

Contents

1	General Introduction	2
I	Multilevel modeling dealing with selection bias and missing data in educational sciences	6
2	Introduction	7
3	Theoretical Framework: Description of multilevel models and methods for selection bias and missing data	10
3.1	Preliminaries and statistical notation	10
3.2	Modeling the data structure: Multilevel models for cross-sectional and longitudinal data	12
3.2.1	Multilevel models as extensions of the regression model	12
3.2.2	Multilevel growth models	15
3.2.3	Cross-classified models and multiple-membership models	16
3.3	Modeling selection bias: Linear regression and Propensity Score models	18
3.3.1	Causal Estimation	18
3.3.2	Regression Models and Selection Bias	20
3.3.3	Propensity Score Analysis	22
3.4	Modeling Missing Data: Multiple imputation	23
3.4.1	Missing data mechanism	23
3.4.2	Multiple imputation	25
4	Simultaneously modeling selection bias and missing data in the context of multilevel data	27
4.1	Selection bias and multilevel data	27
4.1.1	Hypotheses from the selection bias methods in a multilevel context	28
4.1.2	Modeling multilevel data and selection bias	31
4.1.3	Discussion regarding the modeling of multilevel data and selection bias	35
4.2	Multiple imputation and multilevel data	36
4.3	Selection bias, multiple imputation and multilevel data	38
5	Summary and future research	42

II	The influence of Life and Non-Life sciences courses on secondary students' epistemic and self-efficacy beliefs related to sciences	46
6	Introduction and Literature Review	47
6.1	Introduction	47
6.2	Literature Review	49
6.2.1	Science Epistemic and Self-Efficacy Beliefs	49
6.2.2	Science epistemic beliefs, self-efficacy beliefs and change: Influence of academic climates	50
6.2.3	Science university courses for high school students	52
7	Methodology	54
7.1	Sample	54
7.2	Variables: Measures of Sciences Epistemic Beliefs and Self-efficacy beliefs	55
7.3	Procedure	55
7.4	Data Analysis	56
8	Results	58
8.1	Psychometric analysis	58
8.1.1	Confirmatory Factor analysis and reliability estimation of student beliefs	58
8.1.2	Compositional variables to describe class climate	59
8.2	Descriptive analysis	59
8.3	Modeling course effects	61
8.4	Effect of academic climate at the end of the courses	61
8.5	Effect of academic climates some months after the courses	64
9	Discussion and Conclusions	66
III	School effects in academic trajectories of Chilean students	69
10	Introduction and Literature Review	70
10.1	Introduction	70
10.2	Context and Literature Review	72
10.2.1	School effects in students long-term achievement trajectories	72
10.2.2	Chilean school system	76
10.2.3	Comparisons between voucher and public schools in Chile	80
11	Methodology	85
11.1	Sample	85
11.2	Variables	86
11.2.1	Time variable	86
11.2.2	Treatment variable	88

11.2.3	Outcome variables: Student achievement measures	88
11.2.4	Control variables: Student socioeconomic and expectatives measures	89
11.2.5	Treatment variation: School aggregated variables	90
11.3	Data Analysis	93
12	Results	95
12.1	School effects and characteristics of achievement trajectories of Chilean students	95
12.1.1	School effects on students' trajectories	95
12.1.2	Students trajectories in different groups	99
12.2	Comparison of public and vouchers education in students trajectories	104
12.2.1	Background characteristics of the treatment groups	104
12.2.2	Modeling strategies to estimate the effect of public education .	106
12.2.3	Application of propensity score matching	108
12.2.4	Effect of public education: combination of multilevel models, linear adjusting and propensity score matching	110
13	Discussion and Conclusions	115
14	General Discussion	122
	Bibliography	127
	Appendices	137
A	Part II	139
A.1	Epistemic and Self-efficacy Beliefs questionnaire	139
A.2	Multiple imputation procedure	142
B	Part III	144
B.1	Multiple imputation procedure	144
B.2	Definition of multilevel growth models	147
B.3	Results of unconditional multilevel growth models for trajectory de- scription	150
B.4	Common support control variables	158
B.5	Common support treatment variables	161
B.6	Results of unconditional multilevel growth models for treatment effect estimation in the complete sample	163
B.7	Results of unconditional multilevel growth models for treatment effect estimation in the matched sample	167

List of Tables

3.1	Description of the variable notation and terminology in each model. . .	12
3.2	Multilevel models with increasing complexity.	14
7.1	Sample sizes for each measurement time.	54
7.2	Application calendar and procedure.	56
7.3	Sample sizes and drawbacks for the missing data strategies.	56
7.4	Description of variables and percentages of response.	57
8.1	Robust fit indexes for the measurement model.	58
8.2	Principal component analysis for the aggregated beliefs variables. . .	59
8.3	Means and standard deviations of students beliefs.	60
8.4	Description of the estimated multilevel regression models.	62
8.5	Multilevel regression models to determine the effect of academic climates at the post measure.	63
8.6	Linear regression models to determine the effect of academic climates at the follow-up measure.	65
10.1	Mean differences in socio-economic variables and attendance to pre-school.	84
11.1	Sample sizes for each level	86
11.2	Description of variables.	87
11.3	Time coding	87
11.4	Sample sizes for each treatment	88
11.5	Conversion table for educational level.	90
11.6	Conversion table for income.	91
11.7	Conversion table for number of books at home.	91
11.8	School indexes definition	92
12.1	Variances of the random effects from the unconditional mean models for mathematics.	97
12.2	Variances of the random effects from the unconditional mean models for reading.	97
12.3	Fixed effects and variances of the random effects from the unconditional growth models for mathematics.	99
12.4	Fixed effects and variances of the random effects from the unconditional growth models for reading.	100

12.5	Mean differences for background variables.	105
12.6	Mean differences for school indexes.	106
12.7	Models for treatment effect estimation.	108
A.1	Items of Self-efficacy in science	139
A.2	Items of Certainty.	140
A.3	Items of Justification.	140
A.4	Items of Source.	141
A.5	Items of Development	141
B.1	Predictor variables used in the multiple imputation models.	146
B.2	Estimated parameters for the models 3.a, 3.b, 3.c, 3.d, 3.e and 3.f. in mathematics scores.	150
B.3	Estimated parameters for the models 3.g, 3.h, 3.i, 3.j, 3.k and 3.l. in mathematics scores.	152
B.4	Estimated parameters for the models 3.a, 3.b, 3.c, 3.d, 3.e and 3.f. in reading scores.	154
B.5	Estimated parameters for the models 3.g, 3.h, 3.i, 3.j, 3.k and 3.l. in reading scores.	156
B.6	Estimated parameters for the models 4.a, 4.b, 4.c in mathematics scores.	163
B.7	Estimated parameters for the models 4.a, 4.b, 4.c in reading scores.	165
B.8	Estimated parameters for the models 5.a, 5.b, 5.c and 5.d in mathe- matics scores.	167
B.9	Estimated parameters for the models 5.a, 5.b, 5.c and 5.d in reading scores.	169

List of Figures

3.1	Multilevel growth models	17
4.1	Selection bias and clustered data.	34
4.2	Selection bias and clustered data.	35
4.3	Statistical strategies to address selection bias and missing data using multilevel models	41
8.1	Intra-class correlations.	61
10.1	Number of public, voucher and private schools from 2004 to 2013 . .	80
11.1	Normalization of the SIMCE test scores	89
11.2	Statistical strategies to address selection bias and missing data. . . .	94
12.1	Predicted trajectories for students with different characteristics	102
12.2	Predicted trajectories for students with different characteristics	103
12.3	Common support of the estimated propensity scores	109
12.4	Estimated ATE on reading and mathematics scores using different models.	112
12.5	Predicted trajectories for students with different characteristics	113
12.6	Predicted trajectories for students with different characteristics	114
B.1	Histograms for control variables in 4th grade	158
B.2	Histograms for control variables in 8th grade	159
B.3	Histograms for control variables in 10th grade	160
B.4	Histograms for school level variables at 8th grade (treatment variables) . .	161
B.5	Histograms for school level variables at 10th grade (treatment variables) .	162