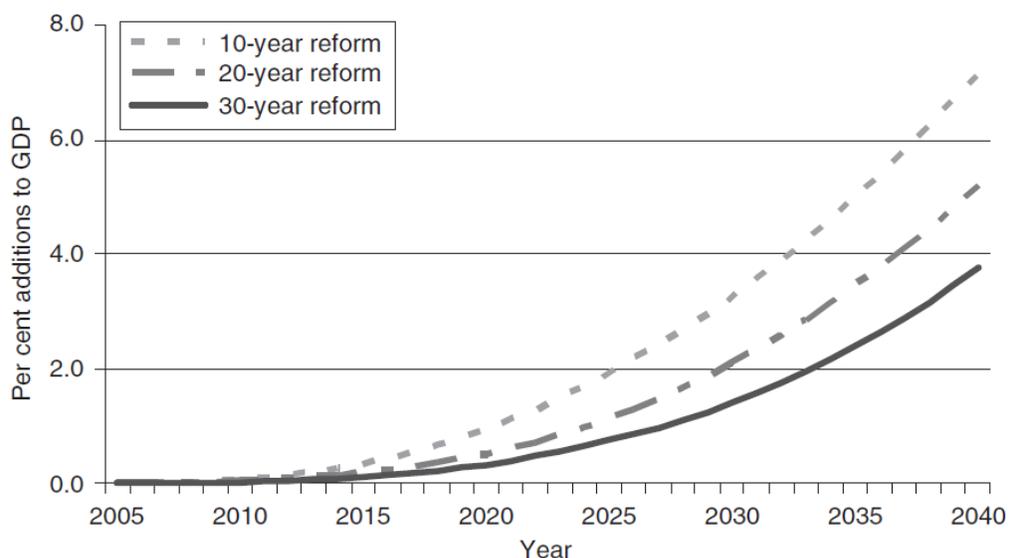
## 2 DETERMINANTES DA QUALIDADE DA EDUCAÇÃO

Apesar dos retornos dos investimentos em Capital Humano serem privados, uma justificativa para o governo investir em Capital Humano é o reflexo do mesmo no bem-estar social. Além disso, à medida que a educação aumenta capacidade produtiva da população, ela passa a ser um investimento não somente para o indivíduo, mas para a sociedade como um todo: uma sociedade mais educada possui maiores taxas de invenção e assimilação de novas tecnologias, fatores determinantes para o crescimento econômico de longo prazo (HANUSHEK, 2006). Dessa forma, investimentos em educação se justificam por seus benefícios – tanto monetários quanto não monetários. Muitos estudos mostram que educação promove saúde, reduz o ato de fumar, aumenta a propensão a votar, melhora o conhecimento a respeito do controle de natalidade e estimula a apreciação de música clássica, literatura e até mesmo tênis.

As políticas e reformas estruturais adotadas no Brasil na década de 90, segundo Bagolin e Pôrto Júnior (2003), resultaram em diversas mudanças na economia e nos indicadores sociais. Uma das políticas que caracterizou fortemente esse período foi a universalização do ensino básico, que atingiu 97% das crianças em idade escolar. Entretanto, além dos aspectos quantitativos, é preciso analisar os aspectos qualitativos, ou seja, melhorar a qualidade da educação que é oferecida para os alunos nas escolas públicas brasileiras é o próximo passo a ser tomado pelos formuladores de políticas públicas. Neste sentido, a relação entre qualidade da educação e resultados econômicos é evidenciada por estudos como o de Hanushek e Kimki (2000) e o de Card e Krueger (1990). Este conclui que a qualidade escolar explica a maior parte da variação da renda dos indivíduos.

Em um estudo posterior, Hanushek (2002) ainda afirma que investimentos públicos em melhorias de qualidade teriam efeitos de longo prazo que compensariam os gastos correntes. Como pode ser visto no gráfico apresentado pelo autor, mesmo que a reforma demore anos para surtir efeito, ela tem efeito significativo de longo prazo sobre o PIB.

**Gráfico 1: Impacto de uma melhora de 0,5 d.p. no desempenho em testes padronizados sobre o crescimento do PIB per capita**



Fonte: Hanushek, 2002.

Além dos benefícios econômicos e diferenciais de renda, a decisão de o aluno permanecer estudando também pode ser influenciada pela qualidade da educação. Utilizando dados de escolas primárias do Egito, Hanushek, Lavy e Hitomi (2006) encontram evidências de que um aprimoramento da qualidade das escolas reduziria a evasão em até dois terços.

A compreensão dos aspectos determinantes da qualidade educacional, todavia, é uma tarefa difícil. Existe grande controvérsia na literatura acerca de quais variáveis de fato impactam na qualidade do ensino. Para Ferreira (2008), o debate tem sido em torno da significância de diferentes variáveis referentes às escolas e aos alunos no aprendizado, tais como salas mais equipadas, laboratórios, menos alunos por sala de aula, o impacto dos professores e do background familiar e socioeconômico.

A discussão sobre insumos escolares e seus impactos no desempenho dos alunos teve seu início provavelmente no ano de 1966, com o Coleman Report, para o qual os insumos escolares não foram relevantes na explicação do desempenho escolar e os principais determinantes do desempenho foram características familiares dos próprios alunos e de seus colegas de escola.

Diversos estudos desde então buscam compreender a relação entre a qualidade da escola e o aprendizado do aluno. Entretanto, mesmo nos casos onde as características da escola se mostraram mais importantes, a magnitude do impacto é bastante inferior ao impacto proporcionado pelas variáveis referentes à família dos estudantes. Barros et al (2001) e Albernaz, Ferreira e Franco (2002), por exemplo, encontraram para o Brasil resultados de que os atributos escolares têm pouco efeito sobre o desempenho escolar se comparados aos impactos das características familiares. Outros trabalhos desenvolvidos por Fuller (1986), Hanushek (1986), Hanushek and Harbison (1992), Hanushek (1995) mostram que os insumos relacionados à escola não impactam sistematicamente nos resultados dos alunos. “...research demonstrates that the traditional approach to providing more quality – simply providing more inputs - is frequently ineffective” (Hanushek, 1995).

Entretanto, Kremer (1995) argumenta que a maioria dos estudos realizados não obtiveram resultados significativos porque ignoraram problemas existentes nos modelos, como viés de variável omitida e endogeneidade. O autor argumenta que os mesmos não são baseados em variações aleatórias, o que acaba gerando os problemas anteriormente citados. O principal problema é que não raro os insumos escolares são correlacionados com outros determinantes do desempenho escolar que não são observados.

Considerando tais problemas para obter estimativas não-viesadas, estudos mais recentes, Case and Yogo (1999), Betts (1996), Card and Krueger (1996), buscam explorar variações exógenas nos insumos escolares a fim de identificar efeitos causais dos mesmos e gerar evidências mais confiáveis sobre este assunto. Dentre os recursos escolares estudados seguindo essa linha de pesquisa, o tamanho da turma tem sido o principal centro do debate. Porém, não há consenso na literatura acerca dos impactos do tamanho da turma sobre o desempenho dos alunos. Existem resultados destes impactos tanto positivos quanto negativos. Por exemplo, Krueger (1999) verifica um melhor desempenho dos alunos em turmas menores, entretanto, Hanushek (2002) ressalta que o valor adicionado para turmas menores é praticamente nulo, exceto para os alunos que frequentam o primeiro ano de ensino. Portanto, na próxima seção é feita uma investigação mais profunda sobre os resultados encontrados na literatura recente de impacto do tamanho da turma.

## 2.1 Tamanho da Turma

De forma geral, pais e professores dos alunos preferem turmas menores. Talvez pelo fato deles acreditarem que classes menores estimulam o aprendizado do aluno, ou talvez simplesmente porque classes menores oferecem um ambiente mais agradável para alunos e professores (Muller, Chase e Walden, 1988). Um estudo realizado por Chatman (1997) mostrou que os tamanhos das turmas podem variar de acordo com o tipo de interação que é necessária na sala de aula, isto é, disciplinas que requerem participação ativa dos alunos são menores. O que remete ao argumento de que em turmas menores o aprendizado é mais eficiente, pois os alunos estão processando ativamente o conteúdo, ao contrário das turmas maiores em que os alunos simplesmente ouvem e leem o conteúdo passivamente (McKeachie, 1990). Entretanto, esse é apenas um ponto de vista no debate sobre a relação entre tamanho da turma e desempenho dos alunos.

Conceitualmente, existem duas abordagens que supõem diferentes relações entre o tamanho da turma e o desempenho dos alunos: a construtivista e a *behaviorista*. De acordo com a teoria construtivista, classes menores são mais eficientes porque possibilitam maior participação dos alunos em aula e mais interação com os colegas. Por outro lado, a abordagem *behaviorista* acredita que classes maiores podem ser eficientes ao passo que o fator principal para o desempenho do aluno é o comportamento do professor.

Segundo estudos citados pelo Banco Mundial, o tamanho da turma não incide ou tem uma incidência pouco significativa sobre o rendimento escolar: acima de vinte alunos por sala não faz diferença se são trinta, cinquenta ou mais (Torres apud Tommasi, Warde, 1998). Porém, para Duso (2009), em uma turma que possui um número elevado de alunos ficam prejudicados o atendimento individualizado, a aprendizagem, a avaliação e a interação professor-aluno e aluno-aluno.

O tamanho da turma normalmente é considerado um insumo escolar mais fácil de manipular. Assim sendo, é uma das principais variáveis a ser considerada no debate acerca de políticas focadas na qualidade educacional e na alocação dos recursos escolares em muitos países: Estados Unidos, Robinson (1990); Reino Unido, OFSTED (1995) e Israel, Moshel-Ravid (1995).

Não obstante, a tarefa de mensurar os efeitos causais do tamanho da turma no desempenho dos alunos não é algo trivial. Embora o nível de insumos escolares varie tanto entre quanto dentro das próprias escolas, essas diferenças estão frequentemente associadas a outros fatores como, por exemplo, background socioeconômico do aluno. Possivelmente esta seja uma explicação para o grande número de estudos inconclusivos na literatura sobre esse tema.

Glass and Smith (1979) e Glass, Cahen, Smith, and Filby (1982) em sua meta-análise de tamanho da turma concluem que classes menores melhoram o desempenho dos alunos nos testes. Além disso, Card and Krueger (1992a) e Card and Krueger (1992b) encontram que classes com menor taxa de alunos por professor estão associadas a maiores salários futuros dos alunos.

Em seu estudo, Krueger (1999) utiliza dados do programa STAR (Student Teacher Achievement Ratio), através do qual é possível analisar o efeito do tamanho da sala de aula baseado em um experimento natural, e conclui que salas de aula menores melhoram o desempenho dos alunos. Mais do que isso, o impacto é ainda maior nos alunos considerados mais vulneráveis: assistir aulas em turmas menores diminuiu o gap no desempenho de alunos brancos e negros. Pode-se pensar, então, que talvez essas políticas de diminuição de turmas devessem ser focadas para escolas que atendem alunos em maior situação de vulnerabilidade social. Ao encontro desses estudos também estão os resultados de Case and Deaton (1999) e Angrist and Lavy (1999) nos quais turmas menores impactam positivamente no desempenho do aluno.

Por outro lado, Hoxby (2000) em seu experimento natural não encontra nenhum efeito significativo do tamanho da turma no desempenho escolar dos alunos utilizando dados para os Estados Unidos. Além disso, o autor ressalta o fato dos estudos que encontram resultados significativos para tamanho da turma analisarem o impacto apenas nos primeiros anos escolares. Para as séries subsequentes os efeitos não são adequadamente pesquisados. Em relação ao efeito do tamanho da turma nos primeiros anos de ensino, Blatchford e Mortimore (1994) realizam uma revisão da literatura sobre tamanho da turma e concluem que *“there is now firm evidence of a link, but only in the early years and only with classes smaller than 20. The evidence supports the reduction of class sizes in the first years of school, especially with disadvantaged pupils, but much still needs to be researched”* (Blatchford e Mortimore, 1994). Outro ponto pouco estudado é o efeito do tamanho da sala em países em desenvolvimento.

Asadullah (2002) investiga os efeitos do tamanho da sala de aula no desempenho escolar dos alunos em Bangladesh e conclui que o impacto da política adotada visando diminuir o número de alunos por sala de aula é estatisticamente insignificante no desempenho escolar dos alunos. *“Our findings thus indicate that the new literature that has claimed to identify a negative effect of class size on student achievement should be interpreted with care”* (Asadullah, 2002). Assim como Alam (2000), que em seu estudo sobre o impacto do tamanho da turma no desempenho dos alunos em Bangladesh conclui que uma redução das mesmas é ineficiente nas séries mais avançadas<sup>1</sup>.

Para Ofsted (1995) o insumo que exerce maior influencia no desempenho dos alunos não é o tamanho da turma, mas sim a qualidade do professor em sala de aula. Nesse sentido, Day *et al.* (1996) afirma que o aprendizado em sala de aula é um processo complexo e que o mesmo é resultado da interação de diversas variáveis *“With so many variables involved in the teaching-learning process it is difficult to be certain that the findings are attributable to class size alone, rather than to the cumulative and aggregative effect of other variables in the teaching-learning process and school environment working in association with each other”* (Day *et al.*, 1996). Portanto, uma dificuldade encontrada na literatura é conseguir isolar o que seria atribuído somente ao efeito do tamanho da turma no aprendizado dos alunos.

Um estudo desenvolvido por Bourke (1986) utiliza dados de turmas do quinto ano de matemática da Austrália e mostra que alunos com maior habilidade estão alocados em turmas maiores. Além disso, o estudo se preocupou em analisar as práticas desenvolvidas pelos professores em diferentes turmas. Ele mostra que a forma com que o professor trabalha os conteúdos em sala de aula varia de acordo com as habilidades dos alunos e com o tamanho da turma. O autor afirma que os maiores efeitos encontrados na investigação realizada por ele são os da habilidade dos alunos e das práticas dos professores, sendo o primeiro ainda mais substancial.

Shapson (1977, 1980) utiliza dados de um experimento realizado no Canada para turmas do quarto e quinto ano de ensino. O autor analisa, assim como o estudo de Bourke (1986), diferentes variáveis relacionadas com a infraestrutura da sala de aula e com as diferentes práticas adotadas pelos professores para transmitir o conteúdo para os alunos. Ele conclui que o tamanho

---

<sup>1</sup> O efeito de salas menores nas series primárias só foi significativo para meninos no teste de matemática.

de sala tem pouca relação com as atividades desenvolvidas em sala de aula e que há pouca relação entre tamanho da turma e aprendizado dos alunos.

Dados os estudos anteriormente realizados envolvendo a avaliação do impacto do tamanho da turma no desempenho dos alunos, algumas reflexões se fazem necessárias: esses efeitos dependem de outros fatores como, por exemplo, a qualidade do professor; é necessário levar em consideração que tal política exige um maior número de professores a serem alocados nas novas turmas (muitas vezes professores inexperientes são contratados para suprir essa demanda e os impactos na proficiência dos alunos acabam sendo negativos) e adequada capacidade física da rede de ensino para acomodar as turmas menores criadas; os diferenciais na proficiência dos alunos que frequentaram turmas menores em relação aos que não frequentaram diminuem com o tempo e a magnitude do impacto varia de acordo com a série em que esta política é implementada; os critérios utilizados para formar as turmas também podem interferir nos resultados da pesquisa – o diretor pode preferir colocar somente alunos com maior dificuldade de aprendizado nas turmas menores, ou então pode preferir arranjar somente os alunos com melhor desempenho nas mesmas; a composição da turma também pode superestimar ou subestimar os efeitos da política, dependendo de como os alunos reagem à interação com os colegas, ou seja, é muito difícil conseguir decompor o efeito e separar o que de fato é devido a redução da turma do que é atribuído aos *peer effects*.

Dito isso, ao contrário da maioria dos estudos desenvolvidos acerca desta literatura, o presente trabalho se propõe a investigar o efeito do acréscimo de alunos na sala de aula; não a redução das turmas. De acordo com Oliveira (2008), o acréscimo de um aluno na classe reduz a quantidade de insumo professor destinado a cada aluno individualmente. Mais do que isso, o aprendizado pode ser prejudicado pelo aumento das interrupções devido à indisciplina, o atendimento das necessidades específicas de cada aluno é praticamente inviável e, além disso, os professores podem se sentir sobrecarregados no trabalho devido ao tamanho da classe, sendo esse um incentivo às faltas e a diminuição do esforço por parte dos mesmos, impactando negativamente a proficiência. Por outro lado, o acréscimo de um aluno na sala de aula, sendo ele diferente de pelo menos um dos demais, pode gerar o que se chama de *learning spillovers*. Ou seja, se de fato existe uma assimetria em que os piores alunos se beneficiam mais da exposição a bons alunos do que o contrário (Hoxby, 2000), as externalidades geradas pelo acréscimo de um novo aluno provavelmente serão positivas.

Não raro os formuladores de políticas públicas se deparam com escassez de recursos e devem tomar decisões acerca de quais insumos priorizar para melhorar a qualidade do ensino, ou simplesmente para conseguir manter a mesma qualidade dado o seu baixo orçamento. Por esse motivo, a preocupação não deve ser apenas identificar se o aumento dos gastos em insumos escolares impacta na proficiência dos alunos, mas também se os benefícios justificam esses custos, assim como se é possível realocar os insumos existentes, considerando que não seja possível aumentar gastos, de forma a conseguir melhor utilizar os recursos financeiros disponíveis e melhorar a qualidade da educação.

A fim de contribuir com a literatura dos determinantes da qualidade da educação, este trabalho investiga o impacto do tamanho da classe no desempenho dos alunos para escolas públicas estaduais no estado do Rio Grande do Sul. As características da política a ser analisada e a estratégia empírica serão detalhadas na próxima seção.

## **2.2 A Política de Enturmação**

O governo do estado do Rio Grande do Sul iniciou o ano de 2007 promovendo, a partir de fevereiro, ações com foco na qualidade da educação. A Secretaria Estadual da Educação (Seduc) iniciou a reorganização da rede de ensino recompondo as Coordenadorias Regionais de Educação (CREs) – “cada coordenadoria é responsável pelas políticas relacionadas às suas regiões, tendo como atribuições coordenar, orientar e supervisionar escolas oferecendo suporte administrativo e pedagógico para a viabilização das políticas da secretaria estadual” (secretaria da educação) – além de remanejar professores que estavam alocados em funções de apoio pedagógico para, efetivamente, a docência. Outra importante medida tomada pela Secretaria foi a reorganização das turmas.

A partir de maio de 2007, as CREs, orientadas pelo Departamento de Recursos Humanos (DRH/Seduc), desenvolveram um estudo para analisar o número de alunos por turma das escolas públicas estaduais. Com dados coletados nas escolas, foi analisado o número de estudantes nas salas de aula, escola por escola, município por município. No documento divulgado pela Secretaria Estadual de Educação “SAIBA TUDO SOBRE A "ENTURMAÇÃO"” foram apresentadas evidências de que seria possível reduzir, na mesma série, de três turmas para duas

ou de duas para uma, por exemplo. Das quase 54 mil turmas no estado, pelo menos 5 mil continham um número pequeno de alunos. Assim, seguindo as recomendações dos atos normativos do Conselho Estadual de Educação (CEED/RS) para a distribuição de alunos por sala de aula, a partir de julho de 2007 as turmas foram rearrumadas. A Secretaria de Educação divulgou em agosto desse ano que, de acordo com o levantamento realizado pelo Departamento de Recursos Humanos, o resultado da “enturmação” foi que das 2.856 escolas, 745 haviam reorganizado suas turmas entre julho e agosto de 2007 - o que representou 26% do total de estabelecimentos de ensino. O processo seguiu os pareceres do Conselho Estadual de Educação (Ceed), os quais determinam o número máximo de 50 alunos por sala de aula. Para o Ensino Fundamental, o número máximo utilizado como critério foi 40 alunos e no Ensino Médio foi de 45 estudantes por sala de aula. Esses mesmos critérios foram mantidos nos anos subsequentes para a organização das turmas, conforme a divulgação do site da Secretaria Estadual de Educação.

Mesmo nas turmas reorganizadas, o Estado mantém em funcionamento: no Ensino Fundamental, 338 turmas com até 20 alunos (12 turmas com até 10 alunos), e, no Ensino Médio, 09 turmas com até 20 estudantes (duas turmas com até 02 alunos). Também, feitas as reorganizações, as turmas ficaram assim: no Ensino Fundamental, 344 turmas com 21 a 30 alunos, 339 turmas com 31 a 35 alunos, e 78 com 36 a 40 alunos; no Ensino Médio, são 180 turmas com 21 a 30 alunos, 731 turmas com 31 a 40 alunos, e 49 turmas com 41 a 45 alunos. Ao todo, foram reduzidas 1.590 turmas (3%) da rede estadual, que passa a ter 51.455 turmas. (Secretaria da Educação do Estado do Rio Grande do Sul, 2007.)

A importância desse novo critério de divisão das turmas para o presente trabalho é o fato do mesmo ser uma variação na regra de alocação dos alunos que é exógena aos diretores das escolas. Como foi dito anteriormente, essa regra foi amplamente divulgada pela Secretaria Estadual de Educação e as CREs acompanharam o processo de implementação, tanto em 2007 quanto nos anos seguintes. Essa mudança de regra gera uma fonte de variação potencialmente exógena no tamanho das turmas que pode ser usada para estimar o efeito do tamanho da classe sobre o desempenho dos alunos. Para melhor explicitar de que forma essa variação ocorre, seguindo os critérios estipulados pela Secretaria de Educação é possível notar que o tamanho da

turma cresce numa proporção de um-para-um em relação ao número de matrículas até que é atingido o número máximo de 40 alunos, porém, quando o número de matrículas passa a ser de 41 em diante há uma descontinuidade nessa relação e pode ser verificada uma queda acentuada no tamanho das turmas, para uma média de (20,5) alunos. O mesmo ocorre quando o número de matrículas chega a 80 alunos por série, o número médio de alunos por turma quando as matrículas passam de 81 estudantes cai para (27) alunos por sala, e assim por diante.

A próxima seção apresenta a estratégia empírica que será utilizada neste exercício para se tentar captar o efeito da “enturmação” no desempenho dos alunos.

### 2.3 Estratégia Empírica

O objetivo deste trabalho é avaliar os impactos do tamanho da turma no desempenho dos alunos. Para isso, é realizada uma tentativa de captar os efeitos que o aumento do número de alunos por turma tem sobre o desempenho dos mesmos. O desempenho dos alunos será mensurado utilizando as notas médias por turma do Prova Brasil de 2007, que será melhor explicado mais adiante.

Os impactos que se pretende avaliar serão modelados de acordo com o modelo causal proposto por Rubin (1974). Assim sendo, para um dado período de tempo, cada indivíduo  $i$  possui dois resultados potenciais que assumem diferentes valores dependendo da participação ou não do indivíduo em um tratamento, que é representado pela variável indicadora  $T_i$ . Dessa forma, os resultados potenciais que o indivíduo  $i$  pode assumir são representados por:

$$\left\{ \begin{array}{l} Y_i(0), \text{ se } T_i = 0 \\ Y_i(1), \text{ se } T_i = 1 \end{array} \right\}$$

É possível, portanto, calcular o efeito causal da intervenção para cada indivíduo através de uma simples diferença entre os resultados potenciais:

$$\tau_i = Y_i(1) - Y_i(0) \quad (1)$$

Assim como é possível, através dessa abordagem, calcular o efeito médio do tratamento (ATE) para toda a população:

$$ATE \equiv E[\tau_i] = E[Y_i(1) - Y_i(0)] = E[Y_i(1)] - E[Y_i(0)] \quad (2)$$

Entretanto, existe uma dificuldade para a implementação prática desse modelo. É trivial notar que um único indivíduo não pode ter a relação de participação e não participação de um tratamento ao mesmo tempo, portanto, é possível observar apenas um dos resultados potenciais do mesmo. Dessa forma, só é possível identificar os dois últimos momentos populacionais em (2) para a subpopulação efetivamente tratada ( $E[Y_i(1)|T_i = 1]$ ) e para a não tratada ( $E[Y_i(0)|T_i = 0]$ ).

Diante de tal limitação, não é possível calcular, em geral, o ATE pela simples diferença de médias, pois, como mostrado abaixo, o resultado não será simplesmente o parâmetro do ATE:

$$\begin{aligned} E[Y_i|T_i = 1] - E[Y_i|T_i = 0] &= E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 0] = \\ &= \underbrace{E[Y_i(1) - Y_i(0)|T_i = 1]}_{ATT} + \underbrace{[E[Y_i(0)|T_i = 1] - E[Y_i(0)|T_i = 0]]}_{\text{viés}} \end{aligned} \quad (3)$$

Dessa forma, ao realizar a diferença de médias de  $Y_i$  entre tratados e não tratados recuperamos o parâmetro do ATT (efeito médio do tratamento sobre a população tratada) mais um segundo termo, que é o chamado viés de seleção. Este está captando diferenças pré-tratamento tanto para os indivíduos selecionados quanto para os não selecionados para o tratamento.

Para que o ATT seja igual ao ATE é necessário que esse viés de seleção seja igual a zero. Ou seja, é preciso que ambos os grupos sejam, em média, iguais.

Na prática, dificilmente esse viés será igual a zero, porém, se a escolha dos participantes do tratamento for independente dos resultados potenciais dos mesmos, ou seja, das suas características pré-tratamento, então é possível recuperar o ATE:

$$\begin{aligned} (Y_i(0), Y_i(1)) \perp T_i &\Rightarrow E[Y_i(1) - Y_i(0)|T_i = 1] + [E[Y_i(0)|T_i = 1] - E[Y_i(0)|T_i = 0]] = \\ E[Y_i(1) - Y_i(0)] + E[Y_i(0)] - E[Y_i(0)] &= E[Y_i(1) - Y_i(0)] = \text{ATE} \end{aligned} \quad (4)$$

Assim, o ATE pode ser identificado por uma diferença de médias dos resultados observados para tratados e não tratados. Sem essa condição, para se estimar o efeito do tratamento são necessárias mais hipóteses relacionadas ao mecanismo de seleção dos indivíduos para o tratamento.

De acordo com Angrist (2004), a hipótese de independência dos resultados potenciais na prática é pouco realista, uma vez que as políticas normalmente são focadas em um grupo específico de indivíduos e as características dos mesmos são levadas em consideração para a participação ou não do tratamento. Nesse caso, duas estratégias são usualmente adotadas para se estimar os efeitos causais na presença de um possível viés de variável omitida. Uma delas supõe que condicional nas covariadas,  $X_i$ , o regressor de interesse,  $T_i$ , é independente dos resultados potenciais. Então, qualquer efeito causal de interesse pode ser estimado a partir de comparações ponderadas condicionais em  $X_i$ . Entretanto, essa é uma suposição forte que parece mais plausível quando os pesquisadores têm uma quantidade considerável de informações prévias sobre o processo de determinação do  $T_i$ . Alternativamente, a segunda estratégia procura utilizar uma variável instrumental que, muitas vezes condicional nas covariadas, é relacionada com  $T_i$ , mas independente dos resultados potenciais.

Este trabalho se baseia na função de tamanho de turma construída por Angrist e Lavy (1999) para construir variáveis instrumentais e, assim, estimar os efeitos do tamanho da turma. De acordo com os autores, embora a função do tamanho da classe e os instrumentos dela derivados sejam uma função dos cortes utilizados para o número de matrículas, elas podem ser utilizadas, uma vez que essas funções são não-lineares e não-monotônicas (essa justificativa será melhor detalhada mais adiante). Dessa forma, utilizando a regra como um instrumento, é possível obter estimativas do impacto do tamanho da turma controlando outros possíveis efeitos que a

matrícula poderia ter sobre o desempenho dos alunos. Esse tipo de estratégia empírica é uma aplicação do método desenvolvido por Campbell (1969), o Desenho de Regressão com Descontinuidade, para avaliações dos impactos do tamanho da turma.<sup>2</sup>

Um ponto interessante é o fato das turmas terem sido reorganizadas na metade do ano letivo. Muitas vezes uma dificuldade em se aplicar o método da regressão com descontinuidade é a possibilidade de manipulação da regra. Ou seja, as pessoas antecipam que haverá uma certa política e manipulam as informações de modo a fazer parte ou não de determinado programa. Assim, como as informações sobre as matrículas já estavam disponíveis, fica mais difícil algum tipo de manipulação da regra por parte das diretoras das escolas ou até mesmo por parte dos pais dos alunos. Poderia ser argumentado que os pais mais preocupados com a qualidade da educação dos seus filhos poderiam transferir os mesmos para escolas com menos matrículas e turmas menores, mas normalmente o critério de aceitação nas escolas públicas é a proximidade do local de residência, além disso, a transferência de uma escola para outra na metade do ano letivo também seria mais complicado. Portanto, é pouco provável que tenha existido manipulação da regra suficiente para comprometer os resultados.

O próximo capítulo detalha as metodologias empregadas neste trabalho.

---

<sup>2</sup> O trabalho de van der Klauww (1996) também utiliza regressão com descontinuidade para investigações econômicas. Assim como Akerhielm (1995), que utilize matrículas como um instrumento para o tamanho da turma.

### 3 METODOLOGIA

O maior desafio da avaliação de impacto é determinar o que teria acontecido aos beneficiários se o programa não tivesse existido. Nesse caso, o resultado do beneficiário sem a existência do programa seria o que a literatura chama de contrafactual. Assim, o problema enfrentado pela avaliação é o fato de que o impacto da intervenção poderia ser calculado simplesmente pela diferença entre o resultado atual do indivíduo e o resultado do seu contrafactual, porém, se o indivíduo participou do programa, o resultado do contrafactual não é observado. Isso porque, como foi apresentado no capítulo anterior, é impossível que um indivíduo participe ao mesmo tempo do grupo que recebe o tratamento e do grupo que não recebe. Ou seja, na prática é possível observar apenas um dos resultados do indivíduo, não sendo possível observar ao mesmo tempo o seu contrafactual. Dessa forma, encontrar um contrafactual apropriado é o maior desafio da literatura de avaliação de impacto.

Esse tipo de estudo é essencialmente um problema de dados faltantes (*missing*), pois não é possível observar os resultados dos contrafactuais. Sem essa informação, a melhor alternativa é comparar os resultados de indivíduos que receberam o programa, chamados de tratados, com os de um grupo de comparação que não tenha sido tratado. Ao adotar essa estratégia é preciso então escolher um grupo de comparação que seja semelhante ao grupo tratado, de tal forma que os indivíduos tratados, caso não tivessem recebido o tratamento, teriam resultados similares aos observados no grupo de comparação.

Um primeiro esforço para se tentar mensurar o impacto de um programa poderia ser uma equação da seguinte forma:

$$Y_i = \alpha X_i + \beta T_i + \varepsilon_i \quad (5)$$

Onde  $T_i$  é uma dummy igual a 1 para aqueles que participam e 0 para os indivíduos que não participam.  $X$  representa características observáveis dos indivíduos, ou até mesmo da família e do ambiente em que os mesmos vivem.  $\varepsilon$  é um termo de erro que reflete características não observadas que também afetam o  $Y$ .

Entretanto, essa abordagem possui um problema grave que é o fato da seleção para o tratamento não ser aleatória: *i*) o programa pode ser propositalmente direcionado a um grupo de indivíduos com características específicas; *ii*) pode haver auto-seleção para o programa. A auto-seleção pode ser baseada em características observáveis; fatos não observáveis ou até mesmo em ambos. No caso em que ela ocorre por fatos não observáveis, o termo de erro da equação (5) irá conter variáveis que também são correlacionadas com a dummy de tratamento  $T$ . Esse fato causa o que na literatura é conhecido como o viés de variável omitida. Ou seja,  $\text{cov}(T, \varepsilon) \neq 0$ . Assim sendo, ocorre a violação de uma das principais hipóteses para a estimação não viesada utilizando Mínimos Quadrados Ordinários – os regressores são independentes do termo de erro  $\varepsilon$ . Como consequência, a estimação do impacto do programa,  $\beta$ , também fica viesada.

Nesse sentido, o modelo de resultados potenciais apresentado no capítulo anterior traz esse problema em uma abordagem mais conceitual. Como fora dito anteriormente, para se estimar o efeito do tratamento e eliminar o problema do viés são necessárias mais hipóteses relacionadas ao mecanismo de seleção dos indivíduos para o tratamento.

A abordagem de Variáveis Instrumentais supõe que o viés de variável omitida é corrigido ao ser utilizada uma variável (ou instrumento) que é correlacionado com a participação no programa, mas não é correlacionado com características não observadas que afetam o resultado,  $Y$ . Esse instrumento é utilizado para prever a participação. Já o Desenho de Regressão com Descontinuidade é considerado uma extensão dos métodos de Variável Instrumental e Experimental. A próxima seção detalha cada um desses métodos e suas respectivas hipóteses.

### **3.1 Variáveis Instrumentais**

O método de Variáveis Instrumentais (IV) consiste em utilizar uma variável que seja altamente correlacionada com a participação no programa, mas que não seja correlacionada com características não observadas que afetam os resultados dos indivíduos. Tais instrumentos podem ser construídos baseados na forma como o programa foi implementado (por exemplo, a seleção para os indivíduos a serem tratados foi aleatória, ou então se regras exógenas foram utilizadas para determinar quais indivíduos seriam os selecionados para participar do programa).

Portanto, a abordagem IV tem como objetivo eliminar a correlação entre  $T$  e  $\varepsilon$ . Para tanto, é preciso encontrar um instrumento, chamado de  $Z$ , que satisfaça as seguintes características:

- 1) Correlacionado com  $T$ :  $cov Z, T \neq 0$
- 2) Não-correlacionado com  $\varepsilon$ :  $cov Z, \varepsilon = 0$

Assim, o instrumento  $Z$  afeta a seleção dos indivíduos para o programa, mas não está correlacionado com fatores que afetam os resultados.

### 3.1.1 Abordagem de Mínimos Quadrados em Dois Estágios

Para isolar a parte da variável de tratamento que é independente de outras características não observadas que afetam o desempenho dos indivíduos é realizada uma regressão utilizando o tratamento como variável dependente e o instrumento como regressor. Esse procedimento é conhecido como o primeiro estágio da regressão.

$$T_i = \gamma Z_i + \phi X_i + u_i \quad (6)$$

Sendo que  $X_i$  representa um vetor de variáveis covariadas utilizadas para controle e  $u_i$  é o termo de erro.

Como resultado desse primeiro estágio é obtido o chamado tratamento previsto,  $\hat{T}$ , que representa a parte do tratamento afetada apenas pelo instrumento,  $Z$ , e que, portanto, incorpora apenas a variação exógena do tratamento. No passo seguinte, esse valor previsto,  $\hat{T}$ , é utilizado no lugar de  $T$  na equação (5) para criar a equação da forma reduzida da regressão:

$$Y_i = \alpha X_i + \beta(\hat{\gamma}Z_i + \hat{\phi}X_i + u_i) + \varepsilon_i \quad (7)$$

Assim, a estimação de IV do impacto do programa é recuperada através do parâmetro  $\hat{\beta}_{IV}$ . Como por hipótese  $cov Z, \varepsilon = 0$ , é possível ver que o  $\beta$  pode ser escrito como:

$$\frac{cov(Y, Z)}{cov(T, Z)} \quad (8)$$

Portanto,

$$cov(Y_i, Z_i) = cov[(\beta T_i + \varepsilon_i), Z_i] = \beta cov(T_i, Z_i) \quad (9)$$

$$\Rightarrow \frac{cov(Y_i, Z_i)}{cov(T_i, Z_i)} = \beta. \quad (10)$$

Dessa forma, se as hipóteses  $cov Z, T \neq 0$  e  $cov Z, \varepsilon = 0$  se mantêm, então IV identifica consistentemente o impacto médio do programa atribuído ao instrumento.

Entretanto, existe um problema que deve ser considerado nessa abordagem. Caso os indivíduos saibam identificar quais são os seus respectivos ganhos esperados ao participar do programa, eles estão antecipando ganhos do programa que os avaliadores não conseguem observar. Dessa forma, ocorre uma seleção na participação do programa que não é observada, uma vez que os indivíduos que antecipam que têm maiores retornos ao participar do programa, dadas as suas características  $X$ , são os mais propensos a participar do mesmo. Como o instrumento  $Z$  afeta a participação, essas características não observadas que influenciam na participação também se correlacionam com  $Z$ , o que torna a estimação viesada. Esse é o problema conhecido como o viés de auto-seleção.

Na próxima seção será então apresentada a abordagem desenvolvida para lidar com esse problema identificado na literatura.

### 3.1.2 Local Average Treatment Effects (LATE)

O problema de auto-seleção em não-observáveis, como citado anteriormente, acaba viesando também as estimativas de Variáveis Instrumentais. Uma solução proposta por Imbens e

Angrist (1994) é a abordagem do *Local Average Treatment Effect (LATE)*. Nesse caso, há heterogeneidade nos resultados dos indivíduos ao programa e o efeito do programa é consistentemente estimado por IV apenas para aqueles indivíduos cuja participação muda devido a mudanças no instrumento,  $Z$ . Ou seja, o LATE é uma estimação do efeito médio do tratamento apenas naqueles indivíduos que decidiram participar do programa devido a uma mudança em  $Z$ .

Assim, o LATE evita o problema do viés de auto-seleção em não-observáveis limitando a análise a indivíduos cujo comportamento é alterado por mudanças locais em  $Z$ , sendo que tais mudanças não estão relacionadas com os resultados potenciais. Em outras palavras, é o efeito médio do tratamento (ATE) para aqueles indivíduos em quem um instrumento  $Z$  afeta apenas numa direção a probabilidade de tratamento.

De acordo com Angrist (2004), as hipóteses de identificação do LATE são:

- 1) Independência –  $[Y_{0i}, Y_{1i}, T_{0i}, T_{1i}] \perp\!\!\!\perp Z_i$ .
- 2) Primeiro estágio –  $P(T_i = 1|Z_i = 1) \neq P(T_i = 1|Z_i = 0)$
- 3) Monotonicidade –  $T_{1i} \geq T_{0i} \forall_i$  ou *vice versa*.

Essas hipóteses capturam a ideia de que o instrumento: *i*) é “tão bom quanto aleatorização” (independência); *ii*) afeta a probabilidade do tratamento (primeiro-estágio); *iii*) afeta a todos da mesma forma (monotonicidade).

Em relação à monotonicidade, Khandker et al (2010) afirma que tal hipótese garante que um aumento em  $Z$  de  $Z = z$  para  $Z = z'$  faz com que algum indivíduo participe do programa, mas não que algum indivíduo saia do programa devido a esse aumento. Assim, a participação  $T$  depende de certos valores de  $Z$ , de tal forma que  $P(T = 1|Z = z)$  é a probabilidade de participar do programa quando  $Z = z$  e  $P(T = 1|Z = z')$  é a probabilidade de participar quando  $Z = z'$ . Essa hipótese faz com que não seja possível ocorrer a existência de *defiers*.

Os quatro possíveis tipos de indivíduos nessa abordagem são:

Tipos =	{	T(1)	T(0)	T	Tipo
		1	1	1	(a) always taker
		1	0	se $Z = 1, T = 1$	(c) complier
		1	0	se $Z = 0, T = 0$	(c) complier
		0	1	se $Z = 1, T = 0$	(d) defier
		0	1	se $Z = 0, T = 1$	(d) defier
0	0	0	(n) never – taker		

Dessa forma, com as hipóteses citadas acima, Imbens e Angrist (1994) mostram que é possível identificar o efeito médio do tratamento para os *compliers*:

$$\text{LATE} = \frac{E(Y_i|Z_i=1) - E(Y_i|Z_i=0)}{E(T_i|Z_i=1) - E(T_i|Z_i=0)} = E(Y_{1i} - Y_{0i} | T_{1i} > T_{0i}) \quad (11)$$

O lado esquerdo da equação,  $\frac{E(Y_i|Z_i=1) - E(Y_i|Z_i=0)}{E(T_i|Z_i=1) - E(T_i|Z_i=0)}$ , é análogo ao estimador de Wald (1940) para regressões com erro de mensuração. O *Local Average Treatment Effect* (LATE) no lado direito da equação,  $E(Y_{1i} - Y_{0i} | T_{1i} > T_{0i})$ , é o efeito do tratamento naqueles cujo *status* do tratamento é alterado pelo instrumento, ou seja, a população para a qual  $T_{1i} = 1$  e  $T_{0i} = 0$ <sup>3</sup>

Entretanto, é preciso observar atentamente o resultado obtido através do LATE e perceber o fato de que em um mundo mais realista, no qual os resultados para cada indivíduo variam, só é possível, através de IV, capturar o efeito em uma subpopulação específica. Em outras palavras, somente naqueles indivíduos cujo *status* do tratamento pode ser alterado ao se manipular  $Z_i$ , como dito anteriormente,  $T_{1i} = 1$  e  $T_{0i} = 0$ , ou ainda  $T_{1i} - T_{0i} = 1$ .

Segundo Angrist *et al.* (1996) os indivíduos representados por  $T_{1i} - T_{0i} = 1$  são a população de *compliers*. Eles são os indivíduos que obedecem o seu *status* de tratamento dado pelo planejador do programa. De forma mais clara, se o indivíduo foi sorteado, por exemplo, para participar do programa ele participará do programa; caso ele não tenha sido sorteado ele não participará. Esses indivíduos obedecem a regra de participação.

<sup>3</sup> A prova desse resultado pode ser vista em Angrist (2004).

Para esses indivíduos específicos os valores médios de  $Y_{1i}$  e  $Y_{0i}$ , assim como a diferença média,  $Y_{1i} - Y_{0i}$ , são identificados. Assim sendo, Abadie (2002) mostra que

$$\frac{E(Y_i T_i | Z_i = 1) - E(Y_i T_i | Z_i = 0)}{E(T_i | Z_i = 1) - E(T_i | Z_i = 0)} = E(Y_{1i} | T_{1i} > T_{0i}) \quad (12)$$

e

$$\frac{E(Y_i (1 - T_i) | Z_i = 1) - E(Y_i (1 - T_i) | Z_i = 0)}{E((1 - T_i) | Z_i = 1) - E((1 - T_i) | Z_i = 0)} = E(Y_{0i} | T_{1i} > T_{0i}) \quad (13)$$

Portanto, as distribuições de  $Y_{1i}$  e  $Y_{0i}$  são similarmente identificadas, fato utilizado pelo autor para estimar o efeito causal do tratamento nos resultados potenciais dos *compliers*.

De forma análoga, é possível considerar  $P(T = 1 | Z = z)$  e  $P(T = 1 | Z = z')$  as probabilidades de participar do tratamento dado o instrumento  $Z$ . Portanto,  $P(z)$  e  $P(z')$ . Consequentemente, é possível identificar o  $\beta_{IV,LATE}$  como:

$$\beta_{IV,LATE} = \frac{E(Y | P(Z) = P(z)) - E(Y | P(Z) = P(z'))}{P(z) - P(z')} \quad (14)$$

Através da equação (14) é possível estimar o LATE utilizando o método de variáveis instrumentais (IV). No primeiro estágio, a participação no programa,  $T$ , é estimada como uma função do instrumento,  $Z$ , para se obter a probabilidade estimada:

$$\hat{P}(Z) = \hat{P}(T = 1 | Z) \quad (15)$$

No segundo estágio é utilizada essa probabilidade estimada,  $\hat{P}(Z)$ , como regressor do produto,  $Y_i = [T_i \cdot Y_i(1) + (1 - T_i)Y_i(0)]$ , em uma regressão linear. A interpretação desse  $\hat{\beta}_{IV}$ , ou seja, do efeito estimado do programa, é que esse parâmetro mostra o quanto varia em média os resultados,  $Y$ , em resposta a uma variação na probabilidade de participação,  $\hat{P}(Z)$ , mantendo fixas outras covariadas observadas,  $X$ .

### 3.2 Desenho de Regressão com Descontinuidade (RDD)

Descontinuidades na implementação de programas, baseadas em critérios de elegibilidade ou em outros fatores exógenos podem ser muito úteis para se estimar os efeitos do programa. A abordagem de Regressão com Descontinuidade explora o fato de serem conhecidas as regras que determinam um tratamento. Em outras palavras, a identificação do Desenho de Regressão com Descontinuidade (RDD) é baseada na idéia de que, em um mundo predominantemente baseado em regras, algumas delas são arbitrárias e, por esse motivo, geram bons experimentos (Angrist; Pischke, 2008). Isto é, as regras de elegibilidade para o tratamento podem ser utilizadas como um instrumento para identificar exogeneamente quem serão os tratados e os não tratados. Para estabelecer tal comparação, são utilizados indivíduos que estejam próximos da vizinhança desse limiar que divide os indivíduos em dois grupos – o de tratados e o de não tratados. Ou seja, a amostra relevante para se estimar o impacto do programa são os indivíduos tratados e não tratados que estejam próximos desse ponto de descontinuidade (*threshold*, ou *cutoff*).

Portanto, indivíduos abaixo e acima do *threshold*, supondo que ambos são similares em características observáveis, podem ser distinguidos em termos de seus respectivos resultados. Entretanto, para que seja possível realizar tal comparação, é necessário que a amostra a ser comparada seja suficientemente próxima do *cutoff*.

Os primeiros desenhos de regressão com descontinuidade foram utilizados por Thistlethwaite and Campbell (1960), com o intuito de estimar o efeito de bolsas dadas aos alunos, tendo como critério o desempenho dos mesmos no exame de admissão, nos seus resultados acadêmicos futuros. Esse estudo explora o fato de que as bolsas são alocadas com base na nota de um teste, que é observada. Estudantes com notas no teste, representadas por  $Z$ , acima ou igual ao valor do *cutoff*,  $c$ , recebem a bolsa, aqueles com nota abaixo do valor do *cutoff* não a recebem.

Esse mecanismo de alocação das bolsas gera uma descontinuidade, que foi explorada pelos autores.

A partir da década de 90, a utilização dessa metodologia se tornou popular e há um número crescente de estudos que a utilizam para avaliar efeitos de programas em diversos contextos econômicos Van der Klaauw (2002), Black (1999), and Angrist and Lavy (1999). Hahn *et al.* (2001) formalizou em seu trabalho a utilização do método de RDD na literatura de avaliação de impacto, incorporando a linguagem da literatura nessa nova estratégia de mensuração do efeito de um programa.

As abordagens utilizando descontinuidade são similares às de variáveis instrumentais (IV) no sentido de que introduzem uma variável exógena que é altamente correlacionada com a participação. Ambos os métodos, portanto, exploram essa variação exógena para mensurar os impactos de uma política ou de um programa, por exemplo.

Porém, se os indivíduos não forem capazes de manipular precisamente a *forcing variable*,  $Z$ , então a variação no tratamento existente próxima ao *threshold* é aleatória, como se fosse uma variação obtida através de um experimento. Isto é, mesmo que algum indivíduo esteja mais propenso a ter valores de  $Z$  mais próximos ao *cutoff*, todo indivíduo terá aproximadamente a mesma probabilidade de ter um  $Z$  que esteja logo acima (sendo considerado tratado) ou logo abaixo (sendo considerado não tratado) do *cutoff*. Portanto, a variação que o RDD isola é aleatória pelo fato de que os indivíduos têm um controle impreciso sobre a *forcing variable*. O que é uma vantagem dessa abordagem sobre a de IV, pois essa última supõe que o instrumento é gerado aleatoriamente, hipótese essa mais difícil de se justificar – a menos que de fato o instrumento seja obtido através de uma loteria – (Lee; Lemieux, 2009).

Dessa forma, as estimações por regressão descontínua (RDD), assim como as de variáveis instrumentais, utilizam o fato da alocação do tratamento ser determinada por uma variável,  $Z$ . Porém, no caso da RDD, a probabilidade do indivíduo ser tratado varia de forma discreta a partir de um ponto de corte,  $Z_i = z_0$ . Caso esse ponto de corte tenha sido fixado exogeneamente, um salto também discreto no valor esperado da variável de resultado pode ser considerado o efeito causal do tratamento (Imbens; Wooldridge, 2003). Assim, em um ponto de descontinuidade  $z_0$  a probabilidade de ser tratado dá um salto discreto.

Tal que:

$$T^+ \equiv \lim_{\delta \downarrow 0} \Pr[T = 1 / Z_i = z_0 + \delta]$$

$$T^- \equiv \lim_{\delta \downarrow 0} \Pr[T = 1 / Z_i = z_0 - \delta]$$

$$T^+ \neq T^-$$

Tal que  $\delta$  é a vizinhança de  $z_0$ . Esse salto discreto será a fonte de variação exógena. O objetivo é possibilitar a identificação do efeito do tratamento médio não mais na população como um todo, mas na vizinhança da descontinuidade. Ou seja, tal metodologia explora o conhecimento de regras exógenas que determinam o tratamento para estimar o efeito do tratamento para uma sub-amostra específica.

Há dois tipos de Regressão com Descontinuidade: *Sharp RD (SRD)* e *Fuzzy RD (FRD)*. O que difere entre os dois é o tipo de relação entre  $Z_i$  e  $T_i$ . No caso da *SRD*,  $Z_i$  determina totalmente a alocação do tratamento; enquanto na *FRD*,  $Z_i$  determina parcialmente o *status* do tratamento.

### 3.2.1 Sharp RD

Sharp RD é utilizada quando o status do tratamento é uma função determinística e descontínua de uma covariada,  $Z_i$  (Angrist; Pischke, 2008). Ou seja:

$$T_i = \begin{cases} 1 & \text{se } Z_i \geq z_0 \\ 0 & \text{se } Z_i < z_0 \end{cases} \quad (16)$$

Onde  $z_0$  é o *cutoff*. Esse mecanismo é uma função determinística de  $Z_i$  porque uma vez que se sabe o valor do  $Z_i$  conseqüentemente também se sabe o valor de  $T_i$ . Tal função também é chamada descontínua porque não importa quão próximo  $Z_i$  esteja de  $z_0$  o tratamento não se altera até o ponto em que  $Z_i = z_0$ .

Nesse caso, o parâmetro populacional é dado por:

$$E[\tau_i^{srd}] = E[Y_i(1) - Y_i(0) | Z_i = z_0] = E[Y_i(1) | Z_i = z_0] - E[Y_i(0) | Z_i = z_0] \quad (17)$$

Como pode ser visto na equação (17) o objetivo é estimar a diferença média dos resultados potenciais entre tratados e controles para o subgrupo de tratados que está exatamente sob o ponto de *threshold*,  $z_0$ . Entretanto, esse valor  $E[Y_i(0) | Z_i = z_0]$  não pode ser observado e, portanto, para estimá-lo é preciso usar valores próximos de  $z_0$ . Para tanto, é necessária a hipótese de que as expectativas dos resultados potenciais sejam funções contínuas nos pontos de corte (Imbens; Wooldridge, 2003). Ou seja:  $E(Y(0)|Z = z_0)$  e  $E(Y(1)|Z = z_0)$ , são contínuas em  $z_0$  (hipótese de continuidade da expectativa condicional dos resultados potenciais).

Caso essa hipótese seja válida, é possível reescrever a equação (17) como:

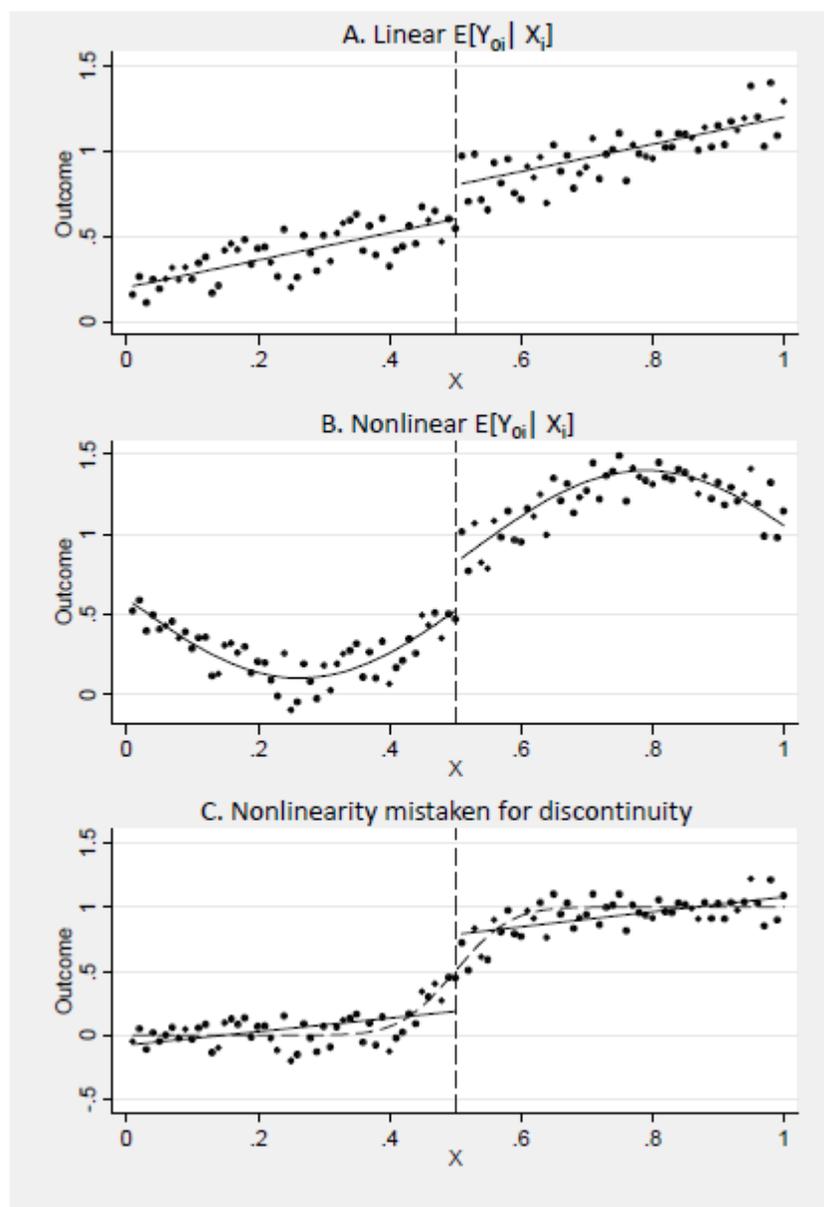
$$E[\tau_i^{srd}] = \lim_{z \rightarrow z_0^+} E[Y_i | Z_i = z_0] - \lim_{z \rightarrow z_0^-} E[Y_i | Z_i = z_0] \quad (18)$$

Essa equação é, portanto, a estimação de duas funções – uma antes e outra depois do *cutoff* – e a inferência da diferença entre essas duas regressões no ponto de corte. É comum também em estimações do ATE por regressão descontínua utilizar a análise gráfica. De acordo com o “*checklist*” desenvolvido por Lee e Lemieux (2009), para a implementação do RD é importante apresentar os gráficos contendo médias locais para diferentes especificações polinomiais.

A Figura 1 mostra um cenário hipotético de *SRD* no qual os indivíduos com  $x_i \geq 0,5$  são tratados ( $x_i$  é a *forcing variable*). No painel A está especificada uma relação linear entre  $Y_i$  e  $x_i$ . No painel B essa relação é não-linear. Em ambos os casos há uma descontinuidade na relação entre  $E[Y_{0i} | x_i]$  e  $x_i$  em torno do ponto  $x_0$ . Já o painel C mostra uma descontinuidade devido a um erro de especificação. Ou seja, o que parece um salto devido ao efeito do programa na verdade é consequência de se ter atribuído de forma equivocada uma relação linear entre  $Y_i$  e  $x_i$ . Por essa razão, na prática, são utilizados valores muito próximos da descontinuidade para

minimizar esse erro de especificação. Além disso, são testados diferentes graus de polinômio para a especificação da forma funcional.

**Figura 1: Exemplos de gráficos com Sharp Regression Discontinuity Design**

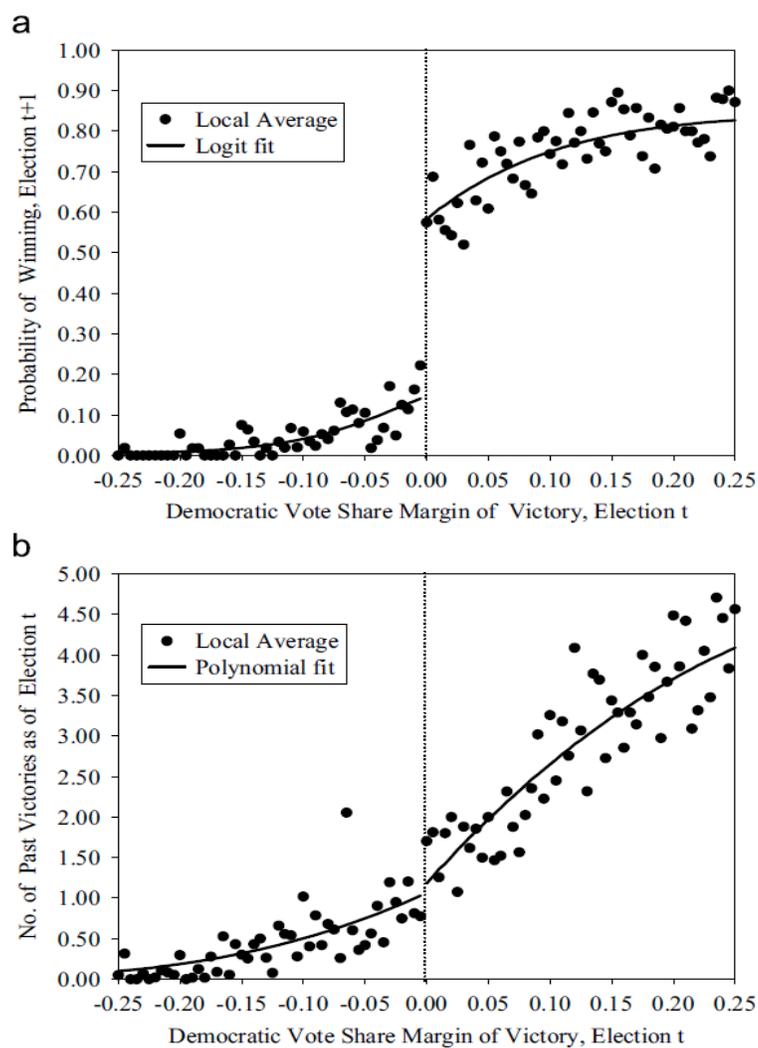


**Fonte:** Angrist; Pischke, 2008

Já a Figura 2 mostra exemplos de implementações práticas do *SRD*. A figura 3.a mostra a probabilidade de um democrata se reeleger, que é uma função crescente da proporção de votos da

eleição passada que o candidato democrata detinha em relação à proporção de votos do candidato republicano. A figura 2.b testa a descontinuidade para a eleição anterior à última eleição. É esperado que não haja descontinuidade, mas esse é um artifício comumente utilizado para se testar se de fato a descontinuidade verificada não tem relação com outros fatores. Em geral também se testa a *SRD* para covariadas para verificar que de fato a descontinuidade não é devido a outros fatores.

**Figura 2: Probabilidade de Vencer a Eleição dada a Eleição Passada**



Fonte: (Lee, 2008)

### 3.2.2 Fuzzy RD

No caso da *Fuzzy RD (FRD)* a probabilidade de receber o tratamento não precisa necessariamente mudar de zero para um no ponto de corte. O desenho permite que haja um salto menor do que de zero para um no ponto de descontinuidade:

$$\lim_{z \downarrow z_0} \Pr(T_i = 1 | Z_i = z_0) \neq \lim_{z \uparrow z_0} \Pr(T_i = 1 | Z_i = z_0) \quad (19)$$

Tal situação pode ocorrer, por exemplo, no caso em que os incentivos para participar de um programa mudam de forma descontínua em um ponto de corte, porém, tais incentivos não são suficientemente eficientes a ponto de fazer com que todos os indivíduos naquele *cutoff* mudem seu *status* de não participante do programa para participante.

Nesse caso o ATE é definido como:

$$\tau_{FRD} = \frac{\lim_{z \downarrow z_0} E(Y_i | Z_i = z_0) - \lim_{z \uparrow z_0} E(Y_i | Z_i = z_0)}{\lim_{z \downarrow z_0} E(T_i | Z_i = z_0) - \lim_{z \uparrow z_0} E(T_i | Z_i = z_0)} \quad (20)$$

Além da hipótese de continuidade, o estimador de *FRD* exige que mais duas hipóteses sejam feitas:

2) Monotonicidade -  $T_i(z)$  é não crescente em  $z_0$  quando  $z_0 = c$ .

Este é um conceito similar ao utilizado no LATE para a definição do *compliers*. Assim:

$$\lim_{z_0 \downarrow z_i} T_i(z_0) = 0 \quad e \quad \lim_{z_0 \uparrow z_i} T_i(z_0) = 1$$

Nesse caso, os *compliers* são os indivíduos que receberão o tratamento se o *cutoff* estiver em um ponto igual a  $Z_i$  ou abaixo dele, e não receberão o tratamento caso o *cutoff* esteja em um ponto acima de  $Z_i$ .

Portanto:

$$\tau_{FRD} = \frac{\lim_{z_0 \downarrow c} E(Y_i | Z_i = z_0) - \lim_{z_0 \uparrow c} E(Y_i | Z_i = z_0)}{\lim_{z_0 \downarrow c} E(T_i | Z_i = z_0) - \lim_{z_0 \uparrow c} E(T_i | Z_i = z_0)}$$

e

$$E(Y_i(1) - Y_i(0) | \text{indivíduo } i \text{ é um complier e } Z_i = c) \quad (21)$$

Dessa forma, o estimador capta o efeito médio do tratamento, porém, apenas para indivíduos próximos à vizinhança do *threshold* e somente para os *compliers*.

Outra hipótese de identificação importante é a de que o ponto de corte  $z_0$  tenha sido escolhido arbitrariamente, chamada de hipótese de *unconfoundedness*, ou seja:

$$Y_i(0), Y_i(1) \perp T_i | Z_i \quad (22)$$

Sob essa hipótese, é possível estimar o ATE em  $Z_i = c$ :

$$E(Y_i(1) - Y_i(0) | Z_i = z_0) = E(Y_i | T_i = 1, Z_i = c) - E(Y_i | T_i = 0, Z_i = c) \quad (23)$$

Para a implementação prática do *FRD*, segundo Angrist e Pischke (2008), a descontinuidade se torna um instrumento para o status de tratamento. Assim,  $g_0(Z_i)$  e  $g_1(Z_i)$  são funções que podem ter qualquer forma contanto que elas difiram uma da outra (quanto mais melhor) no ponto  $z_0$ . No ponto  $z_0$  há um salto na probabilidade de ser tratado, tal que:

$$P(T_i = 1|Z_i) = \begin{cases} g_0(Z_i), & \text{se } Z_i \geq z_0 \\ g_1(Z_i), & \text{se } Z_i < z_0 \end{cases}, \quad \text{onde } g_1(z_0) \neq g_0(z_0).$$

(24)

Assume-se que  $g_1(z_0) > g_0(z_0)$ , portanto,  $Z_i > z_0$  é mais propenso a receber o tratamento. A relação entre a probabilidade de receber o tratamento e  $Z_i$  é dada por

$$E(T_i|Z_i) = P(T_i = 1|Z_i) = g_0(Z_i) + [g_1(Z_i) - g_0(Z_i)]D_i,$$

onde

$$D_i = 1(Z_i \geq z_0).$$

Dessa forma,  $D_i$  é uma variável *dummy* que indica tratamento e o ponto de descontinuidade em  $E(T_i|Z_i)$ . A *Fuzzy RD* assemelha-se muito com a estratégia de estimação de 2SLS.

Supondo que  $g_0(Z_i)$  e  $g_1(Z_i)$  podem ser descritas como polinômios de ordem  $p$ :

$$\begin{aligned} E(T_i|Z_i) &= \gamma_{00} + \gamma_{01}Z_i + \gamma_{02}Z_i^2 + \dots + \gamma_{0p}Z_i^p \\ &+ [\gamma_0^* + \gamma_1^*Z_i + \gamma_2^*Z_i^2 + \dots + \gamma_p^*Z_i^p]D_i \\ &= \gamma_{00} + \gamma_{01}Z_i + \gamma_{02}Z_i^2 + \dots + \gamma_{0p}Z_i^p \\ &+ \gamma_0^*D_i + \gamma_1^*Z_iD_i + \gamma_2^*Z_i^2D_i + \dots + \gamma_p^*Z_i^pD_i \end{aligned}$$

(25)

O primeiro estágio nesse caso é

$$T_i = \gamma_{00} + \gamma_{01}Z_i + \gamma_{02}Z_i^2 + \dots + \gamma_{0p}Z_i^p + \pi D_i + \varepsilon_{1i} \quad (26)$$

E a forma reduzida do Fuzzy RD é dada por:

$$Y_i = \mu + k_1Z_i + k_2Z_i^2 + \dots + k_pZ_i^p + \rho\pi D_i + \varepsilon_{2i} \quad (27)$$

A versão não-paramétrica do *FRD* consiste em se estimar por IV as equações em uma pequena vizinhança em torno da descontinuidade. A forma reduzida da esperança condicional de  $Y_i$  próximo a  $z_0$  é:

$$E[Y_i|z_0 < Z_i < z_0 + \delta] - E[Y_i|z_0 - \delta < Z_i < z_0] \simeq \rho\gamma_0^*.$$

Similarmente, o primeiro estágio para  $T_i$  é dado por:

$$E[T_i|z_0 < Z_i < z_0 + \delta] - E[T_i|z_0 - \delta < Z_i < z_0] \simeq \gamma_0^*$$

Portanto

$$\lim_{\delta \rightarrow 0} \frac{E[Y_i|z_0 < Z_i < z_0 + \delta] - E[Y_i|z_0 - \delta < Z_i < z_0]}{E[T_i|z_0 < Z_i < z_0 + \delta] - E[T_i|z_0 - \delta < Z_i < z_0]} = \rho \quad (28)$$

Assim como o estimador de IV, em (28) temos o estimador de Wald, nesse caso, utilizando  $D_i$  como instrumento para  $T_i$  em uma vizinhança  $\delta$  de  $z_0$ . Da mesma forma que nos casos de IV, o resultado é um *Local Average Treatment Effect* (LATE). O estimador de Wald para o *Fuzzy RD* captura o efeito causal dos *compliers* na vizinhança em que  $Z_i = z_0$ .

Neste trabalho se optou por utilizar uma função de tamanho de turma, baseada na construída por Angrist e Lavy (1999), para construir variáveis instrumentais e, assim, estimar os efeitos do tamanho da turma através do *Fuzzy RD*. O estudo desenvolvido pelos autores observa

que o tamanho da turma nas escolas de Israel tem um número máximo de 40 alunos por turma. Alunos em séries com um número de matrículas que ainda não tenham atingido esse número máximo podem esperar estar em turmas grandes até que se atinja o número de 40 alunos; séries com 41 alunos matriculados, por exemplo, são divididas em duas turmas menores, séries com 81 alunos são divididas em 3 turmas, e assim por diante. Como tal regra foi proposta por um rabino medieval, os autores chamam tal regra do tamanho máximo de “*Maimonides Rule*”.

Como este trabalho também observa uma regra exógena de determinação do tamanho máximo de turma, sendo que também se estabeleceu o número máximo de 40 alunos por turma, também se construirá uma função semelhante à função de *Maimonide*.

Para tanto, remete-se à discussão de Campbell (1969) acerca de métodos não experimentais para a pesquisa em avaliação de impacto. Ele argumenta que se a função utilizada para instrumento é descontínua, ou apresenta não-linearidade, então é possível estimar o efeito do tratamento no ponto de descontinuidade. De acordo com Angrist e Lavy (1999), o padrão *up-and-down* observado nas esperanças condicionais das notas dos exames dado o número de matrícula é interpretado como um reflexo do efeito causal de mudanças no tamanho da turma que são induzidas por mudanças no número de matrículas. Portanto, é plausível que qualquer outro mecanismo ligando matrículas e notas dos testes será muito brando.

A ligação entre o método de Variáveis Instrumentais e o de *Fuzzy RD* foi feita explicitamente por Klauuw (1996) e neste trabalho também é utilizada a idéia de que o *FRD* pode ser analisado numa abordagem de IV. Nesse caso, as estimações de IV se utilizam da descontinuidade ou não-linearidade na relação entre o número de matrículas e o tamanho da turma para identificar o efeito causal do tamanho da turma, ao mesmo tempo que qualquer outra relação entre as matrículas e as notas do teste estão controladas pela inclusão de covariadas.

Para formalizar a regra que será utilizada neste trabalho,  $f_{sc}$  é o tamanho de turma previsto (em uma dada série) para a turma  $c$  e na escola  $s$ . As matrículas dos alunos em determinada série são representadas por  $e_s$ . Supondo que as turmas serão divididas em tamanhos iguais no ponto de corte, o tamanho de turma previsto que resulta da aplicação da regra é dado por:

$$f_{sc} = \frac{e_s}{\text{int}\left(\frac{e_s - 1}{40}\right) + 1}$$

(29)

onde  $int(x)$  é a parte inteira de um número real,  $x$ . Essa função possui um padrão de *sawtooth* com descontinuidades nos múltiplos de 40. Ao mesmo tempo  $f_{sc}$  é uma função crescente das matrículas,  $e_s$ . Ou seja, a equação (29) captura o fato de que a regra adotada pela “enturmação” permite que as matrículas em uma determinada série no coorte de 1-40 sejam agrupadas em uma única turma, mas matrículas no coorte de 41-80 são divididas em duas turmas com tamanho médio de 20,5 a 40 alunos. Matrículas nos coortes de 81-120 são divididas em 3 turmas com tamanho médio que varia entre 27 – 40 alunos, e assim por diante. Portanto,  $f_{sc}$  é fixo intra escola.

A estimação por 2SLS será dada pela seguinte equação:

$$y_{isc} = \alpha_0 + \alpha_1 d_s + \beta_1 e_s + \beta_2 e_s^2 + \dots + \beta_p e_s^p + \rho \pi f_{sc} + \varepsilon_{isc}, \quad (30)$$

onde  $y_{isc}$  é a nota do teste  $i$  na escola  $s$  e turma  $c$ ,  $e_s$  é o número de matrículas na escola  $s$ ,  $f_{sc}$  é o tamanho de turma previsto (em uma dada série) para a turma  $c$  e na escola  $s$ . Nessa equação, fazendo uma analogia com o modelo anteriormente apresentado,  $f_{sc}$  equivale à  $D_i$ ,  $e_s$  corresponde à  $Z_i$  e o tamanho da turma  $n_{sc}$  representa  $T_i$ . A equação ainda contém uma covariada de controle,  $d_s$ , que é uma medida do *background* socioeconômico dos alunos matriculados na escola  $s$ . Uma vez que o *background* familiar tem impacto no desempenho dos alunos, é adicionada essa covariada no modelo. A equação não prevê perfeitamente o tamanho da turma, pois algumas escolas dividem as turmas de determinada série antes que se atinja o número de 40 alunos, este é o motivo pelo qual se utiliza o *RD Fuzzy* para captar o efeito, não o *Sharp RD*.

As hipóteses de identificação por trás desse método são: *i)* qualquer outro efeito de  $e_s$  nas notas dos testes está adequadamente controlado pelo termo  $\alpha_1 d_s$ ; *ii)* os pais não conseguem manipular a regra a fim de alocar seus filhos em escolas com turmas menores. A primeira hipótese, infelizmente, não é testável, por essa razão são realizadas diferentes especificações de modelos; no que se refere à segunda hipótese, como já fora discutido anteriormente, o fato da reorganização ter sido implementada no meio do ano diminui muito a chance de algum pai ter manipulado a regra e mudado seu filho de escola – para que ele fosse alocado em uma escola com turmas menores – pois os alunos são encaminhados para escolas próximas à residência dos mesmos, tornando a mobilidade mais difícil. Teria que haver uma justificativa por parte dos pais

para a transferência do aluno na metade do ano, como mudança de endereço, que necessitaria ser comprovada, e nesse período é mais difícil conseguir uma transferência entre escolas, pois as mesmas já completaram as vagas disponíveis no início do ano letivo. Mesmo que possam ter acontecido alguns casos de transferência por antecipação da regra, acredita-se que não foi grande o suficiente a magnitude desse efeito para impactar nos resultados.

A próxima seção descreve as bases de dados utilizadas para este trabalho e suas características.

### **3.3 Base de Dados**

As bases de dados utilizadas neste trabalho são os microdados do Prova Brasil, coordenado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) e os resultados divulgados por escola do Sistema de Avaliação do Rendimento Escolar do Rio Grande do Sul (SAERS).

O Prova Brasil é um sistema de avaliação do ensino básico composto por dois exames de proficiência: Língua Portuguesa, com foco em leitura, e Matemática, com foco na resolução de problemas. Além disso, também são aplicados questionários socioeconômicos aos alunos, nos quais os estudantes fornecem informações sobre fatores de contexto que podem estar associados ao desempenho, ou seja, sobre o *background* familiar. Outras informações relevantes são coletadas através dos questionários aplicados aos professores e diretores das escolas avaliadas, tais como dados sobre a infra-estrutura da escola, atividades desenvolvidas por professores e diretores, mecanismos de gestão e suas opiniões sobre diversos assuntos de interesse da escola.

A Prova Brasil é aplicada nas escolas a cada dois anos e avalia alunos de 4º e 8º anos do ensino fundamental, da rede pública e urbana de ensino. A avaliação é censitária, e assim oferece resultados de cada escola participante, das redes no âmbito dos municípios, dos estados, das regiões e do Brasil.

As questões do exame estão baseadas na Teoria de Resposta ao Item (TRI), que permite que a escala da nota final seja comparada no tempo e entre as séries. Para este trabalho, serão utilizados os dados de proficiência dos alunos nos testes de Língua Portuguesa e Matemática como variável de resultado. Essa variável é medida numa escala padronizada por uma média de

250 e desvio padrão de 50, sendo que a pontuação máxima em Língua Portuguesa é 350 e em Matemática é 375.

Através dos microdados do Prova Brasil também foi possível obter dados sobre o número de matrículas por série em cada escola e também o número de turmas de cada série e a quantidade de alunos por turma. Assim, a análise do trabalho se inicia estabelecendo a ligação entre a proficiência média de matemática e português para cada turma com dados de características da escola e tamanho da turma.

Além dos dados do Prova Brasil também foram utilizados dados do Sistema de Avaliação do Rendimento Escolar do Rio Grande do Sul (SAERS). A informação obtida através desse banco de dados foi um índice, desenvolvido pela Secretaria Estadual de Educação do Rio Grande do Sul, que identifica o nível socioeconômico atendido pela escola (alto, médio-alto, médio-baixo e baixo).

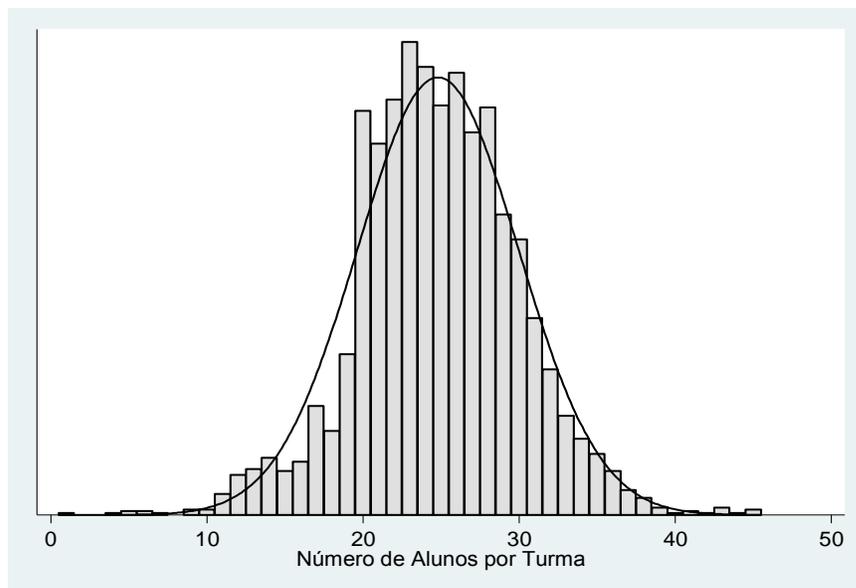
Portanto, foram coletados dados do Prova Brasil e do SAERS para escolas públicas estaduais, da zona urbana do Rio Grande do Sul, referentes ao ano de 2007 (ano da “enturmação”). As informações correspondem aos alunos de 4<sup>a</sup> e 8<sup>a</sup> série do ensino fundamental e foram considerados os resultados de Língua Portuguesa e Matemática.

### **3.3.1 Estatísticas Descritivas**

Os dados utilizados nas estimativas deste trabalho se referem aos alunos da 4<sup>a</sup> e 8<sup>a</sup> série do ensino fundamental da rede pública, de áreas urbanas, pertencentes ao departamento administrativo estadual, do estado do Rio Grande do Sul.

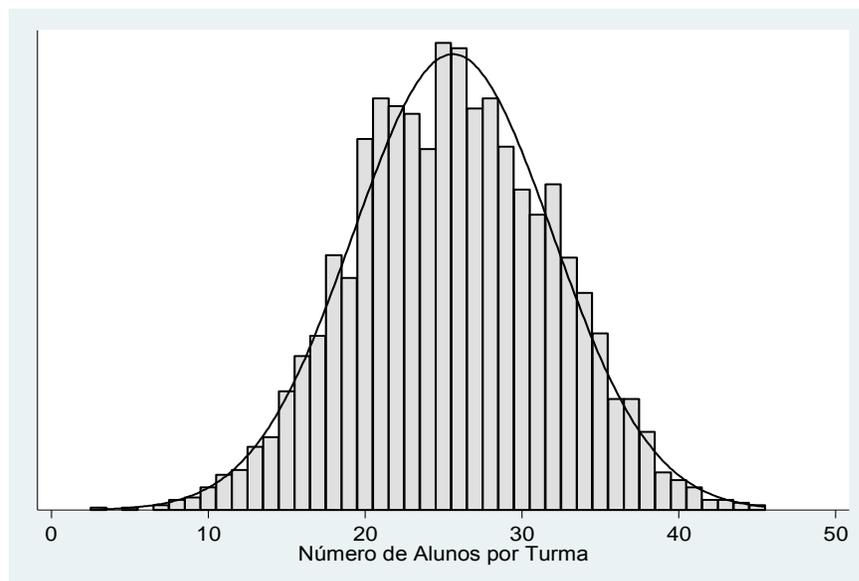
Os Gráficos 2 e 3 mostram a distribuição dos alunos por turma nas escolas analisadas para a quarta e oitava série respectivamente.

**Gráfico 2: Distribuição de alunos por turma para a quarta série – RS**



Fonte: elaborado pelo autor utilizando dados do Prova Brasil de 2007

**Gráfico 3: Distribuição de alunos por turma para a oitava série – RS**



Fonte: elaborado pelo autor utilizando dados do Prova Brasil de 2007

Percebe-se pelos gráficos que a maior concentração está nas turmas que contêm entre 20 e 30 alunos, tanto para a quarta quanto para a oitava série. Entretanto, também é possível notar que há várias turmas contendo entre 30 e 40 alunos, principalmente para a oitava série.

Outra forma de olhar para a distribuição dos dados é através da tabela. A Tabela 1 apresenta mais informações sobre a base de dados, tais como a distribuição da proficiência, do tamanho da turma e do número de matrículas.

**Tabela 1: Estatísticas descritivas**

	<b>Frequência</b>	<b>Média</b>	<b>Desvio-Padrão</b>	<b>Mínimo</b>	<b>Máximo</b>
<b>A. Quarta Série</b>					
Tamanho da turma	2961	24.80	5.17	1	45
Matrículas	2961	58.77	30.86	12	232
Proficiência em Matemática	2961	199.36	17.41	114.20	268.09
Proficiência em Português	2961	181.80	18.07	123.56	287.43
<b>B. Oitava Série</b>					
Tamanho da turma	2877	25.60	6.36	3	45
Matrículas	2877	67.14	39.72	10	294
Proficiência em Matemática	2877	251.41	19.67	139.44	334.61
Proficiência em Português	2877	239.31	18.50	155.47	306.06

Fonte: elaborado pelo autor utilizando dados do Prova Brasil de 2007

É possível observar através da tabela que o tamanho médio das turmas é de aproximadamente 25 alunos por classe para ambas as séries, o que já era esperado pela análise anterior dos gráficos. Para a quarta série, há em média 59 alunos matriculados nas escolas, já para a oitava série, há uma média de 67 alunos matriculados nas escolas estaduais do Rio Grande do Sul.

Em relação à proficiência, na quarta série a proficiência média de Língua Portuguesa é de 181,80 pontos e de Matemática é 199,36. Para a oitava série, as médias são mais elevadas, sendo 239,31 (Língua Portuguesa) e 251,41 (Matemática). Com a mesma base de dados do Prova

Brasil é possível calcular as médias de proficiência para o Brasil a fim de comparar com as médias do Rio Grande do Sul. A proficiência média para o Brasil é de 197, 57 e 212, 40 para Língua Portuguesa e Matemática, respectivamente. Dessa forma, percebe-se que as proficiências médias do Rio Grande do Sul são maiores do que as do Brasil apenas para a oitava série. Além disso, a média nacional é de 32,14 alunos por turma.

Porém, se dividirmos os dados nacionais por série, percebemos que as médias da quarta série nacional para Matemática e Português são 190, 65 e 173, 27, respectivamente, e a média de alunos por turma na quarta série para o Brasil é 30,4. Ou seja, as médias de proficiência do Rio Grande do Sul também são maiores do que as nacionais quando as duas são comparadas considerando a mesma série. Fazendo o mesmo exercício para a oitava série, percebemos que a média nacional para Matemática é 240,33 e para Português é 228,78 – mantendo-se inferior às médias do estado para a oitava série – e a média de alunos por turma é de 34,13.

Apenas olhando para as estatísticas descritivas é possível pensar que há uma relação negativa entre desempenho dos alunos e tamanho de turma, uma vez que as médias de proficiência nacionais são menores que as do Rio Grande do Sul ao mesmo tempo em que as médias de alunos por turma para o Brasil são superiores. Entretanto, existe uma série de outros fatores não considerados que podem estar influenciando essas diferenças no desempenho. Por essa razão, o próximo capítulo apresenta os resultados de exercícios mais rigorosos para que se possa inferir se há impacto do tamanho da turma sobre o desempenho dos alunos.

## 4 RESULTADOS

Os resultados das estimações da forma reduzida apresentada no capítulo anterior serão apresentados neste capítulo. A Tabela 2 reporta os resultados obtidos através das estimações por IV e *FRD* do efeito do tamanho da turma (pós “enturmação”) no desempenho dos alunos da quarta série. Ou seja, as estimações estão tentando captar se o fato das turmas terem sido reorganizadas de forma a aumentar o número de alunos por classe teve algum impacto no desempenho dos estudantes. As variáveis são utilizadas a nível de turma, portanto, os desvios-padrão são estimados em *clusters* de escola.

Os resultados reportados na Tabela 2 correspondem a diferentes especificações de modelos. A primeira coluna reporta os resultados da estimação por Variável Instrumental do efeito do tamanho da turma no desempenho dos alunos da quarta série em Língua Portuguesa em um modelo sem nenhum controle relacionado ao número de matrículas. O único controle neste primeiro modelo é em relação ao nível socioeconômico dos alunos atendidos pela escola, que é o índice elaborado pela própria secretaria da educação. Percebe-se que os coeficientes obtidos no primeiro modelo não são estatisticamente significativos, assim como a grande maioria dos outros coeficientes. Apesar do coeficiente não ser estatisticamente significativo, pode-se notar que sem controlar pelo número de matrículas o sinal do coeficiente entre tamanho da turma e desempenho dos alunos em Português não é negativo (0,239). Nos modelos das colunas 2 e 3 são incluídos controles lineares (2 e 3) e quadráticos (3) para o número de matrículas. Apesar do coeficiente continuar sem significância estatística, percebe-se que com a inclusão de controles o sinal do coeficiente referente ao efeito do tamanho da turma tem uma magnitude maior e passa a ser negativo (-1,267, para o modelo 2; -1,012, para o modelo 3).

O modelo 4 foi construído para testar a validade interna da Regressão com descontinuidade. Ele foi desenvolvido para se tentar verificar se os controles para a variável que gera a descontinuidade estão adequados. Para tanto, é especificado um modelo que se inclui uma *continuous piecewise linear trend* com inclinações idênticas às de  $f_{sc}$  nos segmentos lineares. Portanto, a variação nessa tendência linear é gerada somente pelos saltos devido à regra da “enturmação” nos pontos de descontinuidade. A tendência é definida da seguinte forma:

$$\left\{ \begin{array}{ll} e_s; & e_s \in [0, 40] \\ 20 + (e_s/2); & e_s \in [41, 80] \\ 100/3 + (e_s/3); & e_s \in [81, 120] \\ 130/3 + (e_s/4); & e_s \in [121, 160] \end{array} \right.$$

A idéia por trás desse modelo é que uma vez que os efeitos de tendência da covariada que gera a descontinuidade estão completamente controlados, não há motivos para manter outras covariadas fixas. Assim como nos modelos em que são adicionadas variáveis de controle para o número de matrículas, o sinal do coeficiente nos modelos 4 também é negativo (-0,597), apesar da magnitude do coeficiente ser menor do que a dos outros dois modelos. Da mesma forma que os outros resultados encontrados, esse modelo não é estatisticamente significativo. Os modelos 5 e 6 são as estimações para as turmas pertencentes à amostra no entorno da descontinuidade +5/-5, ou seja, fazem parte dessa amostra as turmas cujo número de matrículas pertença aos intervalos {[36, 45], [76, 85], [116, 125]}. Portanto, os modelos 5 e 6 são as estimações por FRD. Percebe-se que a magnitude dos coeficientes desses modelos são maiores do que as dos modelos anteriores (4,560 para o 5; -2,320 para o 6) – assim como os resultados obtidos por Angrist e Lavy (1999). Entretanto, as estimações por FRD também não são estatisticamente significantes. Novamente, o modelo que não inclui o número de matrículas como covariada tem o sinal do coeficiente negativo para Português e Matemática.

Os mesmos seis modelos foram estimados para os resultados dos testes de Matemática. Basicamente os resultados encontrados seguem o mesmo padrão dos resultados encontrados para Português. Não foram verificados efeitos estatisticamente significantes, porém o sinal dos coeficientes também é negativo para a grande maioria dos modelos. Apenas na especificação do modelo 5 (descontinuidade sem covariada de controle para número de matrículas) percebe-se que o coeficiente, assim como o de Português, apresenta sinal positivo e a maior magnitude do que os coeficientes dos outros modelos especificados (3,291).

Tabela 2: Resultados das Estimções para quarta srie

	Estimções 2SLS para 2007 (quarta srie)													
	Portugués			Matemática										
	(1)	(2)	(3)	(4)	(5)	(6)								
			Amostra completa	+/- 5 Amostra com descontinuidade	Amostra completa	+/- 5 Amostra com descontinuidade								
Tamanho da turma	0.239 (0.559) [0.711]	-1.267 (0.840) [0.229]	-1.012 (0.785) [0.267]	-0.597 (0.868) [0.488]	4.560 (7.311) [0.596]	-2.320 (1.761) [0.279]	-0.286 (0.583) [0.672]	-0.652 (0.827) [0.488]	0.631 (0.522) [0.35]	0.640 (0.532) [0.315]	-0.469 (0.784) [0.582]	-0.296 (0.879) [0.787]	3.291 (6.262) [0.652]	-1.924 (1.713) [0.343]
Nível Sócio-econômico	1.301 (0.499) [0.121]	1.336 (0.539) [0.089]	1.291 (0.525) [0.07]	-1.425 (4.386) [0.776]	2.017 (1.351) [0.232]	2.017 (1.351) [0.232]	0.631 (0.522) [0.35]	0.640 (0.532) [0.315]	0.608 (0.525) [0.312]	0.608 (0.525) [0.312]	0.608 (0.525) [0.312]	0.608 (0.525) [0.312]	-1.170 (3.760) [0.785]	1.440 (1.319) [0.355]
Matrículas		0.077 (0.070) [0.354]	0.213 (0.098) [0.096]		0.238 (0.233) [0.382]	0.238 (0.233) [0.382]		0.019 (0.069) [0.805]	0.116 (0.098) [0.301]	0.116 (0.098) [0.301]	0.116 (0.098) [0.301]		0.180 (0.226) [0.484]	
Matrículas <sup>2</sup> /100			-0.101 (0.039) [0.063]						-0.072 (0.039) [0.14]	-0.072 (0.039) [0.14]				
Tendência linear				0.076 (0.137) [0.547]							0.001 (0.139) [0.968]			
Root MSE	17.35	18.66	18.21	17.81	27.19	19.83	18.23	18.52	18.34	18.20	23.82	19.56		
N	2865	2865	2865	2929	696	696	2865	2865	2865	2929	696	696		

Nota 1: desvios-padrão (corrigidos por cluster nas escolas) entre parênteses. P-valor entre colchetes. Fonte: elaborado pelo autor utilizando dados do Prova Brasil e Saers de 2007.

Tabela 3: Resultados das Estimações para oitava série

	Estimações 2SLS para 2007 (oitava série)											
	Português						Matemática					
	Amostra completa			+/- 5 Amostra com descontinuidade			Amostra completa			+/- 5 Amostra com descontinuidade		
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Tamanho da turma	-0.679 (0.252) [0.114]	-1.032 (1.370) [0.506]	-0.829 (2.384) [0.745]	-1.672 (2.434) [0.563]	-0.133 (0.998) [0.906]	0.730 (8.813) [0.939]	-0.854 (0.282) [0.094]	-0.729 (1.463) [0.653]	0.516 (2.581) [0.851]	-0.618 (2.334) [0.816]	-0.415 (1.155) [0.754]	0.569 (10.036) [0.958]
Nível Sócio-econômico	9.123 (0.628) [0.005]	9.318 (1.089) [0.003]	9.242 (1.41) [0.003]	8.849 (1.227) [0.019]	8.775 (1.666) [0.013]	9.602 (0.703) [0.005]	9.532 (1.164) [0.004]	9.065 (1.530) [0.004]	9.065 (1.530) [0.004]	9.372 (1.429) [0.022]	9.288 (1.909) [0.017]	9.288 (1.909) [0.017]
Matrículas	0.029 (0.097) [0.782]	0.029 (0.097) [0.782]	-0.005 (0.275) [0.987]	-0.005 (0.275) [0.987]	-0.111 (1.011) [0.919]	-0.111 (1.011) [0.919]	-0.010 (0.103) [0.926]	-0.219 (0.298) [0.503]	-0.219 (0.298) [0.503]	-0.127 (1.151) [0.919]	-0.127 (1.151) [0.919]	-0.127 (1.151) [0.919]
Matrículas <sup>2</sup> /100			0.010 (0.056) [0.868]	0.010 (0.056) [0.868]				0.059 (0.060) [0.376]	0.059 (0.060) [0.376]			
Tendência linear				0.277 (0.434) [0.588]						0.029 (0.416) [0.952]		
Root MSE	17.89	18.47	18.11	21.41	17.45	18.10	19.02	18.86	18.59	20.11	18.63	18.89
N	2849	2849	2849	2778	665	665	2849	2849	2849	2778	665	665

Nota 1: desvios-padrão (corrigidos por cluster nas escolas) entre parênteses. P-valor entre colchetes. Fonte: elaborado pelo autor utilizando dados do Prova Brasil e Saers de 2007.

A Tabela 3 apresenta os resultados para a oitava série. A grande maioria dos resultados é semelhante aos resultados da tabela anterior, os coeficientes encontrados são negativos, porém estatisticamente insignificantes. A especificação do modelo 1 para matemática é estatisticamente significativa. Ou seja, o efeito estimado do tamanho da turma no desempenho dos alunos em matemática em um modelo sem nenhum controle para número de matrículas é -0,854, com um erro-padrão de 0,282 a 10% de significância. Ou seja, para os alunos da oitava série, há uma associação negativa e estatisticamente significativa entre o tamanho da turma e o desempenho em Matemática.

Estudos realizados por outros autores utilizando o método de variáveis instrumentais ou *FRD* encontraram associações positivas entre o tamanho da turma e o desempenho dos alunos – Akerhielm (1995), Broozer e Rouse (1995), Angrist e Lavy (1999), Krueger (1999). Entretanto, assim como os resultados obtidos por meio de variáveis instrumentais do estudo conduzido por Hoxby (1996) este estudo não encontrou evidência de relação entre tamanho da turma e o desempenho dos alunos. Ou seja, não há evidência de que a reorganização das turmas ocorrida no meio do ano letivo de 2007 tenha influenciado negativamente o dos alunos nos testes de Língua Portuguesa e Matemática da avaliação do Prova Brasil no final do ano de 2007.

## 5 CONCLUSÃO

Este trabalho teve como objetivo estimar, através dos métodos de Variáveis Instrumentais e Desenho de Regressão com Descontinuidade, o efeito do tamanho da turma no desempenho dos alunos de ensino fundamental nos testes de Língua Portuguesa e Matemática nas escolas públicas estaduais do Rio Grande do Sul. Apesar da maioria dos coeficientes estimados apresentarem sinal negativo, indicando uma associação negativa entre tamanho da turma e desempenho dos alunos, esses coeficientes não são estatisticamente significantes. Portanto, este estudo não encontrou evidências estatisticamente significantes que a reorganização das turmas ocorrida na metade do ano letivo de 2007 (a chamada “enturmação”) tenha impactado no desempenho dos alunos. Entretanto, é possível verificar para os alunos da oitava série, que há uma relação positiva e estatisticamente significativa entre o nível socioeconômico dos alunos o desempenho dos mesmos. Como este índice está construído por nível de escola, talvez escolas que atendam alunos com melhor nível socioeconômico tenham melhores insumos, que impactam positivamente no desempenho dos alunos.

Uma dificuldade encontrada ao se estimar os efeitos do tamanho da turma é que não é possível separar o mesmo dos *peer effects* que ocorrem ao se aumentar o número de alunos por turma. Talvez os *peer effects* positivos tenham se sobressaído aos efeitos negativos do aumento de alunos por turma, como se percebe em alguns coeficientes com sinais positivos encontrados nas estimações, porém, também estatisticamente insignificantes. Isso poderia ocorrer devido à heterogeneidade nas turmas causada pela “enturmação”. Neste caso, os alunos com baixo rendimento podem ter sido expostos a colegas mais habilidosos e se beneficiado dos *learning spillovers* positivos. É necessário notar que mesmo trabalhando com os dados após a reorganização das turmas a média de alunos por sala de aula é aproximadamente 25, isto é, a média continua inferior à nacional. Percebe-se também que a maioria das turmas possui entre 20 e 30 alunos. Dessa forma, podemos dizer que mesmo com a realocação a grande maioria das turmas não possui elevado número de alunos por classe. Talvez por esse motivo não se tenha encontrado efeito estatisticamente significativo. Oliveira (2008) em seu estudo sobre o efeito do tamanho da classe utilizando dados para o Brasil encontrou efeito apenas para turmas a partir de 30 alunos – a redução de 38 para 30 alunos está associada a um aumento de 10,67 pontos, que é equivalente a um movimento de 0,26 desvios padrão na distribuição de proficiência.

Um estudo futuro é poderia ser realizado para se investigar se com um maior intervalo de tempo a nova regra para organização das turmas teria algum efeito nos desempenho dos alunos. Em outras palavras, se o fato dos dados analisados serem os do teste aplicado no final do mesmo ano em que se reorganizaram as turmas pode ter influenciado na falta de evidência de efeito da política no desempenho dos alunos. Pois o intervalo de tempo entre a reorganização e os testes foi pequeno, podendo ainda não ter passado tempo suficiente para se captar o efeito. Portanto, uma possível pesquisa futura seria realizar esse exercício para os dados de 2009, uma vez que os critérios para organização das turmas se mantiveram nos anos seguintes a 2007.

## REFERÊNCIAS

- AKERHIELM, Karen, “Does Class Size Matter?” *Economics of Education Review*, 1995.
- ALAM, Mahmudul. *Development of Primary Education in Bangladesh: The Ways Ahead*. *Bangladesh Development Studies*, v. 26, 2000.
- ALBERNAZ, Ângela; FERREIRA, Francisco H. G.; FRANCO, Creso. *Qualidade e Equidade na Educação Fundamental Brasileiro*. PPE, v. 33 No.3. 2002.
- ANGRIST, J. “Treatment Effect Heterogeneity in Theory and Practice”. *Economic Journal*, 2004.
- ANGRIST, J. D.; PISCHKE, J. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press, 2008.
- ANGRIST, J., IMBENS, G. and RUBIN, D. B. ‘Identification of causal effects using instrumental variables’, *Journal of the American Statistical Association*, vol. 91, p. 444–55, 1996.
- ANGRIST, Josh and LAVY, Victor. *Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement*. *Quarterly Journal of Economics*, May, 1999.
- ASADULLAH, M. Niaz. *The Effect of Class Size on Student Achievement: Evidence from Bangladesh*. *Applied Economic Letters*, v.12, p. 217–221, march 2005.
- ASHENFELTER, Orley; ROUSE, Cecilia. *Schooling, Intelligence, and Income in America: Cracks in the Bell Curve*. NBER Working Paper Series, No. 6902, 1999.
- BAGOLIN, IZETE PENGO; PORTO JUNIOR, SABINO DA SILVA. *A desigualdade da distribuição da educação e o crescimento no Brasil: índice de Gini e os anos de escolaridade*. Texto para discussão, UFRGS, nº 9, 2003.
- BARROS, Ricardo Paes de; MENDONÇA, Rosane. *Education and Equitable Economic Development*. *Economia*, v. 1, n. 1, p. 111-144, 2000.
- BARROS, Ricardo Paes de; MENDONÇA, Rosane. *Os Determinantes da Desigualdade no Brasil*. In: *A Economia Brasileira em Perspectiva*. Rio de Janeiro, IPEA, 1996.
- BARROS, Ricardo Paes de; MENDONÇA, Rosane; SANTOS, Daniel Domingues dos; QUINTAES, Giovanni. *Determinantes do Desempenho Educacional no Brasil*. Texto para Discussão, No. 834, IPEA, Rio de Janeiro, 2001.
- BARROS, RICARDO PAES, HENRIQUES, RICARDO & MENDONÇA, ROSANE. *Pelo fim das décadas perdidas: Educação e desenvolvimento sustentado no Brasil*. Texto para Discussão n. 857, IPEA, 2002.

———. Investimento em educação e desenvolvimento econômico. *A Economia Brasileira em Perspectiva* — 1998. Rio de Janeiro: IPEA, v. 2, p. 605-614, 1998.

———. Investing in human resources. *Economic and social progress in Latin America. Report*, IDB, 1993b.

BECKER, GARY S. *Human Capital: A Theoretical and Empirical Analysis With Special Reference to Education*. 1993 by The National Bureaus of Economic Research.

BEHRMAN, J. R. *Human resources in Latin America and the Caribbean*. Washington: BID, 281 p., 1993a.

BETTS, Julian R. Is There a Link between School Inputs and Earnings? Fresh Scrutiny of Old Literature. In: G. Burtless (Ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adults Success*. Washington, DC: Brooking Institution Press, p. 141-191, 1996.

BLATCHFORD, P. & MORTIMORE, P. The issue of class size for young children in schools: what can we learn from research? *Oxford Review of Education*, 20(4), p. 411–428, 1994.

BLAUG, Mark. *Introdução à Economia da Educação*. Porto Alegre, Globo, 1975.

BOOZER, Michael, and ROUSE, C. “Intraschool Variation in Class Size: Patterns and Implications,” NBER Working Paper No. 5144, June 1995.

BOURGUIGNON, F., FOURNIER, M., GURGAND, M. *Distribution, development and education: Taiwan, 1979-1992*. [S.l.: s.n.], 28 p., Jan. 1998.

BOURKE, S. How smaller is better: some relationships between class size, teaching practices, and student achievement, *American Educational Research Journal*, 23(4), p. 558–571, 1986.

BOWMAN, M. J. The Human Investment Revolution in Economic Thought. *Sociology of Education* (Spring): 111 – 137, 1996.

CAMPBELL, Donald T. “Reforms as Experiments”. *American Psychologist*, XXIV, p. 409–429, 1969.

CARD, David; KRUEGER, Alan B. Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *Journal of Political Economy*, v. 100, No. 1, p1-40, 1992.

CARD, David; KRUEGER, Alan B. Labor Market Effects of School Quality: Theory and Evidence. In: G. Burtless (Ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adults Success*. Washington, DC: Brooking Institution Press, p 97-140, 1996.

CARD, David; KRUEGER, Alan B. School Quality and Black –White Relative Earnings: A Direct Assessment. *Quarterly Journal of Economics*, v. 107, No. 1, p151-200, 1992.

- CARD, David; KRUEGER, Alan B. The Economic Return to School Quality: A Partial Survey. Working Paper, No 334, Princeton University, 1994.
- CASE, Anne and DEATON. School Inputs and Educational Outcomes in South Africa. Quarterly Journal of Economics, May, 1999.
- CASE, Anne and YOGO. Does School Quality Matter? Reruns to Education and the Characteristics of Schools in South Africa. NBER Working Paper 7379, 1999.
- CHATMAN, S. Lower-division class size at U.S. postsecondary institutions. Research in Higher Education, 38 (5), p. 615–630, 1997.
- CHECCHI, DANIELE. Does educational achievement help to explain income inequality? Departmental working paper 2000-11. Department of Economics, University of Milan, Italy, 2000.
- COELHO, A. M. e CORSEUIL, C. H. Diferenciais Salariais No Brasil: Um Breve Panorama. In: Courseuil, C. H. (Ed.). Estrutura Salarial: Aspectos Conceituais e Novos Resultados para o Brasil. Rio de Janeiro. IPEA, 2002.
- COLEMAN, James S. et al. Equality of Educational Opportunity. Washington, 1966.
- DAY, C., TOLLEY, H., HADFIELD, M., PARKIN, E. & WATLING, R. Class size research and the quality of education: a survey of the literature related to class size and the quality of teaching and learning (Haywards Heath, N.A.H.T), 1996.
- DE WALQUE, D. Education, Information, and Smoking Decisions: Evidence from Smoking Histories. Working Paper No. 3362, World Bank, 2004.
- DOPPELHOFER, G., MILLER, R. I. E SALA-I-MARTIN, X. Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. Working Paper No.7750, NBER. 2000.
- DUSO, A. P. ; SUDBRACK, E. M. . Política Educacional: para além da racionalidade econômica - questionando a enturmação. Revista de Ciências Humanas - URI, v. 9, p. 01-50, 2009.
- EHRENBERG, Ronald E.; BREWER, Dominic J. Did Teachers' Verbal Ability and Race Matter in the 1960s? Coleman Revisited. Economics of Education Review, v. 14, 1995.
- EHRENBERG, Ronald E.; BREWER, Dominic J.; GAMORAN, Adam; WILLMS, J. Douglas. The Class Size Controversy. Working Paper, No 14, Cornell Higher Education Research Institute, 2001.
- ENGERMAN, STANLEY e SOKOLOFF, KENNETH. Factor Endowments, Inequality, and Paths of Development among New World Economies. Economia. v. 3, n. 1, 2002.

FERREIRA, Rodrigo A. Desigualdade do desempenho escolar dos alunos do ensino fundamental do Estado de São Paulo. Dissertação de mestrado. Ribeirão Preto, 2008.

FULLER, Bruce. “Raising School Quality in Developing Countries: What Investments Boost Learning”, World Bank Discussion Paper, 1986.

GLAESER, Edward L. Why Does Schooling Generate Economic Growth? *Economics Letter* 44: 333-337, 1994.

HAHN, Jinyong, Petra Todd, and Wilbert van der Klaauw. “Identification of Treatment Effects by Regression Discontinuity Design.” *Econometrica*, v. 69 (1), p. 201– 209, 2001.

HANUSHEK AND HARBISON. *Educational Performance of the Poor*. Oxford University Press, 1992.

HANUSHEK, E. A. School resources. *Handbook of the Economics of Education*, v.2, Elsevier, 2006.

HANUSHEK, E. A.; LAVY, V.; HITOMI, K.. Do students care about school quality? Determinants of dropout behavior in developing countries. Working Paper 12737, National Bureau of Economic Research, Cambridge, MA, 2006.

HANUSHEK, E. Publicly provided education. In: AUERBACH, A.J. e FELDSTEIN, M. (ed.). *Handbook of public economics*. Amsterdam: North-Holland, 2002.

HANUSHEK, Eric A. e KIMKO, Dennis D.. Schooling, Labor-Force Quality, and the Growth of Nations. *The American Economic Review*, v. 90, n. 5, p.1184-1208, December 2000.

HANUSHEK, Eric A.; LUQUE Javier A. Efficiency and Equity in Schools around the World. Mimeo, 2002.

HANUSHEK, Eric. *The Evidence on Class Size*. Rochester, NY.: University of Rochester, W. Allen Wallis Institute of Political Economy, 1998.

HANUSHEK. “Interpreting Recent Research in Developing Countries”, *World Bank Research Observer*, Volume 10, 1995.

HANUSHEK. “The Economics of Schooling: Production and Efficiency in Public Schools”, *Journal of Economic Literature*, 1986, v. 24, 1986.

HECKMAN, James; LAYNE-FARRAR, Anne; TODD, Petra. Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship. In: G. Burtless (Ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adults Success*. Washington, DC: Brooking Institution Press, p192-289, 1996.

HECKMAN, James; LAYNE-FARRAR, Anne; TODD, Petra. Human capital Pricing Equations with an Application to Estimating the Effect of Schooling Quality on Earnings. *The Review of Economics and Statistics*, v. 78, p 562-610, 1996.

HEDGES, L.V., Laine, R.D., and Greenwald, R. Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, 23:5-14, 1994.

HOXBY, M Caroline. The Effect of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics*, November, 2000.

IMBENS, G. and WOOLDRIDGE, J. Regression Discontinuity Designs. NBER Lecture Notes, n. 3, 2003.

IMBENS, G. W. & J. D. ANGRIST. Identification and estimation of local average treatment effects. *Econometrica*, 62, p. 467-475, 1994.

KHANDKER, S. R., KOOLWAL, G. B. AND SAMAD, H. A. Handbook on Impact Evaluation Quantitative Methods and Practices. The World Bank. Washington, D. C, 2010.

KREMER, M. "Research on Schooling: What We Know and What We Don't: A Comment on Hanushek", *World Bank Research Observer*, v.10, 1995.

KRUEGER, Alan B. Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics*, May issue, 1999.

LAM, D., LEVISON, D. Age, experience and schooling: decomposing earnings inequality in the U. S. and Brazil. [S.l.: s.n.], May 1990.

LAU, L. J., JAMISON, D. T., LIU, S., RIUKIN, S. Education and economic growth: some cross-sectional evidence. *Opportunity foregone: education in Brazil*. Washington: BID, p. 83-116, 1996.

LEE, D.; LEMIEUX, T. Regression Discontinuity Designs in Economics. NBER Working Paper Series n. 14723, National Bureau of Economic Research, 2009.

LOPEZ, RAMON ; THOMAS, VINOD ; WANG, YAN. Addressing the education puzzle: The distribution of education and economic reform. World Bank, Policy research working paper n. 2031, 1998.

MCKEACHIE, W. J. Research on college teaching: The historical background. *Journal of Educational Psychology*, 82(2), p. 189-200, 1990.

MARSHALL, ALFRED. *Principles of Economics*. 8th ed. London, 1930. App. E, p. 787-88.

MENEZES-FILHO, N. A. "Os determinantes do desempenho escolar no Brasil". São Paulo: Instituto Futuro Brasil, nº 2, 2007. 30 p. Texto para discussão, 2007.

MENEZES-FILHO, N. A. Educação e Desigualdade. In: Lisboa, M. B. e Menezes-Filho, N. A. (org.). *Microeconomia e Sociedade no Brasil*. Rio de Janeiro. Contra Capa Livraria, 2001.

MILL, STUART. *Princípios de Economia Política. Os Economistas*, vol. 1 São Paulo: Abril Cultural, 1983.

MOSHEL-RAVID, “Learning, Teaching, Education, and Class Size: A Review of the Literature”. Jerusalem: The Henrietta Sczold Institute, the National Institute for Research in the Behavioral Sciences, 1995.

MUELLER, D., C. I. CHASE, and WALDEN, J. D. “Effects of Reduced Class Size in Primary Classes”, *Educational Leadership*, 1988.

OFSTED, “Class Size and the Quality of Education: A Report from the Office of Her Majesty’s Chief Inspector of Schools”. London: Office for Standards in Education, November 1995.

OLIVEIRA, Jaqueline Maria de. *Custo-efetividade de políticas de redução do tamanho da classe e ampliação da jornada escolar: uma aplicação de estimadores de matching*. 2008. Dissertação (Mestrado em Teoria Econômica) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2008.

RAMOS, L. e VIEIRA, M. L. A relação entre educação e salários no Brasil. In: *A Economia Brasileira em Perspectiva*. Rio de Janeiro, IPEA, 1996.

RIVKIN, Steven G.; HANUSHEK, Eric A.; e KAIN, John F. Teachers, Schools, and Academic Achievement. *Econometrica*. vol. 73, No.2, p417-58. 2005.

ROBINSON, G. E., “Synthesis of Research on the Effects of Class Size,” *Educational Leadership*, XLVII, p. 80–90, 1990.

RUBIN, D. Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies. *Journal of Educational Psychology*, v. 66, n. 5, p. 688-701, 1974.

SHAPSON, S. M., WRIGHT, E. N., EASON, G. & FITZGERALD, J. *Effects of class size in the junior grades: a study* (Ontario, Ministry of Education), 1977.

SHAPSON, S. M., WRIGHT, E. N., EASON, G. & FITZGERALD, J. An experimental study of the effects of class size, *American Educational Research Journal*, 17, p. 144–152, 1980.

SCHMIDT, PAULO; SANTOS, JOSÉ LUIZ DOS. *Avaliação de ativos intangíveis*. São Paulo: Atlas, 2002.

SCHULTZ, THEODORE W. Investment in Human Capital. *The American Economic Review*. Vol. 2, Nº 1, Março de 1961.

SEN, A. *Desenvolvimento como liberdade*. São Paulo: Companhia das Letras, 2000.

SMITH, ADAM. A riqueza das Nações. Os Economistas, Vol. 1 São Paulo: Abril Cultural, 1983.

THISTLETHWAITE, Donald L. and Donald T. Campbell, “Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment,” *Journal of Educational Psychology*, December 1960.

TINBERGEN, J. Income differences: recent research. Oxford: North-Holland Publishing, 73 p., 1975.

TOMMASI, Livia DE; WARDE, Jorge, Mirian. O Banco Mundial e as políticas educacionais. 2. ed. São Paulo: Cortez, 1998.

VALLE, Raquel da C. Teoria de Resposta ao Item. In: Heraldo Marelím Viana, *Estudos em Avaliação Educacional*, São Paulo: Fundação Carlos Chagas, p 7-92, 2000.

VAN DER KLAAUW, Wilbert. “A Regression-Discontinuity Evaluation of the Effect of Financial Aid Offers on College Enrollment”. New York University Economics Department, 1996.

VAN DER KLAAUW, Wilbert. Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach. *International Economic Review*, n. 43, p. 1249–1287, November 2002.

WALD, A. ‘The fitting of straight lines if both variables are subject to error’, *Annals of Mathematical Statistics*, vol. 11, p. 284–300, 1940.