Universidade Federal do Rio de Janeiro
Centro de Filosofia e Ciências Humanas
Faculdade de Educação
Programa de Pós-Graduação em Educação

Tiago Lisboa Bartholo

Segregação Escolar na Rede Municipal do Rio de Janeiro:

Causas e Consequências

RIO DE JANEIRO, 2014

Tiago Lisboa Bartholo

Segregação Escolar na Rede Municipal do Rio de Janeiro:

Causas e Consequências

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Educação, como requisito parcial para a obtenção do título de Doutor em Educação pela Universidade Federal do Rio de Janeiro.

Orientador: Marcio da Costa

B287 Bartholo, Tiago Lisboa
Segregação escolar na Rede Municipal do Rio de Janeiro: causas e consequências / Tiago Lisboa Bartholo. 2014.
280 f.

Orientador: Prof. Dr. Marcio da Costa.

Tese (doutorado) - Universidade Federal do Rio de Janeiro, Faculdade de Educação, Programa de Pós-Graduação em Educação, 2014.

1. Segregação na educação. 2. Indicadores educacionais. 3. Burocracia. I. Costa, Marcio da. II. Universidade Federal do Rio de Janeiro. Faculdade de Educação.

CDD: 370.193

Universidade Federal do Rio de Janeiro
Centro de Filosofia e Ciências Humanas
Faculdade de Educação
Programa de Pós-Graduação em Educação

A Tese **"Segregação Escolar na Rede Municipal do Rio de Janeiro: Causas e Consequências"**

Doutorando(a): **Tiago Lisboa Bartholo**

Orientado(a) pelo(a): **Prof. Dr. Marcio da Costa**

E aprovada por todos os membros da Banca Examinadora, foi aceita pela Faculdade de Educação da Universidade Federal do Rio de Janeiro e homologada pelo Conselho de Ensino para Graduados e Pesquisa, como requisito parcial à obtenção do título de

DOUTOR EM EDUCAÇÃO

Rio de Janeiro, 21 de março de 2014.

Banca Examinadora:

Presidente: _____
Prof. Dr. Marcio da Costa

_____
Profª. Drª. Mariane Campelo Koslinski

_____
Prof. Dr. Jose Ignacio Cano Gestoso

_____
Prof. Dr. Gregory Elacqua

_____
Prof. Dr. Fernando Tavares Júnior

Everything good proceeds from enthusiasm

The sense of ‘I really want to know how this turns out’

Will drive you through many, many long nights of no results

Where the feeling of ‘I think I ought to do this’ dries out very quickly

Brian Eno

Esse trabalho foi feito com enorme entusiasmo.

SUMMARY

AGRADECIMENTOS

LIST OF FIGURES……VII

LIST OF TABLES……...IX

ABBREVIATIONS AND ACRONYMS……XX

RESUMO……...XXII

ABSTRACT……XXIII

1. INTRODUCTION……1

1.1 What is Disadvantage in Education? Key Variables in the Study…….10

1.2 What is Segregation and How Should it be Measured? ……15

2. THE ENROLMENT LEGISLATION IN RIO DE JANEIRO PUBLIC SCHOOLS……...23

3. SCHOOL CHOICE IN THE BRAZILIAN PUBLIC SYSTEM: MISUSE OF CONCEPTS……36

4. METHODS…….41

5. DATABASE AND ITS LIMITATIONS……..46

5.1 Who Is Our Missing Data? Testing the "Super Deprived" Hypothesis…….55

6. DESCRIBING SEGREGATION TRENDS OVER TIME……70

6.1 School Segregation across Educational Authorities (CREs) First Segment……..81

6.2 School Segregation across Educational Authorities (CREs) Second Segment……..85

6.3 Clustering Pupils in Rio de Janeiro Public Schools: one or more patterns of segregation? ……...90

6.4 One or More Patterns of School Segregation? Replicating Models across Educational Authorities……...100

7. UNINTENDED IMPACT OF EDUCATIONAL POLICIES ON SCHOOL SEGREGATION……..105

7.1 School Segregation in Rio de Janeiro Public Schools: Impact on Educational Transitions……105

7.2 **Voluntary Transfers in the First Segment…….108**

Discussion……..122

7.3 **"Voluntary" Transfers in the Second Segment…….124**

Discussion……..135

***7.4* Mandatory Transition – Remanejamento……..137**

Cross-Sectional and Longitudinal Designs: Overall Impact of "*Remanejamento*"…….139

Discussion……..143

8. OPEN ENROLMENT SYSTEM IN RIO DE JANEIRO PUBLIC SCHOOLS: ANALYSING THE "HIDDEN-QUASI-MARKET"……146

8.1 Describing the Educational Authorities ……149

8.2 Is Parental Demand Correlated with the Segregation Ratio? …….151

8.3 Patterns of between School Segregation in Educational Authorities: What Do We Know? ……...157

9. SCHOOL SHIFT ALLOCATION…….162

9.1 How to Measure the "Net Effect" of the School Shift Allocation? …….164

9.2 Is There Any School "Shift" Effect? Or, Is the Distribution of Pupils across School "shifts" Random? ……166

Discussion……..171

10. SIMULATION FOR SCHOOL SHIFT ALLOCATION…….173

10.1 The Simulation……175

10.2 Outcomes of the Simulation……..176

Discussion……185

11. SCHOOL "SHIFT" ALLOCATION BY EDUCATIONAL AUTHORITIES……187

11.1 School "Shift" Allocation: Analysing CREs 2, 4 and 10…….194

12. IMPACT OF SPECIAL CLASS POLICY ON SCHOOL SEGREGATION……..197

Discussion……..208

13. EVIDENCE OF SCHOOL COMPOSITION EFFECT……...210

13.1 Research Design: How to Test the School-Mix Effect Theory……...215

13.2 Variable Construction and Missing Data…….218

13.3 Is There any Evidence of School-Mix Effect? …….220

Pupil Level Regression……220

Discussion – Pupil Level Regression……..235

School Level Regression……..237

Discussion – School Level Regression……245

14. FINAL CONCLUSIONS.....247

REFERENCES.....251

AGRADECIMENTOS

Essa é a única parte do documento que está escrita em Português. Esse trabalho é fruto de um esforço coletivo de um grupo de pessoas ao longo de quatro anos. Agradeço aos professores e pesquisadores do Observatório da Educação pelo apoio e ensinamentos.

Agradeço à CAPES pelo financiamento para a realização do estágio sanduíche na Inglaterra – bolsa PDSE. Agradeço também ao Programa de Pós-Graduação em Educação da UFRJ.

Quatro pessoas foram especialmente importantes ao longo do curso de doutorado. Sou imensamente grato a professora Mariane Koslinski pelas inúmeras conversas e conselhos ao longo de todo o processo. Os resultados aqui apresentados são, em grande parte, fruto das suas ideias.

Agradeço também ao professor Ignácio Cano por sua genorosidade e rigor. Principalmente por me relembrar que um bom curso faz toda a diferença. O desenho da tese em grande parte é inspirado nas suas ideias. Os eventuais erros são minha responsabilidade.

Ao professor Stephen Gorard por toda a acolhida na Inglaterra. Aprendi ali, no contato quase diário, o ofício do pesquisador, sem muita porpurina ou parafernalha técnica, apenas trabalho duro e rigor metodológico. Ensinou-me também que eu posso mais. Aos demais colegas da Inglaterra, agradeço por me mostrarem que trabalhar mais e melhor vale a pena.

Por fim ao meu orientador Marcio da Costa. O que dizer? Entrei um e saí outro. Minha carreira começa aqui. Obrigado.

LIST OF FIGURES:

FIGURE 1: Social Development Index Metropolitan Area of Rio de Janeiro. Data Provided by Instituto Pereira Passos.

FIGURE 2: Proportion of the Subnormal Agglomerates (Shantytowns) Metropolitan Area of Rio de Janeiro. Data Provided by Instituto Pereira Passos.

FIGURE 3: Types of transfer in Rio de Janeiro Public Municipal Schools – red arrows represent mandatory transfers; white arrows represent internal transfers (non-mandatory).

FIGURE 4: Four different levels of organization within the educational department of Rio de Janeiro.

FIGURE 5: Proportion of Pupils Retained 1st-5th Grade – Cohort 1st Grade, 2006:

FIGURE 6: Proportion of Pupils Retained 1st-5th Grade – Cohort 1st Grade, 2006.

FIGURE 7: Proportion of Pupils Tracked to "*Realfabetização*" – Cohort 1st grade, 2006:

FIGURE 8: Cross-sectional Design for School Segregation Trends (GS) in the First Segment – 2006-2010.

FIGURE 9: Longitudinal Design 1 for School Segregation Trends (GS) in the First Segment – 2006-2010.

FIGURE 10: Longitudinal Design 2 for School Segregation Trends (GS) in the First Segment – 2006-2010.

FIGURE 11: Proportion of Pupils Retained 6th-9th Grade – Cohort 6th Grade, 2006.

FIGURE 12: Proportion of Pupils Tracked to Special Class – Cohort 6th grade, 2006.

FIGURE 13: Cross-sectional Design for School Segregation Trends (GS) in the Second Segment – 2006-2009.

FIGURE 14: Longitudinal Design 1 for School Segregation Trends (GS) in the Second Segment – 2006-2009.

FIGURE 15: Longitudinal Design 2 for School Segregation Trends (GS) in the Second Segment – 2006-2009.

FIGURE 16: Cross-sectional Design for School Segregation Trends (GS) from the 1st to 6th grade (2006-2011).

FIGURE 17: School "Shift" Effect Controlled by Segment

FIGURE 18: Segregation Index (%) for Distortion 1, Considering Shift, Non-Shift and Random Allocation – First Segment.

FIGURE 19: Segregation Index (%) for Distortion 2, Considering Shift, Non-Shift and Random Allocation – First Segment.

FIGURE 20: Segregation Index (%) for Black Pupils, Considering Shift, Non-Shift and Random Allocation – First Segment.

FIGURE 21: Segregation Index (%) for Distortion 1, Considering Shift, Non-Shift and Random Allocation – Second Segment.

FIGURE 22: Segregation Index (%) for Distortion 2, Considering Shift, Non-Shift and Random Allocation – Second Segment.

FIGURE 23: Segregation Index (%) for Black Pupils, Considering Shift, Non-Shift and Random Allocation – Second Segment.

LIST OF TABLES:

TABLE 1: Simulation of a Segregated Educational System

TABLE 2: School Shift Allocation in Two Different Simulations.

TABLE 3: Proportion of Pupils' Transfers Comparing First and Second Segment Cohorts (2006).

TABLE 4: Comparison of Educational Legislation in Rio de Janeiro and Formal School Choice Policy

TABLE 5: Proportion of Missing Data and Disadvantaged Pupil for Pupil's Colour and Parents' Education.

TABLE 6: Proportion of Missing Data for 1st and 2nd Segment.

TABLE 7: Segregation Index (%) for the First Segment. Data Provided by Municipal Educational Department of Rio de Janeiro.

TABLE 8: Segregation Index (%) for the Second Segment.

TABLE 9: Segregation Index (%) for All Indicators of Potentially Disadvantaged Pupils.

TABLE 10: Proportion of Pupils Living in Poverty.

TABLE 11: Proportion of Missing data for Pupil's colour and Parents' Education for each CRE– First Segment.

TABLE 12: Proportion of Missing data for Pupil's colour and Parents' Education for each CRE– Second Segment.

TABLE 13: Proportion of Disadvantaged Pupils for all Variables 2004-2010.

TABLE 14: Proportion of Disadvantaged Pupils Considering Pupils Missing for the Variable Colour.

TABLE 15: Proportion of Disadvantaged Pupils Considering Pupils Missing for Parents' Education 2004-2010.

TABLE 16: Proportion of Disadvantaged Pupils in *Prova Rio* 2010:

TABLE 17: Mean scores for *Prova Rio* 2010 – 3rd Grade –Missing cases for Pupil's Colour.

TABLE 18: Independent Samples Test Missing for Pupil Colour – Portuguese, Cohort 3rd Grade, 2010 (Missing Type 1).

TABLE 19: Independent Samples Test Missing for Pupil Colour – Portuguese, Cohort 3rd Grade, 2010 (Missing Type 1and 2).

Table 20: Independent Samples Test Missing for Pupil Colour – Mathematics, Cohort 3rd Grade, 2010 (Missing Type 1 and 2).

TABLE 21: Mean scores for *Prova Rio* 2010 – 3rd Grade – Missing Cases for Parents' Education.

TABLE 22: Independent Samples Test Missing for Parents' Education – Portuguese, Cohort 3rd Grade, 2010 (Missing Type 1 and 2).

TABLE 23: Independent Samples Test Missing for Parents' Education – Mathematics, Cohort 3rd Grade, 2010 (Missing Type 1 and 2).

TABLE 24: Mean scores for *Prova Rio* 2010 – 4th Grade –Missing Cases for Pupil's Colour and Parents' Education.

TABLE 25: Mean scores for *Prova Rio* 2010 – 7th Grade – Missing Cases for Pupil's Colour and Parents' Education.

TABLE 26: Mean scores for *Prova Rio* 2010 – 8th Grade – Missing Cases for Pupil's Colour and Parents' Education.

TABLE 27: Segregation Index (%) and Dissimilarity Index (%) for All Available Indicators.

TABLE 28: Correlation between GS and D for All Pupil Characteristics.

TABLE 29: Segregation Index and Dissimilarity Index – Percentage Difference (2004-2010 and 2007-2010) for Pupils' Colour and Parents' Education.

TABLE 30: Segregation Index (%) and Dissimilarity Index (%) for All Available Indicators Considering the First Segment.

TABLE 31: Proportion of Potentially Disadvantaged Pupil for Age/Grade Distortions 1 and 2 – First Segment.

TABLE 32: Proportion of Pupils Enrolled in "*Realfabetização*" and PEJA I.

TABLE 33: Segregation Index (%) and Dissimilarity Index (%) for All Available Indicators Considering the 2nd Segment.

TABLE 34: Segregation Index (%) for Black Pupils Considering All Educational Authorities – First Segment.

TABLE 35: Segregation Index (%) for Poor Families Considering All Educational Authorities – First Segment.

TABLE 36: Segregation Index (%) for Parents' Educations Considering All Educational Authorities – 1st Segment.

TABLE 37: Segregation Index (%) for Age/Grade Distortion Considering All Educational Authorities – First Segment.

TABLE 38: Segregation Index (%) for Black Pupils Considering All Educational Authorities – Second Segment.

TABLE 39: Segregation Index (%) for Poor Families Considering All Educational Authorities – Second Segment.

TABLE 40: Segregation Index (%) for Parents' Educations Considering All Educational Authorities – Second Segment.

TABLE 41: Segregation Index (%) for Age/Grade Distortion Considering All Educational Authorities – Second Segment.

TABLE 42: Proportion of Age/Grade Distortion Pupils All Grades in Fundamental Education.

TABLE 43: Correlation Coefficients for All Available Indicators of Disadvantaged (SR) and the Social Development Index – First Segment 2010.

TABLE 44: Correlation Coefficients for All Available Indicators of Disadvantaged (SR) and the Social Development Index – Second Segment 2010.

TABLE 45: Rotated Component Matrix for All Available Indicators (SR) – Higher Proportion of Disadvantaged – First Segment 2010.

TABLE 46: Rotated Component Matrix for All Available Indicators (SR) – Lower Proportion of Disadvantaged – First Segment 210.

TABLE 47: Rotated Component Matrix for All Available Indicators (SR) – Higher Proportion of Disadvantaged – Second Segment 2010.

TABLE 48: Rotated Component Matrix for All Available Indicators (SR) – Lower Proportion of Disadvantaged – Second Segment 2010.

TABLE 49: Rotated Component Matrix for the First Segment for All Educational Authorities.

TABLE 50: Rotated Component Matrix for the Second Segment for All Educational Authorities.

TABLE 51: Proportion of Pupils in the First Segment Who Have Changed to a Different School – Cohort 1st grade, 2006

TABLE 52: Proportion of Pupils Who Have Changed to a School in a Different Educational Authority – Cohort 1st grade, 2006.

TABLE 53: Proportion of Pupils Who Have Changed to a Different School in the Same Educational Authority and Polo – Cohort 1st grade, 2006.

TABLE 54: Descriptive Statistics for Low, Medium and High Performance School – 1st Graders 2006.

TABLE 55: Figures for Upward and Downward Mobility Considering Pupils' Transfer – Only Type 1 and 3 Schools.

TABLE 56: Pupils' Transfer from 1st to 5th grade Considering Initial Type of School (2006-2010).

TABLE 57: Figures for Upward and Downward Mobility Considering Pupil' Transfers among All Types of Schools.

TABLE 58: GS Relative Proportional Increase for Cross-Sectional and Longitudinal Models – First Segment.

TABLE 59: Proportion of Pupils in the Second Segment Who Have Changed to a Different School – Cohort 6th grade, 2006.

TABLE 60: Proportion of Pupils Who Have Changed to a Different School in the Same Educational Authority and Polo – Cohort 6th grade, 2006.

TABLE 61: Descriptive Statistics for Low, Medium and High Performance School – 6th Graders 2006.

TABLE 62: Figures for Upward and Downward Mobility Considering Pupils' Transfer – Only Type 1 and 3 Schools.

TABLE 63: Figures for Upward and Downward Mobility Considering Pupil' Transfers among All Types of Schools.

TABLE 64: Pupils' Transfer from 6th to 9th grade Considering Initial Type of School (2006-2010).

TABLE 65: GS Relative Proportional Increase for Cross-Sectional and Longitudinal Models – First Segment.

TABLE 66: Segregation Index (%) and Dissimilarity Index (D) for the 5th grade 2009 and 6th Grade 2010.

TABLE 67: Segregation Index for the 5th grade 2008 and 6th Grade 2009.

TABLE 68: Segregation Index (%) in a Longitudinal Design – 5th Grade 2009; 6th Grade 2010.

TABLE 69: Descriptive Statistics for All Educational Authorities – First Segment 2010.

TABLE 70: Correlations between Parental Demand and Mean Score of *Prova Brasil* – First Segment 2010.

TABLE 71: Correlations between Parental Demand for Schools and the Segregation Ratio for All Characteristics of Disadvantaged – First Segment 2010.

TABLE 72: Correlations between Parental Demand for Schools and the Segregation Ratio for All Characteristics of Disadvantaged Considering All Educational Authorities – First Segment 2010.

TABLE 73: Correlations between Parental Demand for Schools and the Segregation Ratio for All Characteristics of Disadvantaged Calculated for CREs 2, 4 and 10 – First Segment 2010.

TABLE 74: Segregation Index (%) for Black Pupils Calculated for CRE2, CRE4, and CRE10 – First Segment:

TABLE 75: Segregation Index (%) for Parents Who Did Not Finish Fundamental Education Calculated for CRE2, CRE4, and CRE10 – First Segment.

TABLE 76: Segregation Index (%) for Poverty Calculated for CRE2, CRE4, and CRE10 – First Segment:

TABLE 77: Segregation Index (%) and Proportion of Disadvantaged for Distortion 2 Calculated for CRE2, CRE4, and CRE10 – First Segment.

TABLE 78: Segregation Index (%) for All available Indicators. Data Provided by Municipal Educational Department of Rio de Janeiro.

TABLE 79: Relative Proportional Increase for the School "Shift" Effect for All Pupils Enrolled in Fundamental Education.

TABLE 80: Relative Proportional Increase for the School "Shift" Effect – First Segment.

TABLE 81: Relative Proportional Increase for the School "Shift" Effect – Second Segment.

TABLE 82: Hypothetical Educational System with No School Segregation.

TABLE 83: Segregation Index (%) for Shift and Non-Shift Calculations – First Segment.

TABLE 84: Segregation Index (%) for Shift and Non-Shift Calculations – Second Segment.

TABLE 85: School "Shift" Effect for Black Pupils for All Educational Authorities – First Segment – 2010.

TABLE 86: School "Shift" Effect for Parents' Education for All Educational Authorities – First Segment – 2010.

TABLE 87: School "Shift" Effect for Poverty (NIS) for All Educational Authorities – First Segment – 2010.

TABLE 88: School "Shift" Effect for Distortion 2 for All Educational Authorities – First Segment – 2010.

TABLE 89: Correlation Coefficients for GS Black Pupil and Distortion 2 with Size Effect "Shift" Allocation.

TABLE 90: Relative Proportional Increase for the School "Shift" Effect (2008) Calculated for CRE2, CRE4, and CRE10 – First Segment.

TABLE 91: Relative Proportional Increase for the School "Shift" Effect (2009) Calculated for CRE2, CRE4, and CRE10 – First Segment.

TABLE 92: Relative Proportional Increase for the School "Shift" Effect (2010) Calculated for CRE2, CRE4, and CRE10 – First Segment.

TABLE 93: Proportion of Pupils Enrolled in Special Class for First and Second Segment.

TABLE 94: Segregation Index (%) and Dissimilarity Index (%) Special Class Effect (2009) – First Segment.

TABLE 95: Segregation Index (%) and Dissimilarity Index (%) Special Class Effect (2010) – First Segment.

TABLE 96: Segregation Index (%) and Dissimilarity Index (%) Special Class Effect (2009) – Second Segment.

TABLE 97: Segregation Index (%) and Dissimilarity Index (%) Special Class Effect (2010) – Second Segment.

TABLE 98: Proportion of Missing Data Comparing Pupils Enrolled in Regular Classes and Special Class – 2009 and 2010.

TABLE 99: Variables used in the Logistic Regressions:

TABLE 100: Logistic Regression Estimating Pupil's Chance to be enrolled in Special Class

TABLE 101: Logistic Regression Estimating Pupil's Chance to be Retained.

TABLE 102: Measurements of *Prova Rio* Used to Test the School-Mix Effect Theory.

TABLE 103: Proportion of Missing Data for Prova Rio, Pupil Colour and Parents' Education.

TABLE 104: Variables used in the Linear Regression – Pupil Level:

TABLE 105: Variables used in the Linear Regression – School Level:

TABLE 106: Linear Regression Estimating Proficiency in *Prova Rio* 2010 – Cohort 3rd Grade 2009.

TABLE 107: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 3rd Grade 2009.

TABLE 108: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 3rd Grade 2009.

TABLE 109: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 3rd Grade 2009.

TABLE 110: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 3rd Grade 2009.

TABLE 111: Linear Regression Estimating Proficiency in *Prova Rio* 2010 – Cohort 7th Grade, 2009.

TABLE 112: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 7th Grade, 2009.

TABLE 113: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 7th Grade, 2009.

TABLE 114: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 7th Grade, 2009.

TABLE 115: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 7th Grade 2009.

TABLE 116: Linear Regression Estimating Proficiency in *Prova Rio* 2011 – Cohort 3rd Grade, 2010

TABLE 117: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 3rd Grade, 2010.

TABLE 118: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 3rd Grade, 2010.

TABLE 119: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 3rd Grade, 2010.

TABLE 120: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 3rd Grade, 2010.

TABLE 121: Linear Regression Estimating Proficiency in *Prova Rio* 2011 – Cohort 7th Grade, 2010

TABLE 122: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 7th Grade, 2010.

TABLE 123: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 7th Grade, 2010.

TABLE 124: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 7th Grade, 2010.

TABLE 125: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 7th Grade, 2010.

TABLE 126: Linear Regression Estimating Proficiency in *Prova Rio* 2010 – Cohort 3rd Grade, 2009.

TABLE 127: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 3rd Grade, 2009.

TABLE 128: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 3rd Grade, 2009.

TABLE 129: Linear Regression Estimating Proficiency in *Prova Rio* 2010 – Cohort 7th Grade, 2009.

TABLE 130: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 7th Grade, 2009.

TABLE 131: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 7th Grade, 2009.

TABLE 132: Linear Regression Estimating Proficiency in *Prova Rio* 2011 – Cohort 3rd Grade, 2010.

TABLE 133: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 3rd Grade, 2010.

TABLE 134: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 3rd Grade, 2010.

TABLE 135: Linear Regression Estimating Proficiency in *Prova Rio* 2011 – Cohort 7th Grade, 2010.

TABLE 136: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 7th Grade, 2010.

TABLE 137: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 7th Grade, 2010.

TABLE 138: Summary Model R Square for Type 3 Regressions

ABBREVIATIONS AND ACRONYMS

CONSORT – Consolidated Standards of Reporting Trials.

CRE – *Coordenadoria Regional de Educação*, refered as Educational Authority.

D – Dissimilarity Index.

Distortion 1 – pupils that have one or more years of age/grade distortion.

Distortion 2 – pupils that have two or more years of age/grade distortion.

EducFS – parents who did not finish fundamental school.

EducFS2 – parents who have finished fundamental school.

EducHS – parents who did not finish high school.

EducHS2 – parents who have finished high school.

First Segment/ 1st Segment – aggregates the five initial grades for fundamental education 1st-5th grades.

Gini Index – is a measure of statistical dispersion intended to represent the income distribution of a nation's residents.

GS – Segregation Index.

GSss – GS calculated with "school shifts" as the unit of analysis.

GSsb – GS calculated with "school building" (disregarding the "school shifts") as the unit of analysis.

IRT – Item Response Theory.

NIS – *Número de Identificação Social/* National Social Registry.

NIS_Max – pupil's family was considered poor when his family and/or the pupil had National Social Registry.

NIS_Sum – pupil's family was considered poor only when both the pupil and his parents had National Social Registry.

PBP – Prova Brasil Padronizada/ Mean Score Prova Brasil (2005-2007-2009)

PEJA – *Programa Educacional de Jovens e Adultos.*

PISA – Program for International Student Assessment.

SAEB – *Sistema de Avaliação da Educação Básica.*

SD – Standard Deviation.

SDI – Social Development Index.

Second Segment/ 2nd Segment – aggregates the last four grades for fundamental education 6th-9th grades.

SR – Segregation Ratio.

UNESCO – United Nations Educational, Scientific and Cultural Organization.

RESUMO

O tema central da tese é a distribuição de oportunidades educacionais na rede pública municipal da cidade do Rio de Janeiro. O estudo usa dados secundários fornecidos pela secretaria municipal de educação para descrever padrões de segregação escolar entre 2004 e 2011. Dois conjuntos principais de dados foram utilizados: a) banco administrativo – banco longitudinal com dados sobre todos os alunos matriculados no ensino fundamental, que contém informações sobre o perfil socioeconomico das famílias e a escola que o aluno está matriculado; b) Prova Rio – banco longitudinal com a informação sobre a proficiência dos alunos nos testes padronizados de matemática e Português. O estudo investiga duas questões centrais sobre o tema da segregaçao escolar. A primeira referente ao papel da burocracia e legislação educacional nos níveis de segregação escolar. A segunda parte trata do impacto da segregação escolar no aprendizado dos alunos. As hipóteses testadas ao longo da pesquisa foram extraídas de estudos anteriores na rede pública municipal, e as interpretações apresentadas ao longo da tese estão alicerçadas em pesquisas que entrevistaram professores, diretores, pais e alunos. Há cinco aspectos principais da tese que são inéditos: a) utiliza três indicadores distintos para medir segregação escolar; b) usa dados para toda a população de alunos do ensino fundamental; c) mede o "efeito líquido" de práticas administrativas e da legislação educacional nos níveis de segregação escolar; d) apresenta evidências iniciais sobre o impacto da segregação escolar nos níveis de aprendizado dos alunos (teoria do "efeito mistura"); e) propõe correções para resultados de pesquisas anteriores, que investigaram temas semelhantes da rede municipal do Rio de Janeiro. Em linhas gerais, os modelos sugerem que: 1) as práticas administrativas e a legislação educational impactam positivamente a segregação entre escolas. Dependendo da variável observada, os modelos explicam até 50% da variação nos níveis de segregação; 2) as evidências iniciais sugerem a plausabilidade da teoria do "efeito mistura" (*school-mix effect*).

Palavras-chave: Segregação Escolar; Indicadores de Segregação; Burocracia Educacional; Aprendizagem.

ABSTRACT

The central theme of the thesis is the distribution of educational opportunities in Rio de Janeiro public municipal schools. Secondary data provided by the educational department were used to describe patterns and analyse the impact of school segregation from 2004-2011. The approach combines different datasets: a) Pupil Profile – longitudinal with information on all pupils, including family profile and school placement; b) *Prova Rio* – longitudinal assessment for all public schools. The study addresses two questions regarding the subject of school segregation. First, presents a detailed analysis of potential impact of educational policies or "administrative practices" on school segregation. The second part presents evidence about the potential impact of school compositional effect on pupils' achievement. The hypothesis tested were drawn from previous studies in the public network, and the interpretations are also grounded in solid evidence from previous research that have interviewed teachers, parents, pupils and school principals. The study is unique in five ways: a) measures school segregation using three different indexes; b) uses figures for the entire population of pupils in a cross-sectional and longitudinal approach; c) captures the "net-effect" of "administrative practices" and educational policies in school segregation; d) presents initial evidence about the impacts of school compositional effect on pupils' achievement; e) proposes corrections from previous publications related to educational opportunities in the public network of Rio de Janeiro. Overall conclusions indicate that: 1) a large proportion of between-school segregation variation (can reach 50% depending on the variable) is associated with elements of the educational bureaucracy; 2) Initial evidence supports the plausibility of school composition effect theory. The school composition effect could explain part of the variance that is usually attributed to school management and teacher performance.

Key-words: School Segregation; Segregation Indexes; Educational Bureaucracy; Pupil's Attainment.

1- INTRODUCTION

The central theme of this thesis is the distribution of educational opportunities in Rio de Janeiro public municipal schools. Secondary data provided by the educational department were used to describe patterns and analyse the impact of school segregation. The hypotheses tested were drawn from previous studies about the public network, and the interpretations are also grounded in solid evidence from previous research in which teachers, parents, pupils and school principals were interviewed (Costa, 2008; 2010; Costa; Koslinski, 2008; Costa; Prado; Rosistolato; 2013; Brito; Costa, 2010). The study is unique in five ways: a) measures school segregation using three different indexes; b) uses figures for the entire population of pupils in a cross-sectional and longitudinal approach; c) captures the "net-effect" of "administrative practices" and educational policies on school segregation; d) presents initial evidence about the impacts of school compositional effect on pupils' achievement; e) proposes corrections from previous publications related to educational opportunities in the Rio de Janeiro public network.

Educational inequalities can take several forms and are measured in many different ways. The subjects of interest can vary: a) distribution of resources; b) pupil's chance to attend a school with more privileged schoolmates; c) achievement gaps across different groups; d) access to the curriculum of their choice (Jencks, 1972). The study addresses two questions regarding the subject of school segregation. The first analyses the potential benefits or deleterious effects of widespread, traditional ways of clustering pupils. There is no consensus about the effects of educational policies that intentionally or unintentionally track pupils based on specific characteristics. Most of the reliable evidence comes from international studies that have observed the impact of school choice or tracking policies.

In Brazil, the incipient use of secondary data for cross-sectional analysis or more robust approaches (longitudinal, experimental or quasi-experimental designs), prevents a

better understanding about the impact of such policies. Rio de Janeiro has a good example of a policy that intentionally clusters pupils: Special Classes. The impacts are unknown and, therefore, it should be asked if the policy is effective. Does it achieve its goals? Are there any adverse effects?

A different question assesses the causes of segregation. Research from different countries suggests that territory explains most of the variation regarding school segregation (EGGRES, 2005). Nonetheless, the role of educational policies should not be underestimated. Admission criteria, the unfettered movement of pupils, and an increase on parental choice are some of the regulations that can influence the overall segregation levels.

Perhaps a less prevalent argument, but equally important, states that the debate about segregation must not abdicate from considering an ethical question: is it fair to deny access to a public school based on ascribed characteristics (for example, parents' educational background, race or residential address)? Should white pupils attend class separated from blacks? These questions raise the issue of social justice and equal access to public goods in a democratic society (Rawls, 1971). If these values are at stake, then policy-makers must consider all possible consequences.

The political conception of justice and fairness developed by Rawls (1971) begins by understanding that persons are free and equal. It affirms the following principles:

> I – Each person has an equal right to a fully adequate scheme of equal basic liberties which is compatible with a similar scheme of liberties for all; II – Social and economic inequalities are to satisfy two conditions. First, they must be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they must be to the greatest benefit of the least advantaged members of society. (Rawls, 1971, p 291)

The second principle – fair equality of opportunity – is a key concept in this study. Rawls argues that, in a modern society, individuals should not only have the 'right' to opportunities, but should have an 'effective' equal chance as another similar natural ability.

Do pupils have the same chance to access any given school in the Rio de Janeiro public network?

From a methodological perspective, this is a difficult question to answer. There are many elements that influence school intake, perhaps the most important being residential segregation. Nonetheless, with robust design, the study will present reliable evidence suggesting that pupils (and their families) are not treated as equal by the educational bureaucracy. School staff can, in specific circumstances, select pupils. This selection process occurs in two separate stages: a) school enrolment; b) school shift allocation.

The concept that all men (persons) are equal is central to understanding the transition between traditional and modern societies. Freedom and equality are presented with the statement that each man is an individual, and, therefore, the whole of humanity is inside each man (Dumont, 1992). If it is true that there is no human experience without hierarchy, the analytical perspective should focus on the underlying mechanisms that legitimize agreements between civil servants and members of the general public, which can mitigate, perpetrate or even increase pre-existing inequalities in a modern, equalitarian society. If schools are publicly funded and all men (children) are equal, what legitimizes the selection of pupils? The segregation patterns observed in public schools in Rio, suggests that the differentiation can benefits the advantage group, violating Rawls's (1971) "difference principle": Inequalities would only be permitted to work to the advantage of the worst-off.

Different pieces of research in sociology and anthropology have shown that the Brazilian culture can be characterized with many adjectives that describe a pre-modern society, such as: hierarchy, familism and patrimonialism. Such interpretations helped understanding of how Brazilians behave, and how social inequalities can be shaped and legitimized. Nevertheless, the country is diverse and it is possible to observe conflicts between different groups that share more or less modern values. The degree to which

individuals share values that are more modern is influenced by many factors, including their schooling level (Almeida, 2007). The argument here is that culture matters and beyond any interpretations of the educational legislations (the 'right' to opportunities as Rawls states), the 'effect' of a real chance to access a school should be analysed (Rawls, 1971). Or, in others terms, despite the fact that there are no specific legislations to cluster poor, black or low-achievement pupils, data shows that disadvantaged pupils in Rio de Janeiro public schools have been tracked in an informal way – without specific legislation to regulate the process.

The segregation patterns described in different models suggest that the "informal tracking" is closely associated with prior educational attainment. Beyond any limitations regarding the educational legislation, public schools in Rio de Janeiro face a real pedagogical challenge. Retention is a major problem in the network (affecting approximately 60% of the population) and it is more prevalent among the disadvantaged group. What strategies should be elaborated by the schools and their teachers to deal with highly heterogeneous groups (in terms of age/grade distortion and attainment levels)?

***

The subject of school segregation should be taken seriously by educational researchers. It is closely linked to the debate about the quality and equity of educational systems and to the matter of social justice. In modern societies, public schools became a central institution that should provide (at least in theory) equal opportunities for all, enabling social mobility. Evidence from different educational systems suggests that clustering pupils with similar characteristic can influence how they are treated at school, the quality of teaching[1], aspiration to advanced education subsequent to the compulsory level and an

[1] There are two main reasons: a) The allocation of teachers across schools; b) Teacher's expectation.

increasing association between academic achievement and socio-economic status. (Haarth et al., 2005; EGGRES, 2005; Jenks et al., 1972; Rosenthal; Jacobson, 1968; Brito; Costa, 2010).

Equity is a central issue in most of the countries that have universalized access to basic education. To be fair and just, an educational system should allow pupils from different backgrounds to achieve high levels of attainment. If intelligence and talent are equally distributed among the population, a low degree of association between social origin and attainment should be expected. However, evidence from different educational systems shows a similar pattern, where socio-economic status is associated with the pupil's achievement. Apparently, what varies is the strength of the association (Boudon, 1981; Gorard; See, 2013).

This is the most traditional finding in the sociology of education field. The social composition of schools appears to be the most important variable to explain their overall performance. Parental characteristics, such as educational level and occupation, may be the best predictors of a children's learning trajectory. This single piece of information has conditioned the focus of educational research since the 1960s – at least in the US.

Since then, it has seemed unfair to judge the quality of schools based only on the mean scores of standardized tests. School intakes should be considered in the model in order to produce a more convincing/fair measurement. School effectiveness studies have started to use multilevel models to estimate the value-added score for each school (Brooke; Soares, 2008). The idea behind these was to control for key variables that could influence the regression coefficients. In this case, the profile of pupils explained most of the variation.

Notwithstanding, the new finding had the advantage of starting a public debate (or at least a concern) about school segregation. How was the clustering happening? Put in different words, what causes between-school segregation? Different studies observed that the territory was a relevant variable to explain school segregation. Pre-existenting social inequalities were somehow being reflected in the social composition of schools. Even where there was relative

spatial proximity between different social groups, high levels of school segregation could be observed. One example is the ethnic segregation in U.S. public schools, separating white from black pupils. Many other examples in different countries can be cited, such as, recent immigrants, poor families and religious faith.

It became clear that keeping track of school segregation was a relevant issue for education researchers. One additional concern became clear when findings by Chicago school authors showed the effects of social isolation in minority groups living in large cities. The attempt was to establish causality between individuals (aspirations, choices, behaviour) and the social context where they live (Wilson, 1987). For educational debate, the contribution of urban sociology is clear. The school composition effect (or "school-mix" effect) should be taken seriously, especially for the opportunities of the most disadvantaged pupils (Thrupp, 1997; Thrupp, Lauder, Robinson, 2002; Gorard, 2006). After all, does it matter who goes to school with whom?

Social distinctions were influencing schools intakes. More than this, the educational legislation could exacerbate or reduce the phenomenon. There is some consensus about key elements related to the educational legislation that can increase school segregation. The first, and perhaps the most important, is related to the admission criteria. Whenever a school has control over its intake, schools segregation grows (West; Hind; Pennell, 2004). The explanation is simple: there is a clear perception that some pupils are easier to teach and, therefore, whenever possible, staff member will select these pupils. A second element associated with school segregation is the existence of different "types" of schools (a non-comprehensive system). There are many examples of policies that created differentiation among schools: faith schools, vocational schools, charter schools; etc. These differences appear to facilitate the clustering of pupils with shared characteristics (Gorard; See; 2013).

The Rio de Janeiro municipal educational system can be described as a comprehensive system, which underwent some minor changes that started in 2009. Nonetheless, it is possible to state that all schools present similar infrastructure, curriculum and teacher qualification. Initial analysis shows that the infrastructure is not associated with school performance. The most important variable to explain the performance is the school intakes. Infrastructure, size, and even location, are not good predictors for school performance.

In Rio de Janeiro, the school composition is influenced by three key elements: parents' choice, residential segregation and selection bias by the school staff. Previous studies in public municipal schools suggested parental competition for the most prestigious schools and the selection of pupils based on specific criteria by the school staff. The phenomenon was characterized as "Hidden-Quasi-Market" and is based on the assumption that: a) school intakes are correlated to school reputation; and b) the potential to select pupils is also correlated to school reputation. The analysis suggested that the transfer of pupils across schools was not random, and the effect was an increase in the overall segregation levels (Bruel; Bartholo, 2012).

The research design and methodological approach used in preliminary studies had two major disadvantages. First, it observed a limited number of schools in very specific areas of the city. The limitation was the inability to extrapolate the results obtained in those areas to other places with distinct characteristic. The main threat was selection bias. The second limitation was the option to divide the schools into groups: low and high-performance based on Standardized Tests – Prova Brasil (Test Brazil). The cutting point was arbitrary, usually considering the top quartile high performance schools. The methodological approach, with the typology of schools, tends to increase any difference between schools and potentially

inflate the outcomes. The thesis proposes important revisions in some interpretations related to educational opportunities in Rio's public schools.

The thesis presents an unprecedented research design to analyse segregation in public schools in Brazil. Three different indexes of segregation[2] were used, and the models combine analysis of the entire population of pupils and more focused observations in all ten local educational authorities. Describing the trends of segregation is the first step in testing the hypothesis about the school compositional effect.

The mechanism that explains the phenomenon is fairly simple. It is clear that schools differ in the proportion of potentially disadvantaged pupils, and these figures are associated with overall attainment levels. Thus, a school with a pupil intake with high prior attainment and low levels of family poverty (for example), generally produces higher outcomes compared to schools with a different profile of pupil intake. The nature of school intake should be considered in a value-added analysis to estimate the "school-effect".

It is reasonable to assume that some schools function better than others and produce better results (pupil's attainment level). The next step is to observe the characteristics of the more effective schools. It is possible that they were disproportionately distributed, with large clusters of students with desirable characteristics. In contrast, the less effective schools might be disproportionately those with large clusters of disadvantaged pupils (perhaps family poverty). This is the claim for school-mix effect.

> A compositional effect is said to exist when a variable (such as SES [socio-economic status]) as an aggregated variable at the school level makes a significant contribution to the explanatory model over and above the contribution of the same variable in the model at an individual level. (Harker, 2004, p. 2)

This "compositional effect" is independent of the school-effect, and basically indicates that the pupil's attainment level can also be influenced by the interaction of pupils

[2] The indices will be presented in detail in the methods session.

clustered in specific schools. In theory, the school-mix effect should not be hard to test. The main problem is the lack of good quality data to produce the analysis. It is necessary to possess reliable comparable scores at pupil level, prior attainment scores, background variables at pupil level and the possibility to track pupil's placement over time in the public network.

1.1 What is Disadvantage in Education? Key Variables in the Study

This chapter presents the key variables used in the thesis and discuss more deeply what disadvantage in education is. It is well-established by research from different countries, that educational opportunities and educational outcomes are heavily stratified by pupil background. The strength of the association between socio-economic status and attainment varies among educational systems, but the correlation is fairly constant.

International studies show that attainment at school is strongly related to pupils' characteristics, such as ethnic background, first language, family structure, parental occupation and educational level, and family income. For the early years, the main factors that influences attainment includes parental education, low income, unemployment and early motherhood. All the studies show a clear pattern: potentially disadvantaged pupils are more likely to perform the worst (Gorard; See, 2013). This is strong evidence of a deeply-rooted pattern of social/educational inequality, which educational systems in many democratic countries are far from overcoming.

This is perhaps the biggest challenge for educational researchers. How can interventions effectively reduce the association between SES and educational attainment? Understanding the "poverty gradient" is particularly relevant for policy-makers in order to find appropriate approaches and suggested behavioural changes. How to overcome disadvantage is clearly a causal question, and, therefore, demands a better approach, with robust research design (Gorard; See, 2013).

There is a wide range of explanations why early disadvantage leads to lower attainment at school. It is reasonable to assume that educational opportunities can be heavily influenced by patterns of residential segregation. Clusters of poverty within the city can produce social isolation and have an impact on the quality of schools (infrastructure and teacher recruitment) and how pupils are treated at school. Concentrating disadvantage in

areas or particular schools may also polarise information about future opportunities and remove role models, thus influencing long-term outcomes, such as levels of aspiration, and endangers students' sense of belonging to society (Gorard; See, 2013). There are also family issues, such as parental involvement, children's early environment, and student motivation and behaviour.

To understand disadvantage in education, one needs to understand its causes. It is urgent that educational researchers, especially that are publicly funded, start to use robust research design to establish causality. There is extensive literature describing criteria for establishing the plausibility of a causal model (Campbell; Stanley, 1963; Cook; Campbell, 1979; Cano, 2006; Gorard; 2013). Educational researchers should be able to separate malleable factors from fixed ones, and suggest interventions that are more promising and have a better cost-benefit.

***

Any attempt to overcome educational disadvantage should not ignore the role of school intakes. Clusters of similar students are an important, universal phenomenon (Logan et al.; 2012). Large-scale international studies show the feasibility of combining high quality (overall attainment) and equity (small standard deviation). Analysis using TIMSS, PISA and PIRLS suggest that lower segregation between schools and more comprehensive, egalitarian systems tend to have higher average attainment, with the highest percentage of very skilled pupils (Condron, 2011; Gorard; See, 2013). On the other hand, there is no clear evidence that highly tracked systems, such as those of Germany or Hungary, produce any overall improvement.

Advocates of tracking policies (early selection of students) claim two main advantages. First, that they increase overall performance (for all students), and, second, that

they provide better opportunities for all students, since schools (and probably classrooms) will be more homogeneous. In theory, less variability would allow more targeted approaches (specific pedagogy), benefiting everybody. Since SES and other student characteristics are correlated with attainment (since very early age), this means that selection based on prior attainment will most likely lead to higher social segregation. The cost-benefit of tracking policies seems to be excessively high. There is evidence suggesting potential damage to attainment, especially to the most disadvantaged (peer effect), and wider society outcomes. Social segregation is strongly linked to wider "social ills", such as delinquency. Students growing up in segregated environments may be less prepared for academic challenges, have less access to information and might feel that they have to rely on luck to be successful.

Educational policies are malleable causes of school segregation (with possible impacts on overall attainment levels) and should be a major concern for policy-makers willing to make educational systems more equitable and fair. More importantly, they are, in most cases, cost-free. It is a political decision that has a direct impact on educational opportunities for all pupils and their future life prospects. Newly designed policies can have a direct impact on school segregation, and decisions should be made based on robust evidence from research. This study aims to make a contribution to this matter.

The thesis uses secondary data provided by the Rio de Janeiro Municipal Educational Department, to measure school segregation. Four variables were chosen: parents' education, pupil's colour[3], poverty and age/grade distortion. This last variable summarizes information on all pupils that have not followed a regular age/grade flow in different educational transitions. All the variables have been used in international studies and have a clear association with attainment.

[3] American or European studies tend to use "ethnic background". We think that "colour" is more appropriate to the Brazilian situation, where it is more difficult to assign the cultural dimension present in the "ethnic background" concept.

Parents' education is perhaps the best predictor of children's lifelong learning trajectory. Even in countries with high rates of social mobility (like Denmark), parental characteristics, such as occupation and educational level, are the best predictors of children's success. Controlling by all other variables that might influence educational outcomes, parents that are more educated tend to have children who are more successful at school (Gorard; See, 2013).

In the database, parental education is an ordinal variable, with five possible outcomes: 1) illiterate; 2) did not complete fundamental school – first 9 years of compulsory school; 3) finished fundamental school; 4) finished high school – first 12 years of schooling; 5) finished advanced education. In order to construct the segregation indices, the variable was summarized, creating two potentially disadvantaged groups: 1) parents who did not finish fundamental school (EducFS); 2) parents who did not finish high school (EducHS). More details about the indices and the limitations on the use of the data will be discussed in the Methods section.

The "poverty gradient" is perhaps the most important subject in the field of education and social justice. Overcoming disadvantage in education is, among other things, closing the gap between pupils brought up in poverty and the rest of the population. Reducing poverty is one of the main goals in democratic societies (including international bodies, such as UNESCO), and schools are expected to play an important role in breaking the vicious circle of poverty. Unfortunately, there is little systematic attempt to understand the causal mechanism between family poverty and pupil underachievement. If a government had to choose one single variable to track school segregation, poverty should probably be the one.

Since 1990, the Brazilian Federal Government, along with State and Municipal administrations, has implemented a number of social policies in an attempt to diminish poverty. The income transfer policies, through the National Social Registry (NIS – *Número*

de Identificação Social)[4], has made it possible to identify the families that are eligible to receive this benefit. This is a simple binary variable yielding the total number of pupils at each school who were living in poverty.

The third variable, pupil's colour, has been used in social sciences to assess social inequalities, not only related to educational opportunities, but also in the labour market, exposure to violence etc. This variable was recoded in two distinctive summary variables for potentially disadvantaged groups: 1) non-white pupils; 2) black pupils.

The last variable is called "distortion" and summarizes information on all pupils that have not followed a regular age/grade flow in any educational transition. In order to detect "distortion", two variables were used: the pupil's date of birth and his/her grade for each school year. The potentially disadvantaged group is composed of any pupil that: a) failed any specific school year(s) – retention; b) started school at or over 7 (6 being the correct age); c) abandoned school and returned after a certain period of time. The variable was summarized as: 1) pupils that have one or more years of age/grade "distortion" (Distortion1); 2) pupils that have 2 or more years of age/grade "distortion" (Distortion2).

[44] "*O Cadastro Único para Programas Sociais do Governo Federal, instituído pelo Decreto 6.135/07, é um instrumento de identificação e caracterização sócio-econômica das famílias brasileiras de baixa renda, entendidas prioritariamente como aquelas cuja renda per capita mensal é de até meio salário mínimo. Famílias com renda superior (até três salários mínimos) também podem ser cadastradas, para o planejamento ou implementação de programas sociais específicos. Dessa forma, o número de famílias cadastradas é maior que a quantidade de famílias beneficiadas pelo PBF*." See http://www.mds.gov.br/falemds/perguntas-frequentes/bolsa-familia/cadastro-unico/beneficiario/cadunico-inclusao

1.2 What is Segregation and How Should it be Measured?[5]

The concept of segregation is widely used both in academia, more specifically in the social sciences, and in publications: newspapers, magazines, books. However, its popularity, instead of making the meaning of the concept clearer, actually has the opposite effect, featuring a multiplicity of definitions and uses.

Massey et al. (1996) identify five dimensions of segregation: evenness, exposure, concentration, centralization and clustering. In this research, the concepts of interest are: evenness and exposure. Segregation here is a measure of an uneven distribution of pupils with similar characteristics across different schools. For the purposes of this study, all the variables/characteristics of the pupils suggest potentially disadvantaged groups, such as pupils living in poverty. The reason for this is simple. In most social sciences, segregation has become almost synonymous with stratification, and the main concern of researchers and policy-makers is to produce accurate information in an attempt to reduce social inequalities. It makes no sense to track segregation in the privileged group, since the main objective is to make the educational system fairer and more equitable.

> "Segregation indices are used, therefore, to measure how various social or ethnic groups of people are distributed across a study region, and whether there is evidence or not that they are separated. In themselves, the indices are not restricted to any singular view of the process which led to the separation, or to whether those separations should necessarily be prevented." (Harris, 2012, p. 671)

One simple example provided by Gorard and Taylor (2002) might help clarify the concept of segregation. Imagine a school system with 2 schools and 200 hundred pupils divided into equal numbers between them. If school "A" had only boys (100 pupils) and school "B" only girls (also one hundred), it would be possible to argue that this school system had total segregation by gender. It is important to note that, if 20 boys from school "A"

[5] Previously published in Bartholo (2013).

simply dropped out, we would still have total segregation by gender. However, if any pupil transfers between schools "A" and "B" occurred, then the segregation levels would decline. Put another way, the index that measures segregation should be able to take account of pupil transfers, as described in the last example.

Which indices are available, and how are they related? Since the concept of segregation has been under debate in the social sciences, it is no surprise that many types of indices have been created to measure the same phenomenon. Thus, it is incorrect to state that a specific index is simply wrong or does not detect segregation. Nevertheless, it is important to highlight some desirable properties that such an index should have in order to provide useful information for public policies or for policy analysis. This is not an exhaustive review of all segregation indices, but, instead, an attempt to present the general debate and inform the reader about the options taken by the authors in this dissertation.

Harris (2012) reinforces that a non-exhaustive typology of indices falls into two categories. The first measures segregation as a function of the difference between each individual observation and some average or expected value in a region. Here it is possible to include the Dissimilarity Index (D) (Duncan; Duncan, 1955) and the Segregation Index (Gorard et al, 2003; Gorard; Taylor, 2002; Gorard, 2009a), also denominated Gorard Segregation (GS). The second type is a function of the product of the observed and expected value. The Isolation Index (Bell, 1954; Shevky; Shevky, 1949) is an example. "Whereas the first type of index is comparative, the second is probabilistic" (Harris, 2012, p. 671).

***

The study presents three different indices of segregation: Segregation Index (GS), Dissimilarity Index (D) and Segregation Ratio (SR). Since this is the first attempt to measure between-school segregation involving all Rio de Janeiro municipal public schools, most of

the longitudinal analysis (patterns over time) will use both D and GS. The reason is the open debate about the strengths and weaknesses of each index (Johnston; Jones, 2010, 2011; Gorard Taylor, 2009; Gorard; 2009; Gorard, 2011; Harris, 2012). D is probably the most known index of segregation in education, and its figures will be correlated with GS to show that both indices present very similar trends over time.

The calculations presented in the thesis will show that measuring segregation in the Rio de Janeiro municipal public schools using GS or D is a little different from measuring temperature using the Celsius or Fahrenheit scales (Gorard; Taylor 2002; Gorard 2009; Bartholo, 2013). In the end, both indicators will give similar answers. What changes is the interpretation: GS indicates the exact proportion of disadvantaged pupils who would have to move schools for there to be no segregation; D represents the proportion of one group or other that would have to move, if there were no segregation (Gorard; Taylor, 2002; Gorard; Cheng, 2011).

Formally, GS can be described as the formula above, where: 1) "Fi" is the number of potentially disadvantaged pupils in school "i", where "i" varies from 1 to the number of schools; 2) "F" is the total number of potentially disadvantaged pupils in the Rio de Janeiro public municipal schools; 3) "Ti" is the total number of pupils in school "i", where "i" varies from 1 to the number of schools; 4) "T" is the total number of pupils in the Rio de Janeiro public municipal schools.

$$\mathbf{GS = 0.5 * \{\sum \left| Fi / F - Ti / T \right|\}}$$

Formally, D can be described as the formula above, where: 1) "Fi" is the number of potentially disadvantaged pupils in school "i", where "i" varies from 1 to the number of schools; 2) "F" is the total number of potentially disadvantaged pupils in the Rio de Janeiro

public municipal schools; 3) “Ni” is the total number of non-disadvantaged pupils in school “i”, where “i” varies from 1 to the number of schools; 4) “N” is the total number of non-disadvantaged pupils in the Rio de Janeiro public municipal schools.

$$\mathbf{D = 0.5 * \{\sum \left| Fi / F - Ni / N \right|\}}$$

Both indicators (GS and D) are appropriate for longitudinal analysis, and provide one summary measure for the unit of analysis chosen by the researcher. The limitation is the inability to specify which schools are more or less segregated. The Segregation Ratio (SR) yields a score for each school, indicating if potentially disadvantaged pupils are over or under-represented. Formally, SR can be described as the formula above (with the same elements presented in GS:

$$\mathbf{SR = (Fi / F) / (Ti / T)}$$

> “A critical element of this measure is that the SR of a school is mutually determined with the relative levels of segregation in other schools. If the SR of one school was equal to 1,5, indicating that this school has 50 percent above its ‘fair share’ of a particular subgroup of children, then there would have to be at least one other school with SR less than one […]” (Gorard; Taylor; Fitz, 2003, p. 38)

Assessing segregation using GS and SR allows a bigger variety of designs, combining analysis over time (GS) and across schools (SR). The indices in themselves do not tell much about the causes of segregation or whether these separations should be prevented. Nonetheless, keeping track of segregation can be relevant for policy-makers and researchers. A fair and equitable educational system should not tolerate high levels of segregation, and democracies should provide information for taxpayers on how pupils are being allocated into

public schools. As an example, England keeps official records of school segregation and considers it a relevant indicator in its educational system. Publicly funded schools should be open to all families. The research describes, for the first time, how segregated Rio de Janeiro public schools are. Causes and consequences of school segregation are also analysed.

***

In this study, the main element of segregation, as previously mentioned, is evenness. Therefore, the index must be a measure of the uneven distribution of pupils with shared characteristics across different schools. Since there are many indices available in the "market", it is important to be concerned with the adequacy (strengths and weaknesses) of each indicator before choosing one. Gorard (2009a), dealing with a specific poverty index, highlights four desirable properties that such indices must present, regardless of the research field:

> "1) organisationally invariant, such that if a school is broken into two, or if two schools merge, with the same proportion of FSM [Free School Meal] pupils in all, then the value of the index remains the same; 2) size or scale invariant, such that if the number of both FSM and non-FSM pupils is multiplied by a constant in all schools, then the value of the index remains the same; 3) compositionally invariant, such that if the number of FSM pupils is multiplied by a constant in all schools, then the value of the index remains the same (equivalent to the margin-free criterion in sex segregation analysis) and; 4) affected by transfers, such that if an FSM pupil moves from a school with more FSM pupils to a school with less, then the value of the index goes down." (Gorard, 2009a, p. 644)

One of the main issues debated by researchers is that the index must not change just by a simple shift in the numbers of potentially disadvantaged pupils in a specific region. This is a crucial element, especially in situations where the researcher is interested in segregation patterns over time – cross-sectional or longitudinal approaches. Since it is very likely that the

number of disadvantaged pupils will fluctuate over the years, it is important to have an index that does not change just because of a simple shift in the total number of pupils in the group of interest.

An example might help clarify this key concept for analysis conducted in the thesis. Recalling the previous example with two schools ("A" and "B"), both with exactly the same number of pupils – one hundred. If school "A" has 30 pupils considered poor and school B only 10, the index should be able to capture some degree of segregation. In this case, GS would show a 25% level of segregation. If the number of poor pupils is multiplied by a constant in both schools, GS will remain the same. Following the example, if school "A" presents 60 pupils considered poor and school "B" a total of 20 poor pupils, GS will still show the same 25% segregation. For all the cross-sectional and longitudinal designs, this desirable property of GS is crucial.

Table 1: Simulation of a Segregated Educational System

	Poor Pupils	**Total**
School A	30 (60)	100
School B	10 (20)	100
Total	40 (80)	200

Gorard and Taylor (2002) make a detailed analysis comparing D and GS, highlighting that the first is "weakly" composition invariant and the second is "strongly" composition invariant. On the other hand, the authors emphasize that GS is not symmetric, meaning that, if the index is calculated inversely, with the privileged group as the focus group, then the outcome given by the index is different, but not contradictory. It could be argued that the lack of symmetry is a smaller problem, if the main/only concern of a segregation index is the disadvantaged group.

Organizationally invariance is another important property that segregation indices must have. One of the designs presented in the thesis, measures the impact of the allocation of pupils across "school shifts" – morning, afternoon and night. In order to measure the "net effect" of this policy, the index of segregation must be organizationally invariant. Using the same example in Table 1, with only two schools, and assuming that schools "A" and "B" have two "school shifts" – morning and afternoon – each shift with 50 pupils, if disadvantaged pupils are equally distributed among shifts, the levels of segregation calculated by GS must not change. In our simulation, the figures would change to 4 schools (or school shifts), as shown in Table 2.

Table 2: School Shift Allocation in Two Different Simulations.

	Poor Pupils	**Total**
School A1	15 (25)	50
School A2	15 (5)	50
School B1	5 (9)	50
School B2	5 (1)	50
Total	40	200

In this hypothetical scenario, GS would not change – 25% segregation – comparing to Table 1. The index correctly captures that, if a school split into two, keeping the same proportion of disadvantage in both, the segregation remains the same. However, if the allocation of pupils across shifts concentrates the disadvantaged group in one specific shift (selection), GS should increase.

A new simulation (also in Table 2) shows a non-random allocation of disadvantaged pupils across shifts. School shifts A1 and B1 have more poor pupils than its expected fair share – A1 with 25 and B1 with 9. In this case, GS increases up to 38%. This is an example of how schools can become more segregated after the shift allocation.

Complete evenness is an unrealistic demand for all educational systems. Therefore, some level of segregation is expected for all calculations presented in the thesis. In fact, Cortese et al. (1976) suggests that segregation should be assessed and interpreted against a random distribution of population characteristics. The study uses simulations to deal with some of these challenges.

2- THE ENROLMENT LEGISLATION IN RIO DE JANEIRO PUBLIC SCHOOLS

The educational system in Brazil is currently divided into three different schooling levels: a) Pre-School – children aged 4-5, not compulsory; b) Fundamental Education – compulsory, catering for pupils aged 6-14, usually divided into 5 initial grades (first segment) and 4 upper grades (second segment); c) High School – not compulsory, for pupils aged 15-17.

Rio de Janeiro city has the largest public school network in Brazil. There are around 1,300 schools providing Pre-School and Fundamental Education. The thesis focuses on the only mandatory educational level, Fundamental School, analyzing the total population – around 700,000 pupils in 900 schools for each school year. Despite the fact that there is data available about Pre-School, they will not be included in this analysis, mainly because this type of education has not reached universal attendance.

Public municipal schools in Rio de Janeiro are administratively grouped into small sets (Polos). There are around 111 sets in the entire network (the number can vary depending on the year observed) that are clustered into 10 Educational Authorities (CREs). In 2011, one more Educational Authority was created. All these Educational Authorities comprise the Rio de Janeiro Municipal Educational Department.

It is also important to mention that around 18% of the pupils in the whole country are enrolled in private schools. In Rio de Janeiro, this number is even higher, reaching 25%. Unfortunately, there is no data available for the private schools to allow a more robust analysis considering the entire pupil population in Fundamental Education. Since the private schools are mainly frequented by the middle and upper class, it is reasonable to assume that the segregation levels presented in this paper are underestimated. The reason is simple. The data used deals with a more homogeneous part of the population and, therefore, it is most

likely that the part of the variation that would enhance the indices used to measure segregation will be left out of the analysis.[6]

Previous studies in Rio de Janeiro city highlighted a singular pattern of residential segregation that combines spatial closeness with social distance. The urban sociology has long shown that the isolation of certain groups (for example, poor families) in distant parts of the city can have a negative impact on individuals from the most deprived neighbourhoods, beyond the simple disadvantage of being poor. Presumably, isolation diminishes the opportunities of interaction among different groups and could have a greater deleterious effect for those at a potential disadvantage (Wilson, 1987). This pattern of residential segregation, called core and periphery, is common in many cities in the U.S. and European countries.

Nonetheless, Ribeiro and Koslinski (2009) showed that the core and periphery model cannot be fully applied to Rio de Janeiro and, presumably, not to other big cities in developing countries. In many cases, the cities present a more complex pattern of residential segregation, with some clusters of poor communities spread all over the city, including the most exclusive neighbourhoods. Figure 1 shows the Social Development Index (SDI) for different areas in the metropolitan region (Cavalliere; Lopes, 2008), and Figure 2 highlights the proportion of the population living in shantytowns (Subnormal Agglomerates) for the same areas.

[6] Very recently, our research team had access to the entire educational data, including private and public schools, but it will be necessary to check the quality of the information and also "clean" the datasets before starting the analysis.

Figure 1: Social Development Index Metropolitan Area of Rio de Janeiro. Data Provided by Instituto Pereira Passos.

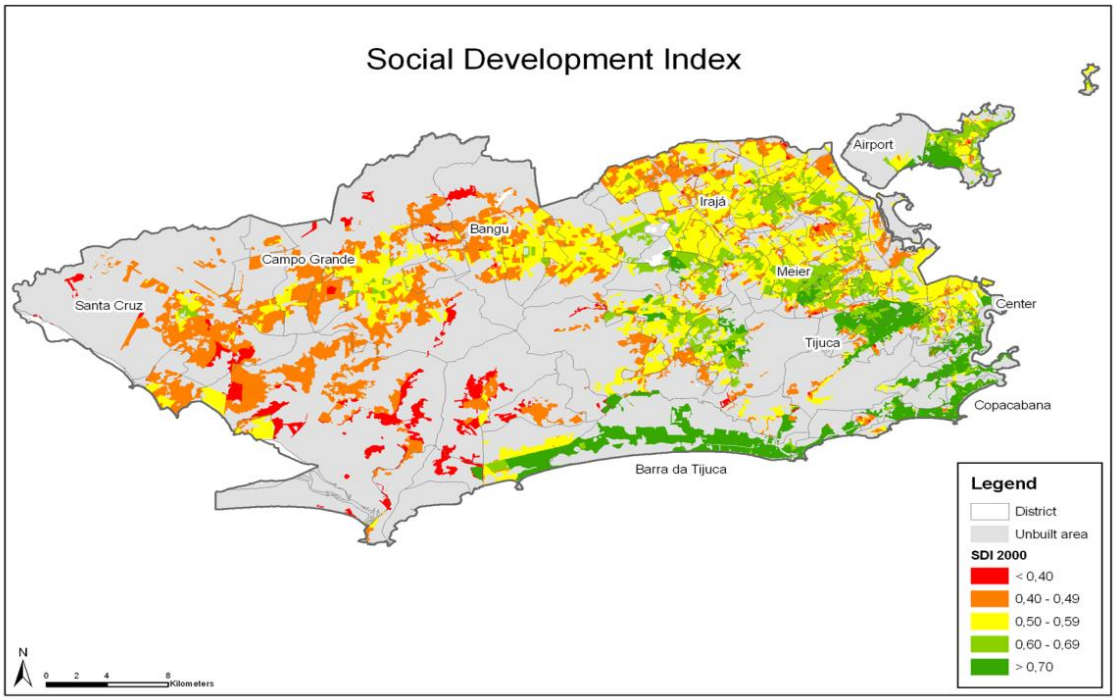
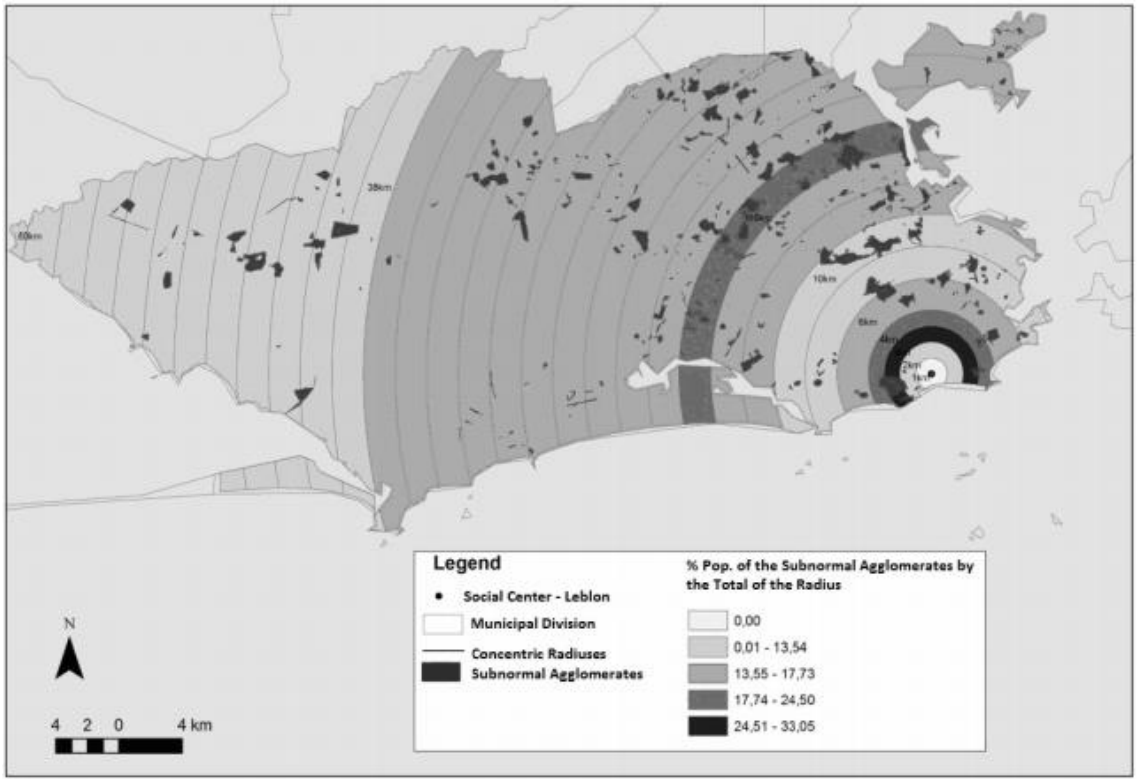


Figure 2: Proportion of the Subnormal Agglomerates (Shantytowns) Metropolitan Area of Rio de Janeiro. Data Provided by Instituto Pereira Passos.



It is possible to identify two concomitant segregation processes. The first, similar to many European cities, shows the most developed area, close to the coast, apart from the less developed. However, the same figure highlights that even areas with a very high Social Development Index (SDI) can be very close to poor neighbourhoods. It is the phenomenon of the shantytowns ("favelas") that characterizes the city. Around 25% of the population that live in the most exclusive neighbourhoods actually live in a shantytown. In fact, it is possible to observe upper, middle and lower class people living very close to each other, but with little social interaction (Ribeiro et al., 2010).

It is reasonable to presume that this type of residential segregation would have a positive impact on between-school segregation. However, previous studies have shown the opposite reality (Ribeiro; Koslinski, 2009; Bruel; Bartholo 2012; Costa; Koslinski, 2011;

Costa, 2008). It is possible to point out at least two hierarchical school stratification levels: 1- regional segmentation, according to the patterns of socio-economic inequality; 2- segmentation within each region of the city. The distribution of students attending public municipal schools seems to have a complex pattern that overlaps/conjugates socio-economic aspects, residential segregation and academic performance. However, it is relevant to notice that previous research analyzed a limited number of schools, with possible implications for selection bias. This is the first attempt to measure segregation in a major Brazilian city using figures for all public schools.

Territory is only one of the elements that can influence the patterns of school segregation. Another key issue is the educational policy, which presumably can increase or prevent the clustering of pupils. In general terms, the municipal public educational system, not only in Rio de Janeiro, but also in other cities of the country, can be described as a comprehensive school system. There are no big differences among schools and, at least in theory, all schools should provide the same curriculum.

More recently, there have been some initiatives in Rio de Janeiro to create different "types" of schools, especially in the most vulnerable areas. One example is the recent policy called "Escolas do Amanhã" (The Schools of Tomorrow), which started in 2009, in 150 schools, and intends to improve the overall attainment levels in schools located in poor neighbourhoods or the so-called "sinking schools" (with very low academic results). The policy provides extra funding so that schools can extend the total study hours per day and offer more diverse activities to the pupils. There is also an economic incentive for teachers who accept to work in these schools. The focus is on hiring more skilled and experienced teachers, as to provide extra school time, not usual in Brazil. Nonetheless, even in the so-called "Schools of Tomorrow", the curriculum does not vary significantly when compared to regular public schools.

A separate issue and central to the analysis of the study is the enrolment legislation. A key element that can influence the social composition of the schools is the fact that parents can choose any school of their preference in the public network. Different international studies have analysed the impact of school choice policies on the levels of school segregation, and the results are not conclusive. There are two explanations for the contradictory outcomes. The first relates to the differences in the policies themselves and the fact that parental choice is enhanced by others elements, such as public transportation and the supply of schools (called "real" choice). A different element refers to the quality of the evaluation studies. Many researches present poor design (the "study case" is one example) to propose any causal mechanism and the results should be analysed with great caution. Another big issue is the lack of studies that use random samples of schools or with secondary data that present information for all the schools. The major treat is selection bias, and the outcomes observed cannot be extrapolated for the entire network.

It is worth mentioning two other elements, linked to the open enrolment policy, that can influence segregation. All pupils enrolled in public schools in Rio de Janeiro have free public transportation, with a student card (RioCard). The system is in full operation and increases mobility, especially for older pupils – second segment age of 11 to 14.

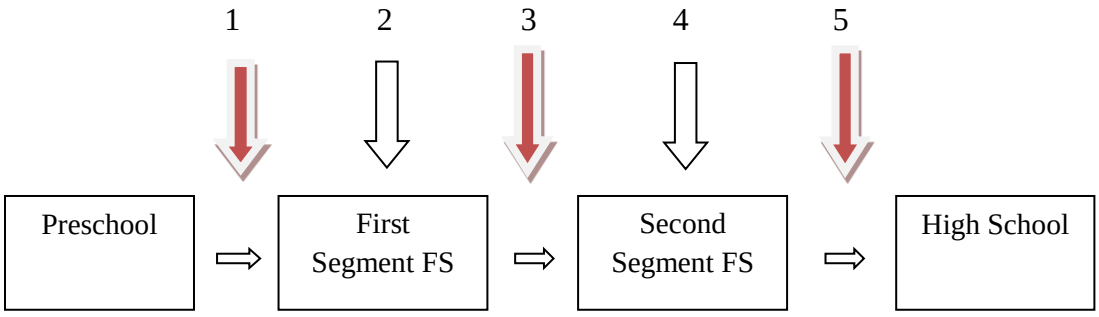
If the aim to increase parental choice (and mobility) is to allow families to choose the 'best' school for their child, even if that school is far from their neighbourhood, it would be desirable to provide as much information as possible about schools. This is absent from the policy. There is very little public information about the schools, and the one available is not easy to find or interpret. Basic relevant information for parents could help families to make better informed decision: graders offered, total number of pupils enrolled; projects in which the school participates; number of drop-outs and retentions (previous years); position in league tables by *Prova Brasil* and *Prova Rio*.

Chile and England have implemented formal school choice policies and presented interesting solutions to increase families' knowledge about schools – especially those close to home. West and Pennel (2002) argue that the effort to provide information about schools in a systematic and pedagogical way is a relevant policy, especially for the potentially disadvantaged pupils. There is evidence from different studies suggesting that upper class families are more capable of collecting relevant information about schools and, therefore, of making a more sensible and informed choice (Thrupp, 2007; Van Zanten, 2005; Gorard; Taylor; Fitz, 2001; 2003).

Rio de Janeiro enrolment legislation indicates three different types of registry: first enrolment; renovation; by transfers. The first enrolment corresponds to pupil's first entry into the municipal public network, and can happen at any given grade. Nonetheless, it is possible to state that it is more prevalent in pre-school and the first grade of Fundamental School. The renovation is an "automatic" enrolment that occurs at the end of each school year, when the pupil will remain in the same school (Deliberation no 08, October 2001, art. 3 and 4).

The transfers of pupils across schools are key to understanding the patterns of segregation. In accordance with the current legislation, there are two types of transfer: mandatory (*remanejamento*/relocation) – red arrows in Figure 3; internal transfer (non-mandatory) – white arrows in Figure 3.

Figure 3: Types of transfer in Rio de Janeiro Public Municipal Schools – red arrows represent mandatory transfers; white arrows represent internal transfers (non-mandatory).



The first type (mandatory transfer) occurs in three situations: transition between Pre-School and Fundamental School (1st grade); transition from 5th grade to 6th grade – both at Fundamental School; transition between Fundamental School (9th grade) and High School. The internal transfers can happen at any given grade. The reason why the norms that regulate the pupil transfers are so important in this study is because previous researches have suggested that the changes are not random and increase school segregation (Bruel; Bartholo, 2012). The thesis focuses on the transfers indicated by arrows 2 and 4 (internal transfers – non-mandatory) and arrow 3 (mandatory transfer).

Table 3 shows the total proportion of pupils who have changed school in different educational transitions. The data is specific to two cohorts: a) 1st graders, 2006; b) 6th graders, 2006.

Table 3: Proportion of Pupils' Transfers Comparing First and Second Segment Cohorts (2006).

	2006-2007	**2006-2008**	**2006-2009**	**2006-2010**
First Segment – Initial Grade 1st	13.5%	21%	30%	33%
Second Segment – Initial Grade 6th	10%	16%	18%	-

The figures indicate that, after 4 transitions (non-mandatory) in the first segment of Fundamental School (1st to 5th, grade), around one third of the pupils have changed school. For the other cohort (6th grade, 2006) the figures show that, after three transitions, around one fifth of pupils change school. These are high figures and the causes (why so many pupils change school?) and their consequences for school segregation should be investigated.

The norms that regulate the first enrolment and all subsequent transfers across public municipal schools in Rio de Janeiro have been changing since the end of 2009. The changes have the potential to affect the school segregation patterns, and will be presented in detail. The new regulation (2009/2010) alters the core of the decision-making process of the school bureaucracy (school staff and the principal) to a higher, and less personalized level. In theory, the new regulations can have a positive impact to prevent patrimonialistic practices by civil servants and make educational opportunities more democratic.

The old regulation of 2001 (Deliberation n.º 08, October de 2001) was vague and unclear about many aspects of the enrolment process. For the first registration, it allowed parents to exercise unlimited choice, but the school staff (represented by the school principal) decided whose preference (parental choice) will be honoured. Interviews conducted with school principals and family members in 2010 and 2011 showed that both the first enrolment and the internal transfers (non-mandatory) were made at the enrolment center (*Pólo de Matrícula*) or directly at the school. A specific calendar stipulated a period for all transfers

and first enrolment (usually around October or November). However, data from the interviews suggest that it was possible to issue a transfer in any given month, directly at the school.

At the enrolment center, the entire process was controlled by the school principal or another staff member appointed by him. Each school filled a spreadsheet (called "vacancy board") indicating the total number of vacancies. The number of open vacancies was not publicized, and there was no external control by a higher authority to audit the veracity of the information presented by each school. This situation allowed the school staff to have almost total control over their intake. In the words of one principal interviewed: "My public school is private" (Interview School Principal, 18/05/2011).

There were no specific criteria to regulate situations of oversubscription – requests of internal transfers or first enrolment – in each enrolment center. This bizarre situation created huge queues on the days stipulated to issue the pupil's registration – called "enrolment calendar". As stated by a principal, who had worked for 20 years in an enrolment center, the only known criterion was the order of arrival in the queue. There were reports of parents who had overnight in the queue trying to have access to more prestigious schools. For the parents, the struggle was not to have access to a public school (Fundamental Education has reached universal coverage in Rio de Janeiro), but the lack of schools with a good reputation.

The absence of regulation triggered two complementary phenomena. On the one hand, it allowed school principals to manipulate the total supply of open vacancy and select pupils. On the other hand, families used different strategies to gain access to more prestigious schools. Costa, Prado and Rissolato (2013) created a typology of families' strategies to proceed with the first enrolment and internal transfers of their children. The first one is using legal and bureaucratic procedures (for example, attending the enrolment center and choosing one of the available school) to acquire the enrolment. The second exclusively uses the

family's network (friends, relatives, teachers or members of the school staff) to gain access to a specific school. Interviews with parents indicated that teachers, principals and even local deputies can help in securing access to more prestigious schools. The third strategy mixes both approaches, using legal and bureaucratic procedures, combined with the family's network to achieve success in the enrolment process.

Interviews with family members, teachers and principals confirmed the existence of patrimonialistic practices in Rio de Janeiro public schools. It is what part of the international literature about school segregation calls principal as "gatekeepers" (Smylie et al., 2004). Perhaps the most symbolic practice was the tradition of the "principal's notebook" (*caderninho*), which represented the power to decide whose preference would be honoured. In any month of the year, parents could approach principals at the school in an attempt to convince them to accept the enrolment. The principle could decide to write down the pupil's name in the "notebook", securing a place for the next school year. Data from the interviews do not suggest that all schools were intentionally selecting pupils, but they all had the potential to do it. This scenario is the opposite of what West and Pennel (2002) described as desirable and fair in an educational system.

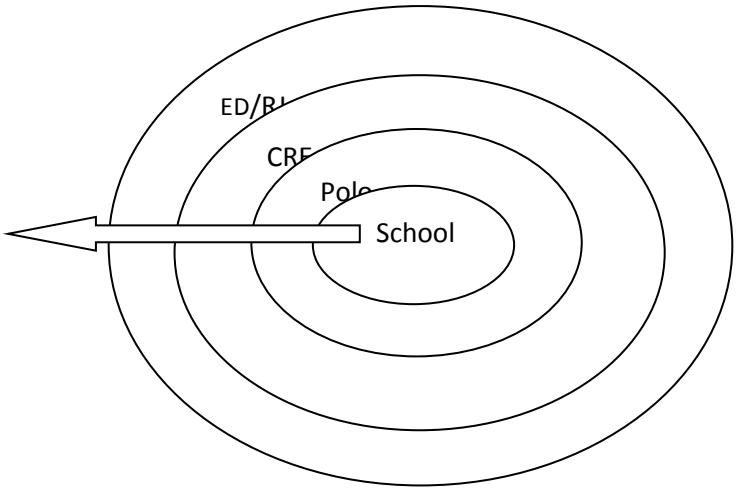
The mandatory transfers (relocation/*remanejamento*) under the 2001 regulations, allowed very little choice to parents. In this case, a predetermined school was assigned to each pupil in a previous agreement among schools – called "*convênio*". These agreements were informal and could be changed without notice.

A new regulation edited in 2009, along further regulations in 2010 (Ordinance E/SUBG/CP no. 24, 22nd October 2010), introduced important changes for the first enrolment and internal transfers. All the changes had the potential to shift the decision-making power from the school staff (usually the principal) to a higher authority. A new system, with online and telephone services, was designed. Parents were asked to choose

between 3 to 5 schools, and pupils would be randomly assigned by software according to their parents' preference. In theory, the new system diminishes the power of the educational staff to intentionally select pupils.

Figure 4 shows four different levels of organization within Rio de Janeiro educational department. The arrow in the figure suggests that the decision-making power is being transferred to a higher, less personalized, authority. In democratic societies, with publicly funded schools, policy-makers should prevent public schools from selecting pupils. Strategies, such as written tests or interviews should not be used in enrolment procedures. The risks of social selection are far too high (West; Pennel, 2002; Gorard; Taylor; Fitz, 2003).

Figure 4: Four different levels of organization within the educational department of Rio de Janeiro.



The random allocation of pupils was regulated by legal priorities such as: 1) candidates with special needs; 2) candidates that have siblings in the school of their preference; 3) a child whose parent is a civil servant working in the school; 4) an adopted

child (Municipal Law no. 2.210, 21st July 1994); 5) a foster child (Ordinance E/SUBG/CP no. 24, 22nd October 2010, art. 7).

The two mandatory transitions (relocation/*remanejamento*) were not included in the new randomized electronic system. If a pupil studies in a school that does not offer the subsequent grade, the registration is automatically renewed to the school closest to his/her home or to the school of the families preference (Ordinance E/SUBG/CP no. 24, 22nd October 2010, § 1 of art. 3).[7]

These are important changes that show the complexity and the urgent demand for more transparency within the public sector in Brazil. The power delegated to principals to select pupils is not democratic and should be a concern for policy-makers. Preventing black, poor or shantytown pupils from attending high performance schools violates Rawl's (1971) principle of fairness and justice in a democratic society. The new regulation appears to correct some important distortions, but the outcomes for school segregation are still unknown. Interviews with members of the Educational Department of Rio de Janeiro indicate that the random allocation, with the electronic system, did not work well in 2009. Apparently the system started at the end of 2010, with potential impact on the composition of schools in the academic year 2011.

[7] "[...] *segmento subsequente ou classe pretendida terão suas matrículas garantidas, por meio de remanejamento, para a unidade escolar mais próxima de sua residência ou para a unidade escolar de opção do responsável*"

3- SCHOOL CHOICE IN THE BRAZILIAN PUBLIC SYSTEM: MISUSE OF CONCEPTS[8]

The educational system in Brazil has experienced important changes since the beginning of the 1990's. Possibly, the two most significant achievements have been the universal attendance of pupils in Fundamental Education and the overall increase in the number of grades completed. Nonetheless, another important accomplishment has been the establishment of a National Pupil Census, along with a National Assessment System of pupil proficiency in Mathematics and Language (with standardized tests). For the first time, researchers have access to reliable comparable data to analyze the overall levels of attainment for both the private and public sectors. Along with these changes, Brazil has also started to participate in international assessment tests (like PISA) that allow international comparison (Veloso, 2009).

All these figures highlight two problems: an "old" one (widely discussed in previous research) and perhaps a "new" one. The achievement gap between the private and public sector was clear in the standardized tests. Good private schools in Brazil have outcomes comparable to those of any of the high performance schools in the top ranking countries in PISA. However, the overall level in the public sector is among the lowest, not only in comparison to South America as a whole, but also to any other developing country. Brazil lies at the bottom in any international assessment, which suggests not only failure of the public schools, but also high levels of social stratification.

Nonetheless, the standardized national tests have allowed researchers to observe a rather underestimated problem within the public sector. Analyzing only public schools, it was also possible to observe big differences in their overall outcomes, which triggered new interest in understanding why some public schools were better than others. Once again, as

[8] Published in Bartholo (2013).

observed previously in the public versus private sector dichotomy, a major part of the variation in the overall achievement levels in public schools could be explained by their pupil intakes. Immediately, researchers started to address two complementary questions: a) How do parents choose their child's school? b) What is the role of the educational legislation in the overall segregation levels?

Most of the research conducted so far has been in major cities like Rio de Janeiro, São Paulo and Belo Horizonte, all urban areas with large, well-established public networks. It is important to note that in Brazil, each State has autonomy to legislate regarding its enrolment policies, which makes comparison more complex. The national legislation just defines general guidance as the compulsory ages of school enrolment.

Recent research regarding stratification in public schools in Rio de Janeiro and São Paulo has used the concept of School Choice and Quasi-Markets in an attempt to analyze the distribution of educational opportunities and different school segregation patterns (Costa; Koslinski, 2011; 2012). Despite the fact that there are no formal School Choice policies in Brazil, some cities present specific legislation that, in theory, allows parents to exercise choice (open enrolment). Some additional policies, such as free public transport for pupils, potentially increase the possibility of attending a school far from home. The results suggest two key elements related to school segregation: 1) first enrolment policies; 2) unfettered movement of pupils among schools (Bruel; Bartholo, 2012).

Admission policies that combine social selection with patrimonialistic practices by civil servants have been analyzed in Rio de Janeiro since 1950 (Consorte, 1959), and, even today, researchers are trying to understand different pupil mobility patterns among public schools (Costa; Koslinski, 2008). Findings by Bruel and Bartholo (2012), analyzing the transition between the first (1st–5th grade) and second (6th–9th grade) segments, of a limited number of pupils in Rio de Janeiro public schools, suggest that variables, such as, parents'

education and pupil's colour impact the chances of access to the most prestigious schools. The same study shows that the previous school in the first segment is the most important variable to predict access to a high performance school (good reputation) in the second segment.

Empirical analyses considering large datasets were interpreted with the theoretical framework of School Choice. The concept of "Hidden-Quasi-Markets" was adopted to characterize Rio de Janeiro educational policy (Costa; Koslinski, 2011). However, it is important to make some comments about the limitations of the use of school choice theory to characterize such educational systems. They do not present a price mechanism (per capita funding) and, most importantly, the school staff (usually the school principal) decides whose preference (parental choice) will be honoured, especially in the case of oversubscription.

The Rio de Janeiro educational legislation allows parents to choose any school. There is a specific calendar with key dates for first enrolment, confirmation of school registration and formal requests to change school. Recently, the Educational Department has provided information about the performance of all public schools in an attempt to better inform parents (schools "performance" table). So far, there is no solid evidence of how parents are using this new information in their decision-making process.

If it is true that parents purportedly freedom of choice, it is also crucial to highlight that school staff (usually represented by the principal) have control over their intake, especially in the case of oversubscription. Since schools have different reputations, it is more likely that oversubscription occurs mainly in two scenarios: a) highly dense regions with a limited supply of public schools. Just a few areas of the city could be put into this category of Fundamental and High School education (Alves; Lange; Bonamino 2010); b) schools that have a good reputation – so-called high performance (Bruel; Bartholo, 2012).

Logic suggests that schools with better reputations would, on average, present higher subscription rates and, therefore, could potentially select their pupils. A vicious circle where: a) school intakes are correlated to school reputation and; b) the potential to select pupils is correlated to school reputation.

The absence of any price mechanism (fees-charging for parents in public schools or funding schools according to the number of pupils) does not stimulate schools to compete for pupils. The only type of funding truly linked to each pupil is the money for lunch, to which all pupils in public schools are entitled. Any other resources are not directly linked to the number of pupils enrolled. This context increases the power of the educational staff to simply deny access to potentially disadvantaged pupils, since there will be no real consequences – including school funding. Table 4 presents a summary comparing the Rio de Janeiro educational legislation with that of other cities that formally implement school choice policies.

Table 4: Comparison of Educational Legislation in Rio de Janeiro and Formal School Choice Policy

	Rio de Janeiro Policy	**Formal School Choice Policy**
Open Enrolment	**X**	**X**
Schools Select Pupils	**X**	
Information About Quality of Schools	**X**	**X**
Price Mechanism (funding per capita)		**X**

Future research should reconsider the use of concepts, such as School Choice or Hidden-Quasi-Markets to characterize the educational system in Rio de Janeiro or São Paulo.

It could be argued that it is the absence of a school market, and not the existence of one, that characterizes such educational systems. Evidence suggests that, in the case of Rio de Janeiro, families compete for schools with good reputations, but schools do not seem to compete for pupils. In reality, the data indicates more of a collaboration process, where schools "exchange pupils" based on specific criteria (Costa; Koslinski, 2012). Perhaps, the concept of "Market Ecology" would be more helpful in characterizing a system where collaboration among schools seems to be more prevalent than competition (Yair, 1996).

4- METHODS

The thesis uses secondary data provided by the educational department of Rio de Janeiro. The analyses combine cross-sectional and longitudinal designs to assess the causes and impacts of school segregation. The main benefits and challenges of this type of data will be discussed in the next section. The analysis can be divided into three parts: a) description of the segregation patterns; b) impact of "administrative practices" and educational policies on school segregation; c) potential impact of school segregation on pupils' achievement – testing the plausibility of the school-mix effect theory.

The empirical analysis presents a great diversity of designs, with much specificity regarding sampling, segregation indexes, and other statistical techniques. For this reason, all the designs will be fully discussed in each specific section. For the moment, the idea is to show an overall view of the designs and main goals for the subsequent chapters.

First Part of the Study

The initial description presents two levels of analysis: a) models considering the entire population for the first and second segments of Fundamental Education; b) models focused on each educational authority – a total of ten. The Segregation Index (GS) and Dissimilarity Index (D) will be used to assess patterns of school segregation over time – 2004-2010. The initial description is followed by Factor Analysis, using the Segregation Ratio (SR), to answer one question: is between-school segregation one single process or should more than one explanatory model be considered?

Second Part of the Study

The second part assesses the causes of segregation. The study focuses on malleable causes, more specifically, on the impact of "administrative practices" or educational policies

on school segregation. There are five independent analyses: 1) first segment educational transitions ("voluntary" transfers) – 1st-5th grade ; 2) second segment educational transitions ("voluntary" transfers) – 6th-9th grade; 3) "mandatory" transfers – 5th-6th grade (relocation/*remanejamento*); 4) Special Class policy ("*realfabetização*"; PEJA I and II); 5) "School shift" allocation.

To assess the impact of "voluntary" transfers on school segregation, the study combines cross-sectional and longitudinal approaches. Two different cohorts were chosen: pupils in the 1st grade/ year 2006; pupils in the 6th grade/ year 2006. For the first segment, a total of four educational transitions are assessed (1st-5th grades) and, for the second segment, three transitions (6th-9th grades). The main hypothesis is that pupils' transfers across schools are not random and could have an impact on school segregation. Cross-sectional and longitudinal approaches will present a slightly different view of the same phenomenon. Because retention is a major issue, affecting almost 60% of all pupils in the public network, the longitudinal approach can help observe the potential impact of retention on school segregation.

The impact of "mandatory" transfers (relocation) will be assessed in cross-sectional design, comparing GS figures in the 5th and 6th grades. The model will be replicated in different cohorts to observe if the outcomes are stable. Bruel and Bartholo (2012) analysed the same transition, considering pupils in one specific educational authority. The authors state that pupils' transfers among schools are not random and increase school segregation. The research design in the thesis shows a different outcome and proposes corrections for the previous conclusions. Two key questions will be answered: 1) What is the overall effect of the "mandatory" transfers on school segregation? 2) Is there any evidence suggesting that pupils are being selected based on key characteristics?

In order to make a causal claim about any impact of relocation on school segregation, two stable measures of GS or D are necessary, before and after the particular event. A cross-sectional design will measure segregation in two different cohorts (5th grade 2008 – 6th grade 2009; 5th grade 2009 – 6th grade2010) to observe if results are stable. If previous studies are correct, the nominal values of GS and D in the 6th grade will be higher compared with those in the 5th grade.

Two educational policies that intentionally track pupils by prior educational attainment will be assessed. The first tracks pupils at age 8, who have not successfully completed literacy – called "*realfabetização*" – after three years in fundamental education. Around three thousand pupils are sent to "*realfabetização*" every year. A second policy tracks pupils that have systematically failed school (retention), or returned to school after a few years of abandonment. Pupils enrolled in PEJA I or II usually have three or more years of age/grade distortion. The policy aims to help students redeem the years that have been "lost" by constant retention or abandonment. A cross-sectional design calculates GS and D, with and without pupils enrolled in "*realfabetização*" and PEJA I and II. Both policies intentionally select pupils based on prior educational attainment, and it is reasonable to assume that disadvantaged pupils will be overrepresented in these classes/schools.

Around 90% of all public municipal schools in Rio de Janeiro have two or more "school shifts". There is no specific regulation to orientate how pupils must be allocated across "shifts". It is reasonable to assume that parents can express their preference for the morning or afternoon shift; nonetheless, the school principal (or another member of the school staff) has total control of the allocation. More rigorously, it could be said that there is not exactly a "shift effect", but the effect of school-bureaucracy-level decisions on segregation, and that this effect can be depicted by analysing segregation coefficients that stem from shift intakes. The expression "shift effect" will be used in this sense.

The study measures, for the first time, the impact of the "school shift" allocation for the entire population. The design captures the "net effect" of the "shift" allocation, comparing two different measures of GS: a) each school as one unit; and b) each "school shift" as a specific unit. The same calculations are replicated for all educational authorities throughout different years to observe how stable the outcomes are. An additional analysis for the "school shift" allocation interprets the outcomes observed in the real distribution against a simulated random distribution of pupils across school shifts. The thesis uses simulations to observe the likelihood of the size effects measured.

Third Part of the Thesis

This last part tests the hypothesis about the school composition effect or school-mix effect. Three complementary models are presented. The first (Type 1) uses *Prova Rio* standardized tests to estimate the pupils' achievement using prior attainment (from the previous year) and other key variables as covariants (pupil's colour, parents' education, poverty, sex, age/grade distortion). The residues of the linear regression will be correlated with the proportion of disadvantaged pupils for each school. This is a simple model to measure the strength and direction of the association between the residues of the regression and the proportion of disadvantaged pupils. In order to corroborate the initial hypothesis, the correlation must be negative and constant among all variables. The same model will be replicated for different cohorts to test for robustness.

The second model (Type 2) aggregates the individual residuals from the same regressions at school level. The mean scores of the individual residuals are once again correlated with the proportion of disadvantaged pupils calculated for each school. In order to corroborate with the initial hypothesis, the correlation coefficients must be negative and the strength of the association higher compared with the previous model – pupil level analysis.

The third model (Type 3) uses school-level figures to replicate the exact same model. The mean scores of *Prova Rio* for one year are used to predict the scores for the next year in a cross-sectional design. Once again, the residuals of the regression will be correlated with the proportion of disadvantaged pupils. In order to corroborate with the initial hypothesis, the correlation coefficients must be negative and the strength of the association higher compared with the previous model – pupil level analysis.

The models are not a definitive test and do not "prove" the existence of school composition effect. They do not rule out one important alternative explanation: school effectiveness. Schools overrepresented with pupils with desirable characteristics could, on average, despite any claim of school-mix effect, be more effective than others. The opposite could also be true: schools overrepresented with disadvantaged pupils could, on average, be less efficient. Nonetheless, if the correlation coefficients are stable for all variables of the disadvantaged, across all models, it could be argued that there is evidence to support the basis for the school composition effect theory. In other words, the model can help to determine if the theory can or cannot be refuted.

***

Each subsequent chapter will present a detailed explanation for each specific design and further discuss its major limitations. Whenever possible, all the models will be replicated using different cohorts or simulations to test for robustness. Falsificationism guides all methodological approaches in the study, and the hypothesis must survive intense testing in order to prove its plausibility.

5- DATABASE AND ITS LIMITATIONS

Longitudinal studies with educational second hand data are something recent and, at least in the case of Brazil, not very common, mainly for two reasons. First, it is worth mentioning that there was not much data available before the second half of 1990 collected by official bodies and local or federal administrations. The second reason is simply the lack of interest or trained researchers to manipulate large databases. The main benefits and challenges of this type of data will be briefly discussed. All empirical analysis presented in the thesis used two independent sets of data: a) Pupil Profile – longitudinal with information on all pupils, including family profile and school placement; b).*Prova Rio* – longitudinal assessment for all public schools.

Undoubtedly, one advantage of second hand data is the relative low cost comparing to research projects that generate primary data. Usually, datasets are likely to be larger in scope and scale and can be often analysed without the use of statistical generalisation. This is relevant because it makes the interpretation easier not only for the researchers, but also for the general public or policy makers that can make future use of the findings. Complex modelling or statistical tools can have the counter effect of making the findings less comprehensive to a more general audience. Another positive aspect of second hand data is the possibility of replication of the analysis using different tools or simply looking at distinctive aspects of the variables. Replication and validity are key elements in good science (Gorard, 2012a).

For some research projects, the use of pre-existing records is inevitable. It's the case of the thesis, which analyses school intakes for eight consecutive years. Many longitudinal studies use second hand data as the only available source – since a long term survey would simply be too expensive. The range, quality and availability of second hand data will most definitely increase, not only in Brazil, but in many other countries. The potential for new

research designs is also growing, with the real possibility of merging distinctive datasets. Educational researchers must take advantage of this new opportunity that can create new sub-disciplines (for example, as it happened to school effectiveness), with relevant impact for policy-making (Gorard, 2012a).

The benefits must not mask the limitations of datasets collected by others for different purposes. Yorke (2011), highlights four problems faced by those who are willing to use such type of data: 1) quality; 2) categorization; 3) limitations (necessary and optional); 4) reformatting. The study briefly discusses two of those challenges: quality and reformatting, which were particularly relevant for the database used.

The Rio de Janeiro Educational Department created in 2003 different datasets for administrative purposes. Originally, the sets of data contained different types of information: 1) pupils and family characteristic; 2) pupils' educational transitions; 3) schools' characteristics; 4) pupils' attainment. The key variables to link all sets of data were "schools' ID number" and "pupils' ID number". The second key variable, Pupil ID, had a major problem, since many times the same pupil appeared with a different ID number. This issue had a big effect in the attempt to follow pupils across different educational transitions. The problem was solved by a matching process using other key variables, such as pupils' date of birth, complete name of the pupils and mothers' name that allowed the routine to detect duplications in the "pupils' ID number".

The next step was to transform pupil level information into school-level. This was accomplished with the split of the datasets in different school years (2004-2012). The pupils' last school in each year[9] was selected in order to construct the school level datasets. For example, if pupil "i" changed school three times in one specific school year, the last school in that year was selected.

[9] The cut-off point was 30th November of each school year.

Another common problem with large databases is missing data. Since it is very difficult to know if missing data is random, there is always the chance of biased interpretation (Vaus, 2011). The longitudinal database utilized in this paper presents better records in the last five years (2007-2011), for variables such as pupil's colour and parents' education. The proportion of disadvantaged pupils and missing data for each variable are presented in the next table. It is possible to observe that missing data declines heavily along the years.

Table 5: Proportion of Missing Data and Disadvantaged Pupil for Pupil's Colour and Parents' Education.

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
Proportion of Missing Pupil's Colour	0.24	0.13	0.06	0.05	0.04	0.04	0.04
Proportion Black Pupils	0.10	0.11	0.12	0.12	0.12	0.12	0.11
Proportion Non White	0.49	0.56	0.61	0.61	0.61	0.61	0.61
Proportion of Missing Parents' Education	0.20	0.18	0.16	0.13	0.12	0.11	0.11
Proportion EducFS	0.18	0.20	0.21	0.22	0.23	0.24	0.24
Proportion EducHS	0.56	0.57	0.59	0.60	0.610	0.61	0.60
Proportion Distortion1	0.57	0.59	0.58	0.59	0.57	0.57	0.57
Proportion Distortion2	0.30	0.31	0.30	0.28	0.26	0.26	0.27

Initial results from Table 5 suggest that: 1) the quality of data is better in the more recent years (2007-2010); 2) the missing data is probably not randomly distributed, with a higher proportion of disadvantaged pupils among the "missing group". There is empirical evidence that, when poverty indicator rise (for example, in economic crises) the segregation levels tend to decline, as an "equality of poverty" effect. The inverse picture is also true. In an economic boom, with the decline of the proportion of poor families, the index tends to rise (Gorard; Taylor; Fitz, 2003). If that is correct, it is most likely that the missing data will

artificially inflate D and GS, leading to the wrong interpretation of a decline in segregation in the first four years (2004-2007) (Bartholo, 2013).

Nonetheless, a more detailed analysis of the dataset allows a deeper understanding of the missing data problem. Breaking down the data, not only by year, but also considering pupils' allocation in the first or second segment of Fundamental School, it is possible to observe that data is better, not only for the most recent years, but also for younger pupils (from 1st to 5th grades) – see Table 6. A plausible explanation can be related to the fact that pupils enrolled in the first segment have entered the public educational system in more recent years, when the protocol to collect data started to work more efficiently.

Table 6: Proportion of Missing Data for 1st and 2nd Segment.

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
1st Segment – pupil's colour	0.13	0.09	0.05	0.04	0.04	0.03	0.03
2nd Segment – pupil's colour	0.40	0.19	0.09	0.06	0.05	0.05	0.05
1st Segment – Parental Education	0.12	0.10	0.09	0.08	0.08	0.08	0.08
2nd Segment – Parental Education	0.29	0.23	0.18	0.14	0.11	0.10	0.10

If it is true that missing data are not randomly distributed and artificially inflate D and GS, then it would be expected that the index values calculated for the first and second segments separately would present different patterns. Tables 7 and 8 show the GS trends pupils' colour and parents' education, for the first and the second segment, respectively.

Table 7: Segregation Index (%) for the First Segment.

	2004	2005	2006	2007	2008	2009	2010
GS Black Pupil	18	18	18	18	17.5	17.5	17
GS Non White Pupil	7.5	7.5	7.5	7.5	7	7	6.5
GS EducFS	27	26.5	25.5	24	23	22	20.5
GS EducHS	10.5	10.5	10.5	10	9.5	10	9.5

Table 8: Segregation Index (%) for the Second Segment.

	2004	2005	2006	2007	2008	2009	2010
GS Black Pupil	20.5	19.5	17	15.5	15	14.5	14.5
GS Non White Pupil	13.5	11.5	9	7.5	6.5	6	6
GS EducFS	34.5	30.5	27	24.5	23	21.5	20.5
GS EducHS	18.5	15	12.5	10.5	9.5	9	8.5

The figures corroborate the initial hypothesis regarding the influence of missing data on GS, and reinforce the idea that the missing data is not randomly distributed. This is an important finding that should be taken into consideration for future interpretation in the thesis. Virtually all longitudinal studies that use secondary data face similar problems, and the real question is not if there will be missing data (as this is most likely), but how researchers approach the problem and take these issues into consideration in their interpretations (York, 2011). In addition, it raises the question: Who is our missing data? Gorard (2012) conducted an analysis with pupils who did not have information for Free School Meals (proxy of poverty) in English state schools. The findings showed that the "missing group" presented lowest attainment when compared to pupils eligible and not eligible to free school meals. They were termed "super deprived". It is our intention to conduct a similar analysis in the next section to observe more closely our missing group.

In order to try to cope with the missing data, two different strategies were used to recalculate GS, depending on the variable – see Table 9. For pupil's colour, GS was

calculated considering: a) GS Black Pupil 1/ GS Non-White 1 – total number of pupils in school "i" equal to the total number of pupils indicated in the database – including missing data cases/ pupils; b) GS Black Pupil 2/ GS Non-White 2 – total number of pupils in school "i" is equal to the sum of all pupils that are not missing from the variable. Example: total number of pupils = sum of all white and all non-white pupils.

For parents' education, two different strategies were used to calculate GS: a) GS EducFS 1 – total number of pupils in School "i" is equal to the total number of pupils indicated in the database – including missing data cases/ pupils; b) GS EducFS 2 – total number of pupils in school "i" is equal to the sum of all pupils that are not missing from the variable. Example: total number of pupils = sum of all pupils whose parents finished fundamental school and all pupils whose parents did not finish fundamental school.

Table 9: Segregation Index (%) for All Indicators of Potentially Disadvantaged Pupils.

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
GS Non White 1	11.5	9	7.5	7	6.5	6.5	6.5
GS Non White 2	5.5	5.5	5.5	5.5	5.5	5.5	5.5
GS Black Pupil 1	18.5	17.5	17	16	15.5	15.5	15
GS Black Pupil 2	16	16.5	16	16	15.5	15.5	15.5
GS EducFS 1	30	28	26	24,5	23	21.5	20.5
GS EducFS 2	27	26	24,5	23,5	22,5	21	20

Calculations suggest that is: 1) missing data is most probably not randomly distributed, with a higher concentration in the potentially disadvantaged pupils; 2) missing data seems to artificially inflate the segregation index, giving the wrong impression that the segregation levels are declining for all indicators in the first 4 years (2004-2007). None of the different strategies to calculate the segregation indices solves the problem of missing data. In reality, there is no definitive solution since the information is simply not available.

The variable "pupils living in poverty" shows a distinctive problem when compared to the other two. Here the issue could be related to sub-notification. As the quality of data improves the proportion of families eligible to participate in the Federal government program increases (2004-2008) – see Table 10. In the last three years, 2008-2010, the curve has changes.

Table 10: Proportion of Pupils Living in Poverty.

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
NIS_Sum	0.19	0.22	0.24	0.26	0.26	0.25	0.23
NIS_Max	0.25	0.27	0.31	0.33	0.33	0.33	0.32

There are two possible explanations for this outcome. The first one is related to a potential risk of sub-notification for this particular information (*Número de Identificação Social*). Let's not forget that the quality of information in the dataset for the first three years (2006-2004) is worst. A second plausible explanation is the growth of the total number of families assisted by the program.

Since it started in 2003, *Bolsa Família* program has grown from 6.5 million families assisted to 14.1 million in 2013, considering the entire country. The government estimates that another half a million families will be included in 2014. These figures indicate that around 25% of the entire population is assisted by the program. It is important to highlight that the increase in the total number of beneficiaries was accompanied by a constant economic growth and historical low rates of unemployment. Data from 2013 indicates that *Bolsa Família* assisted 242.926 families in the municipality of Rio de Janeiro and the dataset used in this research indicates for 2010 a total of: a) 173.555 NIS_Sum; b) 232.636 NIS_Max.

In this variable, the information is recorded in two columns: one indicates the parent's NIS (*Número de Identificação Social*) and the second the pupil's NIS. If the database had no problems regarding the records, both parents and pupils should have the numbers, but this is not true. For this reason, GS was calculated in two ways: 1) NIS_SUM – pupil's family was considered poor only when both the pupil and his parents had NIS; 2) NIS_MAX – pupil's family was considered poor when his family and/or the pupil had NIS. This second measure was an attempt to cope with a potential sub-notification problem.

***

Prova Rio is a standardized test applied annually since 2009 by the Rio de Janeiro Educational Department. The assessment was design and implemented by a private enterprise using Item Response Theory (IRT) and aimed to measure school performance. This particular dataset has a shorter trend (only four years 2009-2012) and the pupil's individual scores can be merged with other datasets. For the first three years, the evaluation assessed proficiency in mathematics and Portuguese for pupils in four different grades – 3rd, 4th, 7th and 8th). In 2012, science was included as another independent measure.

Considering its compulsory character, the fact its results are measured on the same scale as SAEB and *Prova Brasil*, and the series to which it is applied, *Prova Rio* can be regarded as complementary to *Prova Brasil*, generating the possibility of longitudinal analyses, as a cohort can be followed throughout at least three years.

Just as around the world, the information derived from these assessments, almost always the average results of the tests, receive much publicity in the press, above all, when the overall results observed are poor. If there is consensus that the Brazilian educational indicators are hardly bearable (Veloso et al. 2009), incompatible with the country's level of economic development, when it comes to the issue of the relevance of the assessment systems, there is wholesale controversy.

There are two main limitations for this particular dataset. The first one is related to missing data for a key variable: Pupil ID. Without this information it was very hard to merge any useful data. Less than 1% of all cases were excluded from the analysis. The second issue is related to the quality of the measurements. During the course of this study, researchers from *Observatório da Educação* have been in direct contact with staff members from Rio de Janeiro Educational Department in order to obtain more information about the procedures taken in *Prova Rio* and discuss some preliminary results. The interaction allowed a revision in the individual raw scores for two specific years (2009-2010).

All evidence presented with pupil level analysis should be interpreted with great caution. It is not clear at this point, if the total number of questions in each individual test covers the entire scale of IRT. If not, the attempt to estimate individual scores might be misleading.

***

Highlighting the limitations of data is crucial for good science. Arising largely from the medical experimental approaches, the group on Consolidated Standards of Reporting Trials (CONSORT)[10] has suggested a minimum of information that should be made publicly available when reporting results. Unfortunately, it is common in educational research that very basic information such as number of cases, response rate, drop outs, missing data or imputations methods are simply not reported even from high profile peer-reviewed reports (Tooley; Darby, 1998). The initial analysis regarding the missing data suggests that all longitudinal approaches using the datasets provided the municipal educational department of Rio should consider this limitation. Ignoring the problem is the worst "solution".

[10] See http://www.consort-statement.org

5.1 Who Is Our Missing Data? Testing the "Super Deprived" Hypothesis

This chapter further investigates pupils that have one or more missing variable. Initial description have showed that: 1) quality of data is better in the more recent years (2007-2010); 2) quality of data is better for pupils enrolled in first segment (younger pupils); 3) missing data is most probably not randomly distributed, with a higher concentration on potentially disadvantaged pupils; 4) missing data seems to artificially inflate the segregation index, giving the wrong impression that the segregation levels are declining for all indicators in the first 4 years (2004-2007).

Instead of using imputation methods to handle missing data, the thesis presents a simpler, and perhaps, more useful approach. The chapter tries to answer two questions: are missing cases clustered in specific educational authorities? Are missing cases randomly distributed considering disadvantaged and non-disadvantaged pupils? Missing cases will be described in detail and a subsequent analysis will compare the outcomes in *Prova Rio* of pupils who are missing with those who are not. Three different groups are compared: potentially disadvantaged; non-potentially disadvantaged; missing group. The final model presents a T-test, comparing the scores in *Prova Rio* 2010 from missing group with those not missing, in order to observe if the distributions in both groups are significantly different from each other. The model will be replicated for different cohorts (total of four) in two proficient tests (mathematics and Portuguese), for a total of 8 tests, to check for robustness.

The analysis focuses on two variables: pupils' colour and parents' education. The other two variables used to assess segregation will not be considered. Age/grade distortion is the only variable that presents perfect records for all years. The information recorded in the dataset for poverty is somewhat different. In this variable, the information (NIS) is either filled or left blank (missing). In this case, the danger is sub-notification, since the "missing" (blank space) is interpreted as "no" for eligibility for social programs.

An initial description of the missing cases across Educational Authorities shows a peculiar picture. The expected result was that more urban Educational Authorities (closer to downtown), such as CRE 1, 2 and 5, would present better records. Table 11 shows that reality is somewhat different for pupils enrolled in the first segment of fundamental school.

Table 11: Proportion of Missing data for Pupil's colour and Parents' Education for each CRE– First Segment.

	2009		**2010**	
	Pupil's colour	Educ Parents	Pupil's colour	Educ Parents
CRE 1	0.04	0.03	0.04	0.03
CRE 2	0.03	0.11	0.04	0.12
CRE 3	0.03	0.07	0.03	0.08
CRE 4	0.04	0.13	0.04	0.13
CRE 5	0.03	0.09	0.03	0.09
CRE 6	0.04	0.04	0.04	0.04
CRE 7	0.03	0.04	0.03	0.06
CRE 8	0.04	0.07	0.03	0.07
CRE 9	0.02	0.04	0.02	0.04
CRE 10	0.03	0.11	0.03	0.11

Missing data for pupils' colour is fairly stable across all educational authorities. However parents' education shows a larger variance (13% to 4%) depending on the CRE observed. Table 12 presents the figures for second segment pupils and shows an even higher variance, but this time for both variables.

Table 12: Proportion of Missing data for Pupil's colour and Parents' Education for each CRE– Second Segment.

	2009		2010	
	Pupil's colour	Educ Parents	Pupil's colour	Educ Parents
CRE 1	0.10	0.04	0.09	0.04
CRE 2	0.05	0.15	0.05	0.09
CRE 3	0.04	0.12	0.04	0.11
CRE 4	0.07	0.17	0.07	0.17
CRE 5	0.03	0.10	0.04	0.09
CRE 6	0.05	0.04	0.05	0.04
CRE 7	0.04	0.07	0.04	0.07
CRE 8	0.04	0.10	0.05	0.09
CRE 9	0.05	0.04	0.02	0.04
CRE 10	0.03	0.13	0.04	0.12

CRE 9 and 6 present the best records for both segments. This is quite unexpected outcome, since those are less urban and poorer areas in the city. Those are simple descriptive statistics that should be discussed with the member of staff that produces the datasets. The data shows an urgent demand to regulate the protocols for data collection. The fact that the same network presents such high variance, comparing different educational authority, for the proportion of missing should be a concern for policy-makers and educational researchers.

The next analysis compares the total proportion of disadvantaged pupils (considering all pupils enrolled in the public network), with the proportion of disadvantaged pupils considering only pupils missing for one of the two key variables in this chapter. The claim here is that if missing data is not randomly distributed, those missing for pupil's colour or parental education would be more likely to be disadvantaged in other variables. Is that true?

Table 13 is the key table to proceed with the comparison. It presents the proportion of potentially disadvantaged pupils considering all grades for seven years.

Table 13: Proportion of Disadvantaged Pupils for all Variables 2004-2010.

	2004	2005	2006	2007	2008	2009	2010
Poverty NIS Sum	0.18	0.20	0.23	0.26	0.27	0.27	0.26
Proportion Black	0.10	0.11	0.12	0.12	0.12	0.12	0.11
Proportion Non White	0.49	0.56	0.61	0.61	0.61	0.61	0.61
Proportion EducFS	0.18	0.20	0.21	0.22	0.23	0.24	0.24
Proportion EducHS	0.56	0.57	0.59	0.60	0.61	0.61	0.60
Proportion Distortion 1	0.57	0.59	0.58	0.59	0.57	0.57	0.57
Proportion Distortion 2	0.30	0.31	0.30	0.28	0.26	0.26	0.27
Proportion Special Class	0.13	0.13	0.13	0.09	0.09	0.11	0.12

Figures of Table 13 should be compared with Table 14 and 15. The first one presents the proportion of disadvantaged pupils for all variables, considering only cases missing for pupil's colour.

Table 14: Proportion of Disadvantaged Pupils Considering Pupils Missing for the Variable Colour.

	2004	2005	2006	2007	2008	2009	2010
Poverty NIS Sum	0.05	0.09	0.20	0.27	0.30	0.31	0.29
Proportion EducFS	0.13	0.15	0.17	0.19	0.20	0.21	0.19
Proportion EducHS	0.44	0.46	0.53	0.57	0.59	0.58	0.54
Proportion Distortion1	0.68	0.71	0.66	0.66	0.63	0.64	0.63
Proportion Distortion2	0.45	0.48	0.40	0.35	0.32	0.34	0.35
Proportion Special Class	0.20	0.21	0.14	0.09	0.08	0.12	0.16

The figures of Table 14 partially confirm the initial hypothesis. Missing cases for pupils' colour are more likely to be disadvantaged for all other variables. The proportions of pupils with age/grade distortion are consistently higher for all 14 measures across the years. A similar interpretation can be made for poverty: pupils missing colour are more likely to be

poor. In the last four years (2007-2010), the proportions of poor pupils are higher (Table 14) comparing with 13. But what about the first three years? The pattern for those three years is different, showing that missing cases for pupil's colour are less likely to be poor. The previous chapter highlighted how missing was distributed across the years and how it could have an impact on the interpretations. The concern about sub-notification for poverty[11] gains density with Table 14. Apparently, in 2004-05-06, the records for poverty were less reliable, with some poor families being overlooked. If those assumptions are correct, variables, distortion and poverty, corroborate the initial hypothesis.

It should also be considered the intense growth in *Bolsa Família* beneficiaries since 2006. A plausible explanation for the changes in the figures for poverty could be an increase in the total number of families assisted by the federal government.[12]

Nonetheless the variable parents' education does not allow the same interpretations. Missing cases for pupils' colour are less likely to be disadvantaged for parents' education. The figures reject the idea that potentially disadvantaged pupils are overrepresented in the missing group (non-random missing). A possible alternative explanation for the outcome observed is that the total proportion of missing for parents' education increases a lot considering only cases missing for pupil's colour. This means that pupils missing for colour are more likely (on average, 75% increase) to be missing for parents' education. The finding suggests once again that missing data is not randomly distributed and apparently conjugates geographical elements (some educational authorities present better records), but also individual – some families are more likely to have their information missing.

Table 15 presents the proportion of disadvantage pupils considering only pupils missing for parents' education. All 14 measures for age/grade distortion consistently indicate

[11] See previous chapter "Database and its limitations".
[12] See http://veja.abril.com.br/noticia/brasil/numero-de-beneficiarios-do-bolsa-familia-so-cresce

that missing cases for parents' education are more likely to be disadvantaged regarding prior attainment (Distortion).

Table 15: Proportion of Disadvantaged Pupils Considering Pupils Missing for Parents' Education 2004-2010.

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
Poverty NIS Sum	0.10	0.12	0.15	0.19	0.20	0.20	0.19
Proportion Black	0.08	0.11	0.13	0.13	0.13	0.12	0.12
Proportion Non-White	0.36	0.49	0.59	0.60	0.60	0.60	0.59
Proportion Distortion 1	0.69	0.71	0.72	0.73	0.70	0.69	0.67
Proportion Distortion 2	0.47	0.50	0.51	0.50	0.48	0.46	0.45
Proportion Special Class	0.25	0.29	0.32	0.31	0.32	0.32	0.31

However, the figures for pupils' colour and poverty do not corroborate the initial hypothesis. Apparently poor and non-white pupils are not overrepresented when observing only pupils missing for parental education. If this is true, the proficiency in Prova Rio for this particular variable considering both groups (missing and non-missing groups) should be similar. The next model tests these assumptions.

***

Prova Rio is a standardized test applied annually since 2009 by the Rio de Janeiro Educational Department. For the first three years (2009-10-11), the evaluation assessed proficiency in mathematics and Portuguese for pupils in four different grades – 3rd, 4th, 7th and 8th). The model presents a T-test to compare the distributions of the scores in *Prova Rio* 2010 from the missing group (for pupil's colour and parents' education) with those not missing. The initial hypothesis states that missing group will present significant lower scores for all grades in both measures (math and Portuguese). If the outcomes confirm the initial

hypothesis, these can be seen as one more indication that potentially disadvantaged pupils are overrepresented in the missing cases.

This approach is quite different from methods of imputation. Here the main goal is to gather more information about missing cases. The descriptive analysis can be helpful to guide future policy aiming not only the potentially disadvantaged group, but also the "missing cases", which are routinely ignored in most of the studies.

The dataset from *Prova Rio* 2010 has a total of 243.892 cases in all four grades. The first challenge was complete the merge between *Prova Rio* 2010 and the dataset with pupil/family characteristic. The records for Pupil ID (key variable) in *Prova Rio* were not perfect. A total of 2.917 cases were missing for Pupil ID (1%) and another 4407 could not be found in the merge process, adding to a total of 7.324 (3%) of cases excluded from the analysis. Table 16 shows a description of pupils who were absent and those who completed the evaluation.

Table 16: Proportion of Disadvantaged Pupils in *Prova Rio* 2010:

	PROVA RIO 2010	
	ABSENT	PRESENT
Proportion NIS Sum	0.27	0.32
Proportion Black Pupils	0.12	0.11
Proportion Non-White	0.65	0.62
Proportion EducFS	0.30	0.26
Proportion EducHS	0.72	0.67

Data is not conclusive to suggest that both groups (absente and present) are equivalent. The figures for pupils' colour and parents' education suggest that these pupils are

overrepresented in the "absent group". However, the figures for poverty tell a different story. It is not clear if both groups can be considered equivalent.

The next model compares the means scores for mathematics and Portuguese for four different cohorts. The outcomes for all cohorts were very similar. For this reason, the study will only present a more detailed analysis for the cohort 3rd grade, 2010.

The outcomes for pupils' colour present a total of six different groups: a) white pupils; b) non-white; c) black pupils; d) non-black pupils; e) cases missing for pupil's colour; f) mean scores. Table 17 presents the figures for 3rd grade pupils, year 2010.

Table 17: Mean scores for *Prova Rio* 2010 – 3rd Grade –Missing cases for Pupil's Colour.

	***PROVA RIO* 2010**	
	Mathematics	Portuguese
White Pupils	170	164
Non-White Pupils	157	152
Black Pupils	151	147
Non-Black Pupils	163	158
Missing cases	154	148
Average score	162	156

The mean scores for mathematic and Portuguese confirms that potentially disadvantaged pupils present lower attainment levels comparing to non-disadvantaged pupils. White pupils present the highest mean for both measures and black pupils the lowest. Nonetheless, the focus here is the missing group. The figures suggest that pupils missing for colour score below average and present lower attainment than non-whites (one measure of disadvantaged) end a bit higher than black pupils. But, how big are the differences observed?

The next model presents a T-test comparing the scores in *Prova Rio* (Mathematics and Portuguese) from missing group with those not missing. Are the mean scores in both groups significantly different from each other? As mentioned before, there are two "types" of missing and the study tested for both. The first "type" are pupils with Pupil_ID that do not have the information for pupil's colour; the second "type" are pupils in *Prova Rio* 2010 dataset without Pupil_ID. The model will test only for missing "type one" and for both "types" of missing together. Table 18 presents the figures for the T-Test, using Portuguese – "missing type one"; Table 19 for all missing cases.

Table 18: Independent Samples Test Missing for Pupil Colour – Portuguese, Cohort 3rd Grade, 2010 (Missing Type 1).

	Levene's Test Equal Variance		**t-test for Equality of Means**		
	F	Sig.	Sig. (2-tailed)	Lower	Upper
Equal Variance Assumed	9,533	0.002	0.000	-11,681	-5,618
Equal Variance Not Assumed			0.000	-11,527	-5,765

- 99% Confidence Interval of the Difference

Table 19: Independent Samples Test Missing for Pupil Colour – Portuguese, Cohort 3rd Grade, 2010 (Missing Type 1and 2).

	Levene's Test Equal Variance		**t-test for Equality of Means**		
	F	Sig.	Sig. (2-tailed)	Lower	Upper
Equal Variance Assumed	9,297	0.002	0.000	-6,509	-3,180
Equal Variance Not Assumed			0.000	-6,449	-3,329

99% Confidence Interval of the Difference

The figures on the independent samples tests show that the distributions of the mean scores for both groups are statistically different (not equivalent), with a confidence interval of 99%. These are important findings for two main reasons: 1- corroborates the initial hypothesis that missing data is not randomly distributed; 2- reinforce the need for better data collection. Missing cases are often left out of analysis and this could lead to biased interpretations.

Will the T-test for the mathematics scores show a similar outcome? The answer is yes. Table 20 show the outcome for all missing (type 1 and 2), the more conservative result. Once again, the distribution of the mean scores for both groups are not equivalent, highlighting that pupils missing for pupils' colour present lower attainment levels – confidence interval of 99%.

Table 20: Independent Samples Test Missing for Pupil Colour – Mathematics, Cohort 3rd Grade, 2010 (Missing Type 1 and 2).

	Levene's Test Equal Variance		**t-test for Equality of Means**		
	F	Sig.	Sig. (2-tailed)	Lower	Upper
Equal Variance Assumed	0,043	0.835	0.000	-6,814	-3,324
Equal Variance Not Assumed			0.000	-6,807	-3,331

99% Confidence Interval of the Difference

The next step is to observe if the missing group for parents' education present a similar outcome. Table 21 presents the scores for 3rd grade pupils in mathematics and Portuguese, considering six different groups: a) Parents who have finished high school – EducHS2; b) Parents who did not finish high school – EducHS; c) Parents who did not finish fundamental school – EducFS; d) Parents who have finished fundamental school – EducFS2; e) cases missing for parents' education; f) average scores.

Table 21: Mean scores for *Prova Rio* 2010 – 3rd Grade – Missing Cases for Parents' Education.

	***PROVA RIO* 2010**	
	Mathematics	Portuguese
EducHS2	176	171
EducHS	155	150
EducFS	151	147
EducFS2	167	161
Missing cases	158	153
Average score	162	156

The figures confirm that potentially disadvantaged pupils (EducFS and EducHS) present lower attainment levels compared to non-disadvantaged pupils. It is interesting to notice, that black pupils (Table 17) present the exact same mean scores for both measures compared to parents' who did not finish fundamental school (EducFS). Missing cases for parent's education score below average, but higher than potentially disadvantaged pupils (EducFS or EducHS). Is that difference significant?

A T-test comparing the scores (Math and Portuguese) of the missing cases for parents' education, with those not missing was conducted. Figures of the independent samples tests for Portuguese indicate that both groups are not equivalent, with a confidence interval of 99% – see Table 22.

Table 22: Independent Samples Test Missing for Parents' Education – Portuguese, Cohort 3rd Grade, 2010 (Missing Type 1 and 2).

	Levene's Test Equal Variance		**t-test for Equality of Means**		
	F	Sig.	Sig. (2-tailed)	Lower	Upper
Equal Variance Assumed	11,247	0.001	0.000	-4,067	-1,618
Equal Variance Not Assumed			0.000	-4,038	-1,647

99% Confidence Interval of the Difference

The exact same outcome was observed for the Mathematics scores – see Table 23. Missing group scores systematically lower and both groups are not equivalent. All the tests conducted with 3rd grade, 2010 pupils have shown that potentially disadvantaged pupils are very likely to be overrepresented in the missing group. This is an important finding for the future analysis of the thesis, since missing data can have an impact on the segregation indices.

Table 23: Independent Samples Test Missing for Parents' Education – Mathematics, Cohort 3rd Grade, 2010 (Missing Type 1 and 2).

	Levene's Test Equal Variance		**t-test for Equality of Means**		
	F	Sig.	Sig. (2-tailed)	Lower	Upper
Equal Variance Assumed	5,846	0.016	0.000	-4,974	-2,404
Equal Variance Not Assumed			0.000	-4,946	-2,433

99% Confidence Interval of the Difference

The easy access to statistical software should not mask the limitations of data. A common solution in social science is the use of imputation methods to deal with missing cases. There is great danger of propagation error on such approaches that could lead to wrong interpretations. The approach here is simpler and aims to provide useful information for

policy-makers and educational researchers when such data. The analyses presented suggest the need for better data collection in specific educational authorities and also a prudent use of the datasets available. Evidence indicate that missing data is not randomly distributed.

What now? The next challenge was to replicate the same model for all grades – 4th, 7th and 8th. Tables 24, 25 and 26 present the mean scores for mathematics and Portuguese in *Prova Rio* 2010. The replication of the T-test confirmed, once again, that the missing group for both variables of disadvantaged present lower attainment levels compared to non-missing. Both groups are not equivalent (confidence interval of 95%). The results corroborate all the findings highlighted in the 3rd grade, 2010 and reinforce previous interpretation.

Table 24: Mean scores for *Prova Rio* 2010 – 4th Grade –Missing Cases for Pupil's Colour and Parents' Education.

	***PROVA RIO* 2010**	
	Mathematics	Portuguese
White Pupils	197	193
Non-White Pupils	185	181
Black Pupils	180	175
Non-Black Pupils	191	187
Missing cases (Colour)	183	178
Average score (Colour)	189	186
EducHS2	201	198
EducHS	184	179
EducFS	181	176
EducFS2	193	190
Missing cases (Parents' Educ)	186	182
Average score (Parents' Educ)	189	186

Table 25: Mean scores for *Prova Rio* 2010 – 7th Grade – Missing Cases for Pupil's Colour and Parents' Education.

	***PROVA RIO* 2010**	
	Mathematics	Portuguese
White Pupils	224	219
Non-White Pupils	214	208
Black Pupils	209	203
Non-Black Pupils	219	213
Missing cases (Colour)	214	206
Average score (Colour)	218	212
EducHS2	227	223
EducHS	213	207
EducFS	211	204
EducFS2	220	215
Missing cases (Parents' Educ)	216	210
Average score (Parents' Educ)	218	212

Table 26: Mean scores for *Prova Rio* 2010 – 8th Grade – Missing Cases for Pupil's Colour and Parents' Education.

	***PROVA RIO* 2010**	
	Mathematics[1]	Portuguese
White Pupils	200	226
Non-White Pupils	193	216
Black Pupils	190	213
Non-Black Pupils	196	221
Missing cases (Colour)	192	216
Average score (Colour)	196	220
EducHS2	202	230
EducHS	192	215
EducFS	189	212
EducFS2	197	222
Missing cases (Parents' Educ)	194	220
Average score (Parents' Educ)	196	221

[1] The mathematic scores suggest that 8th graders, 2010 present worst records than 7th graders. This is one example of odd outcome for *Prova Rio* raw scores.

6- DESCRIBING SEGREGATION TRENDS OVER TIME

Describing trends of school segregation over time can be useful for policy-makers and educational researchers interested in the theme of educational opportunities. The longitudinal approach can help to observe if segregation is increasing or declining over the years, and identify variables that appear to be associated with these changes. This chapter presents three different descriptive analyses: a) D and GS for all pupils enrolled in the first and second segments – 1st to 9th grades; b) D and GS calculated for each segment separately; c) GS for each Educational Authority separated by segment. A complementary analysis investigates if between-school segregation in Rio de Janeiro is one single process, or it should be considered more than one causal model to explain the phenomenon.

The levels of school segregation considering all available indicators for potentially disadvantaged pupils are presented in Table 27. The thesis uses two similar indexes – the Segregation Index (GS) and the Dissimilarity Index (D) – to describe segregation patterns over time. There is empirical evidence that D and GS will give a similar answer for longitudinal approaches (Gorard, 2012; Bartholo, 2013). Nonetheless, the study show both in order to clarify any doubts about the efficacy of the indexes. We must remember that the "battle of the indexes" is still present.

Table 27: Segregation Index (%) and Dissimilarity Index (%) for All Available Indicators.

	2004	2005	2006	2007	2008	2009	2010
GS Black Pupil	17	16	15.5	14.5	14	13.5	13.5
D Black Pupil	17.5	17	16.5	16.5	16	15.5	15.5
GS Non-White Pupil	10.5	8.5	7.5	7	6.5	6	6
D Non-White Pupil	16.5	16.5	16.5	16	15.5	15.5	15
GS EducFS	29	27.5	25	24	22	20.5	19.5
D EducFS	34.5	33.5	32	31	29.5	28	26.5
GS EducHS	14	12.5	11	10	9.5	9	8.5
D EducHS	26.5	26	26	26	25.5	25	24.5
GS NIS_Sum	24	20	17.5	16.5	16.5	16.5	17.5
D NIS_Sum	29	25	22.5	22	22.5	23	23
GS Distortion 1	8.5	8.5	8.5	8.5	8	8	9
D Distortion 1	20.5	20.5	20	20	19.5	19	20.5
GS Distortion 2	21	21	21.5	24	25	24	23
D Distortion 2	29.5	31	30.5	33.5	34	32.5	31.5

This is the first time that school segregation has been measured for the entire pupil population in a major city in Brazil. GS and D present very similar trends and suggest that missing data, more prevalent in the first two years, have a greater effect on GS. The explanation could be an important detail in the formulas of both indexes. GS considers the total pupil population, including the missing cases. D, on the other hand, disregards the missing cases, since it considerers only the disadvantaged and non-disadvantaged pupils. Table 28 shows the correlations of GS and D for all pupil characteristics.

Table 28: Correlation between GS and D for All Pupil Characteristics.

Indices GS and D	Black	Non-White	EducFS	EducHS	NIS	Distortion1	Distortion2
Correlation	0,97**	0,80*	0,99**	0,86*	0,98**	0,84*	0,94**

- ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

The coefficients are quite high for all indicators. This can be seen as one more indication that both indexes tell a similar story and will provide similar figures for the analyses (Gorard, 2009; Bartholo, 2013). What really changes is the interpretation. GS has a more straightforward interpretation, indicates the exact proportion of disadvantaged pupils who would have to move school for there to be no segregation. D represents the proportion of one group or the other that would have to move, if there were no segregation.

Another way of looking at the data presented in Table 27 is to show just the percentage difference when comparing different time periods. Table 29 shows if GS and D, after a specific period of time, has declined or increased. Because of the apparent influence of the missing data on the trends, two different periods were calculated: a) 2004-2010; b) 2007-2010. The two periods were chosen based on the proportion of missing data previously presented.

Table 29: Segregation Index and Dissimilarity Index – Percentage Difference (2004-2010 and 2007-2010) for Pupils' Colour and Parents' Education.

	2004-2010 % Difference	2007-2010 % Difference
GS Non-White	-43	**- 14**
D Non-White	-9	-6
GS Black Pupil	-21	**- 7**
D Black Pupil	-11	-6
GS EducFS	-33	**- 19**
D EducFS	-23	-15
GS EducHS	-39	**-15**
D EducHS	-8	**-6**

Both indexes give a similar answer. Apparently, between-school segregation is declining for parents' education and pupils' colour in the different periods analyzed (2004-2010; 2007-2010). The reasons for the decline (especially in parents' education) are unknown

and will be explored in further sections. There were no major changes in the educational legislation that could explain the constant decline in segregation for parents' education. The effect of missing data on both indexes could be an alternative explanation. However, the fact that the decline is constant, even when missing data stabilizes (2008-2010), weakens this hypothesis.

Segregation for poverty suggests a distinct pattern, with a decline in the first three years (2004-2006) and, subsequently, a gradual increase as of 2007. One possible explanation for the decline in the overall segregation levels in the first three years, observed in all variables, is the change in the proportion of missing data. As Table 5 shows, the missing data have declined over the years, and the proportion of disadvantaged pupils tends to grow – the only exception is poverty, which shows a proportional decline in 2009 and 2010. There is empirical evidence that, when the poverty indicator rise (for example, in economic crises), the segregation levels tend to decline, as an "equality of poverty" effect. The inverse picture is also true. In an economic boom, with the decline in of the proportion of poor families, the index tends to rise (Gorard; Taylor; Fitz, 2003).

***

A cleverer way of analyzing school segregation patterns for Rio de Janeiro public municipal school is to separate schools that offer the first segment from those of the second. There are two main reasons for this approach: 1) some schools – around 10% – cater for both segments (1st to 9th grades), which can have an effect on the measurement of segregation; 2) the total number of schools in the second segment is a lot smaller compared to first segment – 45% decrease. Such dramatic change in the total number of schools can have an impact on school segregation. Table 30 presents the figures of GS and D for pupils enrolled in the first segment of fundamental school.

Table 30: Segregation Index (%) and Dissimilarity Index (%) for All Available Indicators Considering the First Segment.

	2004	2005	2006	2007	2008	2009	2010
GS Black Pupil	16.5	16.5	16.5	16	15.5	15.5	15
D Black Pupil	18	18	18	18	17.5	17	17
GS Non White Pupil	7	7	7	7	6.5	6.5	6.5
D Non-White Pupil	16	16.5	17	17	16.5	16.5	16
GS EducFS	26	25.5	24.5	23.5	22	21	20
D EducFS	33	33	32.5	31.5	30.5	29	27.5
GS EducHS	10.5	10	10	9.5	9.5	9.5	9.5
D EducHS	26.5	26.5	27	27.5	27	26	25.5
GS NIS	16	15.5	15	15	15.5	16.5	17
D NIS	21.5	21	21	21.5	22	23.5	23.5
GS Distortion 1	6.5	6.5	6.5	6.5	6	5.5	6.5
D Distortion 1	11.5	12	11.5	13	12	10.5	11.5
GS Distortion 2	22	20.5	21.5	22.5	22.5	22.5	19.5
D Distortion 2	24.5	23.5	24.5	26	26	25	22

The figures of GS and D for the first segment reinforce the initial interpretation that segregation for parents' education, especially for parents that have not finished fundamental school (EducFS), is declining in Rio de Janeiro public municipal schools. A very small decline is also detected for black pupils in the last four years. The decline in school segregation for parents' education occurs concomitantly with the drop, observed in every year for the past 12 years, in the Gini coefficient. In addition, the GS and D values are nominally low compared to those for other developed countries. This is a relevant finding since Brazil presents historically high levels of social inequality. How can we begin to interpret the trends?

Perhaps one explanation for the relatively low levels of segregation is the fact that all pupils enrolled in private schools were left out of the analysis. Around 18% of pupils in the

whole country are enrolled in private schools. In Rio de Janeiro, this number is even higher, with a proportion of 25%. Unfortunately, at this point, it is not possible to incorporate data from private schools in order to allow a more robust analysis considering the entire pupil population in fundamental schools.[13]

Since all private schools charge fees and are mainly attended by the middle class and elite, it is reasonable to assume that the nominal levels of segregation presented in the thesis are underestimated. The reason is simple. The data available deal with a more homogeneous part of the population and, therefore, it is most likely that a part of the variation that would influence GS and D has been left out of the analysis. Current data do not allow robust international comparison of Rio de Janeiro (or any other capital in Brazil), with major cities around the world.

The patterns for poverty are different from EducFS and show an increase in between-school, starting in 2007. The most plausible explanation suggests that the changes observed are not related to the educational system. The larger economic growth initiated in 2004[14], leading to historically low levels of unemployment and real increase in household income, led to a decline in the proportion of poor families. There is empirical evidence from different countries suggesting that a decline in poverty is accompanied by an increase in school segregation.

There are perfect records for age/grade distortion for all seven years, presenting different trends depending on the method of calculation – Distortion 1 or 2. Segregation patterns for pupils with one or more years of age/grade distortion show stability over time. Nonetheless, the trends for a distortion of two of more years (more vulnerable pupils) highlight a small decline comparing 2004 and 2010.

[13] Very recently, our research team had access to the entire educational data, including private, but it will be necessary to check the quality of the information and also purge the datasets before starting the analysis.
[14] The figures for Gross Domestic Product (GDP) growth are: 2004 – 5.7%; 2005 – 2.9%; 2006 – 3.8%; 2007 – 5.4%; 2008 – 5.2%; 2009 – -0.2%; 2010 – 7.5%.

In the case of the indicators for Distortion 1 and 2, the change in one educational policy that can interfere with the results presented should be considered. In 2009, the new mayor of the city, Eduardo Paes, on the first day in office, repealed a specific policy called "*Progressão Automática*" (Automatic Approval)[15], which, in theory, would prevent pupils from being retained at the end of the school year, even if the academic results were below the minimum expected for each grade transition. The new policy started in 2009 and gave more power to teachers decide if a pupil with poor academic performance should be held back. The newly shaped policy had the clear potential to influence the proportion of the potentially disadvantaged pupils for Distortions 1 and 2.

Even so, the figures in Table 31 do not fully support this hypothesis, since the proportions of disadvantaged pupils for Distortions 1 and 2 show opposite trends. The effect of the policy on retention is not clear.

Table 31: Proportion of Potentially Disadvantaged Pupil for Age/Grade Distortions 1 and 2 – First Segment.

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
Proport Distortion1	0.44	0.46	0.46	0.49	0.48	0.46	0.44
Proport Distortion2	0.12	0.13	0.12	0.14	0.12	0.10	0.11

Nonetheless, there is an alternative explanation that should be investigated in order to rule out any adverse effects of the policy. It is possible that the repeal of the "Automatic Approval" policy did not increase the total proportions for Distortions 1 and 2, but instead increased the proportion of pupils in "Special Class" – *Realfabetização* and PEJA I – Fundamental Education. Both policies track pupils with persistently low achievement and place them into different classes (or schools), outside the regular grades. For the first

[15] See JB Online: TR – *Eduardo Paes acaba com a aprovação automática na rede de ensino* – 01/01/2009. Página visitada em 2th April 2010.

segment, the most important policy is "*Realfabetização*", which specifically targets pupils that are not proficient in reading and writing (literacy), at the end of 3rd grade. PEJA I tracks older pupils (15 or over) that are not literate – severely delayed.

Table 32: Proportion of Pupils Enrolled in "*Realfabetização*" and PEJA I.

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
Realfabetização	0.09	0.08	0.07	0.01	-	0.04	0.08
PEJA I	0.06	0.07	0.07	0.06	0.07	0.07	0.06

The figures in Table 32 suggest that there is a gradual increase, starting in 2009, in the proportion of pupils enrolled in the "*Realfabetização*" program. This is an important finding that suggests that figures for Distortions 1 and 2 could be heavily influenced by the sudden change in the policy. The pattern from 2004 to 2008 clearly shows a constant decline in the proportion of pupils enrolled in "*Realfabetização*" (at that time, the policy was called *Classe de Progressão*). However, the pattern changes at the beginning of the new administration (2009), with a rapid increase in pupils enrolled in the newly shaped policy. The main issue for the models in the thesis is that pupils enrolled in "*Realfabetização*" are not officially enrolled in regular grades and, therefore, are excluded from the datasets used to calculate GS and D.

The small decline observed in GS and D for Distortion 2 – first segment pupils – (see Table 30) could be related to intense tracking of pupils with low achievement – very likely to be potentially disadvantaged for age/grade distortion. Subsequent chapters will specifically measure the effect of tracking policies on school segregation.

***

The next step is to describe segregation patterns for pupils enrolled in the second segment – 6th to 9th grades. Table 33 presents the figures of GS and D for all variables. The picture for the second segment is not very different from the one described previously. Despite the potential effect of missing data (greater for the second segment) on GS and D, segregation by parents' education seems to be declining in Rio de Janeiro public schools for both segments. A similar interpretation can be made for pupils' colour; however, the decline in the latter variable is less intense.

Table 33: Segregation Index (%) and Dissimilarity Index (%) for All Available Indicators Considering the 2nd Segment.

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
GS Black Pupil	18	17.5	15	14	13.5	13	12.5
D Black Pupil	17.5	17.5	16	15.5	15.5	15	14
GS Non White Pupil	11.5	10.5	8	7	6.5	5.5	5.5
D Non-White Pupil	17	17.5	15.5	15	14.5	14	14
GS EducFS	32	28.5	25.5	23.5	21.5	20.5	19.5
D EducFS	34.5	32	30	28.5	27	26	25.5
GS EducHS	17.5	14.5	11.5	10	9	8.5	8.5
D EducHS	24.5	24	24	23.5	23.5	23.5	23
GS NIS	22	19	17.5	16.5	16	16	16.5
D NIS	24	22	21.5	21.5	21.5	22	23
GS Distortion 1	5.5	6	5.5	4.5	5.5	5.5	5
D Distortion 1	14.5	14.5	14	13	13.5	13.5	12.5
GS Distortion 2	11.5	11.5	11.5	12	12.5	12.5	12.5
D Distortion 2	17	17	16.5	17	17	17	17

Segregation by poverty in both segments presents a similar trend. The observed decline in the first years was interrupted in 2007. A new trend started, with a modest increase in the overall levels of segregation. Sub-notification for this variable in the first four years

(2004-2007) is one probable explanation. Another plausible explanation is the rapid increase in the total number of families assisted by the Federal government. On the other hand, after the intense growth observed in the initial years, a decline in the total proportion of poor families caused by the economic growth could explain the modest increase observed in the indexes since 2007.

Nevertheless, the most important element presented in Table 33 (second segment) is not the segregation patterns themselves (mainly because they match the interpretations regarding first segment pupils). It is possible to notice that the nominal values of GS and D for all variables are smaller in the second segment. This is an important finding that contradicts previous interpretations for Rio de Janeiro public schools.

The explanation for the new interpretation is the new research design presented by the thesis, which measures between-school segregation using two indexes, considering the entire pupil population. Previous studies had two major limitations. First, they analyzed a limited number of schools in very specific areas of the city. The limitation is the inability to extrapolate the results obtained in these areas to other places, with distinct characteristics. The main threat is selection bias. The second limitation was the option to divide the schools into groups: low and high performance based on Standardized Tests – *Prova Brasil* (Test Brazil). The cut-off point was arbitrary, usually considering the top quartile high performance schools.

Costa et al. (2013) and Bartholo and Bruel (2012) analyzed the mandatory transition between the first and second segments (called "*remanejamento*"), of a limited number of pupils in Rio de Janeiro public schools. Similar variables were used in a regression analysis to estimate the chances of access to the most prestigious schools – pupils' colour, parents' education; age/grade distortion. The conclusions of both papers suggest that the transfer of pupils across schools were not random and have the potential to increase school segregation.

Despite the fact that the papers do not explicitly state that the levels of segregation in the second segment are higher, both models suggested that the general effect of "*remanajemento*" would be a nominal increase in segregation. This is a wrong assumption.

A future chapter in this thesis will examine in detail the impact of the mandatory transition between the first and second segments, combining longitudinal and cross-sectional designs. It is possible to anticipate that there will be evidence of pupils selected by a limited number of schools; however, the overall effect is a decline in segregation. The reason is simple: a considerable decline in the total supply (number) of schools – an average decline of 45%. There is empirical evidence that a sudden decline in the total number of schools (for example, policies to close schools) tends to bring segregation down.

Since GS and D are organisationally invariant, what the figures suggest is a simple probabilistic effect of "*remanejamento*". Nonetheless, one question should be asked: if the relocation was presumed to be random, how big would the decrease be? Moreover, would it be similar to the size presented in Tables 30 and 33?

6.1 School Segregation across Educational Authorities (CREs) – First Segment

The chapter describes school segregation patterns for all educational authorities (a total of 10 CREs) in the city of Rio de Janeiro. Table 34 presents figures for first segment pupils considering black pupils as the potentially disadvantaged group. At first, it is possible to observe two interesting points: a) the big variation in the nominal values comparing different regions of the city; b) the decline in segregation, with the exception of CRE1.

Table 34: Segregation Index (%) for Black Pupils Considering All Educational Authorities – First Segment.

	GS Black Pupils									
	CRE1	CRE2	CRE3	CRE4	CRE5	CRE6	CRE7	CRE8	CRE9	CRE10
2007	17,5	20	16,5	18,5	12	11	23	14,5	12,5	14
2008	17	20	16	18	10,5	10,5	19	14	12	13
2009	16,5	20	17	17,5	10,5	10	19	13,5	12	13
2010	17,5	19	15	17	11,5	10	19	12,5	11	12,5

The initial description shows that CRE 2 and 7 have the highest segregation levels by pupil's colour. Located in the touristic part of the city, with many middle class neighborhoods, such as *Tijuca*, CRE2 has key characteristics that can influence the overall levels of segregation: 1) large supply of schools; 2) high demand (population density/ urbanization); 3) good public transportation; 3) clusters of shantytowns (*favelas*) in the middle of affluent neighborhoods. The hypothesis is that all these elements increase parental competition and could have an effect on school segregation.

Table 35 present the figures for all CREs considering poor families. Once again, CRE2 and 7 appear as the most segregated settings in the entire city. However, the segregation trends are opposite. CRE2 has shown a very small decline in the last two years, and CRE7, a constant increase since 2007. In fact, a total of 5 CREs present an increase in

school segregation considering poor families. It is not clear why different regions present opposite patterns in the same period.

Table 35: Segregation Index (%) for Poor Families Considering All Educational Authorities – First Segment.

	GS Poor Families (NIS)									
	CRE1	CRE2	CRE3	CRE4	CRE5	CRE6	CRE7	CRE8	CRE9	CRE10
2007	12,5	23	13	14	17	16	17	12	9	12
2008	15	23	13	15	16,5	17	18,5	11	9,5	11
2009	18	21	13,5	15	17,5	17	19,5	11	11	11
2010	19	21	13	16	18	16,5	21,5	11	12	11

Parents' education shows a clear decline for all educational authorities – see Table 36. The only difference is the intensity of the phenomenon. The figures corroborated previous interpretations that indicated that segregation measured for parents' education is clearly declining in Rio de Janeiro public schools. CRE5, despite the constant decline, presents higher nominal values for all measurements. Future studies should investigate more deeply why one specific region presents such high figures.

Table 36: Segregation Index (%) for Parents' Educations Considering All Educational Authorities – 1st Segment.

	GS Parents Who Did Not Finish Fundamental School (EducFS)									
	CRE1	CRE2	CRE3	CRE4	CRE5	CRE6	CRE7	CRE8	CRE9	CRE10
2007	17	25,5	22	25	32	24	17,5	23	18	21,5
2008	14,5	23	21	24	30	24	16,5	21	18	19,5
2009	13	22	20	22,5	28	23	16	20	17	16,5
2010	12	19,5	18,5	21,5	27	21,5	15	19,5	16	16

The simple decline in school segregation could be seen as a good sign; however, the causes and its effects are unknown. As a matter of fact, before this initial description, there was no clear idea about this phenomenon. It is also intriguing how different educational authorities over the same period can present very different trends in segregation comparing different pupil characteristics. Turning to analysis of CRE1, data from Tables 34, 35 and 36 are revisited. The same educational authority presents three different schools segregation patterns: stability for pupil's colour, a big increase for poor families and a decline for parents' education. It is not clear what explains such outcomes, although it is possible to speculate that the causes of segregation for each variable (e.g. pupil characteristics) are different.

A similar analysis could be conducted for almost all educational authorities. This is an indication that between-school segregation in Rio de Janeiro might not be one single process. This is an important question for policy-makers and educational researchers interested in the phenomenon of school segregation. The next chapter will fully investigate this question, with alternative models.

Table 37 presents figures for all pupils enrolled in the first segment that have two or more years of age/grade distortion. Once again, CRE2 appears as the most segregated setting. Almost all CREs present a small decline over the period. The most likely explanation is implementation of the new policy called "*realfabetização*", which tracks pupils with low attainment.

Table 37: Segregation Index (%) for Age/Grade Distortion Considering All Educational Authorities – First Segment.

	GS age/grade distortion two or more years (Distortion2)									
	CRE1	CRE2	CRE3	CRE4	CRE5	CRE6	CRE7	CRE8	CRE9	CRE10
2007	22,5	25	20,5	20	22	21	22	22	20	21
2008	25	26	19	21	19,5	21	20	23	21,5	21,5
2009	23	25,5	21	20	19	22	18,5	21	22	22,5
2010	20	23	18	17	16	22	17	16,5	18	20

Assessing school segregation considering the educational authorities as the unit of analysis allows a deeper understanding of different processes that can influence the phenomenon. “Segregation is a spatial outcome of spatial processes which, therefore, needs to be measured spatially.” (Harris, 2012, p. 669). This is a major limitation of all indexes used in the thesis (GS, D, SR). These indexes are not spatial and construct “averages” of arbitrary residential or functional areas, such as neighborhoods, educational authorities or electoral wards. Future studies should make an effort to map different areas of the city and investigate the elements associated with school segregation more carefully.

Despite the fact that educational authorities are not a homogeneous area, it is reasonable to regard these areas as a possibility to observe different trends of school segregation more closely. The main findings related to first segment are: 1) a big variation among CREs; 2) on average, CRE 2 emerged as the most segregated setting in the city and CREs 9 and 10 as the most unsegregated; 3) comparing pupils’ characteristics over the same period of time, it was possible to observe different segregation trends. This could be seen as an indication that school segregation in Rio de Janeiro public schools is not one single process.

6.2 School Segregation across Educational Authorities (CREs) – Second Segment

Table 38 presents the figures for second segment pupils considering black pupils as the potentially disadvantaged group. The trends suggest that segregation slowly declines (very modestly) in CREs 1, 2, 3 and 8 for all years. The trends for the other educational authorities suggest stability over the same period. It is not clear what is associated with these changes; however, the nominal values of the changes are quite small and should be interpreted with caution. Once again, CREs 2 and 7 appear as the most segregated areas for this particular variable in the entire city.

Table 38: Segregation Index (%) for Black Pupils Considering All Educational Authorities – Second Segment.

	GS Black Pupils									
	CRE1	CRE2	CRE3	CRE4	CRE5	CRE6	CRE7	CRE8	CRE9	CRE10
2007	13,5	16,5	14	15,5	12	8,5	16	12,5	8,5	13
2008	14	16,5	13,5	15	12	7	16,5	11	8,5	13
2009	12,5	16	13	14,5	10	7	17,5	10,5	8	12,5
2010	10,5	15,5	12,5	15	11,5	7,5	16,5	10,5	8,5	12,5

A very different picture can be observed for poor families – see Table 39. At first, it is possible to observe three different school segregation patterns for all educational authorities: a) increase – CREs 1 and 8; b) stability – CREs 2, 3, 4, 5, 6 and 7; c) decline – CREs 9 and 10. This is an unexpected outcome that should be investigated in further studies. One possible explanation is changes in the economic growth, which could impact the total number of poor families, although its effect should be more or less stable across the entire city. Obviously this is not the case.

Table 39: Segregation Index (%) for Poor Families Considering All Educational Authorities – Second Segment.

	GS Poor Families (NIS)									
	CRE1	CRE2	CRE3	CRE4	CRE5	CRE6	CRE7	CRE8	CRE9	CRE10
2007	10,5	19,5	15,5	14	17,5	18,5	14	11,5	12	17
2008	12	20	15	13,5	19	17	14	10,5	11	15,5
2009	14	21	14	14	18,5	16	14,5	12	10,5	13,5
2010	15	20,5	15	13,5	17,5	16,5	15	13,5	10	13

An alternative explanation (not very plausible) could be changes in the records of the dataset for the NIS (Social Registry Number). However, the figures presented in Table 39 are for more recent years and present better quality. It is important to investigate if any other unknown factor, associated with practices by the educational bureaucracy, could be influencing such outcomes. Since each educational authority has considerable autonomy, it is not totally implausible that specific interventions, such as selection of poor pupils or desegregating strategies, could be happening at the same time in different CREs.

The segregation patterns for parents' education (EducFS), considering only pupils enrolled in the second segment, are quite similar to the one observed in the first segment. The figures for all CREs suggest a decline in the nominal segregation values; the only difference is the intensity of the phenomenon – see Table 40.

It is intriguing why CRE 5 is clearly more segregated by parents' education in relation to all others – in both segments. It would be interesting to observe if these figures are associated with clear selection of pupils by specific schools. More detailed analysis (at school level) could explain some of these processes. Once again, the lack of specific regulation and transparency regarding enrolment and transfers of pupils could allow members of the educational bureaucracy (located in each school) to intentionally select pupils with specific characteristics.

Table 40: Segregation Index (%) for Parents' Educations Considering All Educational Authorities – Second Segment.

	GS Parents Who Did Not Finish Fundamental School (EducFS)									
	CRE1	CRE2	CRE3	CRE4	CRE5	CRE6	CRE7	CRE8	CRE9	CRE10
2007	18,5	24,5	22	21	30,5	24,5	20	22,5	18,5	21,5
2008	16,5	22,5	19,5	19,5	29,5	23	17,5	20,5	19	17,5
2009	15,5	20,5	18,5	19,5	28	21,5	15	18,5	17	16,5
2010	14	19	19	19,5	26,5	22	14,5	17	16	16

It is safe to affirm that segregation by parents' education is declining for all years in both segments. This is an important descriptive finding. All the international literature about educational disadvantage suggests that this is a key variable to explain pupils' future academic performance. The simple fact that schools are becoming more mixed in terms of parents' education could be seen as an interesting factor for the educational system. It is not clear why this is the only variable that shows a constant decline over the years for both segments in all educational authorities.

Future studies should try to understand what variables appear to be associated with such an outcome. Perhaps the enhancement of the datasets accompanied of a general improvement in parents' education – making the population more homogeneous. However, Table 5 shows that the proportion of disadvantaged pupils for EducFS and EducHS is not declining. This finding partially weakens this particular explanations.

Table 41 presents the GS figures for pupils enrolled in the second segment with two or more years of age/grade distortion. It is possible to observe two different patterns: a) increase in segregation – CREs 1, 3, 4, 5, and 10; b) stability – CREs 2, 6, 7, 8 and 9. Comparing figures for the first and second segments, it is possible to observe that the nominal values are smaller in the second segment. The explanation is the exponential increase in the total proportion of disadvantaged pupils, mainly caused by multiple repetitions over the

educational transitions. Since pupils from all schools may be school failures, there is more equality of age/grade distortion across schools.

Table 41: Segregation Index (%) for Age/Grade Distortion Considering All Educational Authorities – Second Segment.

	GS Distortion age/grade two or more years (Distortion2)									
	CRE1	CRE2	CRE3	CRE4	CRE5	CRE6	CRE7	CRE8	CRE9	CRE10
2007	6,5	13	9	11,5	14	10	11	10,5	7	11
2008	6,5	13,5	10,5	12	14,5	12	12,5	10,5	8,5	12
2009	7,5	13	10,5	12	13,5	12	11,5	9,5	8,5	13
2010	10,5	13,5	12,5	13	16	9	12	10,5	8	12,5

The decline in the second segment for Distortion 2 (compared to the figures in Table 37) is not a cause of celebration. A major challenge is ahead of educational researchers: how to overcome the "culture of repetition" (Ribeiro, 1991) and, at the same time, enhance learning? All external assessments (including international comparative research) show a similar picture: the overall levels of proficiency in Brazilian public schools, including those in Rio de Janeiro, are low, especially in mathematics. The strategy to simple pass pupils to the next grade, regardless of their attainment at the end of the school year, is not enough. Some commentators claim that the "automatic approval" (*progressão automática*) is an important policy to help keep pupils at school. However, the international assessments show that Brazilian pupils learn very little, even comparing to other developing countries. It is time to face the pedagogic challenge: why are so many pupils not learning the minimum required for each grade? What strategies are effective in helping the underachievers?

More than half of the pupils enrolled in regular classes (1st to 9th grade) in Rio de Janeiro public schools have at least one year of age/grade distortion. It is not true that all of them have been retained at least once. There are cases of seven-year-old pupils (or about to

turn seven) enrolled for the first time in the 1st grade. It is not clear why such a procedure is in place, since the current legislation indicates that pupils should be enrolled in the 1st grade at age six. Nonetheless, the figures for two or more years of age/grade distortion are still very high (on average 27%) – see Table 42.

Table 42: Proportion of Age/Grade Distortion Pupils All Grades in Fundamental Education.

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
Proportion Distortion1	0.57	0.59	0.58	0.59	0.57	0.57	0.57
Proportion Distortion2	0.30	0.31	0.30	0.28	0.26	0.26	0.27

It is possible to state that around a quarter of pupils enrolled in regular classes have been retained at least once. These are conservative measurements and highlight the size of the problem. If we add to the group pupils in "PEJA I or II", the figures would be even higher: a total of 36% with two or more years of age/grade distortion in 2010.

Future models in the thesis will show that age/grade distortion is a key variable to understand segregation patterns in Rio de Janeiro. Different from other variables, (for example, poverty) that are directly influenced by residential segregation or parental preference, it is mainly influenced by the educational bureaucracy. As will be shown in future chapters, despite the fact that retention is a "democratic" phenomenon, affecting all "types" of pupils and basically all schools, it is more prevalent among disadvantaged pupils and low performance schools. It is a vicious circle where: a) schools present lower overall attainment levels because of their intake; b) pupils enrolled in low performance schools are more likely to be retained and face age/grade distortion. This is a crucial subject for the educational system and pupil's prospects in life.

6.3 Clustering Pupils in Rio de Janeiro Public Schools: one or more patterns of segregation?

After describing the trends of segregation over time for the first and second segment, the next step is to address the issue about the different segregation processes happening at the same time. Do the patterns of each indicator represent one overall process or should different types of segregation be considered? Gorard and Cheng (2011) asked the exact same question analysing the clustering of potentially disadvantaged pupils in public schools in England, and the results showed at least three different simultaneous segregation processes. So, what happens in the "Unregulated Quasi-Market" in Rio?

In theory, it would be possible to answer this question replicating factor analysis, using trends over time (D and GS) and across schools (SR). However, due to major limitations with the short trend over time, the thesis will only present the analysis across schools using the Segregation Ratio (SR) and other variables that can be associated with school segregation: demand and supply of public schools, Social Development Index (SDI) for each school area and proximity to shantytowns (in meters). The calculations for all four variables were described by Alves, Lange and Bonamino (2010).

Two datasets were used for creating the spatial data layers: the official census of the Brazilian Institute for Statistics and Geography (IBGE) and data from the Urban Planning Institute *Pereira Passos* (IPP) of Rio de Janeiro city. The complete census, including the full range of socio-economic variables, is conducted every ten years, and the most recent data from 2010 have been used in this study.

The spatial variables created for this study are as follows: 1) Demand for public schools: Number of children between 6 and 14 in a radius of 1,500 meters, using the IBGE census data; 2) Supply of public schools: Number of public schools in a 1,500 meter radius; 3) Distance to shantytown: Straight line distance to the next shantytown; 4) SDI: Average of

the SDI of the census sectors in a radius of 1,500 meters. The SDI is a synthetic index composed of 10 indicators grouped into four dimensions: urban infrastructure, housing quality, educational level and income of the head of the household (Cavallieri; Lopes, 2008; Alves; Lange; Bonamino, 2010).

***

Table 43 presents the correlation coefficients for all seven indicators of potentially disadvantaged pupils (SR), and the Social Development Index (SDI) for each school in the year 2010, for all pupils in the first segment. The other variables that could influence segregation (demand and supply of public schools and proximity to shantytowns), described previously (demand, supply and proximity to shantytown), showed low correlation (less than 0.3) and were excluded from the table.

Table 43: Correlation Coefficients for All Available Indicators of Disadvantaged (SR) and the Social Development Index – First Segment 2010.

	SR Non-White	SR EducHS	SR Poverty	SR Distort1	SDI
SR Black	0.55**				
SR Non-White	-	0.39**	0.43**		-0.34**
SR EducFS	0.31**	0.83**			
SR EducHS	0.39**	-	0.31**		
SR Poverty	0.43**	0.31**	-		-0.37**
SR Distort 1				-	
SR Distort 2				0.89**	
SDI	-0.34**		-0.37**		-

** Correlation is significant at the 0.01 level (2-tailed);

How should the coefficients be interpreted? The idea behind them is simple: if school segregation is one single process, schools with a bigger share of disadvantaged pupils for one indicator (e.g poverty), should also present high figures for other indicators. Putting this

differently, on average, schools segregated by poverty would be also be segregated by other variables. The low coefficients in Table 17 for almost all indicators are somewhat of a surprise. Whatever causes segregation by distortion does not seem to be related to any of the other indicators of potential disadvantage. The data refute the initial hypothesis, which presumed that this indicator would indirectly be measuring others. The results are very similar, even when we replicate the analysis for the year 2009. The correlation coefficients are still surprisingly low.

The findings shown in Table 43 are relevant, especially in an educational system that presents nearly one third of pupils with two or more years of age/grade distortion. If segregation by distortion is not associated to any of the other indicators of potentially disadvantaged pupils, policy-makers should be concerned with what is causing this phenomenon and, more importantly, what are the possible consequences (impact) are for these pupils' prospects in life.

The only two variables that present higher values for the coefficients are SR Non-White and Poverty. Both are also negative associated with SDI, which can be seen as an indication that residential segregation affects these two variables more. Poorer areas of the city (low SDI) present higher levels of segregation by poverty and non-white.

It is quite surprising that SDI and proximity to shantytowns are not associated with parental education and age/grade distortion. One possible explanation is the method of calculation for SR. Using the entire network as one unit of analysis is not the best approach. The "real choice" of parents is much smaller than the total number of schools.

Gorard, Taylor and Fitz (2003) highlighted these undesired quality of SR, which could be affected by the changes in the basis of the "fair share" – scale. The authors give an example comparing levels of segregation among schools located far away from each other. The main problem is not the distance in itself, but the possibility that both regions present

very different figures in the overall level of poverty or other key variables that may influence segregation. This is a clear limitation of the model presented, and other approaches should consider calculating the SR for each educational authority or other unit of analysis that seems less arbitrary. This model (using educational authorities as a unit of analysis) will be presented later in the chapter to confirm some of the outcomes.

Table 44 presents the coefficients for all seven indicators of potentially disadvantaged pupils (SR), and the Social Development Index (SDI) for second segment pupils (year 2010). A similar outcome to the one described for first segment pupils should be expected. Since proximity to shantytown and supply of schools showed low correlation (less than 0.3), they were excluded from the table.

Table 44: Correlation Coefficients for All Available Indicators of Disadvantaged (SR) and the Social Development Index – Second Segment 2010.

	SR NonWhite	SR EducHS	SR Poverty	SR Distort1	SR Distort2	Demand
SR Black	0.52**			0.35**		0.36**
SR Non-White	-	0.39**	0.43**	0.39**	0.31**	
SR EducFS		0.75**		0.46**	0.41**	
SR EducHS	0.43**	-	0.45**	0.43**	0.33**	
SR Poverty	0.43**	0.45**	-			
SR Distort 1	0.39**	0.43**		-	0.94**	
SR Distort 2	0.31**	0.36**		0.94**	-	
SDI	-0.40**		-0.42**			-

** Correlation is significant at the 0.01 level (2-tailed);

It is clear that the coefficients for second segment pupils are higher compared with Table 44. Nonetheless, the patterns have not changed, with the exception of the variable age/grade distortion, which becomes more associated with non-whites, blacks, EducFS and

EducHS. Another change is the positive association between demand for school (highly dense areas) and black pupils. The coefficients simply show an association, and it is not possible to establish any causal relation. What we do know is that areas with high demand in the city are more likely to be segregated by the pupil's colour (more specifically black pupils).

For the second segment, the association between school segregation for non-white and parental education with Distortion 1 and 2 is clear. The outcome could be seen as an escalation of the segregation process. It is possible to speculate about two different causes of the change. The first would be non-random transfers of pupils across schools. These changes presumably would make the education system as a whole more segregated. A second explanation could be that retention affects schools differently. Specific schools (those with a higher share of potentially disadvantaged pupils) would retain more pupils in each transition. This uneven "retention rate" would lead those schools to also present a higher share of disadvantaged for age/grade distortion.

***

The final step is the Factor Analysis with all indicators of SR and other variables at school level that could be associated with school segregation, such as: demand for school, supply of education, IDS and proximity to shantytowns.

The analysis for each segment considered two models: Model 1 composed of indicators with a higher proportion of potentially disadvantaged pupils – poverty, non-white pupils, EducHS and Distortion1; Model 2 presents the indicators with a smaller proportion – poverty, black pupils, EducFS and Distortion2. The factor analyses were conducted with most recent data (2009 and 2010) that present a better record. Comparing both years, the

results were very similar, and, for this reason, the thesis will only present the figures for 2010.

Both models for first segment pupils present solutions that suggest more than one segregation process occurring simultaneously. The most parsimonious model, with only 8 variables, for the group with a high proportion of disadvantaged pupils shows a 3 Factor solution for the Rotated Component Matrix, with 62% variance explained: non-white, poverty and EducHS as Factor 1 and Distortion1 as Factor 3 – see Table 45.

Table 45: Rotated Component Matrix for All Available Indicators (SR) – Higher Proportion of Disadvantaged – First Segment 2010.

	General Segregation	Demand/ Supply	Segregation by Distortion
SR NIS	0.76		
SR Non-White	0.73		
SR EducHS			0.61
SR Distortion1			0.87
Proximity to Shantytowns			
IDS	-0.71		
Demand for School		0.84	
Supply of School		0.79	

Values lower than 0.5 have been suppressed.

The model with a lower proportion of potentially disadvantaged pupils shows a slightly different outcome for the Rotated Component Matrix, with 3 Factors and 59 % variance explained: Factor 2 as poverty and Factor 3 as Distortion 2, black pupils and EducFS – Table 46. In both Factor Analyses, SR by poverty and Distortion present the highest coefficients (see Tables 45 and 46) and appeared as separate factors.

Table 46: Rotated Component Matrix for All Available Indicators (SR) – Lower Proportion of Disadvantaged – First Segment 210.

	Demand/ Supply	General Segregation	Segregation by Distortion
SR NIS		-0.78	
SR Black Pupil			0.59
SR EducFS			0.69
SR Distortion2			0.71
Proximity to Shantytown			
IDS		0.78	
Demand for School	0.84		
Supply of School	0.77		

Values lower than 0.5 have been suppressed.

The different solutions for each factor analysis were somewhat expected, and were also highlighted in the findings of Gorard and Cheng (2011). Maybe the most important element is the consistency in the suggestion of more than one segregation process. Both models for the first segment, using a higher and lower proportion of indicators for SR, suggested two segregation processes happening at the same time: a) SR Poverty and SDI; b) SR Distortion and SR Parental Education. If these assumptions are correct, it is possible to speculate that there are two independent segregation processes. The first one would be closely linked to patterns of residential segregation in the city, and the second, linked to elements of the educational system that associate segregation by age/grade distortion and parental education. If it is true that parental education is the most important variable to explain proficiency in public schools in Rio, both factors are feasible.

Will the models for the second segment show a similar pattern / factors? Table 47 presents the Factors considering a high proportion of disadvantaged pupils enrolled in the

second segment. The most parsimonious model, with 8 variables, show a 3 Factor solution for the Rotated Component Matrix, with 67% variance explained.

Table 47: Rotated Component Matrix for All Available Indicators (SR) – Higher Proportion of Disadvantaged – Second Segment 2010.

	General Segregation	Segregation by Distortion	Demand/ Supply
SR NIS	0.73		
SR Non-White	0.55	0.57	
SR EducHS		0.76	
SR Distortion1		0.87	
Proximity to Shantytowns	-0.53		
IDS	-0.81		
Demand for School			0.85
Supply of School			0.83

Values lower than 0.5 have been suppressed.

A very similar pattern can be observed. Once again SR Poverty and SDI appear as one Factor and age/grade distortion and EducHS as another completely different. This is the first model in which the variable that measures how close the school is to the nearest shantytown shows coefficient values higher than 0.5. It suggests that the closer a school is to a shantytown, the more segregation by poverty and non-white rise.

The last model presents the values for the smaller proportion of disadvantaged pupils enrolled in the second segment. Table 48 presents a 3 Factor solution, with 64% of variance explained. The outcome is very similar to the one observed previously, with two major patterns of segregation.

Table 48: Rotated Component Matrix for All Available Indicators (SR) – Lower Proportion of Disadvantaged – Second Segment 2010.

	Demand/ Supply	General Segregation	Segregation by Distortion
SR NIS		-0.78	
SR Black Pupil			
SR EducFS			0.78
SR Distortion2			0.82
Proximity to Shantytown		0.52	
IDS		0.80	
Demand for School	0.86		
Supply of School	0.77		

Values lower than 0.5 have been suppressed.

All four models (two for each segment) suggest the same answer. School segregation in Rio de Janeiro public municipal schools is not one single process. Apparently, it is possible to separate two patterns. The first one clearly associates SR Poverty with territorial variables – SDI and proximity to a shantytown. The result is quite plausible and confirms the role of territory in the school segregation patterns (Harris, 2012). A different pattern is associated with SR Distortion (1 and 2) and SR Parental Education (EducFS and EducHS).

The initial findings corroborate the results of similar studies conducted in the U.S. and England, which also pointed out that school segregation is not one single process (Jacobs, 2011; Gorard; Cheng, 2011). The findings have implications for policy-makers, and stress the importance of measuring between-school segregation by multiple pupil characteristics. Another issue is the impact of each type of cluster on pupil attainment and post-compulsory education. Are they equal or are there differences that should be considered? The latter concern is related with the possible interaction effects of different policies that aim to increase or diminish specific types of clustering.

In the case of Rio de Janeiro, the models suggest that one pattern of school segregation is associated with the practice of pupil retention. The concern with retention and its impacts on pupil's life chance have been discussed in Brazil since 1980s (Ribeiro, 1991). The new element here is that age/grade distortion does not appear to be associated with segregation by poverty and pupil's colour. If the assumption is correct, any attempt to address the issue of school segregation in public schools in Rio should consider at least two different approaches.

6.4 One or More Patterns of School Segregation? Replicating Models across Educational Authorities

The index used to test the hypothesis about one or more patterns in schools has one major disadvantage. Segregation Ratio (SR) yields a score for each school indicating if disadvantaged pupils are over or under-represented. Possibly, the biggest limitation of the model using SR is related to the scale of the unit of analysis. Since the score of one school is directly linked to another in the model (the sum of all values is one), the basis of the "fair share" could change if different areas are chosen.

The same limitation could be applied to D and GS. However, for SR the danger is even greater because the correlations and factor analysis will consider the score for each school and not a summary measure for the entire unit chosen – as happens to D and GS.

Despite the fact that Rio de Janeiro has an open enrolment system, parents' "real choice" is much more limited. If this is true, choosing less arbitrary areas could influence the "fair share" for each school and potentially alter the outcomes of the factor analysis. This chapter replicates the models in the previous chapter, but, this time, SR will be calculated separately for each Educational Authority. The main goal is to observe if the Factors will tell a similar story compared with previous models: two separate school segregation patterns.

For each Educational Authority, two Factor Analyses were conducted: 1) variables with higher proportions of disadvantaged pupils – poverty, non-white pupils, EducHS and Distortion1; 2) variables with smaller proportions of disadvantaged pupils – poverty, black pupils, EducFS and Distortion2. In both models, other variables related to territory were included: demand for school, supply of education, IDS and proximity to shantytowns.

For each segment, there are 20 Factor Analyses – 2 for each Educational Authority. Since the main goal is to observe if the outcomes are similar to the ones observed considering the entire network, the results for first segment are summarized in Table 49. The idea is to

point out how many factors were extracted in the Rotated Component Matrix. Factors that only presented territory variables (for example, demand and supply as the only variables in one factor), were not considered in Table 49. Only Factors that presented SR as one of the variables were considered as "valid" Factors to observe segregation processes.

Table 49: Rotated Component Matrix for the First Segment for All Educational Authorities.

	SMALLER PROPORTION			**BIGGER PROPORTION**		
	1 Factor	2 Factors	3 Factors	1 Factor	2 Factors	3 Factors
CRE1			x		x	
CRE2		x			x	
CRE3			x	x		
CRE4		x			x	
CRE5			x		x	
CRE6		x			x	
CRE7		x		x		
CRE8		x			x	
CRE9			x		x	
CRE10			x		x	

The results corroborate the findings from the previous chapter. A total of 13 models (75%) presented a two-Factor solution. It was not realistic to imagine that all models would present the exact same outcome. The idea is to search for clear patterns and the claim here is that between-school segregation in the first segment is not one single process. Only two educational authorities (CRE 3 and 7), considering the model with a higher proportion of disadvantaged pupils, (poverty, non-white pupils, EducHS and Distortion1) showed only one Factor solution (total of 10%).

These are no definitive tests. Two or more factors suggest that schools over-represented for one "type" of disadvantage can be under-represented for another in the same

year. Whatever causes segregation by age/grade distortion in schools, it does not appear to be associated with segregation by poverty. Territorial variables (SDI and proximity to shantytowns) are usually in the same factor as SR poverty.

What about the second segment? Will solutions on the models be similar? Table 50 presents the results of 20 Factor Analyses, two for each educational authority. Using the same criteria described for the first segment, Factors that only presented territory variables (for example, demand and supply as the only variables in one factor), were not considered. Only Factors that presented SR as one of the variables were considered as "valid" Factors.

Table 50: Rotated Component Matrix for the Second Segment for All Educational Authorities.

	SMALLER PROPORTION			**BIGGER PROPORTION**		
	1 Factor	2 Factors	3 Factors	1 Factor	2 Factors	3 Factors
CRE1			x		x	
CRE2		x		x		
CRE3			x	x		
CRE4		x		x		
CRE5		x		x		
CRE6		x			x	
CRE7			x		x	
CRE8		x			x	
CRE9		x		x		
CRE10			x			x

The models calculated for pupils enrolled in the second segment are somewhat different from those observed previously in the first segment. Considering only the models with higher proportions of disadvantaged pupils (poverty, non-white pupils, EducHS and Distortion1), a total of 5 (or 50%) suggested only one Factor solution for the Rotated

Component Matrix. The other models, using variables with smaller proportions of potentially disadvantaged pupils, were very similar to those seen in Table 49 (first segment).

Is it possible to explain the difference observed in Tables 49 and 50? Previous chapters described nominal values of school segregation over the years using D and GS. The figures highlighted that segregation is higher on the first segment. The decline in segregation observed in the second segment will be investigated more deeply later on. Nonetheless, one element associated with the change is the big decline in the total number of schools. Despite the fact that GS and D are organizationally invariant, such a dramatic decline in the total number of schools will most likely make schools more homogeneous. Disadvantaged pupils more equally distributed means, in the definition of the thesis, less segregation.

If it is true that schools are, on average, more homogeneous in the second segment, it is possible that the different school segregation patterns observed in the first segment suddenly became one. It is important to note that, if measuring school segregation with smaller proportions of disadvantaged pupils (variables that discriminate better – poverty, black pupils, EducFS and Distortion2), there is not much difference in each segment.

Adding all Factor Analyses, a total of 40, it is possible to observe a pattern among the Factors extracted. Only 18% suggested one single segregation process. This outcome was only observed when using variables with a higher proportion of disadvantaged pupils – poverty, non-white pupils, EducHS and Distortion 1. All other models indicated two (58%) or three (25%) segregation processes occurring simultaneously. Perhaps the most relevant finding was the clear association between: a) SR Poverty and territorial variables; b) age/grade distortion and parents' education.

It is not possible to be sure if there are two or more different segregation processes happening in Rio de Janeiro public schools. The mere association of the indices is not a clear indication of causation. Nonetheless, the stability of the Factors across educational authorities

is an important indication of the phenomenon of school segregation. Age/grade distortion emerges as an important variable to explain school segregation in Rio de Janeiro public schools. A future chapter will reaffirm the centrality of this variable and suggest the existence of informal tracking in Rio de Janeiro educational system.

7. UNINTENDED IMPACT OF IMPACT OF EDUCATIONAL POLICIES ON SCHOOL SEGREGATION

7.1 School Segregation in Rio de Janeiro Public Schools: Impact on Educational Transitions

This chapter analyzes the transfers of pupils across schools. The general hypothesis is that pupil transfers are not random and have the potential to increase between-school segregation. Previous studies have analyzed one particular mandatory transition named relocation (*remanejamento*) (from the 5th to the 6th grade). This thesis presents unprecedented data for all educational transitions in fundamental education, and proposes a new research design to assess the potential impact of pupil transfers on school segregation.

The results indicate that: 1) around one third (33%) of the pupils change schools in the first four educational transitions (1st – 5th grade) and around 18% in the last three (6th – 9th); 2) the unfettered movement of pupils presents a specific pattern, which suggests a small, but gradual, increase in the overall segregation level for both segments; 3) between-school segregation declines in the only mandatory transition observed (5th-6th grade), mainly due to a massive decline in the total number of schools (45% decline). These findings contradict previous assumptions about school segregation in Rio de Janeiro.

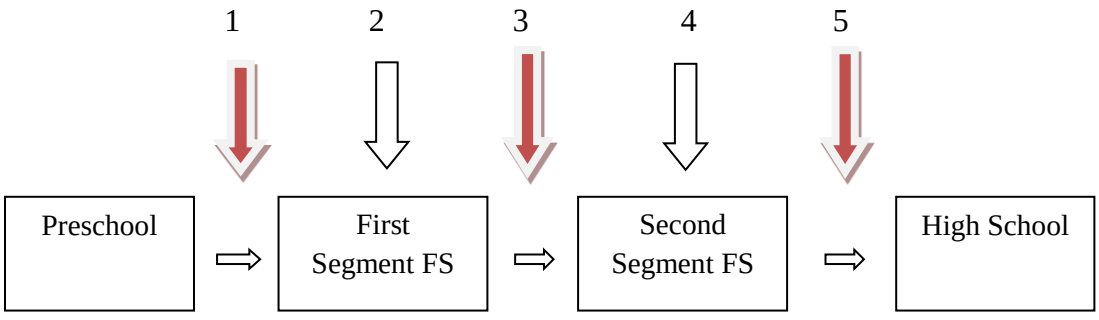
Costa et al. (2013) and Bartholo and Bruel (2012) analyzed the mandatory transition between the first and second segments of a limited number of Rio de Janeiro public schools. The authors used the same datasets presented in the thesis and chose similar variables: parents' education, pupils' colour, age/grade distortion. Schools were divided into two groups – high and low performance – in order to estimate the chances of access to the most prestigious schools. The conclusions of both papers suggested that pupil transfers across schools were not random and increased school segregation. Despite the fact that the papers do not explicitly state that the levels of segregation in the second segment are higher, both

models suggested that the general effect of relocation would be a nominal increase in segregation.

Both papers had two major limitations. The cut-off points to divide schools into groups are arbitrary, usually considering the top quartile high performance schools. Would the regression coefficients be consistent using different cut-off points (taking two equal groups or the top decile as part of the range)? Not being able to answer these questions should raise concern. A second limitation is the criteria to select schools. Intentional sampling, with a very limited number of schools, enables any extrapolation of the results observed. The main threat is selection bias.

This chapter proposes a new research design that takes into account these limitations. The Segregation Index gives stable measurements over the years, without the need to divide schools into arbitrary groups. In addition, the chapter presents two complementary models (cross-sectional and longitudinal) considering all pupils and schools for all educational transitions.

Figure 3: Types of Transfers in Rio de Janeiro Public Municipal Schools – Red Arrows Represent Mandatory Transfers; White Arrows Represent Internal Transfers (Not Mandatory).



Voluntary transfers (see Figure 3 numbers 2 and 4) will be assessed combining a cross-sectional and longitudinal approach. Two different cohorts were chosen: pupils at 1st grade, 2006; pupils 6th grade, 2006. The cross-sectional design measures segregation in

different grades starting in both cohorts of the first and the second segment. For the first segment, a total of four voluntary transfers are assessed (1st-5th grades) and for the second segment three (6th-9th grades).

The longitudinal design follows both cohorts throughout all educational transitions. Because retention is a major issue, affecting almost 60% of all pupils in the public network, the longitudinal approach can help observe the potential impact of retention on school segregation.

The impact of mandatory transfers (relocation, see number 3 in Figure 3) will be assessed in a cross-sectional design, comparing the figures of D and GS on the 5th and 6th grades. The model will be replicated in different transitions to observe if the outcomes are stable.

7.2 Voluntary Transfers in the First Segment

This section focuses on the four initial transitions in fundamental education (1st-5th grade). There are around 800 schools in the first segment with 320,000 pupils enrolled. Table 51 shows the total proportion of pupils who have changed school in different educational transitions 1st-5th grade. The data is specific to one cohort: 1st graders, 2006.

Table 51: Proportion of Pupils in the First Segment Who Have Changed to a Different School – Cohort 1st grade, 2006

	2006-2007	**2006-2008**	**2006-2009**	**2006-2010**
First Segment – Initial Grade 1st	13.5%	21%	30%	33%

Replicating the same model for a different cohort (1st grade, 2007), the proportions are very similar. This is a clear indication that, even when pupils have the option to remain in the same school, many families choose to move their children to a different school.

There are two main reasons for "voluntary" school transfer: a) families that have moved to a different address; b) families that are discontent with their current school choice or believe there is a better option elsewhere – what will be called "pedagogic reasons". For the purposes of the research, it would be important to discriminate between these two types of changes – only type "b" is relevant for the research. However, this is not an easy task with the available data. There are two reasons for this.

The first is related to the quality of the variables related to the family's address. There are many missing cases or invalid zip codes in the dataset (up to 40%). The street names are also problematic with misspelling or other problems that prevent a clear identification of the address. The second problem is that not all changes in address should be a reason to move to a different school. Sometimes, families move to a different house in the same street or a new

street close by, which should not necessarily demand a school transfer. For all of these reasons, it is quite hard to know exactly how many of these changes described in Tables 51 are strictly related to families' discontentment with their previous choice.

An alternative way to describe how many pupils moved to a different school is to "control" for other key variables, such as Educational Authority and Polo. Changing to a school in a different Educational Authority does not necessarily mean that parents have changed address, but it is reasonable to assume that most of these families have moved to a distant address. Table 52 shows the proportion of pupils from the cohort 1st grade, 2006 that has changed to a school in a different educational authority.

Table 52: Proportion of Pupils Who Have Changed to a School in a Different Educational Authority – Cohort 1st grade, 2006.

	2006-2007	**2006-2008**	**2006-2009**	**2006-2010**
First Segment – Initial Grade 1th	4 %	6 %	7 %	8 %

The figures suggest that at least part of the variation observed in Table 51 could be related to matters like parents changing address. A cleverer approach would be to observe how many pupils have changed school in the same Educational Authority and Polo. This ensures that the statistics will only capture transfers to a nearby school – Polos tend to aggregate quite small[16] geographical areas in the city. Table 53 presents the exact proportion of pupils who have moved to a different school in the same CRE and Polo.

[16] Walkable distance.

Table 53: Proportion of Pupils Who Have Changed to a Different School in the Same Educational Authority and Polo – Cohort 1st grade, 2006.

	2006-2007	**2006-2008**	**2006-2009**	**2006-2010**
First Segment – Initial Grade 1th	6 %	9 %	15 %	16 %

The figures in Table 53 correspond to almost half of the variation observed in Table 51. It is quite safe to assume that all these transfers are related to parental preference for a different school – "pedagogic reasons". The real figures are somewhere between the figures in Table 51 and 53. It is quite clear that the phenomenon of pupil transfers across schools is relevant and affects more than 10,000 pupils in each cohort during educational transitions in the first segment (voluntary transfers).

***

After describing the proportion of pupils that have been transferred to a different school, it is important to describe any association between "type of transfer" and pupil's profile. If schools are intentionally selecting pupils based on key characteristics, the patterns of mobility (upward and downward) of disadvantaged and non-disadvantaged pupils should be different. For example, to corroborate the initial hypothesis, potentially disadvantaged pupils should be overrepresented in the group that faces downward mobility.

Previous studies used the mean scores of three consecutive external assessments of *Prova Brasil* (2005-07-09) to create a typology of schools. Usually, the top quartile is named "High Performance" schools and the rest "Low Performance" (Bruel; Bartholo, 2012; Costa et al. 2014). The model in the thesis will discriminate more the schools that are in: a) the top quintile, called high performance schools (Type 3); b) the last quintile, that is, the low

performance or "sinking schools" (Type 1); c) the group in between, termed medium performance schools (Type 2).

The first analysis tracks pupil transfers in the course of four educational transitions during the first segment. The cohort chosen contained pupils in the first grade in 2006 – a total of 68,926 cases. The main hypothesis is that the transfers are not random, with an association between type of mobility (upward or downward) and the pupil's profile. Table 54 presents descriptive statistics of all three types of schools.

Table 54: Descriptive Statistics for Low, Medium and High Performance School – 1st Graders 2006.

	Mean Parents' Education	Proportion Black Pupils	Proportion Poor Pupils	Proportion Boys
Low Performance	**1.86**	**0.13**	**0.37**	**0.52**
Medium Performance	**2.01**	**0.11**	**0.31**	**0.52**
High Performance	**2.32**	**0.10**	**0.23**	**0.51**

The figures simply corroborate what previous studies have shown about the correlation between pupil attainment and socio-economic profile. High performance schools present a higher mean for parents' education and a lower proportion of black and poor pupils. It is interesting to observe that medium performance schools are clearly different from "sinking schools" and high performance schools. Future studies that aim to estimate chance of access to high performance schools should consider using a multinomial regression instead of logistic (dummy variable).

The next step is to observe patterns of mobility. Is there any indication of an association between pupil profile and type of mobility (upward and downward)? Table 55 present the proportion of potentially disadvantaged that have undergone "extreme mobility": a) school Type 1 to Type 3 (upward); b) school Type 3 to 1 (downward). In order to corroborate with the initial hypothesis, potentially disadvantaged pupils should be overrepresented in the group that faces downward mobility and underrepresented in upward mobility. Does it really happen?

Table 55: Figures for Upward and Downward Mobility Considering Pupils' Transfer – Only Type 1 and 3 Schools.

	Mean Parents' Educational	Proportion Black Pupils	Proportion Poor Pupils	Proportion Boys
Upward Type1-3; 1st-3th Grade	**2.09**	**0.07**	**0.29**	**0.46**
Upward Type1-3; 1st-5th Grade	**2.06**	**0.09**	**0.30**	**0.47**
Downward Type3-1; 1st-3th Grade	**2.10**	**0.13**	**0.24**	**0.52**
Downward Type3-1; 1st-5th Grade	**2.04**	**0.15**	**0.24**	**0.53**

The figures in Table 55 should be compared with the initial description in Table 54. It is possible to observe that all variables for both transitions (1st-3rd and 1st-5th) corroborate the initial hypothesis. Pupils that have undergone upward mobility are somewhat different from their peers in low performance schools. If all the transfers were random, a similar proportion of disadvantaged pupils would be expected in both Tables. However, this is not

the case. There are less black, poor and boys in the group that have moved from low to high performance schools. There is also a difference in the mean for parents' education.

The interpretations are quite similar for pupils that have moved from high to low performance schools (downward mobility). Blacks, poor and boys are overrepresented in this particular group. The mean for parents' education is lower (2.04 versus 2.32) comparing groups in both tables. So far, the data confirm the initial hypothesis that pupil transfers across schools are not random. However, it is important to identify how many pupils actually undergo "extreme mobility". If nominal values are small, it could indicate that the potential impact on the educational system is low. Table 56 presents the nominal values and proportions for each type of transfer considering all transitions in the first segment 1st-5th grade (2006-2010).

Table 56: Pupils' Transfer from 1st to 5th grade Considering Initial Type of School (2006-2010).

	Low Performance 5th Grade 2010	Medium Performance 5th Grade 2010	High Performance 5th Grade 2010
Low Performance	**7,928**	**2,985**	**507**
1st Grade 2006	**69.5%**	**26%**	**4.5%**
Medium Performance	**1,897**	**26,185**	**2,275**
1st Grade 2006	**6%**	**86%**	**8%**
High Performance	**345**	**1,636**	**8,130**
1st Grade 2006	**3.5%**	**16%**	**80.5%**

The figures in Table 56 are important to indicate the "size" of the phenomenon analyzed. If it is true that transfers among school types 1 and 3 appear not to be random, it is also true that the total number of pupils that have undergone the so-called "extreme mobility" is very small – less than 1% considering the entire cohort (68,926). This is an important indication that perhaps the propagated selection of pupils is limited to a small number of

schools and pupils. This is important because it redefines the potential impact of pupil selection on between-school segregation considering the entire educational system.

There are two reasonable explanations for the outcomes observed. First, it should be considered that pupil selection could be a phenomenon limited to a small number of schools. Previous studies have suggested that school staff play a key role in the selection process. If true, it could be that only a small number of school principals and staff members actually engage in such practices.

Another possible explanation is that geography "prevents" or "encourages" selection bias. In a scenario where high performance schools are located far away from "sinking schools", the so-called extreme mobility would be less likely. The point here is that, if different if "types" of schools are geographically separated, the chances of pupil exchange are lower. In this case, there is a key variable (school location) which mediates the non-random exchange of pupils. This last hypothesis can be tested in future studies.

***

Another way to observe mobility is to identify all pupils that have moved from: a) school Type 1 to Type 2 or 3 (upward); b) school Type 3 to Type 2 or 1 (downward). Table 57 presents the figures for upward and downward mobility considering two transitions: 1st-3rd grade; 1st-5th grade. To interpret the data, it is important to compare its figures with those in the initial description in Table 54.

Table 57: Figures for Upward and Downward Mobility Considering Pupil' Transfers among All Types of Schools.

	Mean Parents' Educational	Proportion Black Pupils	Proportion Poor Pupils	Proportion Boys
Upward Type1-2or3; 1st -3th Grade	**1.93**	**0.13**	**0.31**	**0.50**
Upward Type1-2or3; 1st -5th Grade	**1.92**	**0.13**	**0.33**	**0.50**
Downward Type3-2or1; 1st 3th Grade	**2.18**	**0.11**	**0.22**	**0.52**
Downward Type3-2or3; 1st-5th Grade	**2.18**	**0.11**	**0.23**	**0.51**

The patterns observed previously, comparing Tables 54 and 55, are not so clear in this new approach. The only variable that apparently sustained the same pattern is parents' education. However, it is crucial to remember that this variable was originally an ordinal variable (of 5 points) transformed into a continuum variable. The differences observed are small, and it should be asked what they really mean in a real life situation. Do these differences mean anything at all in terms of the social composition of schools?

It is not clear, considering all schools and not only the top and bottom quintiles, if pupil transfers have a clear association between type of mobility and pupil profile. Data is not conclusive about the potential impact of pupil transfers across schools. What we do know is that many pupils do change school and at least a considerable number of these changes could be strictly related to parental preference for a different school – see Table 53. Nonetheless, only a very small number of transfers show a clear association between pupil profile and mobility type. Only 852 pupils from the initial cohort of 68,926, can be included in this group

that have undergone “extreme mobility”, with an association between type of mobility and pupil profile.

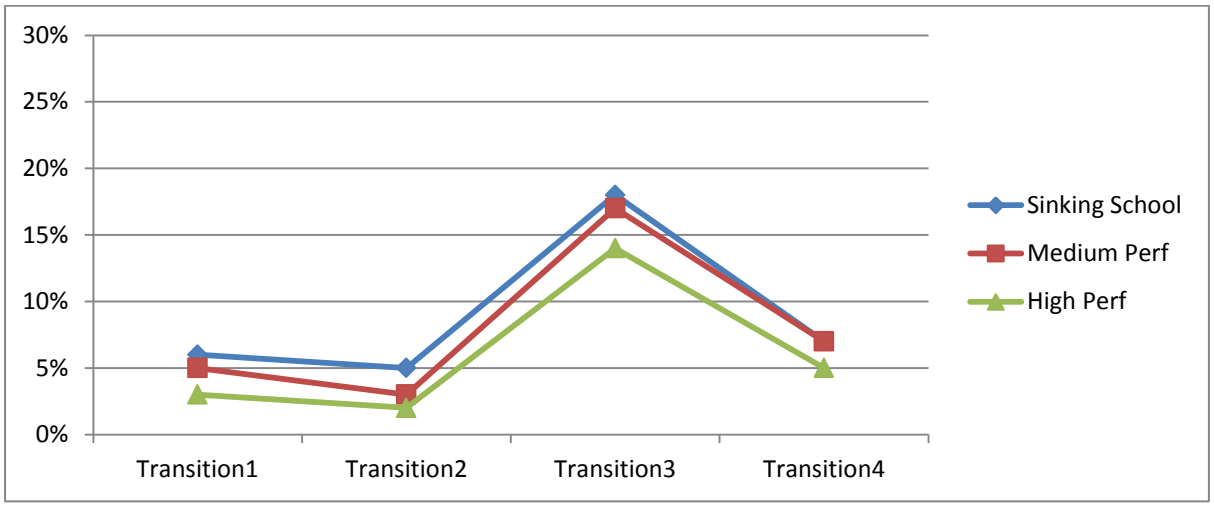
A conservative analysis of the figures in Tables 54, 55 and 57 suggests that previous studies (Bruel; Bartholo, 2012; Costa et al., 2013) might have overestimated that potential impact of pupil selection in school segregation. The arbitrary division of schools into groups, coupled with an intentional sample, are two weaknesses that should be addressed. It is not clear that the propagated selection of pupils occurs in many schools and affects many pupils. The scope of the phenomenon is relevant to assessing its potential impact on the entire educational system.

***

The model using typology of schools has the potential to inflate outcomes and mislead the researcher into making misinterpretations. The thesis proposes a different research design using the Segregation Index to measure school segregation in all educational transitions. The next section presents two different approaches for first segment pupils: a) cross-sectional; b) longitudinal. The hypothesis is that school segregation will rise in the course of different educational transitions.

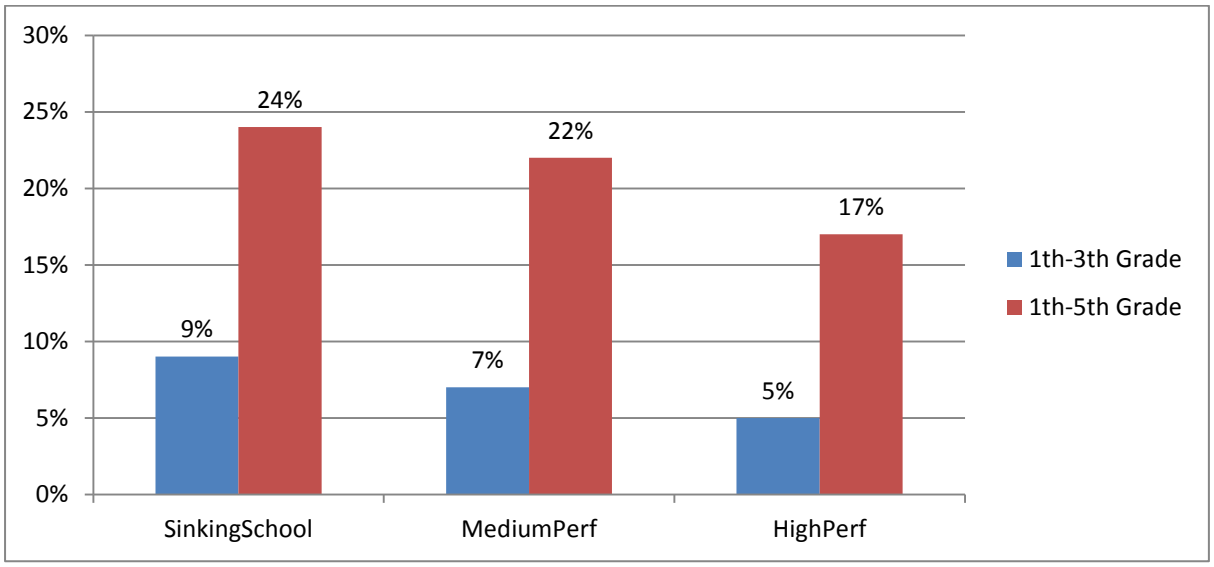
It is important to highlight that, besides the transfers across schools, there are two other elements that can influence GS and should be considered: retention and transfers of pupils to “special class” programs. Repetition is an important phenomenon in Rio de Janeiro public schools and appears to be more prevalent among low performance schools. Figure 5 indicates the exact proportion of pupil retention for each educational transition (1st-5th) grade), considering the typology of schools presented previously.

Figure 5: Proportion of Pupils Retained 1st-5th Grade – Cohort 1st Grade, 2006:



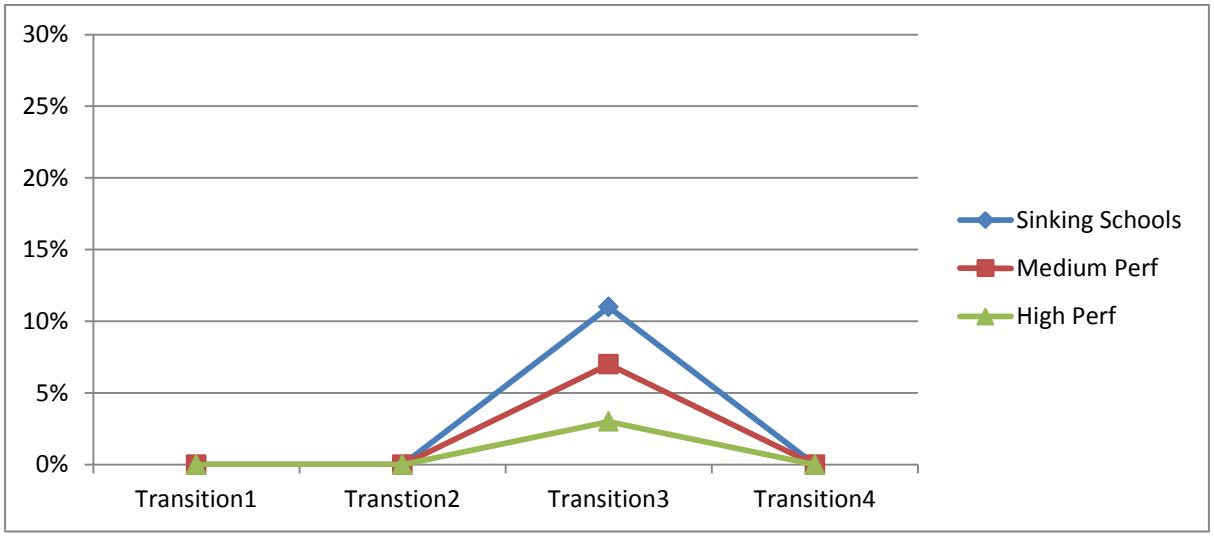
There are two important elements that should be considered. Initially, it is possible to observe that the third transition (3rd-4th grades) concentrates most of the retention in the first segment. It is the end of the first "cycle" of fundamental school (first three years) and many pupils that are still illiterate are retained. A second issue is the fact that "sinking schools" have a higher rate of retention compared to the other "types" of schools. This reaffirms that retention is: 1) a universal phenomenon (affects pupils in all different types of schools); 2) is more prevalent in low performance schools. Figure 6 shows the same data differently.

Figure 6: Proportion of Pupils Retained 1st-5th Grade – Cohort 1st Grade, 2006.



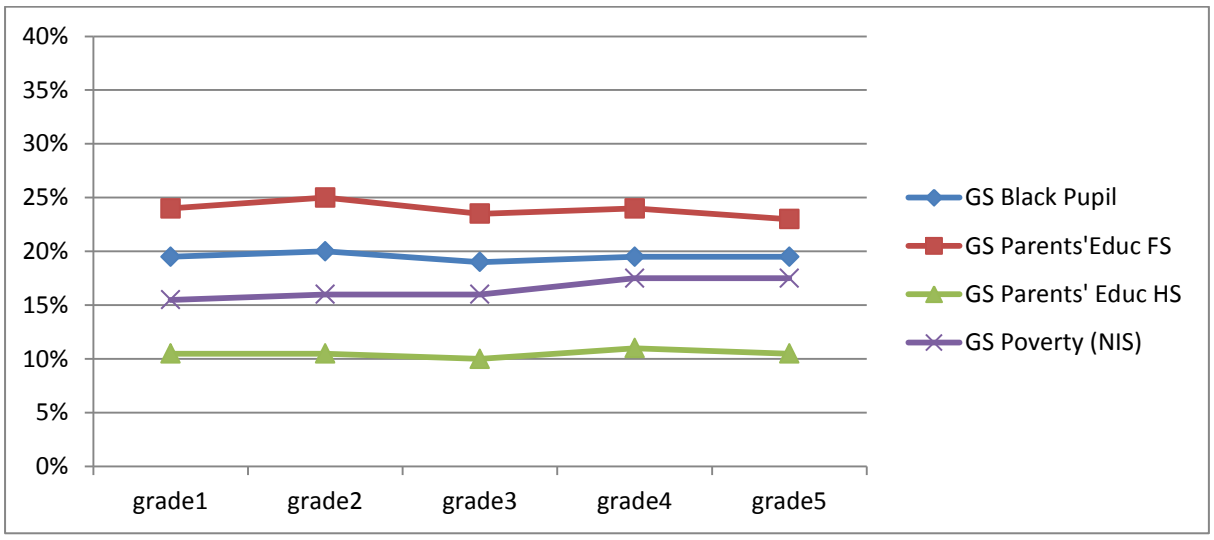
A second element that can influence GS in the cross-sectional design is the program called "*realfabetização*". This particular policy tracks pupils at age 8 (3rd grade), who have not successfully completed literacy. Around 3,000 pupils per year are sent to "*realfabetização*" classes. Figure 7 indicates the exact proportion of pupils tracked by this policy considering each "type" of school.

Figure 7: Proportion of Pupils Tracked to "*Realfabetização*" – Cohort 1st grade, 2006:



Once again, it is clear that pupils enrolled in "sinking schools" are more likely to be tracked into special class programs. "*Realfabetização*" and retention have the potential to change, in a non-linear form, the proportion of disadvantaged pupils across schools and, therefore, influence GS. If this is the case, all figures for the cross-sectional should be interpreted with caution. Figure 8 presents the nominal values of GS in a cross-sectional design for five different grades in the first segment from 2006 to 2010.

Figure 8: Cross-sectional Design for School Segregation Trends (GS) in the First Segment – 2006-2010.

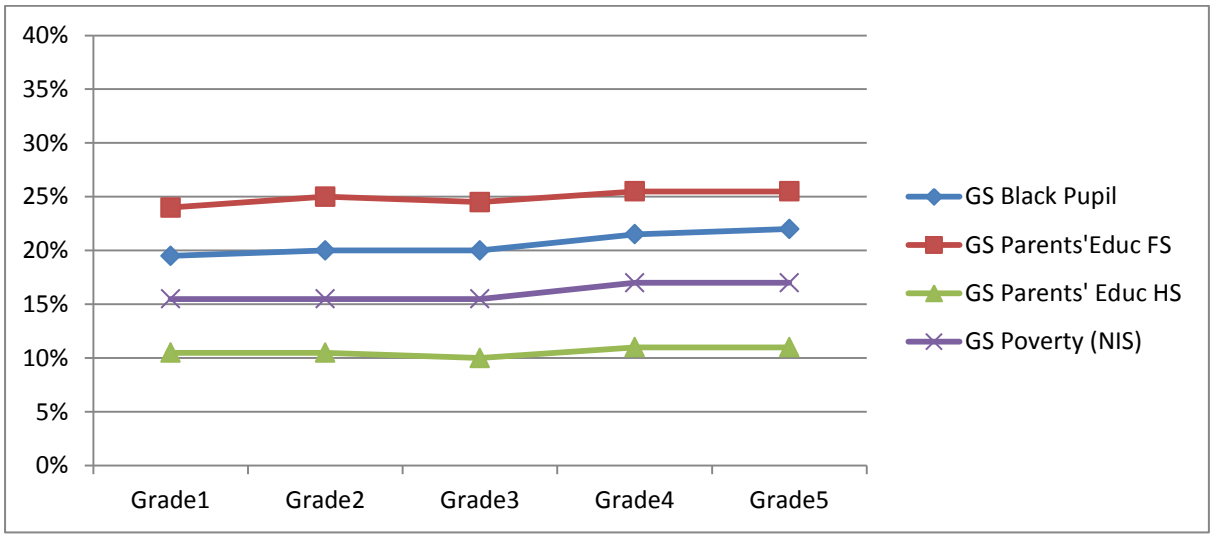


The overall patterns of segregation, considering three different indicators of potentially disadvantaged pupils, suggest stability over time. It is not clear that schools became more segregated when comparing different grades in the first segment. It is important to highlight that GS was measured considering all pupils enrolled in the specific grade for each year. Pupils in "*realfabetização*" were excluded from this particular analysis.

The nominal values of GS are being influenced not only by pupil transfers across schools (the aim here), but also by retention and tracking policies. A different design could help to understand the effects of each one of the potential causes. A longitudinal approach, measuring GS only with pupils that have successfully completed each educational transition,

could help understanding of the segregation patterns in the first segment. Figure 9 presents the nominal values of GS in a longitudinal design for five different grades in the first segment from 2006 to 2010.

Figure 9: Longitudinal Design 1 for School Segregation Trends (GS) in the First Segment – 2006-2010.

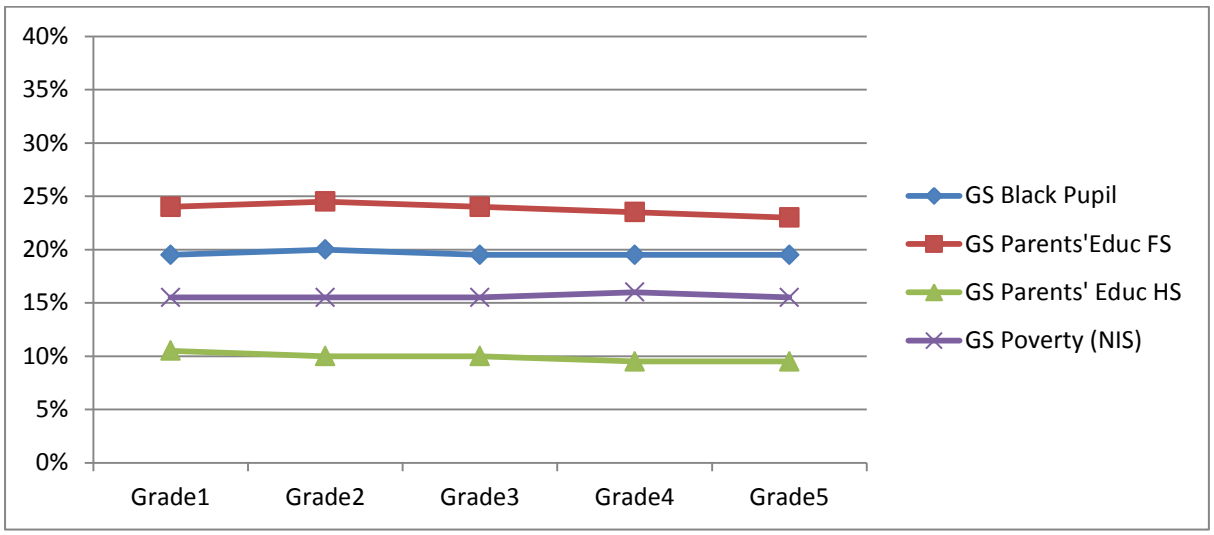


The new approach suggests a slightly different interpretation. All three variables (parents' education has been measured in two different ways) suggest a small, but constant, increase in between-school segregation, which started in the last two transitions (3rd-4th and 4th-5th). Why both designs suggest different trends? How can it be explained?

It is reasonable to assume that at least part of the positive variation in school segregation in the last two transitions of Figure 9 is influenced by retention and "*realfabetização*" – both more prevalent in the third transition. But, it is still not clear what is causing GS to grow. Is it an unwanted effect of the elimination of pupils with specific characteristics in the Longitudinal Design 1 (pupils who were retained or sent to "*realfabetização*")? Or is it really capturing a change in the social composition of schools due to transfers across them?

The last model, Longitudinal Design 2, measures GS considering all pupils in the cohort 1st grade, 2006 for all educational transitions in the first segment, regardless of their outcome. For example, pupils who were retained in any transition or were sent to "*realfabetização*" will still be "valid cases" to assess segregation. The only information that matters is that whether they have changed to a different school. This is perhaps the best design to capture the "net effect" of the voluntary transfers in the first segment.

Figure 10: Longitudinal Design 2 for School Segregation Trends (GS) in the First Segment – 2006-2010.



Once again, the overall school segregation patterns suggest stability over time. It is not clear that GS, measured in a cross-sectional design across different grades in the first segment, is influenced by pupil transfers across schools. If there are any affects at all, it is most likely related to retention and tracking policies, such as "*realfabetização*" – see Figure 9, Longitudinal Design 1.

Discussion

Data presented in the cross-sectional and longitudinal designs refute the initial hypothesis. Pupil transfers across schools apparently do not increase school segregation. This is the first time a study has analyzed the so-called "voluntary" transfers in the first segment of Rio de Janeiro public schools. It is safe to assume that at least 15% of pupils in each cohort change to a different school, very close to its initial choice (same Polo and CRE), in the initial transitions of the first segment. This is a relevant phenomenon that affects around 50,000 pupils, considering the entire population in the first segment.

Table 58 summarizes the findings from all three models highlighting the relative proportional increase in GS using the nominal values of the 5th and 1st grade. When comparing different variables, there is no clear positive pattern. In addition, the Longitudinal 2 model, probably the better to observe any potential impact of unfettered movement across schools in school segregation, does not show any clear effects for all three variables.

Table 58: GS Relative Proportional Increase for Cross-Sectional and Longitudinal Models – First Segment.

	Black Pupil	Parents' EducFS	Parents' EducHS	Poverty
Cross-Sectional	-	- 4 %	- 4 %	+ 13%
Longitudinal 1	+ 13%	+ 8%	-	+ 10%
Longitudinal 2	-	- 4 %	-10%	-

For the first time, there is data available to describe the size of the unfettered movement of pupils across schools. Although it is not clear why so many families choose a different school, it is reasonable to assume that these transfers could influence between-school segregation. However, the models presented in the thesis suggest no positive impact. There are a number of explanations for this outcome. It could be that the majority of the

transfers are indeed random, and, therefore, would not change the segregation patterns – Table 57 corroborated this assumption. The next section presents the same structure, with the same models, using data for the second segment.

7.3 "Voluntary" Transfers in the Second Segment

This section focuses on three educational transitions in the second segment of fundamental education (6th-9th grade). There are around 420 schools in the second segment, with 250,000 pupils enrolled. Table 59 shows the total proportion of pupils who have changed schools in different educational transitions in the second segment. The data is specific for one cohort 6th graders, 2006.

Table 59: Proportion of Pupils in the Second Segment Who Have Changed to a Different School – Cohort 6th grade, 2006

	2006-2007	**2006-2008**	**2006-2009**
Second Segment – Initial Grade 6th	10%	16%	18%

As highlighted in the previous section, it is important to discriminate among different reasons for the "voluntary" transfers. For the purposes of the study, only changes related to parental preference to a better school ("pedagogic reasons") are relevant. In this sense, it is possible to indicate the exact proportion of pupils who have moved to a school near their first pick: the same Educational Authority and same Polo. Table 60 presents the figures for the second segment.

Table 60: Proportion of Pupils Who Have Changed to a Different School in the Same Educational Authority and Polo – Cohort 6th grade, 2006.

	2006-2007	**2006-2008**	**2006-2009**
Second Segment – Initial Grade 6th	4 %	5 %	6 %

Around one third of all "voluntary" transfers in the second segment were to a school very close to the original choice. In the first segment, the same type of transfer corresponded to half of all the variation; in the second segment, only one third. This could be seen as an indication that older pupils have more possibility of searching for a different school, not necessarily close to their original choice. Another explanation is that many Polos will present only one or two schools for the second segment and, therefore, any school change would necessarily demand a change to a different Polo.

It should be remembered that changing to a different Polo in the same Educational Authority does not necessarily mean that pupils have moved to a school far away from their original choice. Polos, as much as Educational Authorities, are arbitrary geographical areas, and these figures should be interpreted with caution. The figures in Table 60 are most likely underestimated, and the proportion of families who have chosen a different school and did not move to a new home address is somewhere between the figures in Table 59 and 60.

***

The next step is to observe if there is any association between "type of transfer" and pupils' profile. If schools are intentionally selecting pupils based on key characteristics, the patterns of mobility (upward and downward) of disadvantaged and non-disadvantaged pupils should be different. The model for the second segment is exactly the same as the one described earlier – the first segment. Using the mean scores of *Prova Brasil* 2005-07-9, schools were divided into three different groups: a) the top quintile, called high performance schools (Type 3); b) the last quintile, low performance or "sinking schools" (Type 1); c) the group in between is termed medium performance schools (Type 2).

The first analysis tracks pupil transfers in the course of three educational transitions in the second segment. The cohort chosen was pupils in the sixth grade in 2006 – a total of

83,555 cases. The hypothesis is that pupil transfers across schools are not random, with an association between type of mobility (upward or downward) and the profile of the pupil. Table 61 presents descriptive statistics of all three types of schools.

Table 61: Descriptive Statistics for Low, Medium and High Performance School – 6th Graders 2006.

	Mean Parental Education	Proportion Black Pupils	Proportion Poor Pupils	Proportion Boys
Low Performance	**1.91**	**0.14**	**0.26**	**0.52**
Medium Performance	**2.09**	**0.12**	**0.25**	**0.53**
High Performance	**2.26**	**0.11**	**0.19**	**0.52**

The variation in the social composition of all three "types" of schools was expected. There is nothing new about this description. The real purpose of Table 61 is to serve as a "key table" to compare with the following tables that will highlight the profiles of the pupils who have undergone upward or downward mobility. Table 62 presents the proportion of potentially disadvantaged who have undergone "extreme mobility": a) school Type 1 to Type 3 (upward); b) school Type 3 to 1 (downward).

Table 62: Figures for Upward and Downward Mobility Considering Pupils' Transfer – Only Type 1 and 3 Schools.

	Mean Parental Educational	Proportion Black Pupils	Proportion Poor Pupils	Proportion Boys
Upward Type1-3; 6th-8th Grade	**2.11**	**0.14**	**0.17**	**0.47**
Upward Type1-3; 6th-9th Grade	**2.14**	**0.15**	**0.16**	**0.48**
Downward Type3-1; 6th-8th Grade	**2.05**	**0.12**	**0.11**	**0.45**
Downward Type3-1; 6th-9th Grade	**2.02**	**0.08**	**0.11**	**0.48**

The figures do not suggest a clear association between type of mobility and pupil profile. Observing the upward mobility group, black pupils are slightly overrepresented in this group (15% against 14%). The figures refute the initial hypothesis. On the other hand, poor pupils are underrepresented in the upward mobility – 16% against 26%. The variable that measures parents' education corroborates the initial hypothesis.

But what about the downward mobility group? Black and poor pupils are underrepresented in this group refuting the initial hypothesis. The only variable that could suggest a selection bias is parent's education. However, the limitations in the difference should be remembered. Originally, parents' education was a categorical variable that was later transformed into a continuum variable. It is not clear what a 0.24 (2.26 – 2.02) difference means.

The figures in Table 63 indicate mobility considering all "types" of schools: a) school Type 1 to Type 2 or 3 (upward); b) school Type 3 to Type 2 or 1 (downward). To interpret the data, it is important to compare their figures with those of initial description in Table 61.

Table 63: Figures for Upward and Downward Mobility Considering Pupil' Transfers among All Types of Schools.

	Mean Parental Educational	Proportion Black Pupils	Proportion Poor Pupils	Proportion Boys
Upward Type1-2or3; 6th -8th Grade	**2.06**	**0.14**	**0.19**	**0.50**
Upward Type1-2or3; 6th-9th Grade	**2.09**	**0.13**	**0.19**	**0.48**
Downward Type3-2or1; 6th 8th Grade	**2.14**	**0.13**	**0.12**	**0.51**
Downward Type3-2or3; 6th-9th Grade	**2.17**	**0.12**	**0.12**	**0.51**

Again, not all variables show a clear association between type of mobility and pupil profile. Surprisingly, the model including Medium Performance schools (Type 2) shows a more defined pattern, with two variables – parents' educational and black pupils – corroborating the initial hypothesis. Poor pupils are underrepresented in both groups of mobility type. In order to corroborate the initial hypothesis, they should only be underrepresented in the upper mobility group. The data from Tables 62 and 63 are not very conclusive.

It is also important to identify how many pupils actually undergo mobility. If nominal values are small, it could suggest a minor impact on the educational system. Table 64 presents the nominal values and proportions for each type of transfer, considering all transitions in the second segment, 6th-9th grade (2006-2009).

Table 64: Pupils' Transfer from 6th to 9th grade Considering Initial Type of School (2006-2010).

	Low Performance 9th Grade 2009	MediumPerformance 9th Grade 2009	High Performance 9th Grade 2009
Low Performance	**9.393**	**1.497**	**286**
6th Grade 2006	**84%**	**13%**	**3%**
Medium Performance	**1.060**	**31.414**	**1.497**
6th Grade 2006	**3%**	**93%**	**4%**
High Performance	**228**	**1.276**	**10.952**
6th Grade 2006	**2%**	**10%**	**88%**

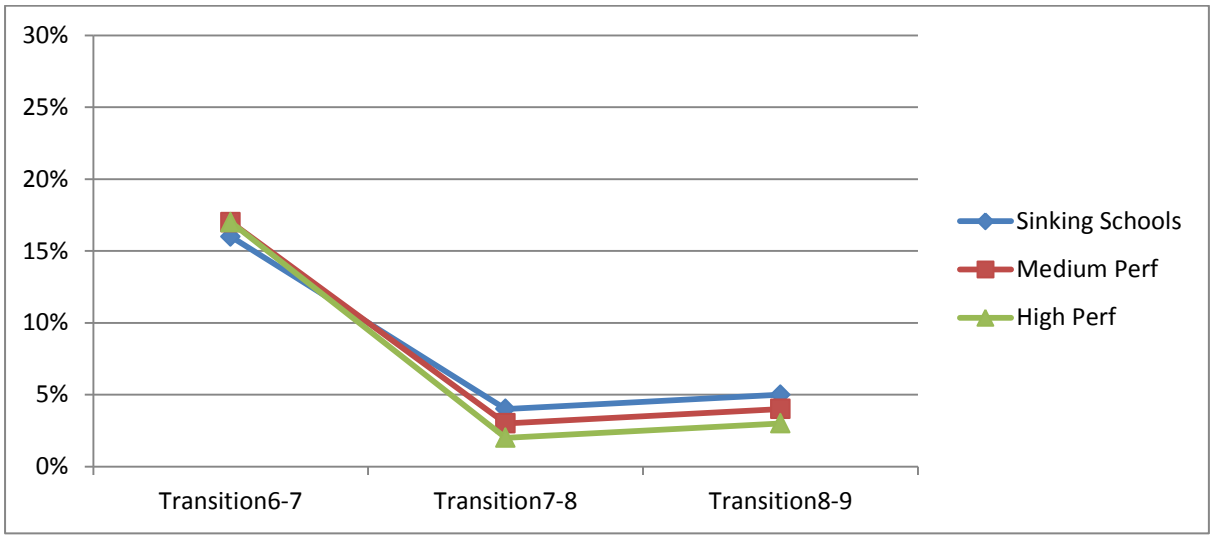
The total proportion of pupils who have undergone "extreme mobility" is very low – around 1%. It is also interesting that not many pupils enrolled in Medium Performance schools end up moving to different "types" of schools. Apparently, it is harder to ascend or descend in the second segment compared to the figures for the first segment.

***

The final models in this section measure school segregation using GS in all educational transitions of the second segment. Once again two different research designs were used: a) cross-sectional; b) longitudinal. The hypothesis is that school segregation will rise in the course of different educational transitions.

It is important to consider two other elements, besides the school transfers, that can influence GS: retention and transfers of pupils to "Special Class" programs. Both have the potential to alter the social composition of schools, because they are more likely to recruit potentially disadvantaged pupils. Figure 11 indicates the exact proportion of pupil retention for each educational transition, considering the typology of schools presented in this chapter.

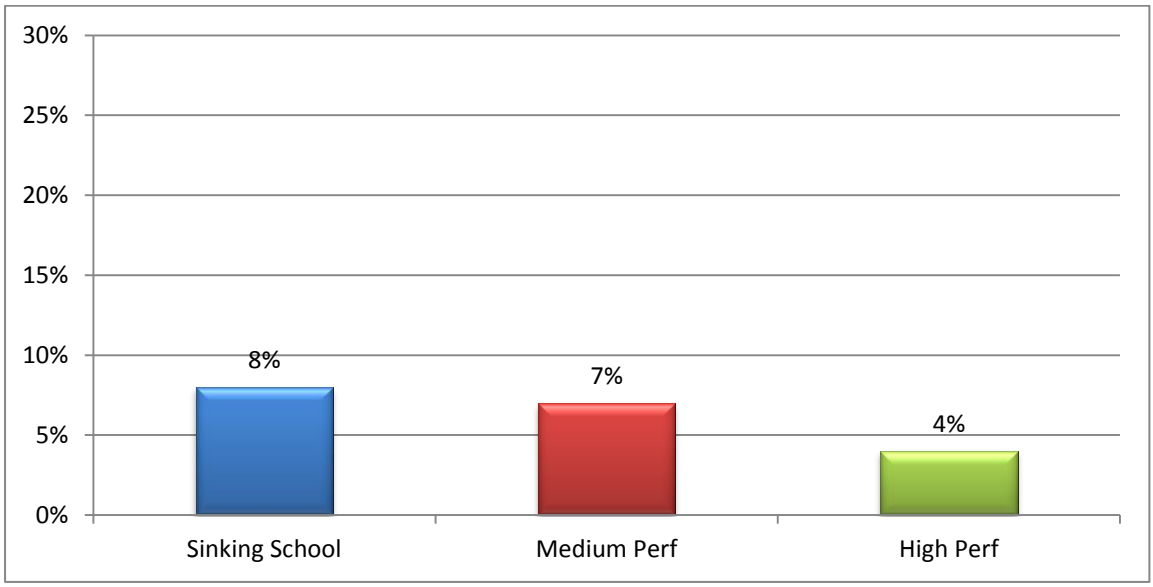
Figure 11: Proportion of Pupils Retained 6th-9th Grade – Cohort 6th Grade, 2006:



It is clear that the first transition in the second segment (6th-7th grades) concentrates most of the retention. This is an important element that should be further investigated regarding the causes and impacts of such a phenomenon. What happens in the 6th grade that explains such high retention rates? Another important element is that retention is more evenly distributed across different "types" of schools. This is an important difference when compared to the first segment data – see Figure 5.

What about the proportion of pupils sent to a "Special Class"? Differently from the pattern observed in the first segment, pupils are tracked during all educational transitions. The main difference is in the total proportion of pupils considering each "type" of school. Pupils enrolled in "sinking schools" are more likely to be sent to these special classes.

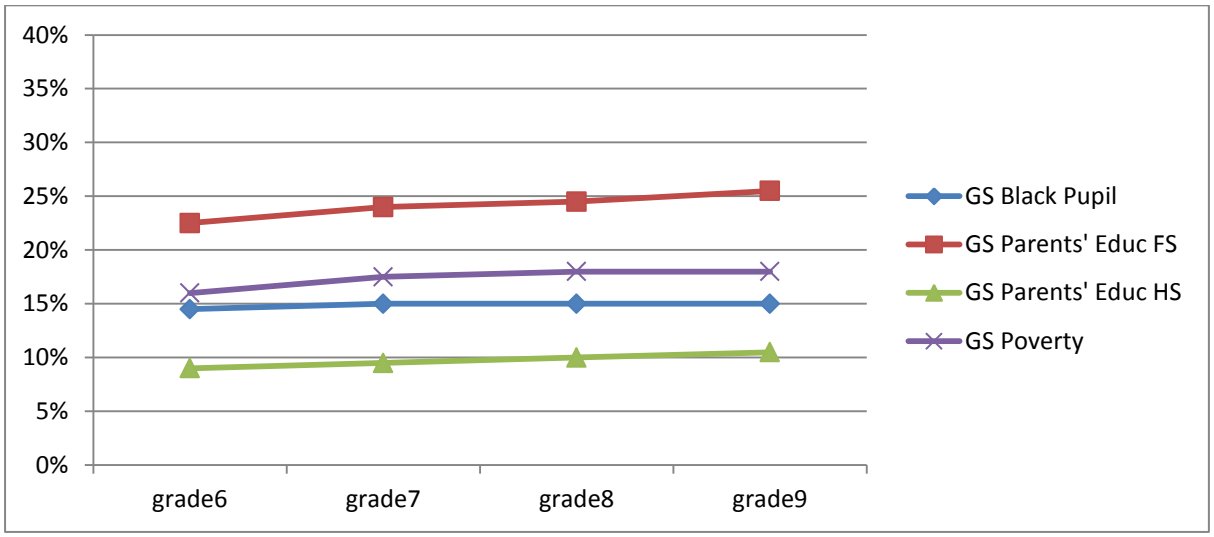
Figure 12: Proportion of Pupils Tracked to Special Class – Cohort 6th grade, 2006:



***

The cross-sectional design considers all pupils enrolled in four different grades (6th-9th) of the second segment in four consecutive years (2006-2009). The overall patterns for all four GS measurements suggest a small, but constant, increase in school segregation. It is important to highlight that GS in Figure 13 was measured considering all pupils enrolled in the specific grade for each year. Pupils in "PEJA I or II" were excluded from this particular analysis.

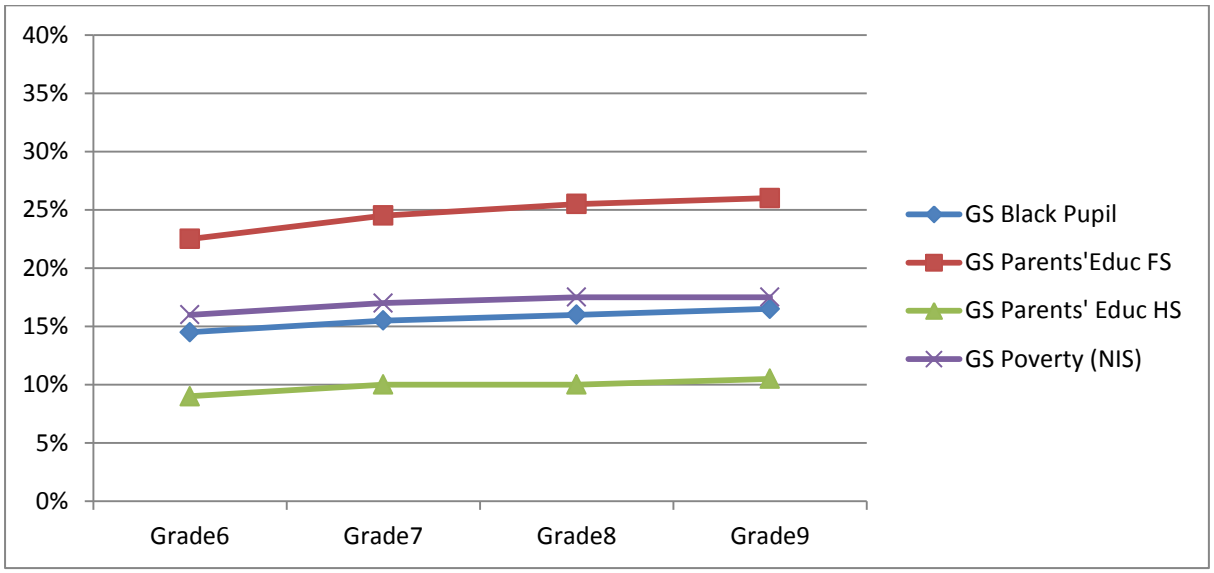
Figure 13: Cross-sectional Design for School Segregation Trends (GS) in the Second Segment – 2006-2009.



The fact that all variables indicate an increase in between-school segregation is an important element. Nevertheless, all changes in the nominal GS values are influenced not only by pupil transfers across schools, but also by retention and tracking policies. A longitudinal design could help to identify the effects of each one of the alternative explanations.

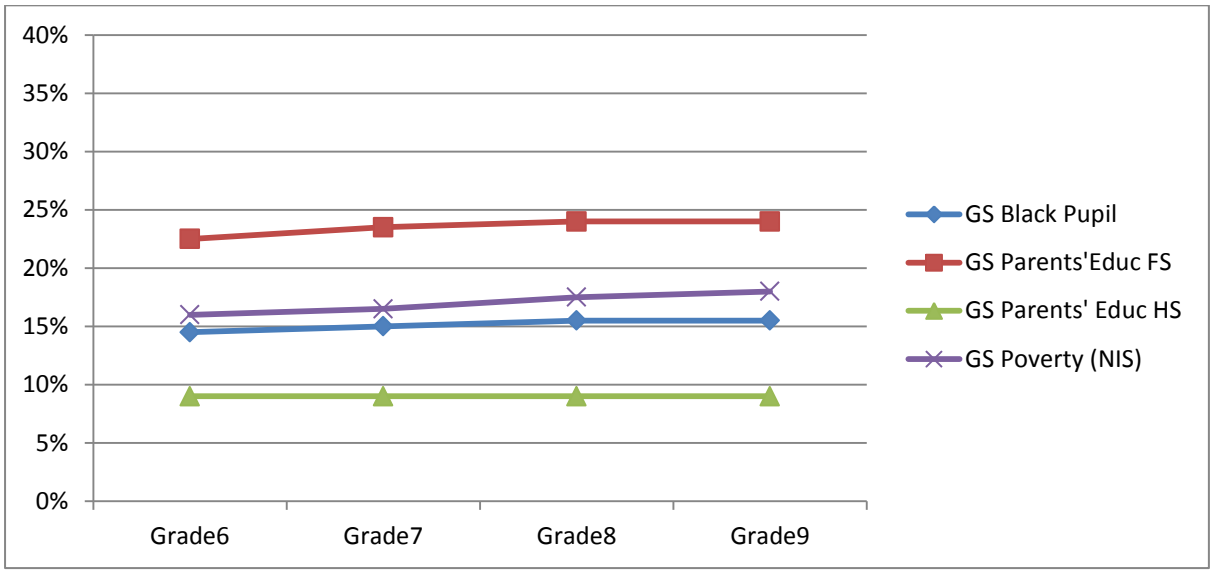
A longitudinal approach, measuring GS only with pupils that have successfully completed each educational transition is presented in Figure 14. This is probably the least conservative model, because it most often tends to exclude potentially disadvantaged pupils. Again, all three variables (parents' education has been measured in two different ways) suggest a constant increase in between-school segregation. The slop is even higher in this model. It should be considered that at least a part of the positive variation in GS is influenced by retention and the "Special Class" policy. Nonetheless, it should be mentioned that both models (Cross-Sectional and Longitudinal 1) suggest the same interpretations: 1) segregation rises during different educational transitions in the second segment; b) at least part of the variation observed could be explained by pupil transfers across schools.

Figure 14: Longitudinal Design 1 for School Segregation Trends (GS) in the Second Segment – 2006-2009.



The last design is probably the most conservative to measure any effect of pupil transfers on GS. Longitudinal Design 2 measures segregation considering all pupils in the cohort 6th grade, 2006 for all educational transitions in the second segment, regardless of their "outcome". For example, pupils who were retained in any transition or sent to "PEJA I or II" will still be "valid cases" to assess segregation. The only information that matters is if they have changed school. Figure 15 presents the calculations of GS for all three variables.

Figure 15: Longitudinal Design 2 for School Segregation Trends (GS) in the Second Segment – 2006-2009.



The overall patterns suggest a very modest, constant increase in school segregation. The only exception is GS EducHS, which indicates stability over time. From all three models, Longitudinal 2 shows the lowest increase for all three variables. Nonetheless, the interpretation did not change: segregation rises during different educational transitions in the second segment.

Discussion

This is the first time a study has analyzed the so-called "voluntary" transfers in the second segment of Rio de Janeiro public schools. Data presented in the cross-sectional and longitudinal designs corroborate the initial hypothesis. Pupil transfers across schools apparently increase school segregation.

The total number of pupils who have changed school is quite significant, and the phenomenon should be further investigated with great caution. The lack of transparency and clear legislation to regulate the "voluntary" transfers should raise concern among policy-makers, researchers and members of the general public.

Table 65 summarizes the findings from all three models, highlighting the relative proportional increase in GS using the nominal values for the 9th and 6th grades. There is a clear pattern suggesting a small, but positive, impact on school segregation. Even the most conservative model, Longitudinal Design 2, suggests some positive impact of pupil transfers.

Table 65: GS Relative Proportional Increase for Cross-Sectional and Longitudinal Models – First Segment.

	Black Pupil	Parents' EducFS	Parents' EducHS	Poverty
Cross-Section	+ 3%	+ 13%	+ 22%	+ 13%
Longitudinal 1	+14%	+13%	+ 22%	+ 9 %
Longitudinal 2	+ 7%	+ 4 %	-	+ 13%

The outcomes for the "voluntary" transfers were different comparing the first and second segments. The total proportion of pupils who have moved to a different school in the first segment was higher. However, the models did not suggest any obvious impact on between-school segregation. For the second segment, despite the fact that a smaller

proportion of pupils actually changed to a different school, all three models suggest a positive impact on school segregation.

The data presented here is unprecedented. For the first time, a study has followed two entire cohorts, used different designs and measured any changes in the overall levels of segregation. The data reinforces the role of educational policies in school segregation. Studies in the U.S. have also shown that the unfettered movement of pupils across schools can have an impact on school segregation (Saporito, 2003). Society should be informed if schools are becoming more segregated in the course of different educational transitions. Admission criteria for first enrolment or regulation for pupil transfers are more than a simple "academic subject". Empirical data from different educational systems show that they can directly affect pupil's educational opportunities and prospects in life. The reasons for the cluster and its potential impact should be of public interest.

7.4 Mandatory Transition – *Remanejamento*

This section revisits one crucial question about the potential impact of the only mandatory transition in Fundamental School called "*remanejamento*": what is the overall effect of the mandatory transfer on school segregation? Future studies should also address a complementary question: Is there any evidence suggesting that pupils are being selected based on key characteristic? Or, in different terms, how far from a random distribution is the allocation of pupil in the only mandatory transition in Fundamental School?

Two studies have analysed this exact same transition and both conclusions were very similar: 1) pupils were being selected based on key characteristics; 2) schools became more homogeneous after "*remanejamento*" – another way to state that the general effect was an increased in between-school segregation (Costa et. al 2013; Bruel; Bartholo, 2012).[17] The thesis proposes a new research design and challenges some conclusions made so far. Perhaps, the main objection to previous interpretations indicates that it is not true that segregation rises after the mandatory transition. As a matter of fact, data shows a consistent decline in segregation among schools after the allocation in "remanejamento". Another relevant finding suggests a change in school segregation patterns (after "*remanejamento*"), with an important increase in within-school segregation – the so called "shift effect" or "shift" allocation.

In order to make a causal claim about any impact of "*remanejamento*" on school segregation, it is necessary two stable measures of D or GS, before and after the particular event. One strategy would be to simply compare the overall levels of GS or D in the first and second segment. This design would give stable measures before and after "*remanejamento*", but it has weaknesses. Saying it different, there are alternative hypothesis that could explain any difference observed.

[17] "*Os resultados indicam que a transferência de alunos no período de transição analisado não é aleatória e cria um cenário de homogeneização das escolas investigadas*" (Bruel; Bartholo, 2012, p. 492).

Last chapter showed that "voluntary transfers" (especially in the second segment) have an impact on school segregation and, therefore, could interfere with the outcome. A better approach would be to simply compare 5th and 6th grades figures – respectively prior and after "*remanejamento*". A cross-sectional design will measure segregation in two different cohorts (2008-2009; 2009-2010) to observe if results are stable. If previous studies are correct, the nominal values of GS and D on 6th grade will be higher compared with 5th grade. A complementary design tracks pupils from one specific cohort in a longitudinal approach. The cohort chosen for this model was the 5th grade, 2009. The plan is to follow all the pupils who have passed to the 6th grade in the following year (2010). From a total of 61.728 pupils enrolled in the 5th grade 2009, only 50.049 made the transition. Around 11% were retained at 5th grade and others left the municipal public network to private or federal network. The design presents two different measurements of school segregation using the 50.049 cases: a) GS for the 5th graders, 2009; b) GS for the 6th graders, 2010.

Cross-Sectional and Longitudinal Designs: Overall Impact of "*Remanejamento*"

The first design aims to answer one simple question: What is the overall effect of the mandatory transfer on between-school segregation? This is perhaps the most relevant question regarding the potential impact of such transition. Nonetheless, most of the research in educational usually ignores questions like this. Study cases or other research with poor design only observe a limited number of cases, invariably with intentional sample that increases the risk of selection bias and enables any overall conclusions for the educational system.

The thesis presents data for the entire population in two specific cohorts: 1) 2009 5th grade – 2010 6th grade; 2) 2008 5th grade – 2009 6th grade. The replication here serves the purpose of observing stability over time. Table 66 presents the figures of GS and D for the cohort 2009 5th grade – 2010 6th grade. The results are quite clear: the overall effect is a decline in nominal levels of segregation. As expected, GS and D showed a very similar answer for this specific transition.

Table 66: Segregation Index (%) and Dissimilarity Index (D) for the 5th grade 2009 and 6th Grade 2010.

	5th Grade 2009	6th Grade 2010	Relative % Decline
GS Black Pupil	19	14	26 %
D Black Pupil	21	16	24 %
GS EducFS	24	18	25%
D EducFS	31,5	24	24 %
GS Poverty	19	16,5	13 %
D Poverty	26,5	22,5	15 %
GS Distortion 2	26	14	46 %
D Distortion 2	30,5	20	34 %

The initial results contradict previous interpretations for this specific educational transition. All four variables indicate a decline in GS and D nominal values. Depending on the variable observed, the relative proportional decline can be higher than 30% – see age/grade distortion for two or more years. Will the outcomes be consistent analysing a different cohort? Table 67 shows the figures of GS for the cohort 2008-2009.

Table 67: Segregation Index for the 5th grade 2008 and 6th Grade 2009.

	5th Grade 2008	6th Grade 2009	Relative % Decline
GS Black Pupil	18	14	22%
GS EducFS	24	19	21%
GS Poverty	17	16	6%
GS Distortion 2	18,5	13	30%

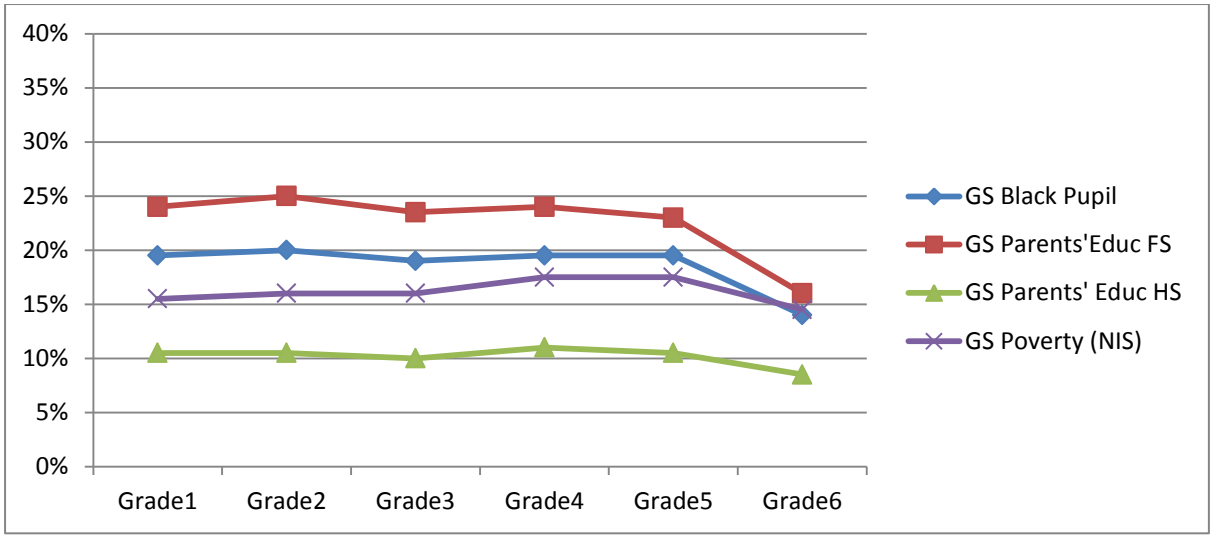
Overall, the results from both cohorts are very similar. The figures suggest a decline in nominal levels of school segregation. Comparing the characteristic of disadvantaged it is possible to observe a pattern in the decline. Poor families indicate the smallest relative proportional decline, almost imperceptible in the cohort 2008-2009, for both cohorts. On the other hand, distortion age/grade of two or more years shows the biggest decline in both cohorts.

It is important to remember that longitudinal analysis of distortion age/grade (especially those that track pupils) should always be analysed with bigger caution, because of its variability. Pupils do not change category for their colour (or at least it is not expected that they do), but many pupils enter the category of distortion along the years. Different changes in the educational legislation (as the "automatic approval", mentioned previously) can affect the total proportion of disadvantaged pupils for distortion age/grade. For the purpose of the

analysis in this section, the most important element is the consistency of the decline after "*remanejamento*" for all four variables.

Another way to observe the same outcome is to measure GS or D in all educational transitions of the first segment, plus the 6th grade, in a cross-sectional design. The index was calculated considering 6 different years 2006-2011.

Figure 16: Cross-sectional Design for School Segregation Trends (GS) from the 1st to 6th grade (2006-2011).



It becomes clear that "*remanejamento*" represents a disruption in a fairly stable trend along "voluntary" transitions in the first segment. How is it possible to interpret such outcome? More importantly, how different models – this one and others presented by Costa et al. (2013) and Bruel and Bartholo (2012) – suggest opposite outcomes?

The figures for the longitudinal approach, following one specific cohort, corroborate the initial findings from the cross-sectional design. Table 68 presents the GS figures for the cohort 5th Grade 2009, considering only pupils who passed to the 6th grade in 2010 – total of 50.049 cases.

Table 68: Segregation Index (%) in a Longitudinal Design – 5th Grade 2009; 6th Grade 2010.

	5th Grade 2009	6th Grade 2010	Relative % Decline
GS Black Pupil	20,5	15,5	24%
GS EducFS	25	19	24%
GS Poverty	20	16,5	17%
GS Distortion 2	29,5	20	32%

The nominal values of GS show minor differences compared with figures from the cross-sectional approach – see Table 66. The overall effect of "*remanejamento*" in school segregation is the same. Once again, it is possible to observe an important decline in segregation for all characteristics of disadvantaged.

Discussion

The limitations of previous studies have been discussed. The designs presented here have the advantaged of using data for the entire population and measured segregation with robust indices. It was possible to observe that the segregation patterns along "voluntary transfers" were stable and any big change should be interpreted with caution.

Does it really mean that the allocation in "remanejamento" is intentionally aiming to desegregate the educational system? Is the hypothesis of pupil's selection wrong? The answer for both questions is most likely no. The hypothesis about school selection bias will be tested in the next section. Nonetheless, it is still relevant to understand such an important decline in segregation. How can it be explain?

The answer is simple and was not envisaged at all by prior studies. "Remanejamento" is not only a period of mandatory transition for more than 90% of pupils. It also reorganizes the educational system, affecting the size of schools and, consequently, the total number of schools. On average, there is a 45% decline in the total number of schools, which increases the chances of schools becoming more mix. The explanation for the sudden decline in school segregation is most likely related to mathematical probability and not to any other element related to educational bureaucracy. There is empirical evidence from different educational systems suggesting that a sudden decline in the total number of schools (for example, policies to close out schools) tends to bring segregation down (Gorard; See, 2013).

***

Before moving to the next section, it is necessary to understand how two distinct models gave opposite answers observing the exact same transition for the same cohort. The models presented in the thesis strongly suggest that the overall effect of "*remanejamento*" is a

decline in school segregation. The motivation here is understand how previous studies miss this main result.

The main concern in previous studies was to analyse pupil's chance to access to the so called "high performance" schools at 6th grade in Rio municipal public schools. Usually, schools were divided into two groups: top quartile as the high performance and the rest classified as low performance. The hierarchy was made using standardized tests from *Prova Brasil* – the same ones used previously in the thesis.

A logistic regression was used to estimate pupil's chance to access the high performance schools. Most of the independent variables in the regression were quite similar to the ones used in the thesis: pupil's colour, age/grade distortion; parent's education. The pupil current school at 5th grade was also used as covariant in the regression. The coefficients were interpreted in a way that highlighted that specific characteristics could increase or lowered pupil's chance to access a high performance school.

The model itself has its logic. If coefficients were significant, indicating, for example, that having one or more years of age/grade distortion lowered someone's chance to access a high performance school. It was reasonable to assume that schools were selecting pupils based on key characteristic. In reality, that was exactly what those papers founded. So where is the problem?

To start, there is a basic limitation with the model itself. None of the papers can tell anything about decline or increase in segregation, basically because they did not measure. They had no prior or after measure of segregation in relation to the transition of interest. It was not possible to establish any causal claims about potential impacts of "*remanejamento*".

A second problem is related to the interpretation of the regression model. Coefficients were considered "valid" in the models if they were significant at .05 or less. Nonetheless, it is important to remember that only a relative small proportion of pupils actually accomplished

mobility. Bruel and Bartholo (2012) shows that around 22% of their sample of 2.219 pupils moved from "types" of schools – low and high performance. Observing the entire cohort (2008-2009 the same in the paper), the figures suggest a very similar number – 20% of mobility. Not only a small fraction of pupils actually accomplish mobility, but also the correlation between type of mobility and pupil's profile might not be that clear.

Maybe that is why the "Model Summary" of the logistic regressions models indicates small figures for the Nagelkerke R Square – between 0.05 and 0.1. Perhaps the model is not very appropriate to estimate pupil's mobility. One question should be answered: if changing the cut-off point to separate schools into groups (for example, two equal groups), would the general interpretation change?

8. OPEN ENROLMENT SYSTEM IN RIO DE JANEIRO PUBLIC SCHOOLS: ANALYSING THE "HIDDEN-QUASI-MARKET"

Previous studies in Rio de Janeiro public municipal schools suggested parental competition for the most prestigious schools and the pupil selection based on specific criteria by the school staff. The phenomenon was characterized as "Hidden-Quasi-Market" (Costa; Koslinski, 2008; 2011; 2012) and is based on the assumption that: a) school intakes are correlated to school reputation and; b) the potential to select pupils is also correlated to school reputation. A few analyses were made utilizing data from the Municipal Educational Department and showed that the transfer of pupils across schools was not random and had the potential to increase between-school segregation (Bruel; Bartholo, 2012; Costa et al., 2013).

One of the main concerns from previous studies was related to the treat of intentional pupil selection. The thesis will explore the same subject presenting a different research design to answer one question: Is demand for school correlated with between-school segregation? The logic behind is simple. Since schools with oversubscription have the potential to select pupils, it is reasonable to assume that: a) schools with very high demand will potentially select pupils and, therefore, present a lower share of disadvantaged pupils; b) areas with greater school choice (high demand and high supply) will present higher levels of segregation. This is a new approach to produce evidence about elements that can be associated with school segregation in Rio de Janeiro.

This section presents two independent models: comparison among schools and comparison among Educational Authorities. A total of three Educational Authorities were selected based on an initial description about key characteristics that can influence between-school segregation. In this section, all analysis will only consider pupils enrolled in the first segment.

The first model considers each school as a unit of analysis and correlates the parental demand for schools with the Segregation Ratio (SR) – an index that provides individual rates for each school. To support the initial hypothesis, the correlation between parental demand and SR should be negative, strong and constant among different characteristics of disadvantaged (parents' education, poverty, age/grade distortion and colour). This model will be replicated in two ways: a) considering all public schools as a unit of analysis; b) comparing three different Educational Authorities. If the hypothesis is correct, the strength of the association for each educational authority should be different. More urban areas, with higher levels of parental dispute for schools, should present a higher correlation compared with less urban areas, with precarious public transportation, which presumably presents lower levels of parental dispute for schools.

The variable for parental demand was constructed using a specific dataset that presents a list of up to five schools for each pupil. This list is filled by the family and used in the enrolment process. The intention here is not to observe the school the child was actually placed, but the family's first choice. If it is true that schools have different reputations, it would be expected that schools with higher reputations (can also be called high performance schools) would present higher parental demand. A continuum variable was constructed aggregating all parental demand for each specific school for the year 2010.

The second model assesses segregation using GS and compares the overall levels of school segregation considering three educational authorities (the same ones used in the previous model). In order to corroborate with the initial hypothesis, the nominal levels of segregation should fluctuate in a pre-determine way. Basically, areas with greater school choice (high demand and high supply of schools) should present higher nominal levels of segregation.

The designs have two major advantages from prior studies. The selection of different Educational Authorities, with very distinct characteristic, allows broader understanding about different elements associated with school segregation. Both indices of segregation (GS and SR) allow a new approach to understand the role of the territory and educational legislation on school segregation patterns.

8.1 Describing the Educational Authorities

The research design compares three different Educational Authorities to test the hypothesis about the association between demand for school and overall levels of segregation. None of the models presented are a definitive test about the existence of pupil selection. The areas were chosen based on an initial description about key characteristics that can influence between-school segregation: a) Demand and supply of schools; b) Proportion of potentially disadvantage pupils; c) Mean and Standard Deviation in *Prova Brasil* scores – years 2005-07-09.

Since the idea is to test the association between parental demand for schools and school segregation, the sampling considered areas that presumably present very different levels of parental dispute for schools. Table 69 shows descriptive statistics from all Educational Authorities considering pupils enrolled in first segment – 1st to 5th grades.

Table 69: Descriptive Statistics for All Educational Authorities – First Segment 2010.

	Mean PBP 4th Grade	**SD PBP 4th Grade**	**Proport Black**	**Proport NIS**	**Proport EducFS**	**Proport Distor2**
CRE 1	5,10	0.45	0.11	0.26	0.35	0.13
CRE 2	**5,42**	**0.60**	**0.12**	**0.18**	**0.28**	**0.12**
CRE 3	5.22	0.51	0.12	0.28	0.27	0.13
CRE 4	**5.10**	**0.46**	**0.10**	**0.31**	**0.28**	**0.12**
CRE 5	5.13	0.51	0.12	0.26	0.18	0.10
CRE 6	4.97	0.47	0.11	0.28	0.27	0.10
CRE 7	5.36	0.45	0.10	0.19	0.31	0.12
CRE 8	5.09	0.44	0.11	0.31	0.22	0.09
CRE 9	5.15	0.36	0.09	0.31	0.22	0.09
CRE 10	**4.85**	**0.31**	**0.10**	**0.34**	**0.25**	**0.10**

Additional information was withdrawn from Alves, Lange and Bonamino 2010. The authors highlight that the city of Rio de Janeiro is quite balanced in terms of demand and supply of schools. There are not clear areas with high demand and low supply or vice-versa. What is clear is that some areas have high demand and high supply and others low demand and low supply.

Based on the information, three Educational Authorities were chosen: CRE 2; CRE 4; CRE 10. Since the legislation is a constant among all areas, it is reasonable to assume that any differences in the overall levels of segregation are most likely influenced by different patterns of residential segregation, proportion of potentially disadvantaged pupils and different levels of parental dispute.

Educational Authority number 2 (CRE 2) is widely known to be a highly dense, urban area, with large supply of schools in some of the most affluent neighbourhoods traversed by shantytowns. It also presents good public transportation network that in theory increases mobility and makes it easier to reach the area. CRE 2 presents the highest levels for mean and SD of *Prova Brasil*. If the hypothesis is correct, this should be the area with the highest levels of between-school segregation.

Educational Authority 10 (CRE 10) seems to be the opposite, with the lowest levels for mean and standard deviation (SD) of Test Brazil in a very low dense area. There are no rural areas in the metropolitan area of the city of Rio de Janeiro, but it is reasonable to assume that the parental dispute for schools in CRE 10 is lower, comparing to CRE 2, mainly due to the long distances between schools and a precarious public transportation. Parents that live in CRE 2 have more options of public schools in a walking distance than families living in CRE 10. This scenario increases parents' "real option".

The last Educational Authority is number 4 (CRE 4), which is "ranked" in the middle of CRE 2 and CRE 10. If the model is correct, the levels of segregation should follow a

gradual increase starting in CRE 10, followed by CRE 4 and ending in CRE 2, supposedly the most segregated area. Does the data support the hypothesis?

8.2 Is Parental Demand Correlated with the Segregation Ratio?

The first model tests the association between parental demand for schools and the Segregation Ratio. Why is this important? Or how can the correlations be interpreted? Since parents have freedom of choice and the schools present different performance/reputation, it is reasonable to presume that parents will choose schools that they perceive as "better". Table 70 presents the correlations coefficients for parental demand and the mean score of *Prova Brasil* for each Educational Authority for the year 2010. The mean score in a standardized test is a measurement of performance and also a proxy of reputation.

Table 70: Correlations between Parental Demand and Mean Score of *Prova Brasil* – First Segment 2010.

	PBP 4th Grade
CRE1 Parental Demand	+ 0.29
CRE2 Parental Demand	**+ 0.49****
CRE3 Parental Demand	+ 0.20
CRE4 Parental Demand	**+ 0.28****
CRE5 Parental Demand	**+ 0.32****
CRE6 Parental Demand	**+ 0.49****
CRE7 Parental Demand	**+ 0.33****
CRE8 Parental Demand	**+ 0.37****
CRE9 Parental Demand	+ 0.03
CRE10 Parental Demand	**+ 0.36****

** Correlation is significant at the 0.01 level (2-tailed).

The values support the initial claim that schools with better performance present higher parental demand. All coefficients are positive and seven figures are significant at 0.01

level – including CREs 2, 4 and 10. It is also logic that schools with greater parental demand are more likely to face the event of oversubscription and, thereby, have the option to select their pupils. If the hypothesis is correct, parental demand for school and the Segregation Ratio (SR) should be negative correlated for all indicators of potential disadvantaged pupils. Table 71 shows the figures of the correlations considering the entire network for the year 2010.

Table 71: Correlations between Parental Demand for Schools and the Segregation Ratio for All Characteristics of Disadvantaged – First Segment 2010.

	White Pupil	Black Pupil	EducFS	EducHS	NIS	Distort1	Distort2
Parental Choice All Schools	**-0.23****	**-0.27****	**-0.16****	**-0.18****	**-0.12****	**-0.12****	**-0.22****

** Correlation is significant at the 0.01 level (2-tailed).

The figures show that all indicators of SR are negative associated with parental demand at the level of 0.01. Pupil colour and Distortion 2 presents the strongest correlations. The data suggests that schools with higher parental demand present a lower proportion of potentially disadvantaged pupils than its expected "fair share". How can we interpret Table 71? One possible explanation is that schools with higher demand are systematically selecting pupils based on key characteristic. The main limitation/problem here is that there is no clear information about oversubscription and, so far, the best "proxy" is the "pure" parental demand. An alternative explanation would be that the schools with good reputation/high performance are clustered in a few neighbourhoods and, therefore, the associations observed in Table 70 and 71 would be influenced by the residential segregation and not necessarily by selection bias mechanism. A third plausible explanation would be that different profile of parents chooses different "types" of schools. This could happen mainly due to inequality of access to information about schools or a self-selection mechanism.

Is it possible to rule all alternative explanations? The first one – the influence of residential segregation – it is hard to test, mainly because the information about the pupils' address presents a large proportion of missing data or incomplete information. Any attempt to automatically georeference the pupils' address fails and presents a large number of missing. Nonetheless, data from the year 2010 indicates that 91% of the pupils enrolled in public municipal schools take no longer than 30 minutes to get to school and 72% walk to school. This is robust evidence that shows that the majority of the pupils study fairly close to their home (in a walking distance).

The same pattern has been observed in other countries with open enrolment system. Parents prefer to enrol their child close to their home for many different reasons, and this phenomenon creates a rationale where "good schools" are relational to the offers they have close to their homes. The "good school" is simply the best one in the neighbourhood and not necessarily the high performance school considering the entire network. Saying in a different way, despite the fact parents can choose any school, their "real choice" is limited to a small number of school close by their houses.

The second alternative explanation (different profile of parents choosing different "types" of schools) will be tested in a different study combining information from two different datasets: a) Parental Demand; b) Pupils and Family Profile.

A different way to test the hypothesis about schools selection bias using information from parental demand and SR is to calculate the correlations considering each educational authority as a separate unit of analysis. In this case, the thesis will compare CRE 2, 4 and 10 as described before. Table 72 presents the correlations between parental demand and SR for all Educational Authorities.

Table 72: Correlations between Parental Demand for Schools and the Segregation Ratio for All Characteristics of Disadvantaged Considering All Educational Authorities – First Segment 2010.

	White Pupil	Black Pupil	EducFS	EducHS	NIS	Distort1	Distort2
Parental Demand CRE1	**-0.4***	**-0.26**	**-0.27**	**-0.21**	**-0.03**	**-0.15**	**-0.16**
Parental Demand CRE2	**-0.43****	**-0.42****	**-0.43****	**-0.47****	**-0.3****	**-0.36****	**-0.38****
Parental Demand CRE3	**-0.14**	**-0.27***	**-0.05**	**-0.11**	**-0.01**	**-0.11**	**-0.23***
Parental Demand CRE4	**-0.26****	**-0.24****	**-0.19***	**-0.26****	**-0.11**	**-0.20***	**-0.28****
Parental Demand CRE5	**-0.26***	**-0.51****	**-0.17**	**-0.28***	**-0.09**	**-0.18**	**-0.20**
Parental Demand CRE6	**-0.34***	**-0.33***	**-0.46****	**-0.58****	**-0.08**	**-0.10**	**-0.23**
Parental Demand CRE7	**-0.37****	**-0.34****	**-0.12**	**-0.08**	**-0.26***	**-0.07**	**-0.25***
Parental Demand CRE8	**-0.20***	**-0.15**	**-0.25****	**-0.25****	**-0.04**	**-0.24****	**-0.32****
Parental Demand CRE9	**+0.05**	**+0.05**	**-0.09**	**-0.02**	**+0.18**	**-0.03**	**-0.19**
Parental Demand CRE10	**-0.15**	**-0.24***	**-0.20**	**-0.24***	**-0.22***	**-0.01**	**-0.11**

** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

To corroborate the hypothesis about school selection bias, the correlation coefficients between parental demand and SR should vary across different educational authorities selected. In order to confirm the initial hypothesis, CRE2 should present a strongest association followed by CRE4 and CRE10. The data support the initial hypothesis. Table 73 summarizes the findings:

Table 73: Correlations between Parental Demand for Schools and the Segregation Ratio for All Characteristics of Disadvantaged Calculated for CREs 2, 4 and 10 – First Segment 2010.

	White Pupil	Black Pupil	EducFS	EducHS	NIS	Distort1	Distort2
Parental Choice CRE2	**-0.43****	**-0.42****	**-0.43****	**-0.47****	**-0.3****	**-0.36****	**-0.38****
Parental Choice CRE4	**-0.26****	**-0.24****	**-0.19***	**-0.26****		**-0.20***	**-0.28****
Parental Choice CRE10		**-0.24***		**-0.24***	**-0.22***		

** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

All correlations not significant at least at 0.05 level were suppressed.

All coefficients in CRE2 are significant at 0.01 level and higher than the figures of CRE4 and CRE10. The data supports the claim that highly dense areas, with greater demand and supply of schools, present higher association between SR and parental demand. CRE2 is the most developed area of the city, with the highest mean and SD for *Prova Brasil*. This is an indication that, on average, schools from CRE2 present not only better outcomes, but also larger differences among schools. It is reasonable to assume that at least part of those differences can be explained by differences in schools intake.

The figures from Table 73 suggests that areas where parents have "real choice" (more than one school in a walking distance), schools with higher demand possess a lower share of disadvantaged pupils. Data support the claim about potential school selection bias. If an increase in parental demand/dispute leads to the more likely event of oversubscription (here there is no real confirmation), it is reasonable to assume that schools could have been systematically selecting pupils based on specific characteristics.

The correlations coefficients in both calculations of SR (entire network and for each Educational Authority) support the claim that an increase in demand for school (or the

possibility of "real choice") is associated with an increase in segregation. A very similar model tested the same hypothesis in public schools in England and the outcomes were different. Urban areas with greater possibilities of parental choice presented, on average, lower levels of segregation.[18] The explanation was the existence of a regulated school market that honoured family's choice. Where chances of "real option" for families were lower (in more rural areas), the influence of residential segregation was higher (Gorard; Taylor; Fitz 2003).

Apparently, the case of Rio de Janeiro is the opposite. Where the families' potential to choose from more than one school is higher, segregation seems to increase. Certainly, part of the variation observed is not related to the educational system (residential segregation should be seen a key variable here). Nonetheless, evidence indicates that it is not possible to rule out the influence of the current legislation in the patterns observed. It is logical to assume that an unclear legislation regarding the first enrolment and pupils' transfers can have an impact on school segregation. The data so far reinforces the role of policy (in this case the absence of regulation) in the patterns of between-school segregation.

[18] One example is the city of London that presented one of the lowest segregation levels in the country, despite the initial claim of being a highly segregated setting (Gorard; Taylor; Fitz, 2003).

8.3 Patterns of between School Segregation in Educational Authorities: What Do We Know?

The second model uses data for the first segment and compares different patterns of segregation over time in three Educational Authorities: CRE2, CRE4 and CRE10. To support the initial hypothesis, the levels of segregation observed in each CRE should vary in a pre-determined order. However, the interpretation of each indicator (poverty, Black Pupils, EducFS and Distortion2) should always consider the proportion of disadvantaged pupils, as stated in the earlier section (see Methods).

Table 74 presents figures of GS for black pupils for three consecutive years (2008-2010). Since there is robust evidence that the missing data is not randomly distributed and appears to artificially inflate GS, the paper presents figures for the most recent years, with lower proportion of missing data.

Table 74: Segregation Index (%) for Black Pupils Calculated for CRE2, CRE4, and CRE10 – First Segment:

	2008	2009	2010
GS BlackPupil CRE2	**20**	**20**	**19**
GS BlackPupil CRE4	**18**	**17.5**	**17**
GS BlackPupil CRE10	**13**	**12.5**	**12.5**

.

All interpretations should consider the proportion of disadvantaged pupils in each Educational Authority. Previous studies have shown that an increase in the proportion of disadvantaged pupils tends to bring the nominal levels of segregation (calculated with GS or D) down (Harris, 2012; Gorard; Taylor; Fitz, 2003).

CRE2 presents the highest levels of GS in all three years. However, if only considering the proportion of disadvantaged pupils for each region (see Table 69), it would be expected a different outcome for the nominal levels of segregation. This confirms that

other elements are strongly influencing GS, to the point of making CRE2 as the most segregated area among the three observed. Despite the fact that CRE 4 and 10 have the exact same proportion of black pupils (total of 10%), CRE 10 shows consistently lower levels of segregation. The variation observed across educational authorities fully corroborates the initial hypothesis: CRE2 has the highest nominal levels and CRE 10 the lowest.

There are two probable explanations. The first one is related to different patterns of residential segregation in all three Educational Authorities. For example, it is possible that the shantytowns in CRE 2, clustered in the middle of affluent neighbourhoods, concentrate the majority of potentially disadvantaged pupils in that area and, therefore, influence GS in a different way. If this is correct, there is the danger of mistaking the impact of residential segregation with possible school selection bias.

A second explanation would be that the greater parental dispute in CRE2 has an effect on GS. Table 70 and 72 supports this claim. If that is correct, it is possible to state that elements associated with the educational legislation are unintentionally influencing the patterns of between-school segregation.

It is relevant to notice that both hypotheses are not mutually exclusive. Saying in a different way, it is possible (actually most likely) that both effects (residential segregation and school selection bias) are happening at the same time and can explain the variation observed in the GS comparing all three Educational Authorities.

Table 75 shows the figures of GS for parents that did not finish Fundamental Education (EducFS). In this variable, CRE2 and CRE4 present the same proportion of disadvantaged pupils (0.28), slightly higher than CRE10 (0.25). If all other variables that can influence school segregation were “controlled” (meaning constant), CRE10 would have, on average, higher figures for GS. Is that the case?

Table 75: Segregation Index (%) for Parents Who Did Not Finish Fundamental Education Calculated for CRE2, CRE4, and CRE10 – First Segment:

	2008	2009	2010
GS EducFS CRE2	**23**	**22**	**19.5**
GS EducFS CRE4	**24**	**22,5**	**21,5**
GS EducFS CRE10	**19.5**	**16.5**	**16**

Once again, data suggests that less urban areas, where parental dispute for schools are potentially smaller, seems to have lower levels of segregation. Despite the fact that CRE10 presents the smaller proportion of disadvantaged pupils, GS is once again smaller compared with CRE2 and 4. In this variable, CRE4 presents slightly higher figures for GS, which can be seen as evidence that partially refute the hypothesis about school selection bias. However, the differences observed along the years are quite small and not very convincing.

The next variable is poverty, which is the only one where CRE2 shows a smaller proportion of disadvantaged pupils. If all other variables that can influence school segregation were controlled, CRE 2 should present, on average, higher levels of segregation. Any odd outcomes in figures of Table 76, such as CRE 2 with lower nominal levels of segregation, should be seen as a strong indication that the initial hypothesis is false.

Table 76: Segregation Index (%) for Poverty Calculated for CRE2, CRE4, and CRE10 – First Segment:

	2008	2009	2010
GS NIS CRE2	**21.5**	**21**	**21**
GS NIS CRE4	**14.5**	**15**	**16**
GS NIS CRE10	**11**	**11**	**11**

Between-school segregation in CRE 2 is almost twice as compared with CRE 10. The outcome does not refute the initial hypothesis about school selection bias. It is possible that at

least part of the difference observed among educational authorities is related to the variation in the proportion of poor families presented in Table 69. Nonetheless, it is worth mentioning that Tables 74 and 75 presented the opposite outcome: Educational Authorities with lower proportion of disadvantaged pupils for blacks or EducFS (CRE10) also presented the lowest levels of between-school segregation.

The last variable is Distortion 2, which presents the highest nominal levels of school segregation considering all indicators. Since this variable presents perfect record (no missing data) for all seven years, GS was calculated from 2004 to 2010. Table 77 shows the figures for Distortion2.

Table 77: Segregation Index (%) and Proportion of Disadvantaged for Distortion 2 Calculated for CRE2, CRE4, and CRE10 – First Segment:

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
GS Distortion 2 CRE2	23	24.5	25.5	25	25.5	25.5	23
Proportion of Disadvantaged CRE2	0.13	0.13	0.13	0.14	0.13	0.11	0.12
GS Distortion 2 CRE4	22	21	20	20	21	20	17
Proportion of Disadvantaged CRE4	0.11	0.13	0.13	0.15	0.13	0.11	0.12
GS Distortion 2 CRE10	17.5	17	18.5	20.5	21.5	22.5	20
Proportion of Disadvantaged CRE10	0.12	0.13	0.12	0.12	0.11	0.09	0.10

For all seven years, the proportion of disadvantaged pupils in CRE10 is consistently lower compared with CRE2. However, GS shows lower figures for all measures. The calculations of GS for all four variables of potentially disadvantage pupils do not allow the researcher to refute the hypothesis about school selection bias. More importantly, the results of GS for black pupils and age/grade distortion strongly suggest that urban areas, with a high

demand and supply of schools have consistent higher nominal levels of segregation. It appears that whenever families have “real choice”, segregation tends to growth. Both models presented (with SR and GS) present similar answers and corroborate the initial hypothesis: schools are potentially selecting pupils based on specific characteristics.

It is important to highlight that all models presented in this section are not causal in nature and only describe or correlate variables that could be associated with a potential increase in school segregation. It is very hard to produce a more reliable test to establish a true causal model to explain school segregation. Key variables that can influence the indexes are simply not available or cannot be fully controlled by the researcher. Perhaps the most significant example is residential segregation.

9. SCHOOL SHIFT ALLOCATION

The school "shift" policy are observed in the majority of cities in Brazil and other developing countries, and basically organizes the allocation of pupils at "school shift" level – morning, afternoon or night. The research design captures the "net effect" of the policy and provides figures to compare the actual segregation levels with the hypothetical scenario with no such policies. There are two possible outcomes for the test: 1) no impact; 2) increase in segregation. Since there is no specific legislation to regulate the allocation of pupils across shifts, a random allocation would be expected. Does it really happen?

More than 90% of all the Rio de Janeiro public schools present two or more "shifts". Basically, the "shift policy" has been the solution for the increasing number of pupils enrolled in Fundamental School in Brazil in the past decades and the lack of new schools (buildings). During the XX century, in only 40 years, the urbanization process rapidly inverted the proportion of the population living in rural areas and cities. There are reports during the 1970s in Rio de Janeiro, with schools that presented four "shifts" in one day, with very few working hours for each "shift". Today, the most common situation is a school with two "shifts" (morning and afternoon), but it is still possible to observe schools that present a third one – known as the "night shift". Undoubtedly, there is an effort to increase the total number of hours pupils spend at school per day, which necessarily demands one "shift" for each school building.

Everything, besides the building and the principal, can change when comparing the "shifts": teachers, staff, even working materials, can vary from one "shift" to another. In some cases the "shift" represents a change in the level of schooling provided (first or second segment of Fundamental School), but, in others, the level can be exactly the same. Maybe the most relevant information for research purposes is the criteria to allocate pupils into "shifts". Since there is no specific regulation or clear criteria, it is the school bureaucracy that makes

the final decision. It is possible to say that the school staff have control over two moments of admission: a) first enrolment and; b) the allocation of pupils to a "school shift". In reality, what characterizes the "shift policy" is the absence of any regulation and transparency regarding the allocation of pupils. It is important to make a distinction between the "shift policy" and the more classical definition of educational policy that presents a collection of laws and regulations with a clear intention to address issues of public interest.

The chapter aims to answer one key question: Does the distribution of pupils across school "shifts" have any impact on the segregation levels? Or rephrasing the question: Is it a random distribution? Previous studies had analysed the allocation of pupils across "shifts" in a very limited number of schools and with a poor research design, with potential selection bias problems. The lack of solid evidence has so far precluded public debate about this issue.

9.1 How to Measure the "Net Effect" of the School Shift Allocation?

The research design compares the levels of GS where every "school shift" is a single unit, differing from the usual approach that considers the "school building" as a single unit (disregarding the allocation of pupils across the "shifts"). Any differences observed in every year should be attributed as the "shift effect". The analysis will present three different approaches to make the results more reliable: 1) all pupils enrolled in Fundamental Education; 2) pupils in the first segment of Fundamental Education (1st to 5th grade); 3) pupils in the second segment of Fundamental Education (6th to 9th grade).

The figures of the "shift effect" are presented as the proportional relative increase for each variable, where: "GSss" is GS calculated with "school shifts" as the unit of analysis; "GSsb" is GS calculated with "school building" as the unit of analysis. This is important because the nominal values of GS for the variables vary greatly.

{ GSss – GSsb } / GSsb

From the methodological point of view, the challenge is to aggregate the pupil data at the different levels: 1) each school building as a unit; 2) each school "shift" as a unit. This is the most appropriate design to measure the impact of the school "shifts" on the overall level of school segregation. Since there is no specific policy that guides the distribution of pupils across "shifts", a random distribution of pupils would be expected.

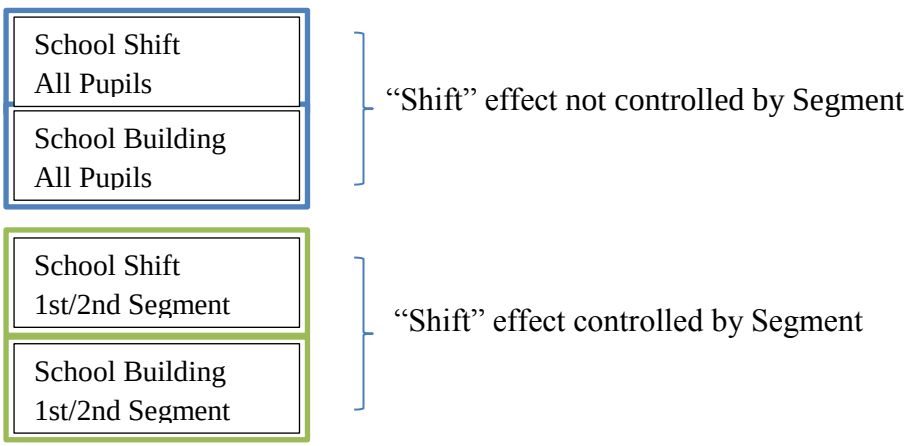
The research design was thought to provide two important pieces of information. The first one is the "net effect" of one specific educational policy. If any difference appears between the GS levels considering the "school building" as one unit and the "school shift", it will be possible to state that the change is due to the policy and nothing else. All the other

elements that can influence school segregation, such as residential segregation or parental choice, are controlled by this design.

The second important point is to rule out any plausible alternative explanations in the case of a positive effect of the policy. The design accounts for this in two ways. First, replicating the outcomes over seven years (2004-2010). If the results constantly showed a similar pattern, it would be very unlikely that this could be due to fluctuation of data (error) in any specific year. Second, by calculating the "shift effect" separating pupils in the first and second segments of Fundamental Education, the design prevents the likely event that the "shift effect" gets mixed with a "segment effect".

This could happen because of two reasons: a) some schools offer all Fundamental Education grades and organize the "shift" considering the grade (for example, younger pupils separated by "school shift" from the older group); b) the proportion of potentially disadvantaged pupils in different segments is most likely to be different (mainly because of drop-outs and pupils retained at the end of each school year). Therefore, the design mostly accounts for the possibility that part of the "shift effect" could be explained by the simple fact that pupils from different grades attend different "shifts" in the same school building.

Figure 17: School "Shift" Effect Controlled by Segment



9.2 Is There Any School "Shift" Effect? Or, Is the Distribution of Pupils across School "shifts" Random?

The main concern with the "school shifts" is that they represent, in practical terms, two or three entirely different schools functioning in the same school building. An attempt to measure the school-mix-effect (Gorard, 2006) should really consider every school "shift" as an independent unit of analysis.

Table 78 presents the GS figures for all variables of disadvantaged pupils considering pupils for all grades in fundamental school (1st to 9th). For each variable, GS was calculated in two different ways: 1) each school building as a unit (NonShift); 2) each school "shift" as a unit (Shift).

Table 78: Segregation Index (%) for All available Indicators. Data Provided by Municipal Educational Department of Rio de Janeiro.

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
GS Black Shift	19	18	17	16.5	16	15.5	15.5
GS Black Non-Shift	17	16	15.5	14.5	14	13.5	13.5
GS NonWhite Shift	12.5	9	8	7.5	6.5	6.5	6.5
GS NonWhite Non-Shift	10.5	8.5	7.5	7	6.5	6	6
GS EducFS Shift	30.5	29	27	25	23.5	21.5	20.5
GS EducFS Non-Shift	29	27.5	25	24	22	20.5	19.5
GS EducHS Shift	15	14	12.5	11.5	10.5	10	9.5
GS EducHS Non-Shift	14	12.5	11	10	9.5	9	8.5
GS NIS Shift	28.5	24	20.5	18.5	19	19	20
GS NIS Non-Shift	24	20	17.5	16.5	16.5	16.5	17
GS Distortion 1 Shift	14.5	12	13	11.5	12.5	11.5	13.5
GS Distortion 1 Non-Shift	8.5	8.5	8.5	8.5	8	8	9
GS Distortion 2 Shift	29	28.5	29	31	33	30.5	30.5
GS Distortion 2 Non-Shift	21	21	21.5	24	25	24	23

It is possible to observe that GS calculated for "Shift" presents higher figures for all variables, with the exception of GS Non-White Shift in 2008. This is a clear indication that the shift allocation has some effect in school segregation. More importantly, the effect seems to be higher for poverty and age/grade distortion. Using the figures presented in Table 78, it is possible to calculate the relative proportional increase of school segregation for each variable. Table 79 presents the "shift effect" considering all indicators of disadvantaged pupils.

Table 79: Relative Proportional Increase for the School "Shift" Effect for All Pupils Enrolled in Fundamental Education.

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
GS Black Pupil	+ 12	+ 13	+ 10	+ 14	+ 14	+ 15	+ 15
GS Non White Pupil	+ 19	+ 6	+ 7	+ 7	-	+ 8	+ 8
GS EducFS	+ 5	+ 5	+ 8	+ 4	+ 7	+ 5	+ 5
GS EducHS	+ 7	+ 12	+ 14	+ 15	+ 11	+ 11	+ 12
GS NIS	+ 19	+ 20	+ 17	+ 12	+ 15	+ 15	+ 18
GS Distortion 1	+ 71	+ 41	+ 53	+ 44	+ 56	+ 44	+ 50
GS Distortion 2	+ 38	+ 36	+ 35	+ 29	+ 32	+ 27	+ 33

It is possible to divide all seven indicators into three different groups considering the overall percentage increase in GS: 1) very low positive impact – parents' education (EducFS; EducHS) and Non-White pupils; 2) medium positive impact – poverty and Black pupils; 3) high positive impact – Distortion 1 and 2.

The research design measures the "net effect" of the school "shift". In practical terms, it means that all other elements that can influence the overall segregation levels are being "controlled" in this model: parental choice, residential segregation, etc. For example, the figures suggest that the "shift" effect explains around 50% of the between-school segregation

for Distortion 1. In theory, if there were no "school shifts", it would be possible to reduce school segregation for all indicators. In some cases the reduction could be up to 50%.

The question now is: Will the results show a similar pattern when the "shift" effect is calculated separately for pupils in the first (1st to 5th grade) and the second segment (6th to 9th grade)? Table 80 shows the figures for first segment pupils for all available indicators for the seven years (2004-2010). The patterns are somewhat different. It is possible to see that the size effect of the school "shift" became smaller for: parents' education (EducFS and EducHS), ethnic background (non-white and black pupils) and poverty (NIS). It is still possible to see a two-digit size effect for GS Black Pupils for the last four years (the most reliable data for this indicator) with a relative percentage increase of 13% in the overall segregation levels.

Table 80: Relative Proportional Increase for the School "Shift" Effect – First Segment.

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
GS Black Pupil	+ 9	+ 9	+ 9	+ 13	+ 13	+ 13	+ 13
GS Non White Pupil	+ 7	+ 7	+ 7	+ 7	+ 8	+ 8	-
GS EducFS	+ 5	+ 5	+ 8	+ 4	+ 7	+ 5	+ 5
GS EducHS	-	+5	+ 5	+ 5	-	+ 5	-
GS NIS	+ 6	+ 6	+ 7	+ 3	+ 3	+ 6	+ 6
GS Distortion 1	+ 131	+ 46	+ 62	+ 31	+ 50	+ 55	+ 100
GS Distortion 2	+ 48	+ 32	+ 30	+ 16	+ 20	+ 16	+ 38

But what happens to age/grade distortion? This indicator still shows the largest impact for the "shift" effect. For Distortion 1, the figures show an average relative percentage increase of around 50%, with the real possibility of the effect doubling the overall level of school segregation. The impact of Distortion 2 is a bit smaller, but still relevant, being responsible for one third of the variation of school segregation. Comparing the outcome of

Table 8 (only first segment pupils) with that of Table 79 (all the pupils), it is possible to state that young pupils are being systematically tracked by their previous academic performance. Future studies should make analyses at pupil level to understand how this tracking is happening. Future questions to be answered: Do schools deliberately change a pupil's "shift" based on his/her attainment in the last academic year? Does the tracking occur at the initial enrolment (1st grade) or does it happen throughout different educational transitions?

Table 81 presents the relative percentage increase for all available indicators for pupils in the second segment (6^{th} to 9^{th} grade) of Fundamental School. Are the results similar to the patterns observed in the first segment? The answer is yes. Once again, comparing Table 81 (only second segment pupils) with Table 79 (all pupils), it is possible to observe that a big part of the variation (size effect) of the school "shift" has disappeared for the indicator of parents' education, ethnic background and poverty. The only indicator of these three that still presents a two-digit size effect is GS Black Pupil (over the seven years).

Table 81: Relative Proportional Increase for the School "Shift" Effect – Second Segment.

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
GS Black Pupil	+ 14	+ 11	+ 13	+ 11	+ 11	+ 12	+ 16
GS Non White Pupil	+ 17	+ 10	+ 13	+ 7	-	+ 9	+ 9
GS EducFS	+ 8	+ 7	+ 6	+ 4	+ 7	+ 5	+ 5
GS EducHS	+ 6	+ 3	+ 9	+ 5	+ 6	+ 6	-
GS NIS	+ 36	+ 16	+ 9	+ 3	+ 6	+ 3	+ 3
GS Distortion 1	+ 136	+ 92	+ 127	+ 100	+ 109	+ 73	+ 120
GS Distortion 2	+ 96	+ 78	+ 87	+ 54	+ 68	+ 52	+ 80

Age/grade distortion once again presented the biggest relative percentage increase for the impact of the school "shift". However, the figures are even higher when compared to Table 80, which suggests that the tracking intensifies throughout the educational transitions.

It appears that the school network consistently track pupils based on prior educational attainment, throughout the year and across different educational transitions. This might explain why all the other indicators still present a positive impact (but a smaller size effect). Despite the fact that there are no formal tracking policies at public schools in Rio de Janeiro, the results clearly indicate systematic "informal tracking" of retained pupils (failed an academic year).

This is the first time a study about Rio de Janeiro public schools (the biggest public school network in South America) presents a robust research design to measure the impact of the "shift" in school segregation. Previous studies only analyzed a very small number of schools with potential problems of selection bias. The implications of the study are important since there is no public debate about a "hidden phenomenon". Some questions should be raised: What are the pedagogical reasons for this systematic tracking of pupils with lower attainment? Is it a phenomenon observed in all educational authorities (total of ten in the city) or is it more concentrated in specific regions? What is the impact on attainment, future education aspirations and the sense of justice of those who are being segregated by school "shift"?

Discussion

The unintentional impact of educational policies on school segregation levels is a relevant subject related to educational opportunities and social justice. There is solid evidence that suggests that between-school segregation can have deleterious effects, with a greater impact on the most disadvantaged pupils, future educational aspirations, the quality of teaching, advanced education subsequent to the compulsory level and an increasing association between attainment and socio-economic status. If it is true that policies can influence the allocation of pupils, researchers and policy-makers should be aware of the risks and "adverse effects" of any legislation that can potentially increase the overall segregation levels.

The measurements presented in the study (total of 147) clearly show a pattern that indicates that the "shift policy" has an impact on the overall segregation levels. Comparing different indicators, it is possible to state that pupils are being systematically selected based on prior educational attainment (variable age/grade distortion). Not only GS Distortion 2 has the highest overall value for GS (see Table 78), but also the "shift" effect for the same variable (Distortion 2) is the highest – see Tables 80 and 81. Apparently, there are two segregation levels within the public network: a) "school building" level; b) "school shift" level. These are two separate stages. The first one is clearly influenced by residential segregation, parental choice, but also by potential selection bias by the school staff. The second level (the net effect of the "shift") is only related to educational bureaucracy.

Despite the lack of specific regulation for the allocation of pupils across "shifts", the regularity of the measurements suggests a constant intentional selection process. Of a total of 147 measurements (see Tables 79, 80 and 81) across seven years, 140 showed that the "shifts" increased the overall segregation levels. This is not the first study to analyse the

impact of the “shifts” in Brazilian public schools, although previous studies had many limitations in terms of research design, with potential selection bias problems that prevented robust analysis. This is the first study to analyse the entire public network in a major Brazilian city, with a robust design that captures the “net effect” of pupil allocation across “shifts”.

It is important to stimulate public debate about the implicit message that such a system announces: Is it fair to separate pupils based on prior educational attainment? The data suggests that the overall segregation levels for other variables, such as poverty and ethnic background, could be inflated by the selection based on prior educational attainment. Poor and black pupils are over-represented in the Distortion group, and are, therefore, being indirectly selected. A different hypothesis is that schools could also be systematically segregating black and/or poor pupils in different “shifts”. Both options seem arbitrary and unfair.

Future chapters will answer two crucial questions: a) is the “shift” effect similar in different educational authorities? b) How far from a random distribution the outcomes observed are? The first question will be answered by replicating the same methods of calculation for each educational authority. As it will be shown, the “shift effect” is not linear and is more prevalent in specific areas of the city. The second question will be answered with a simulation to compare values of GS in a truly random allocation, with the real values observed on Tables 80 and 81. The simulation allows the research to observe how likely the values calculated are.

10. SIMULATION FOR SCHOOL SHIFT ALLOCATION

Previous chapters have shown that school segregation increases as the "school shift" is introduced as the unit of analysis. Measuring each segment separately, there are 98 measurements of the "net effect" of the shift allocation, with 93 showing a positive impact (increase in segregation) for all available indicators of disadvantaged pupils. The initial hypothesis states that the "school shift" allocation is not random and has the potential to increase between-school segregation.

Nonetheless the relative proportional increase for each variable is different. It is possible to separate all indicators into three different groups according to the size of the effect: 1) very low positive impact – parents' education (EducFS; EducHS), Non-White pupils and poverty; 2) medium positive impact – black pupils; 3) high positive impact – Distortion 1 and 2.

This typology is based on simple observation, but it lacks another measurement to really compare the size effect. The real question is: how far are the GS values, calculated considering the "school shift" as the unit of analysis, from a truly random distribution? Generating new datasets with a random distribution for shift allocation might help understanding of how big, or significant, the impacts observed are.

It is not necessarily true that all the increase in segregation by the shift allocation is due to pupil selection. It is probable, even with a random distribution across shifts, to observe positive impacts on school segregation. But how is this possible? The reasons are simple: mathematical probability and the formula of the Segregation Index (GS). A simple simulation might clear any doubts. Taking an imaginary educational system composed of 4 schools, with a perfect distribution of disadvantaged pupils across schools – Table 82 – the hypothetical scenario would show a segregation level of zero.

Table 82: Hypothetical Educational System with No School Segregation.

	Poor Pupils	Number Pupils
School A	3	10
School B	3	10
School C	3	10
School D	3	10
Total	12	40

If all four schools presented two shifts, it would be impossible to make a perfectly even allocation of disadvantaged pupils across shifts. In this case, the index would necessarily show an increase in between-school segregation. This unlikely scenario is only relevant to show that the "shift effect" can only have two outcomes: zero impact or an increase in segregation. Mathematically, it is impossible to observe a decline in segregation when calculating the "shift effect".

A more likely situation is to observe small variations in a random distribution of disadvantaged pupils across schools. Probability will show that it is possible to have an uneven distribution depending on how many allocations (tests) there are. Just like in a card game, there are small chances of receiving a very good hand – an unlikely situation. The simulation allows the researcher to observe how unlikely the outcomes he has calculated are, comparing the nominal GS values in the dataset with those generated in a random distribution.

10.1 The Simulation

In order to run the simulation, it was necessary to create new variables regarding the shift allocation. The chapter used data from 2010 and considered key elements for the simulated allocation. Each school had a particular number of "shifts" that could be from one to three. The "simulated shift" was generated considering the real number of shifts for each school and also the proportion of pupils allocated to each shift for each school. This is the best approach to simulate, as closely as possible, a random allocation of pupils across "school shifts". In each segment, one hundred variables, named "simulated shift 1" through 100, were generated.

One hundred new datasets were aggregated at school level, each one considering the new simulated allocation across shifts. The GS was calculated for three key variables in both segments: black pupils, distortions 1 and 2. The nominal GS values, calculated from the simulations, were compared with the real values obtained in the dataset provided by the Educational Department of Rio de Janeiro. The simulated distribution of GS will provide parameters to establish the likelihood of the measurements for the "shift effect".

10.2 Outcomes of the Simulation

This thesis presents the outcomes of the simulations calculated separately for the first and second segments. Three variables were chosen from testing based on the size effect in Tables 80 and 81: Distortion 1, Distortion 2 and black pupils. In 2010, there were 323,988 pupils enrolled in the first segment – 1st to 5th grade. All pupils participated in the 100 simulations. All possible effects of missing data would have a similar impact on all allocations (real and simulated) and would not interfere with the outcomes.

Table 83 presents the figures of GS considering shift and non-shift calculations, plus the indication of the size effect of "shift allocation", for all three variables. In order to corroborate the initial hypothesis, GS calculated from the simulated shift allocation would have to be consistently smaller than the size effect observed in the real distribution.

Table 83: Segregation Index (%) for Shift and Non-Shift Calculations – First Segment.

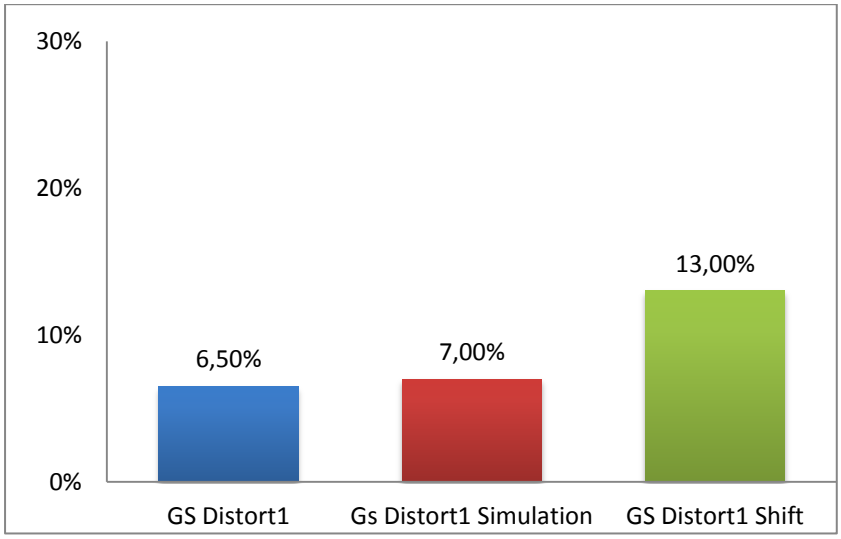
	Non-Shift	Shift	Size Effect
GS Distortion 1	6.5	13	100 %
GS Distortion 2	19.5	27	38%
GS Black Pupil	15	17	13%

Age/grade distortion for one or more years affects 41% of pupils in the first segment for year 2010. Figure 18 presents the data for the simulated distributions of Distortion 1 against the real figures of GS calculated in two different ways: a) each school as one unit of analysis; b) each "school shift" as one unit of analysis. All 100 values in the simulated allocation across shifts indicated a nominal value of 7% for school segregation calculated with GS.[19] This means that 7% of potentially disadvantaged pupils for age/grade distortion for one year or more would have to move to another school in order to achieve an even

[19] The nominal values of GS in the simulations had very small variance – lower than 0.5%

distribution of pupils. The figures suggest a very small increase in segregation with a truly random distribution.

Figure 18: Segregation Index (%) for Distortion 1, Considering Shift, Non-Shift and Random Allocation – First Segment.



Comparing the simulated values of GS (7%) with the real observation (13%) for Distortion 1, it is possible to state that the allocation of pupils across shifts is not random. The nominal values are so different that the existence of an intentional, systematic selection process, based on prior educational achievement, becomes clear. In real numbers, it is possible to state that between 9,000 and 10,000 pupils enrolled in the first segment were tracked and allocated to a different shift from that expected.

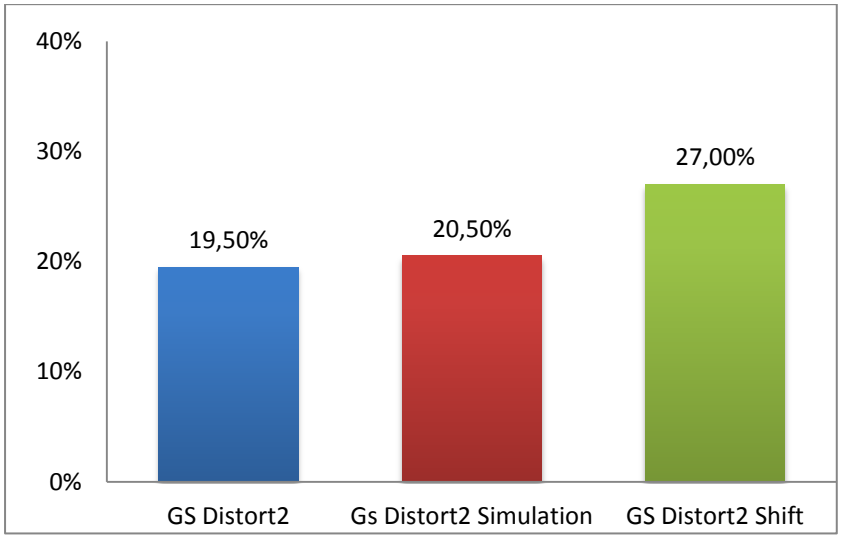
This phenomenon will be called “informal tracking”. It is “informal”, because there is no clear legislation or transparency in the decision-making process to allocate pupils across shifts. The tracking starts at a very early age, and has the potential to affect the educational opportunities of thousands of pupils every year.

What are the pedagogic reasons to cluster pupils with learning difficulties? What evidence do policy-makers have regarding the impact of such “policy”? Does it help to close

the gap between high and low achievers? If there are no clear answers to the questions above, policy-makers, educational researchers and members of the general public should engage in a public debate about the consequences of such "informal tracking". Differentiation should only be permitted if it benefits the worst-off. If any public policy or arrangement among civil servants violates this principal, it could be deemed unfair and unethical (Rawls, 1971).

Figure 19 presents the data for the simulated distributions of Distortion 2 against the real figures of GS. From 100 simulations, 99% presented GS of 20.5 %. It is possible to state that the allocation of pupils across shifts with two or more years of age/grade distortion is not random. The overall impact is an increase in school segregation. This single procedure affects the educational opportunities of around 2,500 pupils enrolled in the first segment every year.

Figure 19: Segregation Index (%) for Distortion 2, Considering Shift, Non-Shift and Random Allocation – First Segment.

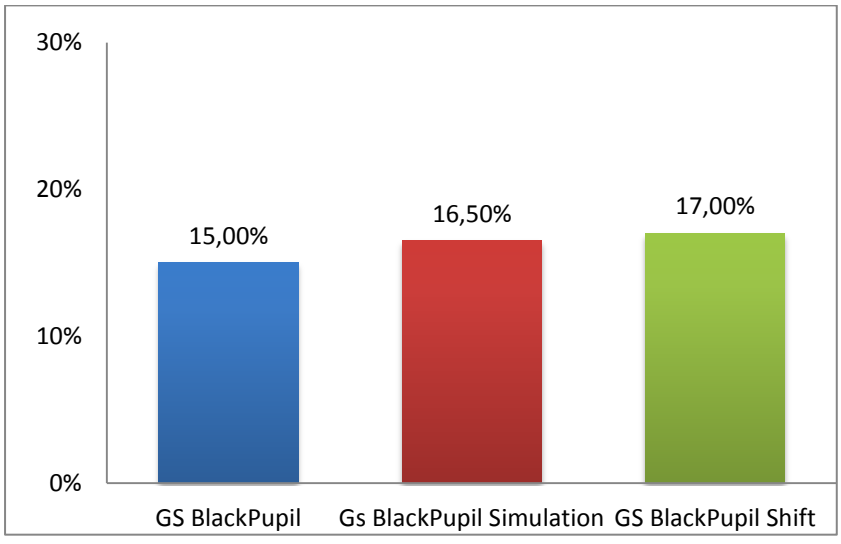


It is reasonable to assume that all pupils with two or more years of distortion have been retained at least once. This is the variable that best discriminates low achievers. The data in Figures 18 and 19 confirm the existence of "informal tracking" in the shift allocation in Rio de Janeiro public municipal schools. The outcomes observed in both simulations

indicate that low achievers are selected and allocated to specific schools shifts. It is interesting to notice that, because the allocation procedures lack clear legislation and transparency, not all pupils with Distortion 1 and 2 are in the same shifts – in this case, they would be the afternoon or night shifts. Some, but not all, pupils with age/grade distortion are being tracked, and sent to different shifts. It is possible that other variables not measured in this study also play an important role in the shift allocation procedure. So far, it is clear that age/grade distortion is an important explanation variable to understand the shift allocation.

The last simulation was performed to observe the distribution of black pupils across shifts. This variable presents a double-digit relative proportional increase – 13% in 2010. How far is the observed valued from a random distribution? Figure 20 presents values of GS for all three calculations. From 100 simulations, 97% presented GS of 16.5%, which suggests that the real distribution (GS of 17%) is just above the expected value.

Figure 20: Segregation Index (%) for Black Pupils, Considering Shift, Non-Shift and Random Allocation – First Segment.



The difference observed between the simulated and real distributions is very small, and it is hard to safely state that black pupils are being systematically tracked and sent to

different shifts. There are around 300 black pupils enrolled in the first segment allocated to a different shift from that expected. There is a clear difference regarding the size effect of the shift allocation when comparing age/grade distortion and black pupils. Apparently, for first segment pupils, distortion is the most important variable to understand school segregation patterns.

The results presented in this chapter are directly linked to procedures of the educational bureaucracy. All other variables that play a key role in school segregation, such as residential segregation, have been "controlled" in this model. The debate about clustering potentially disadvantaged pupils in specific schools or school shifts should be of public interest. Data shows that the "informal tracking" starts at a very early age, and has the potential to affect the educational opportunities of thousands of pupils. What are the pedagogic reasons to intentionally and systematically track pupils with age/grade distortion? What are the impacts of the cluster on attainment and future educational aspirations of the disadvantaged group?

None of these questions have been answered. In fact, the informality and lack of transparency leave an air of mystery about the effects of such policies. Tracking policies should prove the benefits of clustering for the most disadvantaged before full implementation, and not the other way around. The current situation could be interpreted as unfair. In Rio de Janeiro public schools, a potentially disadvantaged pupil becomes, at a very early age, a really disadvantaged one: low achievers are sent to specific school shifts without any guarantee that they will have better educational opportunities.

***

Table 84 presents the figures for the second segment, considering shift and non-shift calculations, plus the size effect of "shift allocation". GS was calculated for three variables:

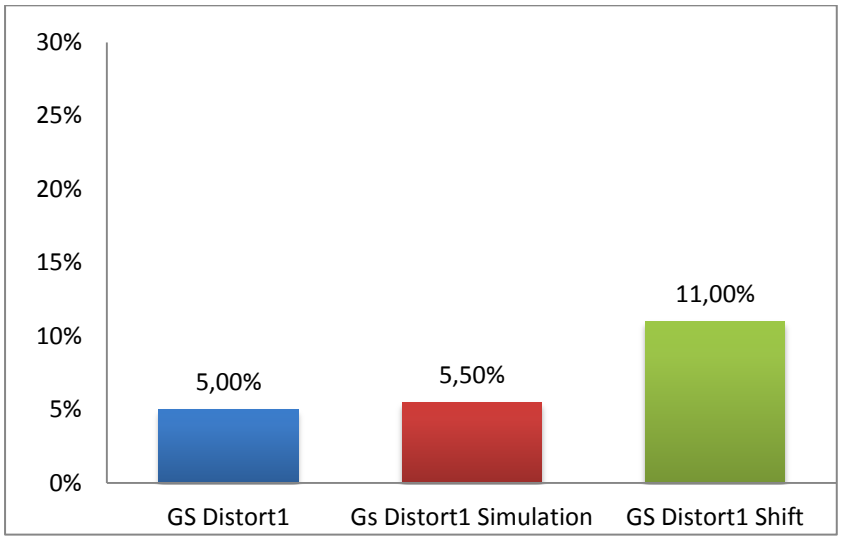
black pupils and Distortion 1 and 2. The outcomes for the second segment are expected to be similar to those presented in the previous section (simulation for the first segment). In order to corroborate the initial hypothesis, the GS calculated from the simulated shift allocation would have to be consistently smaller than the size effect observed in the real distribution.

Table 84: Segregation Index (%) for Shift and Non-Shift Calculations – Second Segment.

	Non-Shift	**Shift**	**Size Effect**
GS Distortion 1	5	11	120 %
GS Distortion 2	12.5	22.5	80 %
GS Black Pupil	12.5	14.5	16 %

Figure 21 presents the data for the simulated distributions of Distortion 1 compared against the real figures of GS calculated in two different ways: a) each school as one unit of analysis; b) each "school shift" as one unit of analysis. All 100 values in the simulated allocation across shifts indicated a nominal value of 5.5 % for school segregation calculated with GS. A truly random distribution would have a very small positive impact on school segregation – changing from 5 % to 5.5 %.

Figure 21: Segregation Index (%) for Distortion 1, Considering Shift, Non-Shift and Random Allocation – Second Segment.

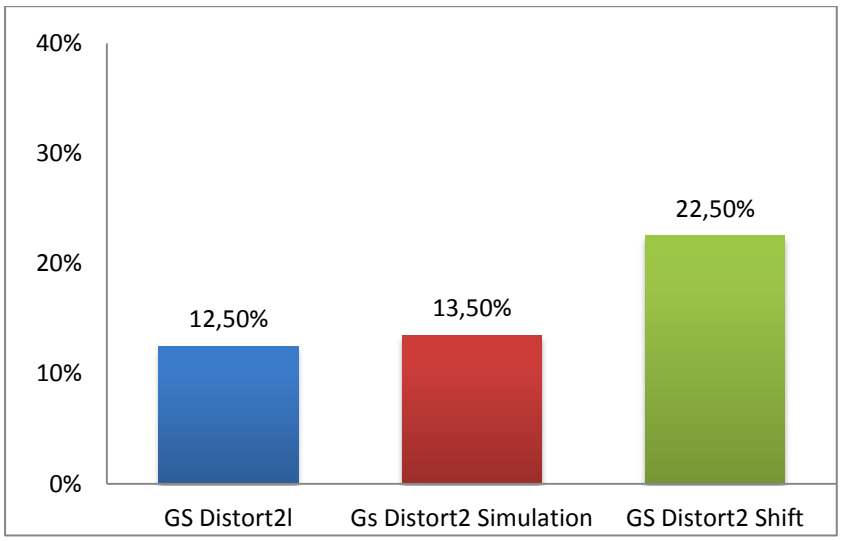


Comparing the simulated values of GS (5.5 %) with the real observation (11 %) for Distortion 1, it is possible to state that the allocation of pupils across shifts in the second segment is far from a random distribution. A very similar outcome was observed in the first segment. The non-random allocation affects the educational opportunities of around 9,000 second segment pupils per year. Apparently, the "informal tracking" is intensified in the second segment – see the size effects in Tables 80 and 81.

Figure 22 presents the data for the simulated distributions of Distortion 2 compared with the real figures of GS for second segment pupils. From 100 simulations, 5 % presented a GS of 13.5 % – the highest value. The other 95% presented a GS of 13 %. Again, it is possible to state that the allocation of pupils across shifts with two or more years of age/grade distortion is not random. The overall impact is an increase in school segregation. This single procedure has the potential to affect the educational opportunities of around 6,000 second segment pupils per year. Comparing figures of the first and second segments, there is an increase in the proportion of pupils with two or more years of age/grade distortion. This

change reflects the increase in the total number of pupils tracked and sent to different schools shifts.

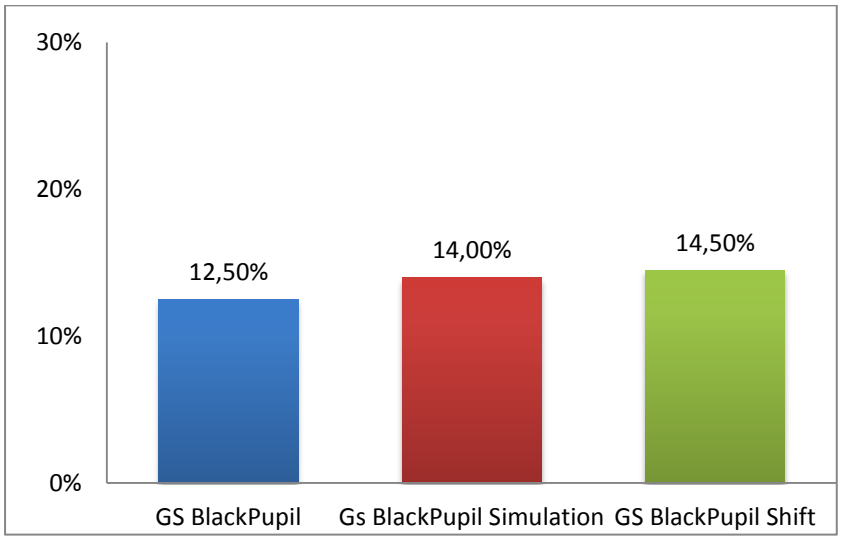
Figure 22: Segregation Index (%) for Distortion 2, Considering Shift, Non-Shift and Random Allocation – Second Segment.



Comparing the figures from both segments, it is possible to identify an interesting pattern. Apparently, the decline in school segregation observed in the transition from the first to the second segment is partially "compensated" by an important increase in the "shift effect". A sudden decline in the total number of schools (5th-6th grade transition) might lead to a small change in the patterns of between-school segregation: segregation between schools declines at the same time the "shift effect" rises.

Figure 23 presents the data for the simulated distributions for black pupils. From 100 simulations, 89% presented a GS of 14 % and the rest (11 %), a GS of 13.5 %. The results are quite similar to those observed in the first segment. The differenced observed is very small and makes it hard to safely state that black pupils are being tracked and sent to different shifts.

Figure 23: Segregation Index (%) for Black Pupils, Considering Shift, Non-Shift and Random Allocation – Second Segment.



Discussion

A total of 6 simulations in both segments, using three different variables, were conducted to help interpret the size effects of the shift allocation described in the previous chapter – see Tables 80 and 81. The simulations were important to clear any doubts, especially about the potential clustering of black pupils in the shift allocations.

Data suggests an intentional, systematic selection process based on prior educational attainment – called "informal tracking". The shift allocation is an important moment in the distribution of education opportunities in Rio de Janeiro public municipal school, and, depending on the variable observed, the "shift effect" can be responsible for more than half of the variation in between-school segregation.

So far, the study has measured the impact of the non-random "shift" allocation on school segregation. A separate study should address a key question: what are the impacts of the "informal tracking" for pupil attainment and future life prospects? This is a crucial question that should raise concern among policy-makers, educational researchers and members of the general public.

International studies suggest that clustering pupils potentially disadvantaged pupils can affect how they are treated at school, the quality of teaching, future educational aspirations and an increasing association between academic achievement and socio-economic status. All these negative impacts (or at least some of them) could be happening at Rio public schools, and should be carefully investigated.

Data shows that the "informal tracking" starts at a very early age in the first segment – pupils from 6 to 10 years old – and intensifies in the second segment. Those who could be seen as potentially disadvantaged pupils when entering the educational system, become, at a very early age, really disadvantaged: low achievers are sent to specific "school shifts", without any guarantee that they will have better educational opportunities. Data suggest that

the non-random allocation can have an impact on the educational opportunities of around 12,000 pupils enrolled in the first segment and 15,000 in the second segment.

The pedagogic reasons for such actions should be made public. Is there any evidence suggesting academic benefits for the disadvantaged group? The lack of clear legislation and transparency in the shift allocation does not encourage public debate about this key stage in the distribution of educational opportunities in public schools.

The simulations have shown that the distribution of black pupils across shifts (the real observation) is very similar to a truly random distribution. A small difference, suggesting selection bias in relation to black pupils, was observed in both segments. However, the variation was minimal– with a potential impact for around 300 pupils in each segment.

11. SCHOOL "SHIFT" ALLOCATION BY EDUCATIONAL AUTHORITIES

One last question regarding school "shift" allocation remains unexplored: is the "shift" effect homogeneous across different educational authorities? Since this is the first "shift" effect that has been measured considering the entire population of pupils, it would be interesting to observe any variation across educational authorities.

This section will present two different models using only data for the first segment – 1st to 5th grade. Initially, it describes the "shift" effect for all educational authorities using data from 2010. Four different characteristics of disadvantaged pupils were used: colour, parents' education, poverty and age/grade distortion. The second model focuses on the analysis of three different educational authorities (CRE 2, 4 and 10), and presents data from three consecutive years (2008-2009-2010). The description of key variables of all educational authorities was presented in Table 69.

Table 85 presents the GS figures for black pupils and the relative proportional increase of the index considering all educational authorities. The overall size effect for this variable, described in the previous chapter, was 13%. However, it is possible to observe that some educational authorities have relative proportional increases up to 25% (CRE 6) or 20% (CRE 8).

Table 85: School "Shift" Effect for Black Pupils for All Educational Authorities – First Segment – 2010.

	Black Pupil 2010	**Black Pupil 2010**	
	Non-Shift	Shift	Size Effect
CRE 1	17.5	19	9 %
CRE 2	19	20.5	8%
CRE 3	14.5	16.5	14%
CRE 4	17	18.5	9%
CRE 5	11.5	13	13%
CRE 6	10	12.5	25%
CRE 7	19	20.5	8%
CRE 8	12.5	15	20%
CRE 9	11	13	18%
CRE 10	12.5	14	12%

It is not clear how far from a random distribution the new calculation for each CRE really is. Future studies should replicate the simulation for the "shift" allocation, considering each educational authority as one independent unit of analysis. Maybe the outcomes will show that the allocation across "shifts" is indeed random for some educational authorities (for example, CREs 1, 2, 3, 4, 5, 7 and 10), and truly not random for others – such as CREs 6, 8 and 9.

Table 86 and 87 show the GS figures for parents' education (EducFS) and poverty (NIS). Both highlight a similar outcome: there is not much variation comparing the size effect across educational authorities. The figures reinforce the previous assumption that the "shift" allocation appears to be random in all educational authorities for both variables – EducFS and NIS.

Table 86: School “Shift” Effect for Parents’ Education for All Educational Authorities – First Segment – 2010.

	EducFS 2010	EducFS 2010	
	Non-Shift	**Shift**	**Size Effect**
CRE 1	11.5	12.5	9 %
CRE 2	19.5	20	2%
CRE 3	18.5	19.5	5%
CRE 4	21.5	22.5	5%
CRE 5	26.5	27	2%
CRE 6	21.5	22.5	5%
CRE 7	15	16	7%
CRE 8	19.5	20.5	5%
CRE 9	15.5	16.5	6%
CRE 10	16	17	6%

Table 87: School “Shift” Effect for Poverty (NIS) for All Educational Authorities – First Segment – 2010.

	NIS 2010	NIS 2010	
	Non-Shift	**Shift**	**Size Effect**
CRE 1	19	19.5	3 %
CRE 2	21	22.5	7%
CRE 3	13	14	8%
CRE 4	16	17	6%
CRE 5	17.5	18	3%
CRE 6	16.5	18.5	12%
CRE 7	21.5	22.5	5%
CRE 8	11	12	9%
CRE 9	12	13.5	12%
CRE 10	11	12	9%

The last variable, Distortion 2, was the most important to understand any impact of the "shift" allocation on school segregation. For 2010, the figures considering the entire population for the first segment suggested a relative proportional increase of 38%.

Table 88 shows the GS figures for each Educational Authority. It becomes clear that there is some variation across educational authorities, which suggests that non-random allocation could be more prevalent in specific regions of the city. The question is why?

Table 88: School "Shift" Effect for Distortion 2 for All Educational Authorities – First Segment – 2010.

	Distort2/ 2010	**Distort2/ 2010**	
	Non-Shift	Shift	Size Effect
CRE 1	20	26	30%
CRE 2	23	28.5	24%
CRE 3	18	24	33%
CRE 4	17	25.5	50%
CRE 5	16	24.5	53%
CRE 6	22	27.5	25%
CRE 7	17	25.5	50%
CRE 8	16.5	24	45%
CRE 9	18	25	39%
CRE 10	20	28.5	43%

The "shift" effect can vary from 24% or 25% (CREs 2, 6) to 50% or 53% (CREs 4, 5, 7). It is also interesting to notice that, apparently, there is a pattern in the relative proportional increase, where CREs with smaller between-school segregation (non-shift column) show a slightly higher "shift" effect (size effect column).

One way to observe this potential difference comparing different educational authorities is to calculate the standard deviation (SD) considering columns, non-shift and shift, in Tables 85 (black pupils) and 88 (Distortion 2). The calculations indicate that standard

deviation (SD) for Distortion 2 non-shift is 2.39 and 1.71 for shift. Again, SD for black pupils non-shift is 3.41 and 3.17 for shift. The differences are not very large, but indicate that the variation across educational authorities is smaller when the school "shift" is considered. This means that, on average, educational authorities with smaller figures for between-school segregation (non-shift) tend to present a higher "shift" effect.

There are two main hypotheses for this outcome. The first, states that these differences in the "shift" effect across educational authorities could simply be related to their size (total number of schools and pupils), which could produce more "noise". There is plenty of evidence showing that a sudden increase in the total number of schools can produce small rises in the index. What if the total number of school "shifts" varied across educational authorities in such a way as to produce different effect sizes?

One way to test the first hypothesis to correlate the size effect observed in Table 88 (Distortion 2) with: a) the total number of school "shifts" in each CRE; b) the nominal increase in the number of units (total number of school "shifts" in each CRE – total number of school units in each CRE); c) the proportional increase in the total number of units.

All correlation coefficients were very low and not significant even at 0.1. The correlation coefficients for the proportional increase in the total number of units, possibly the best variable to correlate with the proportional increase in the GS for each CRE, had a value of 0.007 and not significant. The data do not corroborate the first hypothesis – "statistical trick" or noise from the dataset.

The second hypothesis states that there are two independent circumstances in which school staff can select pupils: a) school enrolment; b) school "shift" allocation. The first stage, school enrolment, is influenced by many factors, including parental preference, and demand and supply of schools. Less urban areas, with less schools nearby (low supply), would present, on average, lower levels of between-school segregation (disregarding the

school “shift) – for example CREs 9 and 10. On the other hand, areas with high demand and supply of schools, such as CRE 2, will be, on average, more segregated. The second stage, school “shift” allocation, would be influenced by the initial stage creating a specific pattern.

The “shift” effect would be lower in areas in which pupils have previously been selected in a more consistent way (for example CRE 2). The explanation is that schools might already be more homogeneous, making it even harder to discriminate at the second level (school shift allocation). On the other hand, in areas where the first stage of the selection did not happen or happened in a very subtle way, the “shift” effect would be stronger. The two separate stages might be inversely associated, highlighting a new pattern of school segregation.

Table 89 shows the correlation coefficients for GS Black Pupils and Distortion 2 calculated disregarding the school “shift” (non-shift column), with the proportional increase (size effect column). The variables were chosen because they present the highest variation and nominal increase for the “shift” effect. The coefficients suggest that educational authorities with lower levels of between-school segregation (non-shift) tend to present bigger size effects for the proportional increase.

Table 89: Correlation Coefficients for GS Black Pupil and Distortion 2 with Size Effect “Shift” Allocation.

	Size Effect Black Pupil	**Size Effect Distortion 2**
GS Black Pupil Non-Shift	-0.85**	
GS Distortion 2 Non-Shift		-0.88**

** Correlation is significant at the 0.01 level (2-tailed).

Data from both variables support the second hypothesis. Apparently, there could be some relation between both stages of pupil selection (first enrolment and school “shift”

allocation). Future studies should replicate the same model presented here for different years to test for robustness. If replication confirms the initial evidence presented here, it could indicate a specific pattern of school segregation that combines demographic characteristics of the city and school selection bias.

11.1 School "Shift" Allocation: Analysing CREs 2, 4 and 10.

This section further investigates the school "shift" allocation, but this time focusing on three different educational authorities – CRE 2, 4 and 10. The description of all educational authorities and the criteria to select these three specific regions were highlighted in previous chapters. The main goal here is to observe in more detail the "shift" effect across three distinct areas in the city. CRE 2 and 10 are quite different areas in relation to the demand and supply of public municipal schools.

In order to corroborate the initial hypothesis, the size effect of the "shift" allocation will have to vary in a predetermined way: always higher in CRE 10 in comparison with CRE 2. This section uses only data for the first segment and replicates the calculations for three consecutive years – 2008-2009-2010.

Table 90 presents the figures of the relative proportional increase in GS ("shift" effect), considering four different characteristics of disadvantaged pupils for 2008. In this section, the variable age/grade distortion was recorded in two different ways: a) one or more years of distortion; b) two or more years of distortion.

Table 90: Relative Proportional Increase for the School "Shift" Effect (2008) Calculated for CRE2, CRE4, and CRE10 – First Segment.

	Black Pupil	**EducFS**	**NIS**	**Distortion1**	**Distortion2**
CRE 2	+ 7%	-	+ 5%	+ 27%	+ 12%
CRE 4	+ 11%	+ 5%	-	+ 42%	+ 26%
CRE 10	+ 19%	+ 5%	+ 5%	+ 82%	+ 23%

The figures corroborated the initial hypothesis. As expected, the variation of the size effect across educational authorities, considering parents' education and poverty, is very

small. However, the figures for age/grade distortion and black pupils show bigger size effects, and the variation suggests that the school "shift" allocation is more selective in CRE 10. Age/grade distortion for one or more years is three times smaller in CRE 2 compared with the figures in CRE 10. These are quite large differences that should be further investigated. Two questions should be asked: what elements are associated with an increase in the "shift" effect? Is there any adverse effect of clustering pupils with shared characteristics in the same school "shift"?

The next step is to analyze the figures for 2009 and 2010, and observe if the interpretation made for 2008 is still valid. Tables 91 and 92 show the size effects of the shift allocation for 2009 and 2010. At least for both measurements of age/grade distortion, it is possible to state that CRE 2 presents consistently lower levels of "shift" effect. Taking the simulation figures presented in the previous chapter as a reference, it is not clear that the allocation of pupils across school "shifts" in CRE 2 for Distortion 2 is far from a random distribution. Perhaps, the non-random allocation considering Distortion 2 is a reality only in specific regions of the city – for example CRE 10. These are important questions that should be fully investigated in future studies.

Table 91: Relative Proportional Increase for the School "Shift" Effect (2009) Calculated for CRE2, CRE4, and CRE10 – First Segment.

	Black Pupil	**EducFS**	**NIS**	**Distortion1**	**Distortion2**
CRE 2	+ 10%	-	+ 7%	+ 19%	+ 10%
CRE 4	+ 8%	+ 5%	+ 7%	+ 40%	+ 20%
CRE 10	+ 11%	+ 9%	+ 14%	+ 80%	+ 22%

Table 92: Relative Proportional Increase for the School "Shift" Effect (2010) Calculated for CRE2, CRE4, and CRE10 – First Segment.

	Black Pupil	**EducFS**	**NIS**	**Distortion1**	**Distortion2**
CRE 2	+ 8%	+ 2%	+ 7%	+ 70%	+ 24%
CRE 4	+ 9%	+ 5%	+ 6%	+ 127%	+ 50%
CRE 10	+ 12%	+ 6%	+ 9%	+ 136%	+ 42%

The "shift" effect for Distortion 1 is consistently higher for CRE 10, and corroborates the initial hypothesis. A non-random allocation across school "shifts" is more prevalent among areas where the initial levels of between-school segregation (disregarding the "shift" allocation) is smaller. Apparently, both stages of pupil selection (school enrolment and school "shift" allocation) are independent, but somehow connected.

12. IMPACT OF SPECIAL CLASS POLICY ON SCHOOL SEGREGATION

The Rio de Janeiro municipal public network has designed a specific policy to track pupils with consistent low achievement over the years. These pupils are placed in different classrooms or school "shifts" and receive different type schooling. It is possible to divide this tracking policy into two types: a) "*realfabetização*"; b) PEJA[20].

"*Realfabetização*" was designed to provide extra help for pupils who have not successfully completed literacy at the end of 3rd grade – 8 years old. This particular policy only affects pupils at the first segment and aims to reintegrate than into regular classes after a certain period of time – usually around 5th grade. PEJA targets pupils with long history of retention or drop outs. It is divided in two types: PEJA I – corresponds to first segment; PEJA II – corresponds to second segment. Most pupils sent to PEJA are enrolled in PEJA II and have struggled to be approved in different grades all along first and second segment or have a history of drop out.

The research design captures the "net effect" of both Special Class policies on school segregation. The idea is quite simple. The nominal levels of segregation calculated for the first and second segment, considering only pupils enrolled in regular classes (1st-9th grade), will be compared with new figures calculated using pupils in regular classes and "*realfabetização*" or PEJA. Any differences in between-school segregation can be attributed to this particular tracking policy. Because, some schools use a specific "shift" to allocate all Special Class kids, all indices were calculated considering each "school shift" an independent unit.

This section replicates all calculations for two consecutive years (2009-2010) to test for robustness. All Special Class modalities were divided into two groups to be clustered with pupils enrolled in regular classes at first and second segment. All pupils tracked to

[20] PEJA –Programa de Educação de Jovens e Adultos. See: http://www.ejario.com/

"*realfabetização*" were clustered with its pairs enrolled in the first segment (1st to 5th grade) and pupils at PEJA II were clustered into the second segment. Table 93 presents the proportion of pupils enrolled in Special Class for each segment in 2009 and 2010.

Table 93: Proportion of Pupils Enrolled in Special Class for First and Second Segment.

	2009	**2010**
First Segment - *Realfabetização*	4%	5%
Second Segment – PEJA	12%	15%

The proportions suggest that the potential impact of Special Class on school segregation can be higher for the second segment. Will the outcomes confirm this premise? Tables 94 and 95 present D and GS figures for the first segment years 2009 and 2010 respectively. It is expected that both tables present a similar outcome.

Table 94: Segregation Index (%) and Dissimilarity Index (%) Special Class Effect (2009) – First Segment.

	First Segment 2009	**First Segment + Special Class 2009**	**Relative % Difference**
GS Black Pupil	17,5	17	-3%
D Black Pupil	19,5	19,5	-
GS EducFS	22	21,5	-2%
D EducFS	30	30	-
GS Poverty	17,5	17	-3%
D Poverty	24,5	24,5	-
GS Distortion 2	26	27	+ 4%
D Distortion 2	29	31	+ 7%

Table 95: Segregation Index (%) and Dissimilarity Index (%) Special Class Effect (2010) – First Segment.

	First Segment 2010	First Segment + Special Class 2010	Relative % Difference
GS Black Pupil	17	17	-
D Black Pupil	19	19	-
GS EducFS	20,5	20,5	-
D EducFS	28,5	28,5	-
GS Poverty	18	18	-
D Poverty	25	25	-
GS Distortion 2	27	27	-
D Distortion 2	30	31,5	+ 5%

The outcomes in both tables are quite similar. The results suggest that the general effect of Special Class on school segregation in the first segment is null. All characteristics of disadvantaged for the year 2010 show the exact same nominal levels of segregation on both methods of calculation – with and without pupils enrolled in "*realfabetização*". The outcomes for 2009 are very similar with minor variations (positive and negative) depending on the variable observed.

There are two reasonable explanations for this outcome. First, it should be remembered that the total proportion of pupils enrolled in "*realfabetização*" is small – around 4 or 5%. The second reason is that, possibly, most of these pupils are integrated in a school "shifts" with other pupils from regular classes – 1st to 5th grade. If that is the case, it should be expected a minor impact on school segregation. Future studies should investigate more deeply the nature of this policy, that appears to cluster pupils with similar abilities into classes – Streaming (Ireson; Hallam, 2001)

The next step is to calculate the "net effect" of Special Class for second segment pupils. Will the outcomes be consistent? Tables 96 and 97 present D and GS figures for the

second segment – 2009 and 2010. The outcomes are different from the ones observed in the first segment.

Table 96: Segregation Index (%) and Dissimilarity Index (%) Special Class Effect (2009) – Second Segment.

	Second Segment 2009	Second Segment + Special Class 2009	Relative % Difference
GS Black Pupil	14,5	14,5	-
D Black Pupil	16,5	16,5	-
GS EducFS	21,5	21	-2%
D EducFS	27,5	28	+ 2%
GS Poverty	16,5	20	+ 21%
D Poverty	23	27	+ 17%
GS Distortion 2	19	27	+ 42%
D Distortion 2	25	41	+ 64%

Table 97: Segregation Index (%) and Dissimilarity Index (%) Special Class Effect (2010) – Second Segment.

	Second Segment 2010	Second Segment + Special Class 2010	Relative % Difference
GS Black Pupil	14,5%	14,5%	-
D Black Pupil	16,5%	16,5%	-
GS EducFS	20,5%	20%	-2%
D EducFS	26,5%	27%	+ 2%
GS Poverty	17%	20%	+ 18%
D Poverty	23,5%	27%	+ 15%
GS Distortion 2	22,5%	27%	+ 22%
D Distortion 2	30%	42%	+ 40%

Two variables (Poverty and Distortion 2) show a consistent growth in school segregation when adding pupils enrolled in PEJA II. Black pupils and EducFS show the same

outcome observed previously in the first segment – no impact. Since PEJA II tracks pupils with many years of age/grade distortion, some effect on school segregation was expected. The surprise is to also observe an impact in the variable that measures poverty. The relative proportional increase is quite substantial, around 18%. In real figures, something close to 2,700 pupils per year.

How can it be explained that D and GS for poverty show an increase and black pupil and EducFS do not? Despite the fact that PEJA tracks pupils with a long history of retention (usually 3 or more years), it is known that there is not a clear cut-off point to select pupils. It means that PEJA I or II can have pupils with 3 or 4 years of distortion, but pupils with the same history could be enrolled in regular classes. The criteria to select pupils are not very clear and selection bias is a potential risk. In an interview, a staff member from the educational department indicated that the decision to send pupils to "*realfabetização*", PEJA I or PEJA II is made by the teachers and school principal (Interview with Rio de Janeiro Educational Department Staff Member – 23 August, 2011).

However, a more plausible hypothesis refers to the quality of data from pupils enrolled in Special Classes – more specifically PEJA. If it is true that missing data is not randomly distributed among regular classes and special classes, it is reasonable to assume that GS and D calculations showed on Tables 94, 95, 96 and 97 could be misleading.

Table 98 shows the proportion of missing data for pupils' colour and parents' education considering four different groups: a) pupils enrolled in regular classes in the first segment; b) pupils enrolled in Special Classes in the first segment ("*realfabetização*"); c) pupils enrolled in regular classes in the second segment; d) pupils enrolled in Special Classes in the second segment (PEJA II).

Table 98: Proportion of Missing Data Comparing Pupils Enrolled in Regular Classes and Special Class – 2009 and 2010.

	2009/ Colour	**2010/ Colour**	**2009/ Educ**	**2010/ Educ**
1st Segment Regular Classes	3,3 %	3,1%	7,8%	8,4%
1st Segment *Realfabetização*	5,8 %	5,9%	8,0%	9,3%
2nd Segment Regular Classes	4,6%	4,7%	10,4%	9,9%
2nd Segment PEJA II	4,2%	4,9%	37,8%	30,1%

Considering the first two rows (1st segment regular classes and "*realfabetização*"), it is possible to observe a small difference in both variables, highlighting that pupils enrolled in "*realfabetização*" present worst records. However, the figures for the second segment suggest even more caution. Pupils enrolled in PEJA II have three times more missing data for the variable parents' education compared with its pairs on regular classes. D and GS calculations for parents' education in Tables 96 and 97 (second segment) should be interpret with great caution. It is also intriguing that the attrition rate does not affect the other variable (pupil colour) the same way.

***

This is the first model in the thesis that has considered pupils enrolled in Special Class for any calculations of the indexes. Another way to observe any potential impact of the key variables of disadvantaged is to follow pupils individually and estimate their chances of being tracked to "*realfabetização*" or PEJA. This model will not only analyse the regression coefficients, but also compare the proportion of missing in two groups: a) pupils that have not being tracked to Special Class; b) pupils that have been tracked to Special Class.

Two different cohorts were chosen: a) 1st grade 2006; b) 6th grade 2006. The idea is to follow the entire cohorts from 2006 to 2009 and create a dummy variable discriminating those who were tracked to "*realfabetização*" or PEJA from others that have not. Both models, for first and second segment, should consider a third outcome that can influence the regression coefficients: retention. Thus, a total of four regressions will be presented, two for each segment. The first model considers only pupils who were sent to Special Class or follow a regular flow along all education transitions (have not been retained). The second model considers all pupils in each cohort (including those that have been retained).

Table 99 presents the variables used in all four regressions. The idea is to use the same variables used in previous models, plus gender. The hypothesis states that potentially disadvantaged pupils will have a higher probability of being tracked to Special Class program.

Table 99: Variables used in the Logistic Regressions:

Variables	Type	Description
Dependent Variables		
Pupil's enrollment first segment	Dummy	Indicate if pupil was enrolled in "*realfabetização*" at 4th grade in 2009. (1=Yes/0=No)
Pupil's enrollment second segment	Dummy	Indicate if pupil was enrolled in PEJA at 9th grade in 2009. (1=Yes/0=No)
Independent Variables		
Parental Education	Dummy	Indicates if the pupil's parents have not finished fundamental education (1=Yes/ 0 =No)
Colour – Black	Dummy	Indicates if the pupil is black (1=yes/0=no)
Poverty (NIS)	Dummy	Indicates if the pupils family participates in government program for extra income (1=yes/0=no)
Sex – Boy	Dummy	Indica se aluno e do sexo masculino (1=Yes/0=No)

Table 100 presents the four models (two for each segment) of the logistic regression. All models estimate the chance to be enrolled in 2009 in a Special Class program: a) "*realfabetização*" for first segment; b) PEJA II for second segment.

Table 100: Logistic Regression Estimating Pupil's Chance to be enrolled in Special Class

	First Segment		Second Segment	
	Model 1	**Model 2**	**Model 1**	**Model 2**
	Exp(B)	**Exp(B)**	**Exp(B)**	**Exp(B)**
Parental Education	1,71**	1,40**	2,16**	1,98**
Colour – Black	1,47**	1.30**	1,54**	1,45**
Poverty (NIS)	1,69**	1,62**	0,46**	0,45**
Sex - Boy	1,75**	1.55**	1,82**	1.70**
Constant	0,045	0,039	0,089	0,077
Cox Nell R Square	0,02	0,01	0,03	0,02
N	40.134	52.322	43.130	51.761

** Coefficient is significant at the 0.01 level;

The regression coefficients for pupils enrolled in the first segment suggest that potentially disadvantaged pupils are more likely to be sent to "*realfabetização*" program. All four variables are significant at 0.01 and the regression coefficients corroborate the initial hypothesis. However, the figures of the Cox & Nell R Square (model summary) suggest that the independent variables chosen (characteristics of disadvantage) are not good predictors. As a matter of fact, the "Classification Table" generated b SPSS shows that all predicted values ("group membership") indicate the value "zero", or pupils enrolled in regular classes. This is an important indication that the models presented are not efficient to estimate chance of access to Special Class program – for both segments.

The figures for the second model (including pupils that have been retained) are more conservative, but still show an impact of all four variables of disadvantaged. The regression coefficients suggest a different outcome for one specific variable – poverty. The figures

indicate that families that are registered in social programs for extra-income (NIS) have a lower probability (-117%) to be sent to PEJA II. This is an unexpected outcome that contradicts most of what is known in the field of education. PEJA was design to track pupils with many years of age/grade distortion and it would be expected that poor pupils would be more likely to get retained and, therefore, be sent to PEJA.

It is important to highlight that, despite the fact that all covariates are significant, (less than 0.01), the model summary suggests that these specific independent variables are not good predictors to estimate pupil's chance to be sent to PEJA II. It should be asked: what the regression coefficients really mean? What are they measuring? The same model for the second segment was replicated with a different cohort (1st grade 2007) and the results were fairly stable. Once again, the coefficients for poverty suggested a negative impact in the chances of been sent to Special Class.

***

A different model could help understanding the odd outcome presented on Table 100. This time, a logistic regression will estimate pupil's chance to be retained at any grade in different educational transitions. The focus here is not on age/grade distortion, but really to track pupils and discriminate those that have been retained one or more years from others that have followed a regular flow. Retention is a key variable here, because only pupils with many years of age/grade distortion are eligible to be sent to PEJA. Observing the potential impact of the independent variable on retention could help to clarify the odd outcome on the previous table.

Table 101 presents the regression coefficients with the same independent variables presented on Table 100, using the same cohorts: a) first segment 1st grade 2006; b) second

segment 6th grade 2006. The only difference is the dependent variable that has changed to: a) retained one or more times = 1; b) not retained = 0.

Table 101: Logistic Regression Estimating Pupil's Chance to be Retained.

	First Segment	**Second Segment**
Parental Education	1,94**	1,54**
Colour – Black	1,53**	1,37**
Poverty	1,18**	1,00+++
Sex - Boy	1,57**	1,50**
Constant	0,227	0,171
Cox Nell R Square	0,04	0,01
N	46.131	47.597

** Coefficient is significant at the 0.01 level;

For the first segment, it is possible to observe that all covariates suggest that being potentially disadvantaged increases the chances of retention after four transitions. It is also possible to observe that poverty is the covariant with the lowest impact in the model, indicating that pupils living in poor families have 18% more chance of getting retained at least once. On the other hand, low parental education[21] is the variable with the biggest impact, with an increase in the probabilities of 94%.

For the second segment, the regression coefficients suggest a smaller impact. Surprisingly, poverty is not significant suggesting that it may not influence pupil's chance to get retained. All other variables are significant at 0.01 and suggest that disadvantaged pupils (black, boys and low parental education) have a higher probability of being retained. This particular outcome, despite the fact that was not expected, corroborates somehow with the figures in Table 100.

[21] Parents that have not finished fundamental education.

It is not clear why the variable poverty is not associated with retention.[22] Future studies should address this question revisiting the specific data. There are two main hypotheses: a) error of measure; b) the variable does not discriminate well between poor and non-poor.

Since it started in 2004, the program called *Bolsa Família* has grown from 6.5 million families assisted to 14.1 million in 2013, considering the entire country. The government estimates that another half a million families will be included in 2014. These figures indicate that around 25% of the entire population is affected by the program. It is important to highlight that the growth in the total number of beneficiaries was accompanied by historical low rates of unemployment and fairly stable economic growth.

Recent study from the *Ministério de Desenvolvimento Social e Combate à Fome* indicates that 72% of the families assisted by the program work. Data from 2013 indicates that *Bolsa Família* assisted 242.926 families in the municipality of Rio de Janeiro and the dataset used in this research indicates in 2010 a total of: a) 173,555 NIS_Sum; b) 232,636 NIS_Max. The figures are not very different from the one observed in the official data from the website[23] and suggests that the first hypothesis (error of measure) is less likely.

Maybe the explanation relies on the intense growth of beneficiaries in the last few years. If it is true that the majority of the beneficiary work and have other sources of income, it would be reasonable to assume that the difference between groups (beneficiary and non-beneficiary) might not be very large – in terms of total income. This could explain why this particular variable does not seem to be associated with retention in the second segment. It could be argued that this particular variable is not discriminating well between poor and non-poor.

[22] If NIS was the only independent variable in the model, the coefficient would be significant at 0.05 and suggest a small positive impact – Exp(B) 1.05.
[23] See: http://aplicacoes.mds.gov.br/sagi/RIv3/geral/relatorio.php#Cadastro Único

Discussion

It could be argued that the concentration of potentially disadvantaged pupils in one area (or in a certain school) can be efficient when seeking to implement focused policies as ameliorative packages (EGGRES, 2005). Special Class policy is the only designed educational policy in Rio de Janeiro that intentionally track and cluster pupils with specific characteristics. All pupils selected have low academic performance or a history of drop out over many years and are deemed "trouble kids" or pupils with learning difficulties.

The main question that should be asked is: what are the benefits of Special Class program for disadvantaged pupils and for the educational system as a whole? Unfortunately, so far there is no hard evidence, or it could be said, no evidence at all about the effects of such policy on pupils' attainment levels and future prospects in life. This is not acceptable, especially if poor pupils, or others with learning difficulties, are being systematically clustered into different class or school "shifts".

Table 93 indicates that PEJA II corresponds to more than 10% of all registrations for the second segment. The high figures suggest that not only pupils with sever learning difficulties are being tracked, but also pupils with minor difficulties that could, in a different scenario, be enrolled in regular classes.

Throughout the thesis, retention has been highlighted as a major problem in basic education in Brazil. It affects kids from all different backgrounds, but it is more prevalent among disadvantaged pupils and those attending class at low performance schools. What is quite unique in Brazil are the retention rates. Rio de Janeiro has one of the largest and more homogeneous public network in the country and, despite all the good characteristics, around 25% of all pupils have two or more years of age/grade distortion.

Retention in Brazil is not an individual problem that simple prevent the pupil to pass to the next grade along with his/her colleagues. In such high rates, it can create problems for

everyday school activities. Dealing with large age/grade discrepancy in the same class can be highly challenging for teachers. Relationship problems and social interactions among children and teenagers can be criticized by the parents, the pupils itself and even educators. So far, apparently, the "solution" has been to separate low achievers (pupils with low academic performance over the years) from others that have followed a regular flow.

Two outcomes in the thesis corroborate this conclusion. The "school shift" allocation and the Special Class policy, highlight that age/grade distortion is the most important variable to understand any potential impact of both policies on school segregation. It is important to highlight that the "shift" allocation is simply an "administrative practice". There is no clear guidance that states any intention to cluster pupils with shared characteristics in different "shifts". Special Class has the clear intention to deal with a pedagogical problem: pupils with many years (usually three or more) of age/grade distortion.

The impact of both policies on school segregation was revelled for the first time in this study. Still, the potential benefits for the most disadvantaged are unclear. Informal conversations and formal interviews with parents and teachers suggest that being sent to PEJA or moved to the afternoon or night school "shift" is a clear indication that pupils will have fewer educational opportunities. Future studies should address this question, using appropriate research design to establish causality.

13. EVIDENCE OF SCHOOL COMPOSITION EFECT

This chapter presents evidence about the potential impact of school composition on pupils' achievement in Rio de Janeiro municipal public schools. This section is divided into three parts. Initially, it presents the basis of the theory about the school composition or school-mix effect and an overall view about previous studies in different countries. The second section discusses in detail the type of data available, their main limitation and the means designed to test the theory. The last part presents the results and highlights the implications for policy-makers and educational researchers.

Since Coleman's (1966) pioneering report, there has been increasing concern about the social composition of schools and its relation to pupils' outcomes. Coleman was most interested in understanding the nature of school effectiveness and appropriateness of policies to raise school effectiveness. The studies that followed that report identified two typically ideal positions. The first claims that school effectiveness is a function of school management and teacher performance, while the latter claims that social factors (social background) determine pupils' outcomes in schools (Trupp; Hirsch, 2006). In this respect, the social composition of schools can be seen as one more factor that can determine pupils' outcomes.

The school composition effect will be taken as an intermediate theory, in between those featured in the schematic model above. Dealing with "school-mix" means taking into account both school management (decisions on enrolment, pupils and teacher allocation, as well as the array of specific activities performed in each school) and pupils' social background (language, motivation and all other cultural and economic characteristics of pupils' social environments). The social composition of a school stems from these two sources.

The subject of school segregation should be taken seriously by educational researchers. It is closely linked to the debate about the quality and equity of educational

systems and to the subject of social justice. In modern societies, public schools have become central institution that should provide (at least in theory) equal opportunities for all individuals, enabling social mobility. Evidence from different educational systems suggests that clustering pupils with similar characteristics can influence how they are treated at school, the quality of teaching, aspiration to advanced education subsequent to the compulsory level, and an increasing association between academic achievement and socio-economic status. (Haarth et al., 2005; EGGRES, 2005; Rosenthal; Jacobson, 1968; Brito; Costa, 2010).

The findings presented in this chapter can be of relevance for policy-makers and educational researchers, since whether pupil composition has a significant impact on school performance assumes a central position in the debate about school effectiveness. Usually, the school composition effect does not enter into official judgments about school performance. This has even more relevance for countries, or cities like Rio de Janeiro, that have implemented high-stake policies. It may be that schools, or even teachers, are being wrongly held responsible for their school's performance. The school composition effect could explain part of the variance that is usually attributed to school management and teacher performance. The evidences and arguments presented here are intended to enhance prudent use of educational assessment.

***

The key consideration in this chapter is whether a disadvantaged pupil in a predominantly high social class school will perform better than one in a predominantly low social class. There is an obvious methodological challenge for researches to test the theory, since it is very hard to control for all other elements that can influence pupils' outcome. For practical and ethical reasons, experimental approaches are not feasible and most of the

research conducted use qualitative observation or secondary data to estimate any impact of social composition of schools.

It is clear that schools differ in their proportions of potentially disadvantaged pupils, and these figures are associated with overall levels of attainment. Thus, a school with pupil intake with high prior attainment and low levels of family poverty (for example), generally produces higher outcomes compared to schools with a less favorable pupil intake profile. The nature of school intake should be considered in a value-added analysis to estimate the "school-effect".

It is reasonable to assume that some schools function better than others and produce better results (pupils' attainment level). The next step is to observe the characteristics of the more effective schools. It is possible that they were disproportionately distributed, with large clusters of students with desirable characteristics. In contrast, the less effective schools might be those with disproportionately large clusters of disadvantaged pupils (perhaps family poverty). This is the claim for a school composition effect, which means that at least part of the school effect can be explained by a particular mix of pupils (Harker, 2004; Gorard, 2006).

The literature about school-mix effect has been extensive in many countries, and, while it is probably correct to state that most evidence favors the existence of such an effect, there is no consensus (Thrupp, 1997; Nash, 2003). Most likely, the basis of the disagreement relies on theoretical and methodological issues.

Not many studies have given substantive considerations on how (the mechanism) school composition might affect pupil outcome. Trupp (1999), in an ethnographic study on working and middle class schools, outlined three different mechanisms: a) peer subculture; b) teaching and the curriculum; c) school policies.

The author claims that peer subcultures might either support school aims and processes or resist them. Schools with a disproportionate share of disadvantaged pupils (for

example, pupils living in poverty) are more likely to face classroom disruption. In turn, teaching and curriculum might be changed to seek to capture their interest, or deal with daily problems in school. At the policy level, more time is spent on issues of discipline and ways of funding activities that are not related to traditional disciplines. Trupp (1999) believes that school composition has a significant impact on school and individual performance.

Nevertheless, other authors claim that there might also be an indirect effect of clustering disadvantaged pupils in specific schools. Perhaps the most important point is the relationship between teacher allocation and school composition (Lauder et al. 2010). Some might argue that "ghetto schools", or the so-called "sinking schools", might recruit less experienced or less motivated teachers. Often, and this is the case of Rio de Janeiro public network, more experienced teachers have priority in choosing their school of preference, and this creates the basis for an association between teacher profile and school composition. This is called an "indirect" effect of school composition on pupils' performance, because it impacts other elements (in this case, the quality of teaching) that are also associated with school effectiveness.

There are two major issues with respect to the methodology: sampling and techniques. Thrupp, Lauder and Robinson (2002) have established what they agree would be ideal with respect to modelling and sampling for school-mix effect. It is reasonable to assume that at least part of the disagreement regarding the effects of school composition can be attributed to differences in the techniques used in previous studies.

The authors highlight a list of desirable criteria to estimate the school-mix effect. First, the sample should include schools from both ends of the socio-economic spectrum. This is relevant because school composition effects are more likely to appear in more segregated schools. For example, the peer subculture effect is more probable in schools that concentrate a large proportion of pupils with shared characteristics.

Second, a full set of individual variables, including prior attainment, need to be included in the model. The authors suggest measurements of social class and not simply measurements of Free School Meal, which has been typically used in the UK and the U.S. Kounali et al (2007) have shown that Free School Meals Measurement is not a good predictor of subsequent performance.

Third, the authors highlight that there should be measurements to capture the possible correlations among the theoretical dimensions of school composition (for example, school organizational and management processes, peer group processes, teaching, etc.). Unfortunately, there are no available measurements in the datasets for these dimensions. Fourth, the studies should, whenever possible, be longitudinal. Lastly, the studies should conduct their analysis using appropriate statistical methods that respects the dependence structure characterizing such data (for example, multi-level modelling).

The next section will discuss in detail the data used to test the hypothesis about school-mix effect and their main limitations. Not all the criteria described here are met in the study for two main reasons. First, there is the limitation of the variables available. There is not much information about how schools function and the teaching in the classroom. In addition, there is a limitation regarding the data quality, more specifically, the large number of missing data.

13.1 Research Design: How to Test the School-Mix Effect Theory

The variables available to test the school-mix-effect theory are limited. There are three major sources of problems. First, there is no good information about school management and classrooms teaching. Second, despite the fact that all pupils in a specific cohort are invited to participate in *Prova Rio*, there is large proportion of missing data – around 18% – referring to the number of pupils who did not take the test or there is lack of or no information at all on their family background. Because the model is longitudinal, the total number of pupils excluded from the analysis can reach 30%. Third, the standardized measure for pupil's attainment (*Prova Rio*) was designed using Item Response Theory to assess school performance. It is not clear at this point, if the total number of questions in each individual test covers the entire scale of (IRT). If not, the attempt to estimate individual scores might be misleading.

The limitations impose simpler models that will search for associations between the regression residuals (difference between the predicted and observed scores) for pupils' achievement and the proportion of disadvantaged pupils in each particular school.[24] Three distinct characteristics of potentially disadvantaged pupils were chosen based on previous research in Rio de Janeiro public schools: a) non-white pupils; b) pupils' parents that have not finished high school; c) families that participate in the federal program, *Bolsa Família* (family income supplement – proxy of poverty) (Bartholo, 2013).

Three complementary longitudinal models are presented. The first (Type 1) uses standardized tests of *Prova Rio*, for mathematics and language (Portuguese), at pupil level to estimate the pupils' achievement using prior attainment and other key variables as covariants: 1) sex; 2) colour; 3) parental education; and 4) poverty. The scores for mathematics and

[24] Another valid approach is to use the Segregation Ratio (expected fair share of potentially disadvantaged in each school) in the correlations. The index (SR) was calculated considering the entire network as one unit of analysis and each educational authority as one independent unit of analysis. The correlation coefficients between the regression residuals and SR are very similar to the ones observed using the proportion of disadvantaged pupils in each school.

Portuguese will be estimated separately. The individual residuals from the linear regressions will be correlated with the proportion of disadvantaged pupils in each school. In order to corroborate the initial hypothesis, the correlation must be negative and constant among all variables of the potentially disadvantaged.

The second model (Type 2) aggregates the individual residual from the same regressions at school level. The mean scores of the individual residuals are once again correlated with the proportion of disadvantaged pupils calculated for each school. In order to corroborate with the initial hypothesis, the correlation coefficients must be negative and the strength of the association higher compared with the previous model – pupil level analysis.

The third model (Type 3) uses school-level figures to replicate the exact same model. The mean scores of *Prova Rio* for one year are used to predict the scores for the next year in a cross-sectional design. Once again, the residuals from the regression will be correlated with the proportion of disadvantaged pupils. In order to corroborate with the initial hypothesis, the correlation coefficients must be negative and the strength of the association higher compared with the previous model – pupil level analysis (Type 1).

The models presented are not a definitive test and do not "prove" the existence of school-mix effect, mainly because they do not rule out one important alternative explanation: school effectiveness. Schools overrepresented with pupils with desirable characteristics could, on average, despite any claim of school-mix effect, be more effective than others. The opposite could also be true, schools overrepresented with disadvantaged pupils could be less effective. The models presented here provide the basis to demonstrate the plausibility of the school-mix effect theory.

All models will be replicated in four different cohorts, in a total of eight independent test, considering mathematics and language to search for robustness. Table 102 presents the characteristics of the data used.

Table 102: Measurements of *Prova Rio* Used to Test the School-Mix Effect Theory.

	Year 2009	**Year 2010**	**Year 2011**
Cohort 3rd Grade 2009	First Measure	Second Measure	
Cohort 7th Grade 2009	First Measure	Second Measure	
Cohort 3rd Grade 2010		First Measure	Second Measure
Cohort 7th Grade 2010		First Measure	Second Measure

A quick review of the criteria highlighted by Thrupp, Lauder and Robinson (2002) shows that the model proposed here fulfil part of the demands. There are data available for all schools that enhance the validity of the study. More importantly, the study replicated the analysis for four different cohorts, which can be of interest when observing the consistency of the outcomes.

The design is longitudinal, but only tracks pupils for one year. Future studies will track pupils for longer periods (for example, from the 4th to the 7h grade). The downside is the lack of measurements related to school management and classroom teaching. This limitation should not be underestimated, since it makes it really hard to assess school effectiveness. The choice here was to present a more conservative model and replicate throughout different cohorts to observe the consistency of the results. Are the residuals from the linear regressions correlated with the proportion of disadvantaged pupils?

13.2 Variable Construction and Missing Data

The regressions use data from *Prova Rio* and pupils' characteristics such as colour, parents' education, sex and poverty. These variables were described in detail by Bartholo (2013) and had been used in previous research to measure between-school segregation in Rio de Janeiro. In order to minimize the problem of missing data, all pupils who were retained at the end of school year (in the first measure of Prova Rio)[25] were tracked and incorporated into the sample. The main source of missing data is the *Prova Rio* dataset. Table 103 presents the proportion of missing data for key variables in each cohort.

Table 103: Proportion of Missing Data for Prova Rio, Pupil Colour and Parents' Education.

	First Measure *Prova Rio*	Second Measure *Prova Rio*	Colour	Parental Education
3rd Grade 2009	17 %	25%	3%	7%
7th Grade 2009	18%	37%	4%	9%
3rd Grade 2010	19%	25%	5%	10%
7th Grade 2010	24%	35%	6%	10%

The large proportion of missing data is a real problem. There is no way to ensure that pupils who did not take the test are equivalent to those who have successfully completed the assessment. Entering the independent variables in the model, the coefficients will be estimated using between two thirds and three quarters of the total population. Because of this, one strategy would be to simply use prior attainment – by far the best predictor in the model – to estimate the second measurement. For this reason, all the regressions for Models Type 1 and 2 (pupil level analysis) were conducted twice, that is, first ignoring and then using pupils'

[25] Retention is still a major problem in public education in Brazil. As an example, 11% of all pupils enrolled in the 3rd grade, 2009, were retained at the end of the school year.

characteristics (sex, colour, parental education and poverty). Tables 104 presents the variables used in Type 1 and 2 models and Table 105 the variables used in Type 3 model.

Table 104: Variables used in the Linear Regression – Pupil Level:

Variables	Type	Description
Dependent Variable		
Prova Rio 2nd Measurement	Continuum	Indicate pupil's achievement in the 2nd measurement of *Prova Rio*
Independent Variables		
Prova Rio 1st Measurement	Continuum	Indicate pupil's achievement in the 1st measurement of *Prova Rio*
Parental Education	Dummy	Indicates if pupil's parents have not finished High School (1=Yes/ 0 =No)
Colour Non-White	Dummy	Indicates if pupil is non-white (1=Yes/0=No)
Poverty	Dummy	Indicates if pupil's family participates in government program for supplementary income (1=yes/0=no)
Sex – Boy	Dummy	Indicates if pupil is a boy (1=Yes/0=No)

Table 105: Variables used in the Linear Regression – School Level:

Variables	Type	Description
Dependent Variable		
Mean Score *Prova Rio* Second Measure	Continuum	Indicate the mean score of *Prova Rio* – 2nd measurement
Independent Variables		
Mean Score *Prova Rio* First Measure	Continuum	Indicate the mean score of *Prova Rio* – 1st measurement

13.3 Is There any Evidence of School-Mix Effect?

Pupil Level Regression

The results will be presented in a specific order. The outcomes of Type 1 and 2 models are presented together. Table 106 presents the regression coefficients for the cohort 3rd grade, 2009. It becomes clear that prior attainment (first measurement in 2009) is the best predictor in the model. All variables referring to the disadvantaged were significant at 0.01 and suggest that potentially disadvantaged pupils scores are significantly lower in *Prova Rio*. Even so, it is worth mentioning that none of the characteristics of the disadvantaged prove to be good predictors in the model. The preliminary results corroborate previous findings from international studies about similar variables regarding the disadvantaged – for example, Free School Meals in England (Kounali et al, 2007).

Table 106: Linear Regression Estimating Proficiency in *Prova Rio* 2010 – Cohort 3rd Grade 2009.

	Language		**Mathematics**	
	Step 1	**Step 2**	**Step 1**	**Step 2**
	Standardized Coefficients			
Prova Rio First Measure	0,565**	.525**	.602**	.575**
Parental Education	-	-.096**	-	-.08**
Colour Non-White	-	-.056**	-	-.043**
Poverty	-	-.062**	-	-.048**
Sex – Boy	-	-.090**	-	-
Constant	78,1	103.95	90,64	105,46
N	44.970	40.604	45.086	40.699

** Coefficient is significant at the 0.01 level; All coefficients not significant at 0.05 level were excluded

For the purpose of the research, the regression coefficients are important only to observe if they are significant and have a potential impact (for example, disadvantaged pupils present lower overall attainment levels). The real interest is to correlate the residuals from the

regressions with the proportion of disadvantaged pupils. The correlation coefficients can indicate, for example, if pupils with consistent positive residuals (the predictions were underestimated) are disproportionally enrolled among schools with lower proportions of disadvantaged pupils. In order to corroborate with the initial hypothesis, the correlation coefficients must be negative and constant among all variables referring to the disadvantaged. Table 107 presents the correlation coefficients for the residuals of four regressions presented in Table 106 (cohort 3rd grade, 2009), with the proportion of disadvantaged pupils for all schools.

Table 107: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 3rd Grade 2009.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.10**	-.10**	-.10**
Language Step 2	-.08**	-.08**	-.09**
Mathematics Step 1	-.08**	-.09**	-.08**
Mathematics Step 2	-.06**	-.07**	-.07**

** Correlation is significant at the 0.01 level (2-tailed);

All correlation coefficients are significant at 0.01 and corroborate the initial hypothesis about a potential impact of school composition effect on pupil attainment. Nevertheless, the strength of the association is low, suggesting that the residuals in the predictions are not very associated with the proportion of disadvantaged pupils. Perhaps, the most important evidence is the fact that all correlations were negative and significant, suggesting a pattern.

Thrupp, Lauder and Robinson (2002) highlight that the school-mix effect is more likely to influence schools that concentrate high proportions of pupils with shared characteristics (disadvantaged or non-disadvantaged). In this sense, a complementary test will

correlate the residuals from the regressions and proportions of disadvantaged pupils, considering only the more segregated schools.

The choice here was to select the top and bottom quartile of a descriptive analysis regarding the proportion of poor families in each school. Previous calculations using the Segregation Ratio for all characteristics of disadvantaged indicated that poverty was the variable with the highest correlations among all.[26] It means that schools segregated by poverty are more likely to be segregated by other characteristics of disadvantaged. In order to corroborate with the hypothesis, the association should be higher (still negative) compared with the figures in Table 107 (cohort, 3rd grade 2009 for all schools). Table 108 presents the figures for the new correlations considering only the more segregated schools (top and bottom quartile).

Table 108: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 3rd Grade 2009.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.14**	-.14**	-.13**
Language Step 2	-.11**	-.11**	-.11**
Mathematics Step 1	-.11**	-.13**	-.12**
Mathematics Step 2	-.09**	-.10**	-.10**

** Correlation is significant at the 0.01 level (2-tailed);

Once again, the coefficients corroborate the initial hypothesis. The residuals of a particular set of schools are more associated with the proportion of disadvantaged pupils. The schools were intentionally selected based on their segregation levels. All the evidence so far

[26] See previous chapter "Clustering Pupils in Rio de Janeiro Public Schools: one or more patterns of segregation?"

suggests that at least part of the school effect could be explained by a "peer effect". Since there is no information to discriminate school management or classroom teaching, it is not possible to completely rule out the school effectiveness hypothesis. Once again, the stability across all correlation coefficients (they are all negative and significant) reinforce the need for further investigation about the potential effect of school composition on pupils' attainment.

The final model for the cohort 3rd grade, 2009 aggregates the individual residuals at school level. The mean scores of the residuals are correlated with the proportion of disadvantaged pupils calculated for each school. In order to corroborate with the initial hypothesis, the correlation coefficients must be negative and the strength of the association higher compared with that in previous model – pupil level residuals.

Table 109: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 3rd Grade 2009.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.39**	- .40**	-.41**
Language Step 2	-.31**	-.31**	-.37**
Mathematics Step 1	-.28**	-.33**	-.31**
Mathematics Step 2	-.20**	-.25**	-.27**

** Correlation is significant at the 0.01 level (2-tailed);

The total number of cases (schools) included in this correlation was 774. As expected, the nominal values of the coefficients are higher compared with the figures in Table 107 and the coefficients were negative and significant at 0.01 level. Individual residuals aggregated at school level can be interpreted as a plausible "school effect", and its association with the proportion of disadvantaged pupils, evidence of a plausible school composition effect.

Table 110 presents the correlation coefficients, considering only a particular set of schools – the most segregated. Using the same criteria described above, the top and bottom quartile of a descriptive analysis regarding the proportions of poor families in each school was used as a cut-off point. Only half of all schools (total 387) were included.

Table 110: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 3rd Grade 2009.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.50**	- .51**	-.47**
Language Step 2	-.41**	-.40**	-.42**
Mathematics Step 1	-.40**	-.43**	-.41**
Mathematics Step 2	-.30**	-.33**	-.35**

** Correlation is significant at the 0.01 level (2-tailed);

The data for the cohort 3rd grade 2009 fully corroborate the initial hypothesis. The association between the aggregated residuals for a specific group of schools (highly segregated) and the proportion of disadvantaged pupils is higher. Comparing the correlation coefficients across different methods of calculations, it is possible to observe that all measurements are quite stable: a) always negatively associated; b) all significant at the 0.01 level; c) the association between the residuals and the proportion of disadvantaged pupils varies across different "types of schools" (more or less segregated).

***

The next section replicates all calculations for the cohort 7th grade 2009 – Table 111 presents the regression coefficients. As expected, prior attainment in the 7th grade 2009 is the best predictor. It is not clear why pupils on 8th grade (and others retained) have, on average,

lower scores in mathematics. This is something that should raise concern for researchers using secondary data from *Prova Rio* – especially this particular cohort measurement for mathematics. There are many uncertainties regarding the quality of the tests.

During this study, there has been direct contact with staff members from Rio de Janeiro Educational Department and the enterprise who has conducted the assessment in order to understand some unusual outcomes. The inquiry resulted in two revisions (made by the enterprise) of the raw individual scores for the year 2009 – 3rd and 7th grades. It is not clear the procedures conducted in the revision or the "mistakes"[27] corrected. What is clear is that the datasets have been updated after specific comments made by researchers involved in this study. This is one more indication that the *Prova Rio* raw scores should be interpreted with great caution.

Table 111: Linear Regression Estimating Proficiency in *Prova Rio* 2010 – Cohort 7th Grade, 2009.

	Language		**Mathematics**	
	Step 1	**Step 2**	**Step 1**	**Step 2**
	Standardized Coefficients			
Prova Rio First Measure	.608**	.574**	.442**	.426**
Parental Education	-	-.059***	-	-.048**
Colour Non-White	-	-.040**	-	-.031**
Poverty	-	-.045**	-	-.026**
Sex – Boy	-	-.011**	-	-.010+
Constante	99,38	117,18	116,33	123,47
N	32.167	27.889	31.161	27.876

** Coefficient is significant at the 0.01 level; + Coefficient is significant at 0.1 level; All coefficients not significant at 0.10 level were excluded

[27] We are assuming there has been some mistake, due to changes on the raw scores.

All characteristic of disadvantaged were significant and suggested that disadvantaged pupils presented lower attainment levels. The only exception is gender (boy) for mathematics, which was only significant at .10. This confirms that the achievement gap between boys and girls is only constant across the first and second segments for language (Portuguese).

The next step is to observe the correlations between the individual residuals from the regressions and the proportions of disadvantaged pupils – see Table 112. Once again the coefficients suggest a pattern: all correlations are negative and significant at .01. It is possible to notice that the strength of the association for the second segment cohort is slightly smaller compared with the first segment cohort. This outcome partially refutes Thrupp, Lauder and Robinson (2002) and Kounali et al. (2007) hypothesis regarding one of the main mechanism of school composition effect – "peer subculture". The nominal levels of the correlation coefficients in both segments are small and the differences between Tables 107 and 112 are even smaller. Nonetheless, what becomes clear is the pattern that negatively associates the individual residuals and the school social composition.

Table 112: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 7th Grade, 2009.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.08**	-.07**	-.07**
Language Step 2	-.07**	-.06**	-.06**
Mathematics Step 1	-.08**	-.06**	-.08**
Mathematics Step 2	-.07**	-.05**	-.07**

** Correlation is significant at the 0.01 level (2-tailed);

The smaller correlation coefficients could also be related to poor measurement in *Prova Rio* considering this particular cohort. The average scores for mathematics suggest that

pupils in the 8th grade 2010 know less about the subject compared to the previous year. Since the model is truly longitudinal (tracks the same pupils), it is hard to believe that pupils would actually forget part of what have been learned previously.

The hypothesis about poor assessment in *Prova Rio* would only be applied for mathematics and not necessarily for language that shows "credible"[28] scores for the same pupils. However, the correlation coefficients on Table 112 indicate that the strength of the association in the second segment is somewhat smaller, for both mathematics and language, compared with previous figures of first segment. This evidence weakens this particular hypothesis.

The next step is to select a particular set of schools for the cohort 7th grade 2009 in order to retest the correlations. Once again, the top and bottom quartile were chosen based on a descriptive analysis regarding the proportion of poor families in each school. Table 113 shows the correlation coefficients for all variables of disadvantaged. In order to corroborate with the initial hypothesis, the associations must be higher compared with the last table for the same cohort. Data confirms the hypothesis for all variables.

Table 113: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 7th Grade, 2009.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.10**	-.10**	-.10**
Language Step 2	-.08**	-.08**	-.09**
Mathematics Step 1	-.10**	-.09**	-.11**
Mathematics Step 2	-.09**	-.07**	-.09**

** Correlation is significant at the 0.01 level (2-tailed);

[28] The mean scores for language suggest that pupils in 8th have learned a little more about the subject.

So far, data from both cohorts in 2009 corroborate the initial hypothesis. The individual residuals of a particular set of schools are more associated with the proportion of disadvantaged pupils. The final model for the cohort 7th grade 2009 aggregates the individual residuals at school level. The mean scores of the residuals will be correlated once more with the proportions of disadvantaged pupils. In order to corroborate with the initial hypothesis, the correlation coefficients must be negative and the strength of the association higher compared with that in previous model – pupil level residuals.

Table 114: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 7th Grade, 2009.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.34**	-.27**	-.26**
Language Step 2	-.29**	-.21**	-.23**
Mathematics Step 1	-.33**	-.27**	-.31**
Mathematics Step 2	-.28**	-.22**	-.29**

** Correlation is significant at the 0.01 level (2-tailed);

The total number of cases (schools) included in this correlation was 377. As expected, the nominal values of the coefficients in Table 114 are higher compared with the figures in Table 113 and all the coefficients are negative and significant at 0.01 level. This is perhaps the best model to observe the plausibility of the school composition effect theory. The mean scores of the individual residuals can be interpreted as a plausible "school effect", and its association with the proportion of disadvantaged pupils, evidence of a plausible school composition effect.

Table 115 presents the correlation coefficients, considering only a particular set of schools – the most segregated. Using the same criteria described above, the top and bottom

quartile of a descriptive analysis regarding the proportions of poor families in each school was used as a cut-off point. Only half of all schools (total 189) were included.

Table 115: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 7th Grade 2009.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.35**	- .35**	-.39**
Language Step 2	-.30**	-.27**	-.35**
Mathematics Step 1	-.40**	-.34**	-.43**
Mathematics Step 2	-.36**	-.28**	-.40**

** Correlation is significant at the 0.01 level (2-tailed);

The association between the aggregated residuals and the proportion of disadvantaged pupils is higher analysing only a particular set of schools (highly segregated). The data from the first two cohorts (3rd grade 2009; 7th grade 2009) indicates that: a) prior attainment is by far the best predictor in the models; b) characteristics of disadvantaged are not very good predictors to estimate pupil's achievement in *Prova Rio* – for both segments; c) all variables of disadvantaged are significant in the regression coefficients suggesting that potentially disadvantaged pupils scores significant lower; d) there is an association between the regression residuals and the proportion of disadvantaged pupils – both variables are negatively correlated; e) Evidence support the initial hypothesis. It is plausible to presume that at least part of the variation usually interpreted as a school effect could attribute to a particular mix of pupils.

The data presented in the thesis do not tell anything about the mechanisms that might help understanding the results. Nonetheless, the so called "indirect effect" of school

composition on teacher recruitment (Lauder et al. 2010) should be highlighted as one plausible explanation. The current legislation in Rio de Janeiro public municipal schools provide choice and mobility for teachers with a few years working in the network. It means that less experienced teachers usually choose from a smaller pool of schools. Future studies should investigate the potential association between teachers profile and the social composition of schools. This could help understanding the results presented here.

***

After detailed presentation of the first two cohorts, the next step is to replicate the models Type 1 and 2 in the last two cohorts: a) 3rd grade 2010; b) 7th grade 2010. The general conclusions for both cohorts in 2010 are quite similar to those previously presented. The correlation coefficients between the regression residuals and the proportion of disadvantaged pupils are negative and significant at 0.01. The thesis presents all the Tables with the outcomes for the new cohorts. Preliminary conclusions are presented before moving to Type 3 model – school level analysis.

Table 116: Linear Regression Estimating Proficiency in *Prova Rio* 2011 – Cohort 3rd Grade, 2010

	Language		**Mathematics**	
	Step 1	**Step 2**	**Step 1**	**Step 2**
	Standardized Coefficients			
Prova Rio First Measure	.627**	.591**	.601**	.572**
Parental Education	-	-.082**	-	-.080**
Colour Non-White	-	-.052**	-	-.064**
Poverty	-	-.051**	-	-.051**
Sex – Boy	-	-.055**	-	.015**
Constante	69,12	91,1	77,95	123,47
N	41.940	37.471	41.905	37.432

** Coefficient is significant at the 0.01 level.

Table 117: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 3rd Grade, 2010.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.09**	- .09**	-.07**
Language Step 2	-.07**	-.07**	-.07**
Mathematics Step 1	-.09**	-.09**	-.07**
Mathematics Step 2	-.06**	-.07**	-.06**

** Correlation is significant at the 0.01 level (2-tailed);

Table 118: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 3rd Grade, 2010.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.13**	- .13**	-.08**
Language Step 2	-.10**	-.09**	-.07**
Mathematics Step 1	-.12**	-.13**	-.09**
Mathematics Step 2	-.09**	-.09**	-.07**

** Correlation is significant at the 0.01 level (2-tailed);

Table 119: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 3rd Grade, 2010.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.37**	-.38**	-.31**
Language Step 2	-.28**	-.28**	-.27**
Mathematics Step 1	-.33**	-.33**	-.25**
Mathematics Step 2	-.24**	-.23**	-.22**

** Correlation is significant at the 0.01 level (2-tailed);

Table 120: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 3rd Grade, 2010.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.51**	- .48**	-.31**
Language Step 2	-.41**	-.36**	-.27**
Mathematics Step 1	-.41**	-.41**	-.29**
Mathematics Step 2	-.31**	-.29**	-.24**

** Correlation is significant at the 0.01 level (2-tailed);

***

Table 121: Linear Regression Estimating Proficiency in *Prova Rio* 2011 – Cohort 7th Grade, 2010

	Language		**Mathematics**	
	Step 1	**Step 2**	**Step 1**	**Step 2**
	Standardized Coefficients			
Prova Rio First Measure	.618**	.587***	.533**	.515***
Parental Education	-	-.048***	-	-.068***
Colour Non-White	-	-.040***	-	-.047***
Poverty	-	-.038***	-	-.038***
Sex – Boy	-	-.088***	-	-
Constante	79,92	99,01	107,77	119,88
N	34.300	30.039	34.295	30.036

** Coefficient is significant

Table 122: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 7th Grade, 2010.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.06**	- .05**	-.05**
Language Step 2	-.05**	-.03**	-.05**
Mathematics Step 1	-.08**	-.06**	-.07**
Mathematics Step 2	-.07**	-.05**	-.06**

** Correlation is significant at the 0.01 level (2-tailed);

Table 123: Correlation Coefficients between Individual Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 7th Grade, 2010.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.09**	- .07**	-.07**
Language Step 2	-.07**	-.07**	-.06**
Mathematics Step 1	-.11**	-.09**	-.07**
Mathematics Step 2	-.09**	-.06**	-.06**

** Correlation is significant at the 0.01 level (2-tailed);

Table 124: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 7th Grade, 2010.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.28**	-.21**	-.22**
Language Step 2	-.22**	-.14**	-.20**
Mathematics Step 1	-.40**	-.31**	-.31**
Mathematics Step 2	-.34**	-.21**	-.26**

** Correlation is significant at the 0.01 level (2-tailed);

Table 125: Correlation Coefficients between Mean Scores of Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 7th Grade, 2010.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.40**	- .31**	-.29**
Language Step 2	-.33**	-.22**	-.26**
Mathematics Step 1	-.51**	-.41**	-.35**
Mathematics Step 2	-.44**	-.29**	-.29**

** Correlation is significant at the 0.01 level (2-tailed);

Discussion – Pupil Level Regression

This is the first attempt to produce evidence about school compositional effect using secondary data from Rio de Janeiro Educational Department. The design has the advantage of being longitudinal and the data covers all schools. Nonetheless, many limitations have been highlighted regarding the available data. Perhaps, the three most important are: a) missing data for *Prova Rio* – referring to the number of pupils who did not take the test; b) lack of information about school management and classroom teaching; c) uncertainty regarding the use of *Prova Rio* for pupil level analysis, and the consistency of measurements over the years.

The models presented are quite simple and aims to observe the association between the regression residuals and the proportion of disadvantaged pupils in each school. The school-mix effect theory claims that pupil composition can be seen as one factor that that may be significant in determining pupil outcomes.

The first section used regressions at pupil levels to obtain the residuals. There are many doubts about the use of *Prova Rio* for pupil level analysis, since the assessment was designed to measure school performance and uses Item Response Theory (IRT). It is not clear at this point, if the total number of questions in each individual test covers the entire scale of (IRT). If not, the attempt to estimate individual scores might be misleading. The next section will replicate the same models at school level to search for robustness.

All correlations coefficients (a total of 192 calculations) for Type 1 and 2 models suggest that the regression residuals and the proportion of disadvantaged pupils are negatively associated. This means that pupils enrolled in schools with a higher proportion of disadvantaged pupils presented, on average, negative residuals (the predicted scores were overestimated). The association between the social composition of schools and the residual

from the regressions (at pupil level) demonstrates the plausibility of the school-mix effect theory.

School Level Regression

This section presents evidence for model Type 3 – school level regression. The variables have been described in Table 105 and the model presents two main differences: a) the regressions are estimated with aggregated data (school-level) from *Prova Rio*; b) the design is cross-sectional and not longitudinal. The choice for a cross-sectional design was based on the criteria to incorporate a larger number of cases (students) for the second measurement of *Prova Rio* – dependent variable.

Table 126 presents the regression coefficients for the cohort 3rd grade 2009. It becomes clear that aggregating *Prova Rio* scores at school levels leads to a better model. The R Square in the "Summary Model" from SPSS indicates an improvement. Pupil level regressions presented R Squares from 0.25 to 0.40; the school level regressions show an increase – from 0.45 to 0.6. This was expected and indicates that the unexplained part is smaller in Type 3 models. This can have an impact on the correlations between residuals and proportion of disadvantaged. It also reinforces the need to interpret all pupil-level analysis (Type 1 specially) with great caution. At this point, it is not clear how safe it is to estimate *Prova Rio* scores at individual level.

Table 126: Linear Regression Estimating Proficiency in *Prova Rio* 2010 – Cohort 3rd Grade, 2009.

	Language	**Mathematics**
	Step 1	Step1
	STANDARDIZED COEFFICIENTS	
***Prova Rio* 1st Measurement**	.676**	.594**
Constant	67,63	101,31
N	703	703

** Coefficient is significant at the 0.01 level.

The next step is to correlate the residuals from the regression with the proportion of disadvantaged pupils. In order to corroborate with the initial hypothesis, the correlation coefficients should be negative and constant across all characteristics of disadvantaged. Table 127 presents the coefficients.

Table 127: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 3rd Grade, 2009.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.31**	- .31**	-.28**
Mathematics Step 1	-.29**	-.26**	-.21**

** Correlation is significant at the 0.01 level (2-tailed);

All correlation coefficients are significant 0.01 and negative. It is also relevant to notice that the strength of the association is higher compared with previous models using pupil-level residuals. The next step is to select the most segregated schools considering one specific indicator: poverty. Using the same criteria highlighted previously, only the top and bottom quartile were selected. Table 128 presents the correlation coefficients.

Table 128: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 3rd Grade, 2009.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.41**	- .40**	-.36**
Mathematics Step 1	-.37**	-.35**	-.28**

** Correlation is significant at the 0.01 level (2-tailed);

The correlation coefficients corroborate the initial hypothesis. Schools with a higher proportion of disadvantaged pupils present, on average, more negative residuals. It means that the predicted scores were overestimated than the actual observations in the second measurement of *Prova Rio* (dependent variable). This can be interpreted as a plausible evidence of school composition effect.

***

The next model uses data for the cohort 7th grade, 2009 and replicates the exact same analysis – Table 129 presents the regression coefficients. In the previous models (Type 1 and 2), the association between the individual residuals and the school social composition was lower considering the second segment cohort (7th grade). Will the residuals from the aggregated scores of *Prova Rio* tell a similar story?

Table 129: Linear Regression Estimating Proficiency in *Prova Rio* 2010 – Cohort 7th Grade, 2009.

	Language	**Mathematics**
	Step 1	Step1
	STANDARDIZED COEFFICIENTS	
***Prova Rio* 1st Measurement**	.758***	.794***
Constant	70,58	54,26
N	375	375

** Coefficient is significant at the 0.01 level.

Once more, it becomes clear that aggregating pupil score from *Prova Rio* at school level leads to a better R Square. It means that the model is more appropriate, with a direct impact on the residuals. Table 130 presents the association between the residuals and the proportion of disadvantaged pupils.

Table 130: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 7th Grade, 2009.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-	- .14*	-
Mathematics Step 1	-	-	-.12*

* Correlation is significant at the 0.05 level (2-tailed).

This is the first time, considering all correlation tables (pupil and school level analysis), where regression residuals are not associated with all characteristic disadvantaged. Considering the estimations for language, only the proportion of poor families seems to be negative associated (0.05 significant) with the regression residual. For mathematics, only parents' education (parents who did not finish high school) is associated (.05 significant) with the residuals.

It is not clear why there are such differences comparing cohorts in different segments (3rd grade 2009 and 7th grade 2009). Previous chapter highlighted that school segregation is higher in the first segment. The severe decline in the total number of schools leads to a decline in nominal levels of segregation. It is hard to make any speculations regarding the differences in the regression coefficients; nonetheless, the figures suggest that the potential impact of the school composition on attainment is higher in the first segment (3rd grade).

It is hard to explain why the correlation coefficients in different segments show different patterns. There are a few possible explanations for this odd outcome, none of them very convincing. First, it is important to highlight that the total number of cases (schools) in the second segment is smaller (375) compared with first segment (703). This could help explain why most correlations were not significant. However, 375 cases is not a very small number and do not fully explain the outcome.

Another possibility is related to the smaller segregation nominal levels in the second segment. Previous chapters in thesis have shown that segregation is higher in the first segment. This means that schools are more homogeneous in the second segment, which, in theory, diminishes the chances of observing any school-mix effect.

Table 131 presents the figures for the new correlations for the cohort 7th grade 2009, considering only the more segregated schools (top and bottom quartile). A total of 187 schools were included in the calculations.

Table 131: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 7th Grade, 2009.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-	- .17*	-16*
Mathematics Step 1	-	-	-.16*

* Correlation is significant at the 0.05 level (2-tailed);

The coefficients show an increase in the associations between the residuals and the proportions of disadvantaged pupils. This particular result refutes the first alternative explanation, which suggested that the small number of cases was the reason for the low correlations. If that was true, the correlation coefficients for the most segregated schools (using only half of the total number of schools) should have been even smaller. However the figures in Table 131 suggest the opposite outcome. The results for the cohort 7th grade 2009 were not so convincing regarding the plausibility of the school composition effect on pupils' attainment. For the first time, not all characteristic of disadvantaged were associated with the regression residuals.

***

After the presentation of the data for the first two cohorts, the next step is to replicate the Type 3 model in the last two cohorts: a) 3rd grade 2010; b) 7th grade 2010. The overall interpretations for both cohorts in 2010 are basically the same. For this reason, the thesis will only present the Tables with the regressions and correlation coefficients for: a) Cohort 3rd grade, 2010; b) Cohort 7th grade, 2010. Evidence for school-level regression will be followed by discussion.

Table 132: Linear Regression Estimating Proficiency in *Prova Rio* 2011 – Cohort 3rd Grade, 2010.

	Language	**Mathematics**
	Step 1	Step1
	STANDARDIZED COEFFICIENTS	
***Prova Rio* 1st Measurement**	.708***	.644***
Constant	76,77	91,74
N	697	697

** Coefficient is significant at the 0.01 level.

Table 133: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 3rd Grade, 2010.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.27**	- .30**	-.24**
Mathematics Step 1	-.28**	-.29**	-.21**

** Correlation is significant at the 0.01 level (2-tailed);

Table 134: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 3rd Grade, 2010.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.40**	- .40**	-.26**
Mathematics Step 1	-.38**	-.38**	-.24**

***

Table 135: Linear Regression Estimating Proficiency in *Prova Rio* 2011 – Cohort 7th Grade, 2010.

	Language	**Mathematics**
	Step 1	Step1
	STANDARDIZED COEFFICIENTS	
***Prova Rio* 1st Measurement**	.758***	.794***
Constant	70,58	54,26
N	375	375

** Coefficient is significant at the 0.01 level.

Table 136: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for All Schools – Cohort 7th Grade, 2010.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.13*	-	-
Mathematics Step 1	-.16*	-	-

** Correlation is significant at the 0.01 level (2-tailed);

Table 137: Correlation Coefficients between Regression Residuals and Proportions of Disadvantaged Pupils for Highly Segregated Schools – Cohort 7th Grade, 2010.

	PROPORTION OF DISADVANTAGED PUPILS		
	Non-White	Poverty	Parental Education
Language Step 1	-.19**	-	-
Mathematics Step 1	-.22**	-	-

** Correlation is significant at the 0.01 level (2-tailed);

Discussion – School Level Regression

Aggregated analysis at school level presented a slightly different outcome compared with models Type 1 and 2 (pupil level regressions). The estimations considering the cohorts in the first segment (3rd grade 2009; 3rd grade 2010) fully corroborates the initial hypothesis. The mean scores of *Prova Rio* considerably improved the models, increasing the R Square. Even with a smaller unexplained part, the association between the residuals and the proportion of disadvantaged was constant among all characteristic of disadvantaged. All correlation coefficients for both cohorts in the first segment were: a) always negatively associated; b) all significant at the 0.01 level; c) the association between the residuals and the proportion of disadvantaged pupils varied across different "types of schools" (more or less segregated). Evidence support the claim that at least part of the variation usually interpreted as a school effect could be attributed to a particular mix of pupils.

Nonetheless, data for second segment cohorts (7th grade 2009; 7th grade 2010) do not fully corroborate the initial hypothesis. Not all characteristics of disadvantaged are associated with the regression residuals. From a total of 24 calculations, only 9 correlations coefficients were significant at 0.05 or 0.01 level. The characteristics of disadvantaged associated with the residuals varied, depending on the cohort analysed, and did not highlight a clear pattern.

One possible explanation for the outcomes observed is the differences in the nominal levels of segregation comparing first and second segment. It is reasonable to assume that a decline in segregation would diminish the chance to observe any school composition effect. If it is true that schools are more homogeneous across second segment schools, it could be argued that the school-mix effect would be more prevalent among first segment schools. Future studies should retest the data used in this research and also create new clever designs to understand the differences observed between the first and second segment.

An alternative explanation highlights that there is no clear patterns in the association between the residuals and the proportion of disadvantaged pupils, because the models are more robust (independent variable explains a larger share of the total variance) in the second segment regressions. Data from the models support this claim. Table 138 presents the R Square figures for all Type 3 model regressions.

Table 138: Summary Model R Square for Type 3 Regressions

	R Square Language	**Mathematics**
Cohort 3rd Grade 2009	0.48	0.42
Cohort 3rd Grade 2010	0.50	0.41
Cohort 7th Grade 2009	0.57	0.63
Cohort 7th Grade 2010	0.59	0.64

It is plausible to assume that the larger variance explained in second segment regressions is the cause of the smaller association between the residuals and the proportion of disadvantaged. It should be added that the standard deviation from the regression residuals in the second segment are consistently smaller compared with figures from the first segment. All this information recommends a caution interpretation of all school composition effect evidence. All the associations observed in the models Type 1, 2 and 3 might not be related to an alleged school effect or school composition effect, but, instead, it could be related to problems in the measurements of *Prova Rio* (quality of the data). Nevertheless, it should be remembered that even for the most conservative outcomes (in this case, second segment Type 3 models) a total of 9 correlations coefficients were significant (at least at 0.05).

14. FINAL CONCLUSIONS

The data presented in this thesis provides a better understanding about the educational opportunities in Rio de Janeiro municipal public schools. The analytical models can be divided into two groups: a) causes of school segregation; b) potential impact of school segregation on pupil attainment. It becomes clear that the current educational legislation, along with some administrative practices, plays a key role in determining school intake. The models presented indicate that a large proportion of between-school segregation variation (can reach 50% depending on the variable) is associated with elements of the educational bureaucracy.

Initial evidence supports the plausibility of school composition effect theory. The findings presented in this thesis can be of relevance for policy-makers and educational researchers. It may be that schools, or even teachers, are being wrongly held responsible for their school's performance. The school composition effect could explain part of the variance that is usually attributed to school management and teacher performance. The results presented here are no definitive tests and future studies should replicate to test for robustness.

By depicting segregation between and within schools, and advancing in research on its possible effects on pupils' achievement, we presume to be drawing attention to a phenomenon usually not recognized in the realms of our academia and policy-making. The results are not striking, partly because private schools have not been taken into account. However, they are not negligible, as they refer to the whole of a major educational system, and are stable and convergent in all tests performed. Moreover, our analyses indicate the need for discussion about regulatory mechanisms that could prevent public bureaucracy from becoming an extra source of educational inequality in a country persistently listed amongst the most unequal in the world.

Throughout the thesis, age/grade distortion was highlighted as a major problem in basic education. It affects children from all different backgrounds, but the data indicate that it is more prevalent among disadvantaged pupils and those attending low performance schools. What is quite unique in Brazil is the retention rate. Rio de Janeiro has one of the largest and most homogeneous public networks in the country, and, despite all the good characteristics, around 25% of all pupils have two or more years of age/grade distortion.

Dealing with large age/grade discrepancy in the same class can be highly challenging for teachers. Relationship problems and social interaction among children and teenagers can come under criticism from parents, the pupils themselves and the educators. So far, apparently, the "solution" has been to separate low achievers (pupils with low academic performance over the years) from others that have followed the regular flow.

Two outcomes in the thesis corroborate this conclusion. The school "shift" allocation and the Special Class policy, suggests that age/grade distortion is the most important variable to understand any potential impact of both policies on school segregation. It is important to highlight that the "shift" allocation is simply an "administrative practice". There is no clear guidance that states any intention to cluster pupils with shared characteristics in different "shifts". Special Class has the clear intention to deal with a pedagogic problem: pupils with many years (usually three or more) of age/grade distortion.

It is crucial that policy-makers start to address the issue of evaluating educational policy more seriously. There are very few studies that have tried to analyse any potential impact of such policies, and even these usually present very limited research designs (sometimes study cases) that do not allow any causal claims. If the benefits for the most disadvantaged are not clear at all, what legitimizes such policies?

Society should not allow around 25% of its pupils to have two or more years of age/grade distortion. Clustering disadvantaged pupils in particular "school shifts" should

raise concern. Tracking policies (specifically aimed at the most disadvantaged) should be proven to be effective before implementing them in the entire educational system. These statements have not been fully discussed. The debate about quality should not ignore the issue of equity. International data indicate that the most successful educational systems (high overall averages) also have a low variation among schools (low standard deviation).

The challenge to improve the quality of public education in Brazil demands a decline in the proportion of low achievers. "Administrative practices" or formal policies that separate low achievers from other pupils could be ineffective in helping to close the gap between these two groups. International evidence suggests that clustering disadvantaged pupils can increase the association between socio-economic status and academic achievement (EGGRES, 2005).

This is perhaps the core of the debate in the thesis. International comparative studies are unanimous in suggesting that public education in Brazil is of very low quality. Pupils are not learning what they should at the appropriate age. More government resources should be focused on understanding the main causes of such failure, and implement changes. Since many countries have overcome our current stage, it should not be very hard to learn from successful examples and implement similar policies in Brazil. Nonetheless, there is no clear sign of educational reforms that could change the current scenario for the next generation.

Since 2009, Rio de Janeiro Educational Department has implemented many changes in its educational system. Some of these changes are connected with the very latest thinking and debates in education, such as, the demand for standardized assessment, high-stake policies, clear curriculum guidance and compensatory policies.[29] Notwithstanding, the efforts can be wasted (at least most of the resources) due to a lack of robust assessment. It is not absurd to state that policy-makers know very little (being optimistic) about the impact of such

[29] One good example of recent compensatory policies is the *Escolas do Amanhã* (The Schools of Tomorrow) program.

policies. The claim for better education cannot escape the urgent demand for serious evaluation of current educational policies.

REFERENCES

Almeida, A. C. (2007) A cabeça do brasileiro. São Paulo: Editora Record.

Alves, F. Lange, W & Bonamino, A. (2010) A Geografia Objetiva de Oportunidades Educacionais na Cidade do Rio de Janeiro. In: Ribeiro, L. C. Kolisnski, M. Alves, F & Lasmar, C. (Org.). *Desigualdades Urbanas, Desigualdades Escolares*. Rio de Janeiro: Letra Capital.

Bartholo, T. L. (2013) Measuring between-School Segregation in an Open Enrolment System: The Case of Rio de Janeiro. Journal of School Choice, 7: 353-371.

Bell, W. (1954). A Probability Model for the Measure of Ecological Segregation. Social Forces, (32) 357–364.

Boudon, R. (1981) *As Desigualdades de Oportunidades*, Editora Universidade de Brasília, Brasília.

Brito, M.; Costa, M. (2010) Teachers' practices and perceptions and their relations with prestige and school environment in public schools in the municipality of Rio de Janeiro. Revista Brasileira de Educação (Impresso), v. 15, n.45, p. 500-510.

Bruel, A. L; Bartholo, T. L. (2012) Inequality of Educational Opportunities in Rio de Janeiro Public School System: Transition between Segments of Elementary School. *Revista Brasileira de Educação*, 17 (50), 303–328.

Campbell, D. & Stanley, J. (1963). Experimental and quasi-experimental designs for research. Chicago, IL: Rand-McNally.

Cano, I. (2002) Introdução à Avaliação de Programas Sociais. São Paulo: FGV.

Cavallieri, F & Lopes, G. (2008) Índice de Desenvolvimento Social - IDS: comparando as realidades microurbanas da cidade do Rio de Janeiro. COLEÇÃO ESTUDOS CARIOCAS. Nº 20080401 Instituto Pereira Passos. Rio de Janeiro.

Coleman, J. S. (1961) *Social Climates in High Schools.* U.S. (Dept. of Health, Education, and Welfare, Office of Education).

Consorte, J. G. (1959) A criança favelada e a escola pública. *Educação e Ciencias Sociais*, 5 (11), 45–60.

Condron, D. (2011) Egalitatianism and educational excellence: Compatible goals for affluent societies? Educational Reseacher, 40, 2, 47-55.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston, MA: Houghton Mifflin Company.

Costa, M. (2008). Prestígio e hierarquia escolar: estudo de caso sobre diferenças entre escolas em uma rede municipal. Revista Brasileira de Educação, (13) 455–469.

Costa, M., & Koslinski, M. C. (2011). A Hidden-Quasi-Market: Dispute for Common Schools in Brazil. Cadernos de Pesquisa, (41), 246–266.

Costa, M., & Koslinski, M. C. (2012). Public schools: choice, strategy and competition. Pró-Posições, 23 (2), 1-19.

Costa, M. ; Prado, A. P.; Rosistolato, R. (2013) Talvez se eu tivesse algum conhecimento... : caminhos possíveis em um sistema educacional público e estratificado. Interseções (UERJ), v. 14, p. 165-193.

Costa; M.; Koslinski; M. Bruel, A. Bartholo, T. L (2013) Movimentação de estudantes em um sistema educacional – padrões de "tracking", rotulagem e reprodução da estratificação social. Revista Universidade Federal de Juiz de Fora (PRELO).

Costa, M., Alves, M. T. G., De Sá, T. C. D. & Moreira, A. M. (2013). Oportunidades e escolhas: famílias e escolas em um sistema escolar desigual. In: Nogueira, Zago e Romanelli (orgs.) Família e Escola 2. Editora Vozes Rio de Janeiro.

Dumont, L (1992). Homo Hierarchicus: O Sistema das castas e suas implicações. São Paulo: Editora da Universidade de São Paulo.

Duncan, O. D., & Duncan, B. (1955). A Methodological Analyses of Segregation Indexes. American Sociological Review, 20 (2), 210–217.

European Group for Research on Equity in Educational Systems (2005) Equity in European Educational Systems: a set of indicators, European Educational Research Journal, 4(2), 1–151.

Goldhaber, D. (1999) School Choice: An Examination of the Empirical Evidence on Achievement, Parental Decision Making and Equity. Educational Researcher, 28, 16–25.

Gorard, S. (2006). Is there a school mix effect? Educational Review, 58 (1) 87–94.

Gorard, S. (2009a). Does the index of segregation matter? The composition of secondary schools in England since 1996. British Educational Research Journal, 35 (4) 639–652.

Gorard, S. (2009b). Serious doubt about school effectiveness. British Educational Research Journal, 36 (5) 745-766.

Gorard S. (2011) Measuring Segregation – Be aware of the cautionary tale by Johnston and Jones. Environment and Planning A 43: p: 1-7.

Gorard, S., & Taylor, C. (2002). A comparison of segregation indices in terms of strong and weak compositional invariance. Sociology, 36 (4), 875–895.

Gorard, S., & Cheng, S. C. (2011). Pupil clustering in English secondary schools: one pattern or several?, International Journal of Research & Method in Education, 34 (3), 327–339.

Gorard, S., & Smith, E. (2010). Equity in Education: An international comparison of pupil perspectives. London: Palgrave Macmillan.

Gorard, S., See, B. H. (2013) Overcoming Disadvantage in Education. London: Routledge Falmer.

Gorard, S., Taylor, C., & Fitz, J. (2003). Schools, markets and choice policies. London: Routledge Falmer.

Gorard, S., See, B. H., & Davies, P. (2011). Do attitudes and aspirations matter in education?: A review of the research evidence, Saarbrucken: Lambert Academic.

Gorard, S. (2012a) The Increasing Availability of Official Datasets: Methods, Limitations and Opportunities for Studies of Education. *British Journal of Educational Studies*, 60(1), 77-92.

Gorard, S. (2012b) Who is eligible for free school meals?: Characterising FSM as a measure of disadvantage in England, *British Educational Research Journal*. 38(6) 1003–1017.

Gramberg, P. (1998). School Segregation: The Case of Amsterdam. Urban Studies, 35 (3) 547–564.

Haarh, J., Nielsen, T., Hansen, E., & Jakobsen, S. (2005). Explaining student performance: evidence from the international PISA, TIMSS and PIRLS surveys, Danish Technological Institute. Available online at: www.danishtechnology.dk

Harker, R. (2004) Compositional effects in school effectiveness studies: a New Zealand case study, paper presented at the AERA Annual Conference, San Diego, CA, April.

Harris, R. (2012). Local Indices of Segregation with Application to Social Segregation between London's Secondary Schools. Environment and Planning, 44 669–687.

Hill, P. T.; & Lake, R. J. (2010). The Charter School Catch-22, Journal of School Choice: Research, Theory, and Reform, 4 (2), 232–235.

Ireson, J. Hallam, S (2001). Ability Grouping in Education. London: Paul Chapman Publishing.

Jacobs, N. (2011). Racial, Economic, and Linguistic Segregation: Analyzing Market Supports in the District of Columbia's Public Charter Schools. *Education and Urban Society.* XX(X) 1–22. DOI: 10.1177/0013124511407317

Jencks, C (1972). Inequality: A Reassessment of the Effect of Family and Schooling in America. London: Lowe & Brydone.

Johnston R, Jones K, (2010). Measuring segregation “a cautionary tale”. Environment and Planning A 42: p: 1264-1270.

Johnston R, Jones K, (2010). A brief response to Gorard. Environment and Planning A 43: p: 8-9.

Kounali, D.,Robinson, A., Goldstein, H., and Lauder, H., (2007) The Probity of Free School Meals as a Proxy Measure for Disadvantage, *Bristol University, Bristol.* http://www.bristol.ac.uk/cmm/publications/fsm.pdf

Lauder, H.; Kounali, D.; Robinson, T.; Goldstein, H. (2010) Pupil Composition and Accountability: and analysis in English primary schools. International Journal of Educational Research, 49 (2-3), 49-68.

Logan, J., Minca, E.; Adar, S. (2012) The geography of inequality: Why separate means unequal in American public schools. Sociology of Education, 85, 3, 287-301.

Massey, D., White, M., & Phua, V. (1996). The Dimensions of Segregation Revisited. Sociological Methods and Research, 21 (2), 281–292.

Merrifield, J. (2008). The Twelve Policy Approach to Increase School Choice. Journal of School Choice, 2 (1), 4-19. DOI: 10.1080/15582150802007267

Nash, R. (2003) Is the School Composition effect Real? A Discussion with Evidence from the UK PISA Data, *School Effectiveness and Improvement*, 14, 4, 441-457.

Ravitch, D. (2010) The Death and Life of the Great American School System: how testing and choice are undermining education. New York, Basic Books.

Rawls, J. (1971) *A Theory of Justice*. Harvard University Press.

Ribeiro, L. C & Koslinski, M. C. (2009) A cidade contra a escolar? O caso do Municipio do RJ. *Revista Contemporanea de Educacao*, (4) 351–378.

Ribeiro, L. C. Q., Koslinski, M. C., Alves, F., Lasmar, C. (2010). Desigualdades Urbanas, Desigualdades Escolares. Rio de Janeiro: Letra Capital.

Rosenthal, R & Jacobson, L. (1968) *Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development*. (New York. Holst, Rinehart & Winston).

Shevky, E., & Shevky, M. W. (1949). The Social Areas of Los Angeles: Analyses and Typology. University of California Press, Berkeley.

Sikkink, D.; & Emerson, M. (2008). School choice and racial segregation in US schools: The role of parents' education, Ethnic and Racial Studies, 31 (2), 267–293.

Saporito, S. (2003). Private Choices, Public Consequences: Magnet School Choice and Segregation by Race and Poverty. Social Problems, 50 (2) 181–203.

Smylie, M. A. Robert L. Crowson, V. C & Rebekah A. L. (2004) The Principal and Community-School Connections in Chicago's Radical Reform. *Educational Administration Quarterly* 30: 342.

Thrupp, M. (2007) School Admissions and the Segregation of School Intakes in New Zealand Cities. Urban Studies, Vol. 44, No. 7.

Thrupp, M. (1997) The school Mix effect: The History of an Enduring Problem in Educational Research, Policy and Practice, British Journal of Sociology of Education, 16,2, 183-203.

Thrupp, M. (1999) *Schools Making a Difference: Let's be Realistic*, Buckingham, Open University Press.

Thrupp, M., Lauder, H., and Robinson, T. (2002) School Composition and Peer Effects, International Journal of Educational Research, 37 (5), 483-504.

Thrupp, M., and Hirsch, D., (2006), The Limits of Managerialist School Reform: The Case of Target setting in England and the USA, in Lauder, H, Brown, P, Dillabough J-A and Halsey, A.H. (eds.) (2006) *Education, Globalization and Social Change*, Oxford, Oxford University Press.

Van Zanten, A. (2005). Efeitos da Concorrência sobre a Atividade dos Estabelecimentos Escolares. Cadernos de Pesquisa, v 35, n. 126, p.565-593.

Vaus, D. A. (2011) Research Design in Social Research. (London: SAGE Publications).

Veloso, F. (2009). 15 anos de avanços na educação no Brasil: onde estamos? In: Veloso, F.; Pessôa, S.; Henriques, R.; Giambiagi, F. Educação Básica no Brasil: construindo o país do futuro. Rio de Janeiro: Elsevier, p. 3-24.

West, Anne and Pennell, Hazel (2002) *How new is New Labour? The quasi-market and English schools 1997 to 2001*. British Journal of Educational Studies, 50 (2). pp. 206-224. ISSN 1467-8527

West, A.; Hind, A.; Pennell, H. (2004) School admissions and "selection" in comprehensive schools: policy and practice. Oxford Review of Education, 30 (3). pp. 347-369. ISSN 0305-4985

Wilson, W. J. (1987) *The Truly Disadvantaged: the inner city, the underclass and public policy*. (Chicago: University of Chicago Press).

Yair, G. (1996). School Organization and Market Ecology: a Realist Sociological Look at the Infrastructure of School Choice. British Journal of Sociology of Education, 7, (4) 453-471.

Yorke, M. (2011) Analysing existing datasets: some considerations arising from practical experience. International Journal of Research & Method in Education, 34(3), 255–267.

Sites:

http://aplicacoes.mds.gov.br/sagi/RIv3/geral/relatorio.php#Cadastro Único

http://www.ejario.com/

http://veja.abril.com.br/noticia/brasil/numero-de-beneficiarios-do-bolsa-familia-so-cresce

http://noticias.r7.com/brasil/sete-em-cada-dez-beneficiarios-do-bolsa-familia-trabalham-18102013

http://www.mds.gov.br/falemds/perguntas-frequentes/bolsa-familia/cadastro-unico/beneficiario/cadunico-inclusao

http://portalpbh.pbh.gov.br/pbh/ecp/comunidade.do?evento=portlet&pIdPlc=ecpTaxonomiaMenuPortal&app=politicassociais&lang=pt_BR&pg=5567&tax=8963